

Evaluation of the Hierarchical Correspondence between the Human Brain and Artificial Neural Networks: A Review

Trung Quang Pham¹, Teppei Matsui² and Junichi Chikazoe^{1,*}

- ¹ Araya Inc., Tokyo 101-0025, Japan; trung_pham@araya.org
- ² Graduate School of Brain Science, Doshisha University, Kyoto 610-0321, Japan; matsui.teppei@gmail.com
- * Correspondence: chikazoe_junichi@araya.org

Simple Summary: Artificial neural networks, inspired by the human brain, have demonstrated human-level performance across multiple task domains, raising the exciting possibility of them returning insights to neuroscientists about the human brain. However, artificial neural networks cannot be directly compared to the brain due to inherent differences in structure and computation. Here, we review the variety of approaches that researchers have thus far taken to evaluate the correspondence between the two, organized across multiple levels of analysis (node, layer, network, and behavior). In surveying these approaches, we note some of the insights uncovered, their limitations, and future directions in a domain of research that is developing quickly and with few established standards and practices. Our aim is to provide a systemized overview and guidance toward establishing a firmer theoretical and methodological framework in this emerging field.

Abstract: Artificial neural networks (ANNs) that are heavily inspired by the human brain now achieve human-level performance across multiple task domains. ANNs have thus drawn attention in neuroscience, raising the possibility of providing a framework for understanding the information encoded in the human brain. However, the correspondence between ANNs and the brain cannot be measured directly. They differ in outputs and substrates, neurons vastly outnumber their ANN analogs (i.e., nodes), and the key algorithm responsible for most of modern ANN training (i.e., backpropagation) is likely absent from the brain. Neuroscientists have thus taken a variety of approaches to examine the similarity between the brain and ANNs at multiple levels of their information hierarchy. This review provides an overview of the currently available approaches and their limitations for evaluating brain–ANN correspondence.

Keywords: hierarchical correspondence; artificial neural networks; neuroscience

1. Introduction

Understanding the information processing of the human brain is one of the biggest challenges for neuroscientists. In recent years, the artificial neural network (ANN) has become a powerful tool, performing at or better than human levels in several domains, including image classification (AlexNet [1]), conversation (ChatGPT [2], LaMDA [3]), games (Go [4], Starcraft II [5], and biological science (i.e., protein folding [6,7]). Growing interest has thus emerged as to the degree to which the information processing of ANNs can inform that which occurs in the brain.

Studies have shown that the processing of human perceptions is hierarchically distributed over the brain ([8–13]). In the visual domain, for instance, the V2 neuron appears to be sensitive to naturalistic texture stimuli [14], V4 neurons increase selectivity for the conjunction of features representing the surface shape (i.e., non-Cartesian gratings [15]), and IT neurons show stimulus selectivity, sensitive to the specific combinations of features, i.e., the face [16]. A similar hierarchy can be found in language processing [17–19], music processing [20–22], and tactile processing [23–25]. Taking a broader perspective, converging



Citation: Pham, T.Q.; Matsui, T.; Chikazoe, J. Evaluation of the Hierarchical Correspondence between the Human Brain and Artificial Neural Networks: A Review. *Biology* **2023**, *12*, 1330. https://doi.org/10.3390/ biology12101330

Academic Editor: Julian Keenan

Received: 14 August 2023 Revised: 22 September 2023 Accepted: 10 October 2023 Published: 12 October 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/).



evidence alerts us to the brain's global hierarchical organization beyond a collection of independent sensory hierarchies. At the cellular level, Murray et al. [11] found different decay rates of a single-unit activity in early sensory areas and distant cortical regions. Whereas sensory areas may need to respond more rapidly to environmental changes to reflect faster decay rates, regions involved in more complex, integrative cognitive tasks exhibit longer decay rates, suggesting a hierarchical ordering of a measure as intrinsic as single-neuron spiking timescales. Neuroimaging evidence of a global, sensorimotor-to-transmodal gradient supported this hierarchy of temporal dynamics, as well as other converging evidence such as increasing intracortical myelination, functional connectivity, and semantic processing along the gradient [26,27].

Given the intrinsic hierarchical architecture of ANNs, it becomes natural to wonder if they can capture the information processing that occurs in the human brain, thus serving as a framework for its understanding [28,29]. While both the human brain and the ANN are "black-boxes", the latter is easier to customize and analyze. ANNs may provide a useful model for understanding the former, akin to how the atomic model can usefully convey the interaction between protons and electrons. As statistician George Box once said, "All the models are wrong, but some are useful".

Relationships between the human brain and modern ANNs have been found since the early stages of ANN development. Studies have revealed the similarity between cognitive processing, such as vision and audition, and the hidden layers of ANNs [28,30–36]. The similarity was not limited to well-known "supervised" learning, but also "unsupervised" learning and "self-supervised" learning [37,38]. The growing number of studies in this area offers promise toward improving our understanding of the brain, as ANNs rapidly grow in sophistication and performance across problem domains. However, comparing ANNs to the brain to arrive at meaningful references is not a straightforward process. ANNs are inspired by the brain but are not replicas. Not only do they differ in substrate, but there are also vastly fewer ANN nodes than neurons. The principal algorithm that discovers the hierarchical circuitry of most modern ANNs is unlikely to exist in the brain [39].

Conventionally, evaluation of the similarity between the ANN and the human brain has been based on their performance in "intelligent" tasks (e.g., object detection, object classification, text generation, image generation, game playing, etc.). However, just this high-level comparison is inadequate for determining whether the ANN under the hood is undergoing comparable information processing to the brain. Neuroscientists have thus taken a variety of indirect approaches to evaluate the correspondence between ANNs and the brain. Here, we review the current approaches toward evaluating brain–ANN correspondence, considering their insights, limitations, and future directions.

2. Levels of Correspondence

Brain–ANN correspondence can be examined at multiple levels of the information hierarchy. A typical ANN consists of nodes and layers, where nodes constitute a layer, and layers constitute a network. These levels analogously map to neuron, region, and network levels of the brain.

A recent review of convolutional ANNs and human vision [40] considered two levels at which ANNs correspond to the brain: at the lowest level of neural activity and at the highest level of human behavior. But an examination of hierarchical correspondence would have to consider intermediate levels of hierarchy in the brain, such as those found in the visual system [28,31–36]. While the brain has served as inspiration for modern ANNs that achieve human-level performance, ANNs are not yet bound to the same global structural and functional hierarchy of the brain. We are indeed on the brink of large-scale multimodal ANNs (e.g., processing image, audio, and text); however, their emergence is driven more by engineering insights than natural selection or the neuroscience of brain correspondence. In other words, with the current and future likelihoods of divergence between ANN models and the brain, we will be better equipped to evaluate correspondence at multiple levels.

