The Dissertation Committee for Jamieson Warner
certifies that this is the approved version of the following dissertation:

Developmental Generative Models of Brain Connectivity

Committee:

Risto Miikkulainen, Supervisor

Laura Colgin

Robbe Goris

Christiane Linster

Samantha Santacruz

Xuexin Wei

Developmental Generative Models of Brain Connectivity

by

Jamieson Warner

Dissertation

Presented to the Faculty of the Graduate School
of the University of Texas at Austin
in Partial Fulfillment
of the Requirements
for the Degree(s) of

Doctor of Philosophy

The University of Texas at Austin
May 2025

Developmental Generative Models of Brain Connectivity

by

Jamieson Warner, PhD
The University of Texas at Austin, 2025
SUPERVISOR: Risto Miikkulainen

Recent advances in connectomics have produced large-scale datasets aligning brain connectivity with genetic expression, and these resources are poised to grow dramatically with ongoing technological and institutional investment. In response, researchers have developed computational models linking gene expression to brain wiring, often using regression techniques that identify genes implicated in neural development. While these models offer valuable biological insights, they are inherently constrained: they explain connectivity only as a function of known gene expression patterns, limiting the kinds of hypotheses that can be formulated about how connectivity arises.

This dissertation introduces a generative modeling framework that reverses the traditional direction of explanation. Instead of predicting connectivity from genetics, it constructs connectivity from first principles, using unsupervised latent variables that are fit to explain observed wiring patterns. The model reveals strong statistical associations between learned latent representations and genetic expression, especially under sparse and low-complexity assumptions, suggesting that gene expression encodes the structural motifs of connectivity.

Building on this foundation, a developmental model is introduced based on the theory of morphogenesis. This model generates latent variables through a sequential process governed by local signaling rules, capturing the constraints of biological development. It is found that the developmental model explains even more of the observed genetic expression than the static model, providing evidence that gene expression contains disambiguating signals that support pattern formation, cell identity, and positional inference.

Together, the results support the theory that brain wiring arises from a low-dimensional, structured genetic program optimized under biological constraints. These models not only advance understanding of neural development, but also open new directions for engineering applications in AI, regenerative medicine, and biologically inspired design.

Contents

List of Figures

Chapter 1

Introduction

Did you know that a tree tells a story? Look closely, and you'll notice a branch scar from the time a branch was pruned during a drought, now hardened; reaction wood grown after strong winds off the lake buffeted the trunk; and a wide growth ring from when the tree flourished in an especially sunny season. The branches take numerous bends and turns, attesting to once-flexible shoots that explored multiple directions before lignifying into strong, structural supports. From the many emerging buds, only a select few will receive the tree's growth resources, reflecting the direction it has chosen to invest in for long-term sunlight capture and structural stability—decisions now solidified in hardened wood. Because every investment in growth can be a matter of life and death, the tree's form is the product of millions of years of environmental challenges that have molded its developmental choices—its final structure encoding the facts of evolution. There is so much information about the tree's life and evolution clearly present in its form, if only one knows how to look. Where did the structure come from? With the right tools to analyze the mature form, one could see its story—the environmental factors and developmental decisions that steered its growth to produce the observed structure.

A similarly data-rich system is found in brain connectomics. During development, neurons proliferate, migrate, extend growth processes, and form intricate electrical networks. Unlike the biology of trees, which is relatively well-characterized (Zhu and Roeder, 2020), the origins of the brain's intricate directed connectivity are only beginning to be understood (Gilmour et al., 2017; Tiberi et al., 2012). Developmental biology offers system-level models that explain the chemical and genetic dynamics driving these processes (Dickson, 2002). These models illuminate how the brain might self-organize and reveal the guidance cues that determine diverse neural fates across different regions (Sánchez-Camacho and Bovolenta, 2009). Evidence suggests

that a complex interplay between genetic expression, chemical signals, and neuronal activity collectively shapes the destiny of each cell (Flavell and Greenberg, 2008).