Here, we introduce two intermediate levels toward a more comprehensive evaluation of the brain–ANN correspondence. Our review is thus structured along four levels of brain–ANN correspondence as follows (Figure 1a–d).

- Node level: Correspondence between one node of the ANN and the smallest unit of the brain (a single-unit recording, an electrode, or a voxel, depending on the measuring instruments).
- Layer level: Correspondence between an ANN hidden layer and a brain region (e.g., V1, V2, V4, IT).
- Network level: Correspondence between the overall information flow inside an ANN and inside the brain (e.g., hierarchical representation, multimodal network integration).
- Behavioral level: Similarity of ANN behavior, output, or performance metrics (e.g., image classification and generation, conversation, folding protein, and playing games) compared to the human counterparts.
- a. Node level



Figure 1. Cont.



behavioral metrics

Figure 1. Four levels of correspondence evaluation. (a) The node level evaluates the correspondence between one node of the ANN and the smallest measured unit of the brain (e.g., a single unit recording, an electrode, a voxel). When computing correspondence via encoding, brain units are modeled using ANN nodes, and vice versa in decoding. (b) The layer level evaluates the correspondence between an ANN hidden layer and a brain region. Illustrated approaches show encoding via many nodes of a layer to a single voxel activity in a known brain region (top), decoding via many voxels to one node (middle), and correspondence between many nodes and many voxels via RSA (representational similarity analysis) (bottom). (c) The network level evaluates the correspondence between the overall information flow inside an ANN and across multiple regions, cortices, or the whole brain. Illustrated approaches sum the number of voxels in each region most associated with a given ANN layer (top), and examine the relationship between voxels' scores (e.g., principal gradient score [26]) and their associated layers (bottom). (d) Behavioral level correspondence compares the ANN output, which includes qualitative assessments such as the Turing test in conversation, or compares the performance of behavioral metrics (e.g., classification accuracy, response time, error pattern, playing games) against their human counterparts. Brain illustrations were generated using the Connectome Workbench visualization software v1.4.2 [41] with Gordon's parcellation [42].

3. Node Level

At the node level, correspondence is examined between one node of the ANN and the activity of a neuron, or the smallest unit of neural measurement, such as an electrode or a voxel. The idea is that if the information at one node can effectively predict the activity of a single neuron, or vice versa, it is likely that they are associated. Modeling can be bidirectional, where "encoding" refers to an ANN node predicting a neuron, and "decoding" refers to a neuron predicting a node.

3.1. Encoding/Decoding

In an encoding study, Yamins et al. [30] obtained the visual response of every node in a visual categorization ANN and used them to predict the neural recordings from V1, V4, and IT sites in rhesus macaques, collected via implanted multielectrode arrays. Linear regression was used for identifying the best combination of ANN unit outputs, ultimately accounting for more than 54% of the variance of the image-evoked brain responses. Bashivan et al. [32] used an ANN-driven image synthesis method to generate texture-like patterns that pushed the predicted spiking activity of neural sites (V4) beyond the range of naturalistic stimuli (a firing rate that was 39% higher than the maximal naturalistic firing rate). In human subjects, Zhuang et al. [37] extended the evaluation at node level correspondence along the ventral visual stream with an unsupervised learning method. Their ANN trained with unsupervised objectives using deep contrastive embeddings beforehand exhibited a neural predictivity comparable to its supervised counterpart.

Correspondence can also be demonstrated in the opposite direction, such as by Horikawa et al.'s [35] decoding approach, where a single node was predicted using voxel information from the visual cortex. Then, based on which features were predicted, the authors were able to identify which object categories participants saw or imagined. The predicted feature values were shown positively correlated with the true values across test images.

Studies outside the visual domain have illustrated the potential for creatively analyzing ANNs at the node level. In the auditory domain, Drakopoulos [43] used ANNs to model mechanical (inner-hair-cell transduction) and neural (auditory-nerve fiber) processing of sound that accurately simulated single-unit auditory nerve responses. Further, Nakai and Nishimoto [44] examined fMRI data of participants performing mathematical problems, training ANNs on distributed representations of quantity and mathematical operations that predicted activity in the intraparietal sulcus.

3.2. Limitations

An evident limitation here is interpretability, whether encoding or decoding. The single neural response is noisy and the weight of the encoder is not straightforward for interpretation [45]. Even in the ideal case where a node could perfectly model a neuron, the model would be meaningless without understanding what the ANN is trained to perform and in which functional brain region the neuron is situated. Once positioned in an appropriate framework, it can model and potentially facilitate insight into otherwise opaque information like neural activity in primates.

Another limitation of node-level evaluation is its scalability and difficulty of training. Comparison of every node in the deep neural network with the brain neuron becomes impractical with an increasing depth of the typical ANN. However, we note a collection of studies of spiking neural networks (SNNs), a type of ANN that takes a set of spikes as input and produces a set of spikes as an output that closely mimics the behavior of biological neurons [46]. The idea is that the temporal difference between spikes and their timing may contribute to the functionality. By turning the real-valued signals into binary ones, SNNs reduce the costs of signal transmission and computational complexity. While SNNs currently lag behind other ANNs in terms of performance, they highlight for us the potential for inventive ANN approaches in this domain.

4. Layer Level

Correspondence at the layer level examines the information within a hidden ANN layer to a brain region, such as V1 or IT cortex. Compared to the node level, the correspondence of a layer to a brain region can offer a better interpretability based on the well-established hypothesis of the brain region's function, based on the accumulated history of anatomical and functional studies. However, we do not know the precise hierarchical boundaries of brain regions across the variety of information that it encodes, nor do we know whether or how these regions align with the relatively well-defined layers of an ANN. This has resulted in a variety of different approaches to evaluating the different possibilities of correspondence between different ANN layers against different brain regions.

4.1. Encoding/Decoding

Encoding and decoding approaches can be extended to the responses of a whole brain region instead of a single unit. Eikenberg [47] tested layer-level correspondence using a linear predictive model to encode each voxel's activity in the human visual system based on reduced sets of layer features (e.g., subsampling $\sim 10^6$ features down to $\sim 10^4$ from the first ANN layer). They found a lower-to-higher-layer hierarchical mapping across dorsal and ventral visual stream regions-of-interest (ROIs). Two studies examined layer-level correspondence in non-human primates, applying encoding to single-unit activity. Cadena et al. showed the capacity for a pre-trained image classification ANN to predict V1 responses to images in rhesus macaques [48]. They found ANN correspondence to V1 spiking activity across multiple early hidden layers, and better than prior Gabor filter bank models, suggesting a more complex picture of nonlinear processing in early visual regions. In a creative encoding study employing different types of ANNs, Kar et al. [49] showed evidence suggesting that the ventral visual stream may not be just one-way hierarchical

but may rely on recurrent, feedback connections. They modeled each electrode recording site in the V4 and IT cortex of rhesus macaques using layer nodes (linear combination of the first 1000 principal components), which showed that shallower recurrent ANNs could explain more variance in IT activity than "regular-deep" ANNs (8 layers), and comparably to deeper ANNs (more than 20 layers).

The visual system's hierarchical correspondence across layers has been mirrored in decoding studies [33,34]. These findings were demonstrated by Horikawa and Kamitani, decoding not just images seen but also dreamt [34] and imagined [35], where decoding accuracy, evaluated by correlation between the category feature and decoded feature, was greater in deeper ANN layers and corresponded to higher visual cortical regions implicated in object recognition (lateral occipital complex, fusiform face area, and parahippocampal place area).

The limitations of these encoding/decoding approaches are their computational cost and the need for large amounts of training data, growing with the number of layers constituted inside an ANN.

4.2. Representational Mapping

The above techniques largely employed mapping many-to-one techniques, such as encoding many nodes of a layer to single-unit recordings in monkeys, then interpreting that activity in the context of a particular brain region. However, researchers may want to evaluate the many nodes of a layer to the many measured voxels of a region. A popular technique for evaluating the correspondence between the human brain and ANN that emerged from analyzing multivariate brain activity patterns [50] is representational similarity analysis (RSA) [51–53]. RSA facilitates the comparison of two disparate information sets by converting each set to a unitless space of "distances" that characterizes its geometry (i.e., a representational dissimilarity matrix; RDM) [52] in order to test the similarity of the two datasets' underlying structures. This enables evaluating the correspondence of two datasets that differ in units and dimensions, such as computational models, EEG, MEG, fMRI, neural recording, or nodes in a layer versus voxels in a region [52], regardless of whether the data originate from across species, modalities, models, or condition [54].

For example, Cichy et al. [55] combined temporal (magnetoencephalography; MEG) and spatial (fMRI) information in response to images. They showed a hierarchical correspondence between ANN layers and human visual processing in early V1 to farther upstream areas both ventral (IT) and dorsal (intraparietal sulcus regions 1 and 2), with RDMs computed at each MEG frame yielding dynamic timecourses of ANN correspondence. Similarly, Khaligh-Razavi et al. [51] examined supervised and unsupervised ANN models to responses toward object images in human IT (measured via fMRI) and in monkey IT (measured via cell recording). Comparing RDMs, they found correspondence of human and monkey IT RDMs to those from the last, fully connected layers of supervised ANNs but not in any layers of unsupervised ANNs.

The use of RSA is not without limitations [56]. When exposed to novel stimuli, correspondence could be vanished. In RSA, scores can be a biased product of the structure of the dataset rather than the actual data [57] and may benefit from approaches such as featureand voxel-reweighting [56] and voxel-wise-encoding RSA [38] to improve fit and correspondence. There are also structural concerns at the study level, such as those raised by Xu and Vaziri-Pashkam [58], who conducted an RSA-driven visual hierarchical correspondence study using block design fMRI and functionally defined ROIs (whereas prior studies [51,55] used a potentially noisier event-related design and anatomically defined ROIs). Using this approach, the authors evaluated 14 different ANN architectures and confirmed that lower layers of several ANNs could fully capture lower-level visual representations. However, none could achieve this for higher-level neural representations of real-world objects, suggesting the current limits of ANN–brain correspondence in visual representation.

5. Network Level

Network-level correspondence examines the overall information flow inside an ANN to a comparable network in the brain, such as the hierarchical representations across a single modality, or the multimodal integrative network across the whole brain. In relation to the layer-level correspondence, a straightforward approach is to quantify the alignment between the sequence of ANN layers and sequential processing expected in the brain. For example, one can compute the correlation between the two and count the nodes of layer that are most associated with each ROI in order to test if there is a shifting of distribution from low-level to high-level cortices. Given the intrinsic feed-forward characteristics of the ANN, a sequential alignment between the brain and the ANN would indicate a hierarchical network-level correspondence.

5.1. Sequential Alignment Approach

Early examinations of network-level correspondence have been conducted for sensory networks (visual network, auditory network) due to their interpretability [59]. For the visual network, the distribution of a model-explained variance of neural activity from Yamins et al. [30] shows a clear shift from V1 to IT as the layers changed from the first to the top layer. A similar correspondence across layers was found for information extracted along the visual ventral pathway [60] as well as the dorsal pathway [61]. A recent study from Mineault et al. [62] confirmed a similar correspondence between the ANN and the visual dorsal pathway in non-human primates. Using the ANN decoding approach, Horikawa and Kamitani showed that dreaming recruits visual feature representations that correlated hierarchically across the visual system [33]. For the auditory network, Kell et al. [36] found that an ANN trained on speech and music correlated with the auditory processing hierarchy in the brain with different layers processing different aspects of sound. In another study using ANNs trained to classify music genres, Guclu et al. [63] showed a representational gradient along the superior temporal gyrus, where anterior regions were associated with shallower layers and posterior regions with deeper layers.

Evaluating large-scale networks, such as across modalities or the global brain hierarchy, poses an additional problem. For instance, the actual hierarchical correspondence between the human auditory system and visual ANNs remains unclear, as other studies have raised the suggestion of parallel organization [64]. Spatial locations like brain coordinates may provide an intuitive correspondence but not concrete evidence of the brain's structural–functional organization. For instance, not all of the many functional networks of the brain may adhere to a clear posterior-to-anterior hierarchy.

5.2. Gradient-Based Approach

Brain–ANN correspondence at the network level should also account for the sequence of chosen ROIs and the design of the ANN, such as the features that each node processes, whether they are processed sequentially or in parallel, how multiple modalities are integrated, and so on. A promising approach here is to use the principal gradient (PG) [26] as a reference. The PG is a global axis of brain organization that accounts for the highest variability in human resting-state functional connectivity (Figure 2a). Its arrangement begins with multiple satellites of unimodal sensory information that converge transmodally and integrate with the default mode network (DMN). A meta-analysis using the NeuroSynth database [65] has reinforced the relationship between cognitive function and position along the PG (Figure 2b), with sensory perception and motion exhibiting lower positions, and higher-order, abstract processes such as emotion and social recognition exhibiting higher positions [26]. The implications of PG on the hierarchical organization of functionality are further supported by clinical evidence, such as the compression of the principal motor-tosupramodal gradient in patients with schizophrenia (96 patients with schizophrenia vs. 120 healthy controls) [66] and the decrease in PG values in a neurodegenerative condition like Alzheimer's disease [67].



Figure 2. (a) Principal gradient (PG) of connectivity spans the brain, with lower scores at the start of the PG occupying sensorimotor regions and higher scores occupying higher cognitive and associative regions in a generally posterior-to-anterior gradient. (b) The distribution of brain functions along the PG begins at sensorimotor functions that converge transmodally toward higher cognition and affective processing. Figures adapted from [26].