In this dissertation, a data-driven approach is employed to infer the properties of the system responsible for the observed connectivity patterns, advancing a computational approach and understanding. The core contribution lies in identifying signs that neuronal interconnectivity is encoded within the genetic expression. Building on this foundation, a developmental model is applied—offering a novel way to interpret the data—to uncover the origins of these connections. This approach not only proposes a candidate growth hypothesis for how the entire brain might self-organize but also suggests that the developmental model can serve as a robust and efficient framework for describing complex architectures. Looking forward, these models offer a foundation for building brain-like computational systems, advancing regenerative medicine, and enabling the self-assembly of biological machines.

1.1 Motivation

"What I cannot create, I do not understand." Widely attributed to physicist Richard Feynman, this quote encapsulates the idea that true comprehension comes from the ability to construct or recreate something from first principles. As connectomic datasets expand—mapping the neuronal connectivity of larger brains at ever-higher resolutions—the ability to understand and explain the data using traditional methods, such as hand-made mathematical models, struggles to keep pace. While a handful of key genetic transcription factors and their interactions have successfully explained some broad developmental patterns, these approaches are unlikely to scale for deciphering an entire connectomic map. Consequently, this dissertation, along with prior work, leverages machine learning for the automated discovery of candidate developmental hypotheses. By generating the observed data, these generative models offer a scalable pathway to understanding whole-brain development. Therefore, generative models of data are the only scalable solution to whole-brain developmental models. Although exploring the detailed roles of specific transcription factors or complex interaction networks falls beyond the scope of this work, the high-level focus maintained here lays the groundwork for future investigations into these specific mechanisms.

This automated inference of the developmental program is essential for the large-scale application of developmental models into practical domains. In medicine, these models provide a framework for simulating both normal and aberrant developmental processes, thereby offering insights into the origins of developmental disorders and guiding the design of targeted interventions. They pave the way for predictive diag-

nostics, personalized treatment plans, and even strategies in regenerative medicine. In the field of bioengineering, automated developmental models inspire the creation of semi-organic machines that integrate synthetic components with biological tissues. Such systems, which mimic the self-organizing and self-repairing properties of living organisms, hold promise for developing adaptive, resilient technologies. Ultimately, these high-level models act as a bridge between theoretical insights and real-world applications, enabling scalable solutions that address complex challenges in both medicine and engineering.

1.2 Approach

This dissertation analyzes data from mouse brains, focusing on region-to-region connectivity and genetic expression patterns. The approach involves training models that generate the observed connectivity, which is then compared to the genetic expression data. Various methods for constructing connectivity are tested, each representing a hypothesis for the organic growth of neural networks. By comparing multiple hypotheses based on how well they match the genetic expression data, a better understanding is formed about where the connectivity comes from, how it is encoded by the genetic expressions, and how the genetic expression patterns form.

Two primary models are tested as hypotheses. The first, the static model, simplifies the system by treating each region independently, assigning a numerical "barcode" to each location to represent its unique and group-level identity. These barcodes are then combined in a generative model to approximate the observed connectivity. The static model evaluates various coding principles, such as the preferred barcode distribution and its role in encoding connectivity, thereby elucidating the relationship between genetic expression and neuronal wiring. The second, the developmental model, introduces a more biologically grounded hypothesis by generating the barcodes through a sequential process constrained by local interactions. Inspired by Alan Turing's theory of morphogenesis, it captures the spatial and temporal dynamics of development, producing structured patterns that better reflect how neural connectivity may emerge in real organisms. The end result is a computational model that represents a theory of how gene expressions form and determine connectivity.

1.3 Results

The core findings of this dissertation demonstrate that genetic expression in the mouse brain carries structured information predictive of connectivity patterns. La-

tent variables derived from actual brain connectivity data show strong correlations with genetic expression, significantly exceeding those obtained from randomized baselines. These results indicate that gene expression does not merely reflect general spatial organization but encodes meaningful, region-specific wiring features. Furthermore, testing across multiple prior distributions reveals that gene predictivity remains stable regardless of regularization strategy, with no significant evidence that sparsity-inducing priors enhance the interpretability or specificity of gene-to-latent correspondences.