For evaluating correspondence at the global brain level, the PG provides an independent, quantifiable metric of its hierarchy, anywhere from sensorimotor and transmodal to higher cognitive and affective information processing. Examining how subjective value emerges in the brain, ANNs individually trained to output subjective value from visual input have been shown to hierarchically correspond to the PG in the brains of those same individuals experiencing a similar value during fMRI [68], whereas Nonaka et al. [31] showed that most ANNs tend to have similar representations to the lower portion of the higher visual cortex (divided by the PG), but not the middle and higher ones, suggesting that findings of detailed correspondence in local areas could be more complex than simply an adherence to a global hierarchy.

6. Behavioral Level

The behavioral level is arguably the most important, as it serves as the entry point for ANN interest. It is the most important standard for commercial ANN use, but also critical for research since researchers ignored potential insights about the brain conferred by ANNs when they were not performing at human levels. At the same time, the behavioral level is the least scrutable. Usefully objective standards for similarity have been difficult to come by due to the open-ended nature of ANN domains and variance in their output.

6.1. Performance-Based Approach

Previous studies comparing ANN outputs directly to human performance showed that a higher categorization performance might be related to more explained variance at the high-level regions of the brain [30]. One may also consider other behaviorally relevant measures such as reaction times [69], error patterns [70], and out-of-distribution testing [71–73]. However, in more recent examinations, the most accurately categorizing ANNs do not necessarily compute the highest benchmark scores, such as brain score [74] and brain hierarchy score [31], which we will discuss in the next session.

More creative evaluations can be found in domains outside of sensory systems. For example, in language processing, Goldstein et al. [75] evaluated the three shared behavioral principles between the brain and ANN (called the deep language model): prediction

before word onset, calculation of post-word-onset surprise from pre-onset predictions, and contextual embedding of words. While their results are supported by neural evidence derived from the recordings of electrocorticography (ECoG), employing similar principles does not assure similar hierarchical correspondence. Using fMRI, Caucheteux and King [76] also demonstrated a correlation between the true brain responses and ANN-predicted responses decreasing with the language performance.

Overall, a performance-based evaluation at behavioral level is currently underdeveloped and insufficient for providing comprehensive insight into brain–ANN correspondence.

6.2. Turing Test

In terms of a qualitative evaluation of behavior, perhaps the most considered has historically been the Turing test [77]. Here, the test of similarity is whether a person in conversation finds machine output indistinguishable from human output. This can extend beyond the conversational domain, such as the embodied Turing test, which challenges AI models to interact with the sensorimotor world at skill levels akin to their living counterparts [78]. However, as Searle [79] noted, a limitation is that the Turing test is about observation, not an actual measurement of "understanding" or "consciousness", which further underscores the challenge of evaluating machine behavior for brain correspondence.

Interestingly, recent advances in large language models like ChatGPT [2] and LaMDA [3] appear to have either passed or sidestepped the Turing test. But beyond convincing contents of conversation, ChatGPT may not appear human because it replies too quickly, with too few apparent spelling or grammatical errors, or because we intuit the quality of its incoherence over time as different from our fellow humans. Machine performance has rapidly developed in other domains as well, including image generation (DALL-E by OpenAI), games closed and open (AlphaGo [4], AlphaStar [5]), and biological science (protein folding [6,7]). Once human performance within these domains were attained, questions of human similarity shifted smoothly from a basic "Turing test" to one of generality across domains or of increasing abstraction (e.g., reasoning and logic puzzles ConceptARC test [80]). For better or worse, these moving targets of behavioral similarity reiterate the complexity and open-endedness of brain–ANN correspondence at this level.

7. Composite Evaluation

While brain–ANN correspondence at each level may be evaluated independently, the ultimate goal is not to understand each level in isolation. Thus, studies have often examined more than one level of correspondence [30,33,34,37,48,51,60,61]. Recently, researchers have attempt to summarize brain–ANN correspondence in multiple levels as a single composite score.

For example, the brain score computed by Schrimpf et al. [74] is the mean of three scores: two at the node level (neural predictivity scores at V4 and IT) and one at the behavioral level. However, the brain score may not adequately describe the hierarchical correspondence between the information flow across an ANN's layers and its analogous network regions of the brain. In a composite evaluation that includes intermediate levels, the brain hierarchy score proposed by Nonaka et al. [31] computes the degree of hierarchical homology across ANN's layers and brain areas. This score utilizes both the encoding/decoding capability (node/layer level) and the Spearman rank correlation between the alignment of the ANN's layer and brain areas (network level).

Composite scores can quickly convey an intuitive correspondence, especially when computed across many ANNs. However, due to the head start of image classification ANNs, such large-scale examinations have been limited to the correspondence of the visual ventral stream of the brain, owing to the current preponderance of image processing ANNs.

8. Summary and Outlook

In this paper, we focused on how researchers have approached the problem of empirically evaluating the hierarchical correspondence between ANNs and the brain, which is at a high level separate from reviews such as those charting the history of ANN development. Surveyed studies were organized into four levels of evaluation (node, layer, network, and behavioral levels) in an attempt to present the available approaches systematically. As a new field, the methodologies and analyses that researchers have taken have been more creative and less established as standards. Thus, our aim here was that such organization could serve in part as guidelines for future studies that may be contextualized in a firmer methodological framework. This domain of brain–ANN correspondence poses many challenges and opportunities for innovative solutions moving forward.

Our survey found that the layer-level and network-level evaluations have largely been confined within well-established sensory networks (i.e., vision, audition). We anticipate that this will shift as ANNs expand in function, size, and complexity. For example, recent ANNs have encroached into the somatosensory networks [81], as well as hippocampal and entorhinal systems [82]. With a growing industrial interest and development of multimodal ANNs, network-level evaluations should rise in prominence. Or, with greater interest in ANNs that model our emotional or value systems, evaluations may include a wider range of networks such as the default mode network and limbic network. This may in turn highlight global hierarchies such as the PG for the architectural inspiration. Recent studies have also extensively identified other functional connectivity gradients within the isocortex, at the cerebellum, and at the hippocampus (for details, see [83]). As ANNs grow in variety and complexity, researchers may come to rely on broader composite evaluations where experimental results from many laboratories are combined into a single suite, including all domains of information such as vision, language, and motor control. The growth of online repositories sharing neural data, such as OpenNeuro or Neurovault, may provide infrastructural support for such endeavors.

There is room for innovation nearer the structural foundations of ANNs as well. For example, currently available ANNs assign the same learning behaviors to all nodes, largely ignoring the anatomical and functional varieties of neurons, such as network differences in the wiring between networks or the different activities between cell types. Taking a broader structural perspective, recent studies [84] have also found that not only brain connectivity but the brain geometry also contributes to its functionality. Thus, future novel approaches may incorporate properties like spatial connectivity and geometry.

Conventional approaches employed correlation to evaluate hierarchical correspondence, which may be susceptible to an incorrect assumption that brains and ANNs are similar in variance and linearity. To overcome such limitations, one may consider using canonical correlation analysis (CCA [85]) or mutual information (MI [86–89]), both of which better explain non-linear relations. CCA finds the best correlation coefficient between two transformed variables, whereas MI of two random variables measures the mutual dependence between them (i.e., the amount of information obtained about one random variable by observing the other). Along with entropy, MI is an important concept in information theory [90], more closely resembling what we observe in the brain.

It should be noted that the correlation between the ANN and the brain need not be solely feed-forward and one-on-one. A single brain region may be associated with a stack of an ANN's hidden layers and vice versa [48]. Sexton et al. [91] found in reassessing the previous studies [34,92,93] that the neural interface does not show the early-to-early and late-to-late pattern any more. The correspondence may be regulated by feedback/top-down processing. The current state-of-the-art feed-forward architecture is not robust to partial visibility, which could be caused by physiological delays [94]. The consideration of top-down processing may enhance the explainability and robustness of the ANN in future work. Another characteristic of the human brain that would be problematic in comparing the human brain and ANN is its connectivity and functional dynamics [95], which poses meaningful concerns such as individual flexibility in processing stimulus novelty and relevancy, particularly in the higher cognitive areas.

Optimistically, we are reminded that the arrow of progress can point in more than one direction. While the invention of human flight may have taken inspiration from the animal kingdom, the subsequent advancements in aerodynamics have cooperatively informed our understanding of both natural and mechanical systems in unforeseeable ways. Analogously, even as the industrial applications of ANNs appear to diverge from neuroscience, they continue to be inspired by the systems inside the brain, such as attention, working memory, episodic memory, and continual learning [96]. Thus, moving forward, brain-inspired ANNs may inform our understanding of intelligence and the brain's information processing, which may in turn reinforce the development of novel, more efficient, and comprehensive ANNs.

Author Contributions: Conceptualization, T.Q.P., T.M. and J.C.; investigation, T.Q.P. and T.M.; writing—original draft preparation, T.Q.P.; writing—review and editing, T.Q.P., T.M. and J.C.; visualization, T.Q.P.; supervision, J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the following funds: JSPS KAKENHI grant (23K17182 to T.Q.P., 21H0516513 to T.M., 21H05017 and 21H05060 to J.C.), Core Research for Evolutionary Science and Technology (CREST) grant (22717254 to T.M.), Japan Agency for Medical Research and Development (AMED) grant (JP19dm0207086 to J.C.).

Institutional Review Board Statement: This study did not require institutional review board approval.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: Special thanks to Daniel H. Lee for his edits and feedback.

Conflicts of Interest: T.M. declares no conflict of interest. T.Q.P. and J.C. are employed by Araya Inc., Japan.

References

- Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. In Proceedings of the NIPS'12: 25th International Conference on Neural Information Processing Systems, Red Hook, NY, USA, 3–6 December 2012; Volume 1, pp. 1097–1105.
- 2. OpenAI. GPT-4 Technical Report. arXiv 2023, arXiv:2303.08774.
- 3. Thoppilan, R.; Freitas, D.D.; Hall, J.; Shazeer, N.; Kulshreshtha, A.; Cheng, H.T.; Jin, A.; Bos, T.; Baker, L.; Du, Y.; et al. LaMDA: Language Models for Dialog Applications. *arXiv* 2022, arXiv:2201.08239.
- Silver, D.; Huang, A.; Maddison, C.J.; Guez, A.; Sifre, L.; van den Driessche, G.; Schrittwieser, J.; Antonoglou, I.; Panneershelvam, V.; Lanctot, M.; et al. Mastering the game of Go with deep neural networks and tree search. *Nature* 2016, 529, 484–489. [CrossRef] [PubMed]
- Vinyals, O.; Babuschkin, I.; Czarnecki, W.M.; Mathieu, M.; Dudzik, A.; Chung, J.; Choi, D.H.; Powell, R.; Ewalds, T.; Georgiev, P.; et al. Grandmaster level in StarCraft II using multi-agent reinforcement learning. *Nature* 2019, 575, 350–354. [CrossRef]
- 6. Sudha, P.; Ramyachitra, D.; Manikandan, P. Enhanced Artificial Neural Network for Protein Fold Recognition and Structural Class Prediction. *Gene Rep.* 2018, 12, 261–275. [CrossRef]
- Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* 2021, 596, 583–589. [CrossRef] [PubMed]
- 8. Kiebel, S.J.; Daunizeau, J.; Friston, K.J. A Hierarchy of Time-Scales and the Brain. PLoS Comput. Biol. 2018, 4, e1000209. [CrossRef]
- 9. Hasson, U.; Chen, J.; Honey, C.J. Hierarchical process memory: Memory as an integral component of information processing. *Trends Cogn. Sci.* **2015**, *19*, 304–313. [CrossRef] [PubMed]
- 10. Hasson, U.; Yang, E.; Vallines, I.; Heeger, D.J.; Rubin, N. A Hierarchy of Temporal Receptive Windows in Human Cortex. *J. Neurosci.* **2008**, *28*, 2539–2550. [CrossRef] [PubMed]
- 11. Murray, J.D.; Bernacchia, A.; Freedman, D.J.; Romo, R.; Wallis, J.D.; Cai, X.; Padoa-Schioppa, C.; Pasternak, T.; Seo, H.; Lee, D.; et al. A hierarchy of intrinsic timescales across primate cortex. *Nat. Neurosci.* **2014**, *17*, 1661–1663. [CrossRef] [PubMed]
- 12. Burt, J.B.; Demirtas, M.; Eckner, W.J.; Navejar, N.M.; Ji, J.L.; Martin, W.J.; Bernacchia, A.; Anticevic, A.; Murray, J.D. Hierarchy of transcriptomic specialization across human cortex captured by structural neuroimaging topography. *Nat. Neurosci.* 2018, 21, 1251–1259. [CrossRef]