Extending this framework, a developmental model was introduced to simulate the process by which latent connectivity codes might arise through sequential, locally constrained interactions. The developmental model achieved markedly higher gene expression predictivity compared to the static model, indicating that gene expression encodes not only the final wiring outcomes but also disambiguating signals necessary for developmental pattern formation. This effect was strongest when the complexity of the model was appropriately constrained, highlighting the importance of a simple, locally informed process—limited in both perceptual scope and representational capacity—in accurately capturing the inference demands faced by developing cells. Finally, the developmental encoding was shown to provide greater efficiency at low complexity and the capacity for self-repair through local updates, highlighting its potential utility for scalable, resilient architecture design.

1.4 Guide to the Thesis

This dissertation applies generative modeling techniques to explore how genetic expression encodes brain connectivity in the mouse. Two main modeling approaches are employed: a static model, adapted from prior latent-variable approaches in connectomics, and a developmental model, inspired by morphogenetic processes and implemented using neural cellular automata.

Chapter 2 provides the conceptual foundation, synthesizing insights from connectomics, developmental biology, and computational neuroscience. Chapter 3 outlines the dataset preparation, the modeling architectures, and the evaluation metrics used to assess the biological plausibility of each hypothesis. Chapter 4 introduces the static model, which demonstrates a clear correspondence between connectivity and gene expression. Chapter 5 presents the developmental model, showing how incorporating sequential, spatial constraints yields a more biologically grounded representation. Chapter 6 explores the computational properties of the developmental framework, including its efficiency and self-repair capabilities. Chapter 7 addresses the limitations of the current models, discusses emerging opportunities in large-scale connectomics,

and outlines the broader implications of developmental modeling across artificial intelligence, regenerative medicine, and biological engineering. Finally, Chapter 8 reflects on how generative models can illuminate developmental processes and guide future scientific and technological efforts.

Chapter 2

Background

Prior studies have identified statistical associations between gene expression and brain connectivity, suggesting that genetic patterns play a role in shaping neural circuits. Building on this foundation, this dissertation introduces introduces a statistical and generative modeling framework that aims not only to replicate these associations but to explain them — revealing how gene expressions encode wiring and how they emerge from the developmental process. The approach integrates insights from three domains: statistical connectomic modeling, developmental biology, and computational neuroscience.

Section 2.1 reviews prior models linking genetic expression to connectivity, highlighting the need for more expressive, latent variable approaches to test competing hypotheses about how connectivity is encoded. This motivates the use of generative models that can reconstruct wiring while revealing underlying coding principles. Section 2.2 introduces a developmental perspective, reviewing classical theories of morphogenesis and modern computational models—especially neural cellular automata (NCA)—that formalize development as a self-organizing, spatially constrained process. This provides the conceptual foundation for the developmental model introduced later in the dissertation.

The approach also draws from computational neuroscience, particularly the use of latent variable models to explain connectivity as arising from spatially distributed "barcodes." Section 2.3 reviews how representational learning is used in neuroscience to infer internal variables from observed behavior that explain brain activity. This provides a conceptual foundation for the present work, which similarly models development as an inference process—recovering how genetic expression might solve the task of wiring the brain. The connection is made explicit by appealing to the Free Energy Principle of developmental biology, which argues that development itself is an

inference task. This perspective motivates the statistical toolkit used throughout the dissertation and and clarifies how methods from computational neuroscience—such as brain-predictive metrics—can be extended to developmental biology.

2.1 Computational Models Linking Genetic Expression and Connectivity

Recent advances in high-throughput imaging have made it possible to collect large datasets describing both the connectivity and genetic expression of animal brains. A critical feature of some of these datasets is that both data types are mapped to a common spatial coordinate reference frame. This alignment makes direct spatial comparison possible, allowing statistical models to assess how patterns of gene expression relate to patterns of brain connectivity. It is thus possible to computationally derive candidate hypotheses for how genetic factors may influence neuronal wiring. These computational models have found that across several species much of the connectivity can be predicted from the genetic expressions distributed in the brain, supporting the idea that chemical guidance cues serve as an organizing principle for axonal targeting. The computational models can further identify specific genes and how they may influence wiring in the model.