- Demirtaş, M.; Burt, J.B.; Helmer, M.; Ji, J.L.; Adkinson, B.D.; Glasser, M.F.; Van Essen, D.C.; Sotiropoulos, S.N.; Anticevic, A.; Murray, J.D. Hierarchical Heterogeneity across Human Cortex Shapes Large-Scale Neural Dynamics. *Neuron* 2019, 101, 1181–1194.e13. [CrossRef] [PubMed]
- 14. Freeman, J.; Ziemba, C.M.; Heeger, D.J.; Simoncelli, E.P.; Movshon, J.A. A functional and perceptual signature of the second visual area in primates. *Nat. Neurosci.* 2013, *16*, 974–981. [CrossRef] [PubMed]
- 15. Gallant, J.L.; Connor, C.E.; Rakshit, S.; Lewis, J.W.; Van Essen, D.C. Neural responses to polar, hyperbolic, and Cartesian gratings in area V4 of the macaque monkey. *J. Neurophysiol.* **1996**, *76*, 2718–2739. [CrossRef] [PubMed]
- Gross, C.G.; Rocha-Miranda, C.E.; Bender, D.B. Visual properties of neurons in inferotemporal cortex of the Macaque. J. Neurophysiol. 1972, 35, 96–111. [CrossRef] [PubMed]
- 17. Moro, A. Dynamic Antisymmetry; MIT Press: Cambridge, MA, USA, 2000.
- Friederici, A.D. The brain basis of language processing: From structure to function. *Physiol. Rev.* 2011, 91, 1357–1392. [CrossRef]
 [PubMed]
- 19. Friederici, A.D.; Gierhan, S.M. The language network. Curr. Opin. Neurobiol. 2013, 23, 250–254. [CrossRef] [PubMed]
- Thompson-Schill, S.; Hagoort, P.; Dominey, P.F.; Honing, H.; Koelsch, S.; Ladd, D.R.; Lerdahl, F.; Levinson, S.C.; Steedman, M. Multiple Levels of Structure in Language and Music. In *Language, Music, and the Brain: A Mysterious Relationship*; MIT Press: Cambridge, MA, USA, 2013.
- 21. Patel, A.D. Language, music, syntax and the brain. Nat. Neurosci. 2003, 6, 674–681. [CrossRef]
- 22. Koelsch, S.; Rohrmeier, M.; Torrecuso, R.; Jentschke, S. Processing of hierarchical syntactic structure in music. *Proc. Natl. Acad. Sci. USA* 2013, *110*, 15443–15448. [CrossRef] [PubMed]
- Bodegård, A.; Geyer, S.; Grefkes, C.; Zilles, K.; Roland, P.E. Hierarchical Processing of Tactile Shape in the Human Brain. *Neuron* 2001, *31*, 317–328. [CrossRef]
- 24. Sathian, K. Analysis of haptic information in the cerebral cortex. J. Neurophysiol. 2016, 116, 1795–1806. [CrossRef] [PubMed]
- 25. Bola, Ł.; Matuszewski, J.; Szczepanik, M.; Droździel, D.; Sliwinska, M.W.; Paplińska, M.; Jednoróg, K.; Szwed, M.; Marchewka, A. Functional hierarchy for tactile processing in the visual cortex of sighted adults. *NeuroImage* **2019**, 202, 116084. [CrossRef]
- Margulies, D.S.; Ghosh, S.S.; Goulas, A.; Falkiewicz, M.; Huntenburg, J.M.; Langs, G.; Bezgin, G.; Eickhoff, S.B.; Castellanos, F.X.; Petrides, M.; et al. Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proc. Natl. Acad. Sci. USA* 2016, 113, 12574–12579. [CrossRef]
- 27. Huntenburg, J.M.; Bazin, P.L.; Margulies, D.S. Large-Scale Gradients in Human Cortical Organization. *Trends Cogn. Sci.* 2017, 22, 21–31. [CrossRef]
- 28. Yamins, D.L.K.; DiCarlo, J.J. Using goal-driven deep learning models to understand sensory cortex. *Nat. Neurosci.* 2016, 19, 356–365. [CrossRef] [PubMed]
- Richards, B.; Lillicrap, T.; Beaudoin, P.; Bengio, Y.; Bogacz, R.; Christensen, A.; Clopath, C.; Costa, R.P.; de Berker, A.; Ganguli, S.; et al. A deep learning framework for neuroscience. *Nat. Neurosci.* 2019, 22, 1761–1770. [CrossRef]
- Yamins, D.L.K.; Hong, H.; Cadieu, C.F.; Solomon, E.A.; Seibert, D.; DiCarlo, J.J. Perfomance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc. Natl. Acad. Sci. USA* 2014, 111, 8619–8624. [CrossRef]
- Nonaka, S.; Majima, K.; Aoki, S.C.; Kamitani, Y. Brain hierarchy score: Which deep neural networks are hierarchically brain-like? iScience 2021, 24, 103013. [CrossRef] [PubMed]
- 32. Bashivan, P.; Kar, K.; DiCarlo, J.J. Neural population control via deep image synthesis. Science 2019, 364, e453. [CrossRef]
- 33. Horikawa, T.; Tamaki, M.; Miyawaki, Y.; Kamitani, Y. Neural Decoding of Visual Imagery During Sleep. *Science* **2013**, *340*, 639–642. [CrossRef]
- 34. Horikawa, T.; Kamitani, Y. Hierarchical Neural Representation of Dreamed Objects Revealed by Brain Decoding with Deep Neural Network Features. *Front. Comput. Neurosci.* 2017, 11, e4. [CrossRef] [PubMed]
- 35. Horikawa, T.; Kamitani, Y. Generic decoding of seen and imagined objects using hierarchical visual features. *Nat. Commun.* 2017, *8*, 15037. [CrossRef]
- Kell, A.J.E.; Yamins, D.L.K.; Shook, E.N.; Norman-Haignere, S.V.; McDermott, J.H. A task-optimized neural network replicates human auditory behavior, predicts brain responses, and reveals a cortical processing hierarchy. *Neuron* 2018, 98, 630–644.e16. [CrossRef] [PubMed]
- 37. Zhuang, C.; Yan, S.; Nayebi, A.; Schrimpf, M.; Frank, M.C.; DiCarlo, J.J.; Yamins, D.L. Unsupervised neural network models of the ventral visual stream. *Proc. Natl. Acad. Sci. USA* **2023**, *118*, e2014196118. [CrossRef] [PubMed]
- Konkle, T.; Alvarez, G.A. A self-supervised domain-general learning framework for human ventral stream representation. *Nat. Commun.* 2022, 13, 491. [CrossRef]
- Lillicrap, T.P.; Santoro, A.; Marris, L.; Akerman, C.J.; Hinton, G. Backpropagation and the brain. *Nat. Rev. Neurosci.* 2020, 21, 335–346. [CrossRef] [PubMed]
- Lindsay, G.W. Convolutional Neural Networks as a Model of the Visual System: Past, Present, and Future. J. Cogn. Neurosci. 2021, 33, 2017–2031. [CrossRef]
- Marcus, D.; Harwell, J.; Olsen, T.; Hodge, M.; Glasser, M.; Prior, F.; Jenkinson, M.; Laumann, T.; Curtiss, S.; Van Essen, D. Informatics and Data Mining Tools and Strategies for the Human Connectome Project. *Front. Neuroinform.* 2011, 5, 4. [CrossRef] [PubMed]