This dissertation takes computational models a step further. While previous work found genes that correlate with wiring patterns, this dissertation learns an entire spatial profile of latent variables that explain the connectivity, allowing the comparison of different coding hypotheses. Therefore, this work offers a richer explanation of how wiring is encoded in genetic expressions and where the genetic expression patterns are formed. The section will begin with historical context linking wiring to chemical guidance cues and then review prior work.

2.1.1 Foundation: Chemical Guidance of Axon Targeting

The chemoaffinity hypothesis suggests that developing neurons rely on chemical signals to form precise neural connections with their intended targets. This hypothesis originated from early experiments conducted by Sperry (1963), who investigated the chemical basis underlying neural circuit formation through studies involving amphibian visual systems. In these experiments, Sperry surgically rotated the eyes of frogs and observed that regenerating optic nerve fibers reconnected to their original, chemically defined target regions rather than adjusting to the altered spatial orientation. Consequently, frogs exhibited persistent visual disorientation, demonstrating that ax-

ons follow chemical cues rather than adapting solely through functional experience. The chemoaffinity hypothesis was an affront to the predominant functionalist view at the time, which considered behavior to be too complex to be explained by connective patterns, and thus there was no need for neurons to have selective synaptic targets.

The chemoaffinity hypothesis proved to be controversial as time went on, and evidence was compiled for the role of experience-based plasticity in determining wiring rules. However, after decades of discussion, the chemoaffinity hypothesis was refined into several widely accepted postulates (Meyer, 1998): (1) axons carry distinct molecular markers; (2) target cells and pathways express corresponding complementary markers; (3) these markers are produced as a consequence of cellular differentiation; and (4) axonal growth is actively guided by these chemical markers to establish specific neuronal connections. Collectively, these observations have formed the basis for understanding how chemical signaling directs accurate axonal targeting during neural development.

Building on the understanding that chemical cues guide axons toward synaptic targets, researchers have sought to piece together the molecular and genetic mechanisms underlying wiring behavior. Axon-guidance studies have identified families of attractive and repulsive cues that regulate the behavior of the developing axonal growth cone (Dickson, 2002). In several cases, large-scale wiring patterns have been successfully explained by the spatial expression of morphogens—signalling proteins that are expressed in subsets of cells and trigger concentration-dependent cellular responses, thereby coordinating spatially patterned cellular differentiation (Sánchez-Camacho and Bovolenta, 2009). Together, these findings support the view that signalling molecules play a central role in organizing neural connectivity.

The studies on chemical guidance of neural targeting and the associated genetic factors lay the groundwork for the computational models discussed in this section. These computational models extend the analysis of connectivity and chemical guidance patterns to large-scale data, thus accelerating the discovery of gene transcription factors relevant to synaptic targeting.

2.1.2 Neuron-level Computational Models

Computational models linking genetic expression and wiring were first attained for *C. elegans*, for which a wiring diagram has been constructed as far back as 1986 (White et al., 1986), and decades of genetic data has been collected in the *WormBase* archive (Sternberg et al., 2024). Hence, early evidence came from *C. elegans*: the landmark study by Kaufman et al. (2006) found that a neuron's gene expression "signature" carries significant information about its synaptic connections.

At the single-cell level, recent models use single-cell transcriptomics coupled with synaptic wiring data to decode connectivity rules. Barabási and Barabási (2020) proposed a connectome model (CM) to explain how genes encode neural wiring. They hypothesized that each neuron's gene expression profile guides synapse formation, predicting the existence of specific biclique motifs (complete bipartite subgraphs) in neural networks. Indeed, they identified large, statistically significant bicliques in the connectomes of three species, within which neurons shared common gene expression patterns and morphological features. This statistical observation supports the idea that genetic compatibility rules (i.e. specific gene-gene interactions) drive circuit formation. The CM provides a generative framework linking an organism's genome to the architecture of its connectome. In a detailed case study, the model was applied to identify the genetic rules of wiring in *C. elegans* (Kovács et al., 2020). The method incorporated biological realities like data noise and spatial contact constraints (neurons must be in physical contact to connect). Applying this to *C. elegans* uncovered a network of 19 gene-gene interactions that likely govern the formation of electrical synapses. Notably, five of these predicted interactions were supported by prior experimental evidence, lending credibility to the model's predictions.

Following Barabási's work, Qiao (2024) introduced a bilinear modeling approach to link neuronal gene expression profiles to a *neuronal type connectome*. The method takes two sets of inputs: gene expression data for each neuron (or cell type) on the presynaptic side and on the postsynaptic side. It then learns a matrix factorization that predicts the strength of the connection between any pre-post pair as a function of their gene profiles. The method is likened to a recommendation system: genes of the source and target neurons interact, like "user" and "item" features, to predict a "rating", i.e. the connectivity likelihood. Qiao tested this model on the *C. elegans* connectome (neuron-level) and on a mouse retina connectome at the level of cell types. In *C. elegans*, the model slightly outperformed a previous state-of-the-art (the "spatial connectome model", which incorporated neuron locations) in reconstructing the network of electrical synapses. Notably, it rediscovered all the gene interactions that the earlier model had identified and found additional candidate interactions, demonstrating improved sensitivity. When applied to the mouse retina, it produced known wiring motifs—for example, connectivity between bipolar cell types and retinal ganglion cells. Notably, the model output was biologically interpretable: it pointed out specfic genes associated with particular motifs, many of which have plausible roles in synapse formation.

As the *C. elegans* connectome is highly consistent across different specimens, it has proved to be a good testing ground for models linking genetic expression to wiring. One study to highlight is Barabási's biclique model, which, using connectivity alone,

was able to predict group-membership of individual neurons that correlated with neuron-level genetic expressions. This study stands out among other studies which explicitly predict the connectivity from the genetics. This dissertation builds upon such connectivity-first analysis. However, while the wiring rules such as Barabási's "connectome model" and Qiao's "recommender" have focused *C. elegans*, this dissertation uses high-dimensional mouse brain data to test wiring principles. Applying generative models to a more complex dataset, the approach tests whether the same principles hold in mammalian neural networks. The following section will introduce other models that studied mammalian wiring and genetic expressions and will contrast them with this work.

2.1.3 Structure- and Region-level Computational Models

In mammalian systems, researchers lacked neuron-level data, so they turned to region-level analysis. Two early studies relied on the Allen Brain Atlas (ABA) for genetic data while turning to other sources available for connectivity. Using the ABA in conjunction with the Brain Architecture Management System (BAMS) for connectivity, French and Pavlidis (2011) demonstrated that regions sharing gene expression profiles tend to share connectivity patterns. They identified genes specifically implicated in neural development and axon guidance as strong predictors of connectivity, utilizing Mantel correlation (correlation of correlations) to analyze the various brain regions. Concurrently, Wolf et al. (2011) took a predictive approach, using a linear SVM classifier to predict the connectivity from the genetic data across 146 regions using cross-validation. They relied on the ABA mice data but used a rat connectome, for which data was available, and which mapped onto the mouse brain at the structural level. They were able to attain a high predictive power of the connectivity, with mean Area Under the ROC Curve (AUC) between 0.73 and 0.74. These studies were the first to integrate whole-brain rodent transcriptomic and connectomic data.

Wiring prediction studies were extended to the Allen Mouse Brain Connectivity Atlas (MBCA) upon its final release in 2014. Ji et al. (2014) performed an integrative study of the mouse brain by combining the Allen Mouse Brain gene expression atlas with the ABA. They built machine learning models to predict the presence of a projection between any two brain regions from their gene expression patterns. The key finding was that gene expression is strongly predictive of mesoscale connectivity in the mouse brain. In fact, using the expression patterns of about 4,000 genes across the brain, their model could predict region-to-region connections with about 93% accuracy (after binarizing connectivity strength). Remarkably, they achieved about 91% accuracy using only a small subset of 25 top predictive genes. These genes were

implicated in roles in neural development and synaptic function, suggesting a meaningful biological signal. Ji et al. (2014) demonstrated that at a coarse resolution (i.e. brain areas as nodes), the molecular signature of a region encodes its wiring. Further methodological advancements by Roberti et al. (2019) incorporated neural network-based models (multi-layer perceptrons), achieving improved prediction performance compared to previous linear methods, achieving an AUC of 0.943.