- 42. Gordon, E.M.; Laumann, T.O.; Adeyemo, B.; Huckins, J.F.; Kelley, W.M.; Petersen, S.E. Generation and Evaluation of a Cortical Area Parcellation from Resting-State Correlations. *Cereb. Cortex* 2014, *26*, 288–303. [CrossRef]
- 43. Drakopoulos, F.; Baby, D.; Verhulst, S. A convolutional neural-network framework for modelling auditory sensory cells and synapses. *Commun. Biol.* **2021**, *4*, 827. [CrossRef]
- Nakai, T.; Nishimoto, S. Artificial neural network modelling of the neural population code underlying mathematical operations. *NeuroImage* 2023, 270, 119980. [CrossRef]
- 45. Kriegeskorte, N.; Douglas, P.K. Interpreting encoding and decoding models. *Curr. Opin. Neurobiol.* **2019**, 55, 167–179. [CrossRef] [PubMed]
- Yamazaki, K.; Vo-Ho, V.K.; Bulsara, D.; Le, N. Spiking Neural Networks and Their Applications: A Review. *Brain Sci.* 2022, 12, 863. [CrossRef] [PubMed]
- 47. Eickenberg, M.; Gramfort, A.; Varoquaux, G.; Thirion, B. Seeing it all: Convolutional network layers map the function of the human visual system. *NeuroImage* **2017**, *152*, 184–194. [CrossRef]
- Cadena, S.A.; Denfield, G.H.; Walker, E.Y.; Gatys, L.A.; Tolias, A.S.; Bethge, M.; Ecker, A.S. Deep convolutional models improve predictions of macaque V1 responses to natural images. *PLoS Comput. Biol.* 2019, 15, e1006897. [CrossRef] [PubMed]
- 49. Kar, K.; Kubilius, J.; Schmidt, K.; Issa, E.B.; DiCarlo, J.J. Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. *Nat. Neurosci.* **2019**, *22*, 974–983. [CrossRef]
- 50. Norman, K.A.; Polyn, S.M.; Detre, G.J.; Haxby, J.V. Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. *Trends Cogn. Sci.* 2006, *10*, 424–430. [CrossRef] [PubMed]
- Khaligh-Razavi, S.M.; Kriegeskorte, N. Deep Supervised, but Not Unsupervised, Models May Explain IT Cortical Representation. PLoS Comput. Biol. 2014, 10, e1003915. [CrossRef] [PubMed]
- 52. Kriegeskorte, N.; Mur, M.; Bandettini, P. Representational similarity analysis—Connecting the branches of systems neuroscience. *Front. Syst. Neursci.* **2008**, *2*, 4. [CrossRef] [PubMed]
- Kriegeskorte, N.; Kievit, R.A. Representational geometry: Intergrating cognition, computation, and the brain. *Trends Cogn. Sci.* 2013, 17, 401–412. [CrossRef] [PubMed]
- 54. Xie, S.; Hoeh, S.; Moeskops, M.; Kayhan, E.; Kliesch, C.; Turtleton, B.; Koster, M.; Cichy, R.M. Visual category representations in the infant brain. *Curr. Biol.* 2022, *32*, 5422–5432. [CrossRef]
- 55. Cichy, R.M.; Khosla, A.; Pantazis, D.; Torralba, A.; Oliva, A. Comparision of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. *Sci. Rep.* **2016**, *6*, 27755. [CrossRef]
- 56. Kaniuth, P.; Hebart, M.N. Feature-reweighted representational similarity analysis: A method for improving the fit between computational models, brains, and behavior. *Br. J. Philos. Sci.* 2022, 257, 119294. [CrossRef] [PubMed]
- 57. Dujmovic, M.; Bowers, J.S.; Adolf, F.; Gupta, A.; Tarr, M.J.; Malhotra, G. The pitfalls of measuring representational similarity using representational similarity analysis. *bioRxiv* 2022, *6*, e49. [CrossRef]
- 58. Xu, Y.; Vaziri-Pashkam, M. Limits to visual representational correspondence between convolutional neural networks and the human brain. *Nat. Comm.* **2021**, *12*, 2065. [CrossRef] [PubMed]
- 59. Kell, A.J.; McDermott, J.H. Deep neural network models of sensory systems: Windows onto the role of task constraints. *Curr. Opin. Neurobiol.* **2019**, *55*, 121–132. [CrossRef] [PubMed]
- 60. Güçlü, U.; van Gerven, M.A.J. Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream. *J. Neurosci.* **2015**, *35*, 10005–10014. [CrossRef]
- 61. Güçlü, U.; van Gerven, M.A.J. Increasingly complex representations of natural movies across the dorsal stream are shared between subjects. *NeuroImage* **2017**, *145*, 329–336. [CrossRef]
- 62. Mineault, P.; Bakhtiari, S.; Richards, B.; Pack, C. Your head is there to move you around: Goal-driven models of the primate dorsal pathway. *bioRxiv* 2021. [CrossRef]
- 63. Güçlü, U.; Thielen, J.; Hanke, M.; van Gerven, M.A.J. Brains on Beats. *arXiv* 2016, arXiv:1606.02627.
- Hamilton, L.S.; Oganian, Y.; Hall, J.; Chang, E.F. Parallel and distributed encoding of speech across human auditory cortex. *Cell* 2021, 184, 4626–4639.e13. [CrossRef] [PubMed]
- 65. Yarkoni, T.; Poldrack, R.A.; Nichols, T.E.; Van Essen, D.C.; Wager, T.D. Large-scale automated synthesis of human functional neuroimaging data. *Nat. Methods.* **2011**, *8*, 665–670. [CrossRef]
- Dong, D.; Luo, C.; Guell, X.; Wang, Y.; He, H.; Duan, M.; Eickhoff, S.B.; Yao, D. Compression of Cerebellar Functional Gradients in Schizophrenia. *Schizophr. Bull.* 2020, 46, 1282–1295. [CrossRef] [PubMed]
- 67. Hu, Q.; Li, Y.; Wu, Y.; Lin, X.; Zhao, X. Brain network hierarchy reorganization in Alzheimer's disease: A resting-state functional magnetic resonance imaging study. *Hum. Brain Mapp.* **2022**, *43*, 3498–3507. [CrossRef]
- 68. Pham, T.Q.; Yoshimoto, T.; Niwa, H.; Takahashi, H.K.; Uchiyama, R.; Matsui, T.; Anderson, A.K.; Sadato, N.; Chikazoe, J. Vision-to-value transformations in artificial neural networks and human brain. *bioRxiv* **2021**. [CrossRef]
- Spoerer, C.J.; Kietzmann, T.C.; Mehrer, J.; Charest, I.; Kriegeskorte, N. Recurrent neural networks can explain flexible trading of speed and accuracy in biological vision. *PLoS Comput. Biol.* 2020, 16, e1008215. [CrossRef]
- Mnih, V.; Kavukcuoglu, K.; Silver, D.; Graves, A.; Antonoglou, I.; Wierstra, D.; Riedmiller, M.A. Playing Atari with Deep Reinforcement Learning. arXiv 2013, arXiv:1312.5602.
- Geirhos, R.; Rubisch, P.; Michaelis, C.; Bethge, M.; Wichmann, F.A.; Brendel, W. ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. *arXiv* 2022, arXiv:1811.12231.