Building on structure-level studies, researchers have pushed to finer resolutions where local variations in gene expression and connectivity can be captured. In a 2015 follow-up, Fakhry and Ji (2015) extended their analysis to the voxel level in the mouse brain. They divided the brain into small voxels (grid elements 200um in size) and predicted voxel-to-voxel connectivity based on gene expression in those voxels. Even at this high resolution (tens of thousands of voxels), gene expression retained strong predictive power, again about 93% accurate in classifying connected vs. unconnected voxel pairs. Using only about 100 genes still gave >80% accuracy. Notably, the most predictive genes at voxel-scale were related to synaptic transmission, aligning with the theory that neurons' gene expression profiles determine their projection patterns. This fine-scale model implies that connectivity is imprinted even in local gene expression variations.

Subsequent studies confirmed and extended these findings, finding links between different species models. For example, Fornito et al. (2019) identified transcriptomic signatures distinguishing hub regions in mouse and human networks. Therefore, conserved transcription factors can play similar organizational roles between multiple species. Studies in humans, though lacking detailed data available in simpler model animals, also reveal a link between genetics and connectivity as evidenced by heritability (twin studies), association studies (gene-trait correlations), and spatial transcriptomics (genetic expression and connectivity imaging) (Arnatkeviciute et al., 2021).

Many previous studies of the rodent brain predicted connectivity directly from gene expression data using supervised learning, mapping observed gene expression patterns to region-to-region connectivity (genetics $\rightarrow$ connectivity). In contrast, this dissertation employs latent variable modeling, which provides greater flexibility. Rather than learning a direct mapping, the approach begins by extracting latent spatial variables from the observed connectivity using a generative model. These latent factors are then used to predict and explain gene expression profiles (connectivity $\rightarrow$ latent variables $\rightarrow$ genetics).

The latent variable approach is more complex than direct supervised methods, but it also provides a richer and more informative framework for understanding brain wiring. Supervised models constrain the hypothesis space to functions of the

observed genetic expressions, limiting the types of explanations they can offer for connectivity. In this dissertation, the latent variables instead discover spatial distributions—termed "barcodes"—that explain the connectivity structure. Learning these underlying variables is a higher-dimensional and more expressive task than simply learning readout weights from measured genes.

The latent variable approach makes it possible to address a different scientific goal. Whereas earlier work primarily focused on identifying individual genes involved in neural wiring, this dissertation investigates how genetic expression might encode connectivity. By generating different hypothetical codes through generative models of connectivity and comparing them against measured gene expression, it is possible to evaluate the plausibility of various coding hypotheses. If a hypothetical code aligns well with observed gene expression, it provides a stronger explanation of the biological process. This generative framework thus enables the comparison of more complex models of genetic coding, offering deeper insights into both how connectivity is encoded and how gene expression patterns arise.

This reformulation of the modeling approach—from a direct supervised mapping to a latent variable framework—enables greater flexibility in the types of generative models that can be applied to explain brain connectivity. This dissertation introduces, in particular, a developmental generative model inspired by morphogenesis. Unlike earlier models that imposed only static spatial constraints, such as a preference for short connections, the developmental approach generates latent variables through a sequential process in which each region updates its state based on local interactions with its neighbors. These local updates reflect developmental constraints, such as diffusible signaling and spatially restricted interactions, that shape biological tissues. By incorporating these dynamics, the developmental model relaxes several assumptions of previous work and offers a more biologically grounded explanation of how gene expression patterns and connectivity co-emerge during development.

2.2 Computational Models of Pattern Formation

How do cells assemble into coherent tissues, organs, and bodies? Developmental biology seeks to answer this question by studying how organisms grow and how their cells and tissues differentiate over time, guided by chemical signals exchanged between cells. These signaling gradients create spatial variation in cellular environments, driving the emergence of structure and identity across the developing organism. This section reviews classical models of morphogenesis—the foundational theory describing how such spatial patterns arise—as well as recent work that uses machine learning to rediscover or extend these models. This background provides

the theoretical basis for the developmental modeling framework used in this dissertation. Building on these ideas, this dissertation applies a machine learning system inspired by morphogenesis (the developmental model) to mouse brain connectivity data, offering a new hypothesis about the developmental processes that shape neural circuitry.

2.2.1 Classical Models of Morphogenesis

The field of morphogenesis models the system of development as a spatially distributed process with local signalling and interaction, giving rise to complex patterns and structures. This view treats development as a kind of computation performed by cells communicating locally. Classic work by Turing (1952) introduced the idea of reaction-diffusion systems to explain how an organism's genes could generate ordered patterns (like stripes or spots) without encoding each detail explicitly. Turing showed that two diffusible substances (morphogens) reacting with each other could spontaneously form stable periodic patterns, providing a morphological blueprint for where cells should differentiate. This work was foundational for theoretical developmental biology: it demonstrated how simple chemical rules, encoded in the genome, might produce complex spatial organization in tissues. Building on this, Wolpert (1969) French Flag model introduced the concept of positional information, in which cells read their position in a morphogen gradient and differentiate accordingly. This framework illustrated how embryonic cells interpret genetic and molecular cues to reliably create spatial patterns, supporting the idea that local gene expression gradients can yield global structure.

Over the decades, such morphogen-based models of development have been refined and expanded. They form the basis of many computational models of pattern formation in biology. A famous example from 1988 is a reaction-diffusion model of skin patterning, which explains how patterns like stripes and spots emerge on animal coats (e.g. zebras, leopards) due to the interaction of morphogens like activators and inhibitors (Murray, 1988). A similar model explained the skin patterning of angelfish (Kondo and Asal, 1995). The reaction-diffusion model provides the mathematical framework for how local activators and inhibitors can form stable spatial patterns. The activators increase the concentration of the inhibitor in their vicinity, leading to feedback loops that generate spots or stripes in the skin. These reaction-diffusion models are classic examples of abstract patterning processes in biological systems.

Models involving locally interacting morphogens continue to explain a variety of anatomical processes. In the developing limbs of vertebrates, digit formation is influenced by several signalling molecules, such as the morphogen Sonic hedgehog

(Shh) (Riddle et al., 1993). A model detailed how a gradient of Shh in the zone of polarizing activity (ZPA) at the posterior margin of the developing limb bud controls the number and identity of digits in a gradient-dependent fashion, so that cells closer to the ZPA expressed higher levels of Shh and resulted in more posterior fates (like the little finger), while those at the front (anterior) of the limb bud, where Shh concentration was lower, formed more anterior digits (like the thumb). This Shh-based patterning demonstrates how morphogen gradients coordinate complex spatial patterns in limb development. This model supports the concept that differential gradients of morphogen signalling can give rise to patterned variation downstream.

These examples illustrate the versatility of morphogen-based models in explaining various forms resulting from development. They provide a computational framework for understanding how simple molecular interactions can lead to the precise creation of biological structures. This computational perspective emphasizes that development is not simply a process of following genetic instructions but is actively shaped by feedback and local interactions among the molecular signals present in the tissue.

The model of locally interacting species of molecules described in morphogenesis is used as a foundation for the *developmental model* of this dissertation. The developmental model learns rules defining how local interactions occur to explain the brain connectivity data, thereby finding a Turing mechanism that is the best fit for the data. While the rules for dynamics in the model are not put into one-to-one comparison with chemical interactions—such a correspondence is not supported by the architecture—modeling the data using such a process model is demonstrated to improve the predictivity of genetic expression data. Therefore, morphogenesis helps explain the observed pattern of genetic expression.

2.2.2 Neural Cellular Automata (NCA)

Modern computational approaches have revisited these ideas with new tools — namely, deep learning. Through the use of learnable pattern generation schemes, the parameters of an arbitrary reaction-diffusion-type system can be learned in such as way to produce a *target pattern*, which is a known two-dimensional image or three-dimensional structure, which the model will approximate. The hallmark design is the Neural Cellular Automata (NCA) (Mordvintsev et al., 2020), a differential cellular automata that learns to "grow" a target image by iteratively updating cells based on local rules. The use of the NCA in this dissertation, along with more technical details, are outlined in Section 3.2.4. This model, inspired by Turing pattern reaction-diffusion, highlights the power of simple local signalling rules to create complex patterns and creates an opportunity for automated developmental model

learning. The NCA has been extended in several directions with an aim to extend it to more dimensions, to impose additional biological constraints, and to test the ability for small, simple rules to generate complex patterns.