- 72. Geirhos, R.; Temme, C.R.M.; Rauber, J.; Schütt, H.H.; Bethge, M.; Wichmann, F.A. Generalisation in humans and deep neural networks. *arXiv* 2020, arXiv:1808.08750.
- 73. Singer, J.J.D.; Seeliger, K.; Kietzmann, T.C.; Hebart, M.N. From photos to sketches how humans and deep neural networks process objects across different levels of visual abstraction. *J. Vis.* **2022**, *22*, 4. [CrossRef]
- 74. Schrimpf, M.; Kubilius, J.; Hong, H.; Majaj, N.J.; Rajalingham, R.; Issa, E.B.; Kar, K.; Bashivan, P.; Prescott-Roy, J.; Schmidt, K.; et al. Brain-Score: Which artificial neural network for object recognition is most brain-like? *bioRxiv* 2018. [CrossRef]
- 75. Goldstein, A.; Zada, Z.; Buchnik, E.; Schain, M.; Price, A.; Aubrey, B.; Nastase, S.A.; Feder, A.; Emanuel, D.; Cohen, A.; et al. Shared computational principles for language processing in humans and deep language models. *Nat. Neurosci.* 2022, 25, 369–380. [CrossRef]
- 76. Caucheteux, C.; King, J.R. Brains and algorithms partially converge in natural language processing. *Commun. Biol.* **2022**, *5*, 134. [CrossRef] [PubMed]
- 77. Turing, A.M. Computing Machinery and Intelligence; Oxford University Press: Oxford, UK, 1950; Volume 59, pp. 433-460.
- 78. Zador, A.; Escola, S.; Richards, B.; Ölveczky, B.; Bengio, Y.; Boahen, K.; Botvinick, M.; Chklovskii, D.; Churchland, A.; Clopath, C.; et al. Toward Next-Generation Artificial Intelligence: Catalyzing the NeuroAI Revolution. *arXiv* 2022, arXiv:2210.08340.
- 79. Searle, J.R. Minds, brains, and programs. *Behav. Brain Sci.* 1980, 3, 417–424. [CrossRef]
- Moskvichev, A.; Odouard, V.V.; Mitchell, M. The ConceptARC Benchmark: Evaluating Understanding and Generalization in the ARC Domain. arXiv 2023, arXiv:2305.07141.
- Lucas, A.; Tomlinson, T.; Rohani, N.; Chowdhury, R.; Solla, S.A.; Katsaggelos, A.K.; Miller, L.E. Neural Networks for Modeling Neural Spiking in S1 Cortex. *Front. Syst. Neurosci.* 2019, 13, 13. [CrossRef]
- 82. Banino, A.; Barry, C.; Uria, B.; Blundell, C.; Lillicrap, T.; Mirowski, P.; Pritzel, A.; Chadwick, M.J.; Degris, T.; Modayil, J.; et al. Vector-based navigation using grid-like representations in artificial agents. *Nature* **2018**, *557*, 7705. [CrossRef]
- Katsumi, Y.; Zhang, J.; Chen, D.; Kamona, N.; Bunce, J.G.; Hutchinson, J.B.; Yarossi, M.; Tunik, E.; Dickerson, B.C.; Barrett, K.S.Q. Correspondence of functional connectivity gradients across human isocortex, cerebellum, and hippocampus. *Commun. Biol.* 2023, 6, 401. [CrossRef]
- Pang, J.C.; Aquino, K.M.; Oldehinkel, M.; Robinson, P.A.; Fulcher, B.D.; Breakspear, M.; Fornito, A. Geometric constraints on human brain function. *Nature* 2023, 618, 566–574. [CrossRef]
- Zhuang, X.; Yang, Z.; Cordes, D. A technical review of canonical correlation analysis for neuroscience applications. *Hum. Brain* Mapp. 2020, 41, 3807–3833. [CrossRef] [PubMed]
- Tsai, A.; Fisher, J.W.; Wible, C.G.; Wells, W. Analysis of Functional MRI Data Using Mutual Information. *Lect. Notes Comput. Sci.* 2004, 1679, 473–480. [CrossRef]
- Afshin-Pour, B.; Soltanian-Zadeh, H.; Hossein-Zadeh, G.A.; Grady, C.L.; Strother, S.C. A Mutual Information-Based Metric for Evaluation of fMRI Data-Processing Approaches. *Hum. Brain Mapp.* 2011, 32, 699–715. [CrossRef] [PubMed]
- Wang, Z.; Alahmadi, A.E.; Zhu, D.C.; Li, T. Brain functional connectivity analysis using mutual information. In Proceedings of the 2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Orlando, FL, USA, 14–16 December 2015; pp. 542–546.
- 89. Zhang, W.; Muravina, V.; Azencott, R.; Chu, Z.D.; Paldino, M.J. Mutual Information Better Quantifies Brain Network Architecture in Children with Epilepsy. *Comput. Math. Methods Med.* **2018**, 2018, 6142898. [CrossRef] [PubMed]
- 90. Cover, T.M.; Thomas, J.A. Elements of Information Theory; John Wiley & Sons: Hoboken, NJ, USA, 1991.
- 91. Sexton, N.J.; Love, B.C. Reassessing hierarchical correspondences between brain and deep networks through direct interface. *Sci. Adv.* **2022**, *8*, eabm2219. [CrossRef]
- 92. Chang, N.; Pyles, J.; Marcus, A.; Gupta, A.; Tarr, M.J.; Aminof, E.M. BOLD5000, a public fMRI dataset while viewing 5000 visual images. *Sci. Data* 2019, *6*, e49. [CrossRef] [PubMed]
- Majaj, N.J.; Majima, K.; Aoki, S.C.; Kamitani, Y. Simple Learned Weighted Sums of Inferior Temporal Neuronal Firing Rates Accurately Predict Human Core Object Recognition Performance. J. Neurosci. 2015, 35, 13402–13418. [CrossRef]
- Tang, H.; Schrimpf, M.; Lotter, W.; Moerman, C.; Paredes, A.; Caro, J.O.; Hardesty, W.; Cox, D.; Kreiman, G. Recurrent computations for visual pattern completion. *Proc. Natl. Acad. Sci. USA* 2018, 115, 8835–8840. [CrossRef]
- 95. Ju, H.; Bassett, D.S. Dynamic representations in networked neural systems. Nat. Neurosci. 2020, 23, 907–917. [CrossRef]
- 96. Hassabis, D.; Kumaran, D.; Summerfield, C.; Botvinick, M. Neuroscience-Inspired Artificial Intelligence. *Neuron* 2017, 95, 245–258. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.