One straightforward extension of the NCA model was to produce three-dimensional shapes (Zhang et al., 2021). This implementation uses a voxel lattice rather than a pixel lattice to represent the data, and the underlying transformations (convolution and recurrence) remain the same, albeit in three dimensions. The voxel grid is larger to store in memory, so the size of appropriate patterns must be appropriately downscaled.

Neural CA models have also been extended to enforce isotropy and rotational equivariance in their local update rules. Recent work draws inspiration from classic reaction-diffusion systems to ensure that pattern formation is independent of any global orientation cue. For example, Mordvintsev et al. (2022) found that by removing the orientation-specific Sobel filters, leaving only an isotropic Laplacian filter in the perception kernel — thus forming the *isotropic NCA* (IsoNCA) — the system could still generate complex asymmetric target patterns by breaking symmetry through the initial conditions or the loss function. An alternative approach to rotational invariance is found in Steerable NCA models, which give each cell an internal degree of freedom representing its orientation. In this formulation, cells can locally adjust their orientation ("turn") as part of the dynamics, achieving a form of rotation equivariance without an external reference frame (Randazzo et al., 2023). The Steerable NCA can be trained to reliably grow oriented patterns. These developments show that NCAs can grow grow the directional context that they need to produce patterns, as in biology; chemical reactions are not typically orientation-dependent.

Parallel efforts have focused on making NCAs more compact. Mordvintsev and Niklasson (2021) eliminates some redundant perceptual filters to reduce the NCA model to a very small size, resulting in the μ-NCA (micro-NCA). In this approach, the entire NCA update rule is parameterized by only a few hundred weights – the smallest variant uses just 68 parameters – yet it can generate intricate, varied textures comparable to those produced by much larger models.

The NCA model is the core of the developmental model, to be introduced in Section 3.2.4. It serves to explain the developmental origin of morphogen patterns that guide the neural wiring, as contrasted with the static model, which explains wiring in terms of gene expressions but does not account for where the gene expressions come from. While morphogenesis provides the conceptual foundation for the NCA developmental model, the next section reviews the broader latent variable framework that underpins this work and presents a biological rationale for the relevance of representational learning to development.

2.3 Representations in the Brain and in Development

The previous section introduced computational models of development that build structured patterns from local rules and interactions. This section now turns to the question of how such developmental processes can be understood in informational terms. In neuroscience, neural activations are commonly interpreted as encoding information about behaviorally relevant variables—those that help the brain make predictions or guide actions. Similarly, developmental systems require internal variables that carry information needed to coordinate tissue growth, resolve ambiguity, and enforce structural constraints. This analogy supports treating developmental signals as task-relevant representations, or "developmental codes," enabling the use of tools from computational neuroscience to model biological development in this dissertation.

Subsection 2.3.1 develops this idea by treating development as an inference process, in which morphogens and gene expression patterns carry task-relevant information. In this framework, spatially distributed signals can be interpreted as internal representations that help organize the system. The analogy to neural coding becomes especially powerful when combined with tools from machine learning. As introduced in Section 2.3.2, deep learning models are often used to extract structured representations from complex data. This provides a computational lens for understanding how internal signals—whether in brains or in development—can emerge to solve problems.

2.3.1 Development as Inference

A thermodynamic, or statistical, view of development considers the inferential or decision-making process that a developing system performs to achieve complex form. In this view, life fights entropy by actively maintaining order. A classic perspective comes from Schrödinger (1944), who argued that organisms avoid decay into equilibrium by feeding on "negative entropy," exporting disorder to keep their own structure intact. This thermodynamic insight that living systems resist the universe's tendency toward disorder laid the groundwork for later theoretical biology. Modern theoretical biologists have formalized this idea. Karl Friston's Free Energy Principle (Parr et al., 2022) generalizes this notion by treating living systems as probabilistic inference machines that minimize their internal "surprise" over time. Under this principle, development can be seen not as an entropy-minimizing process, but rather as a free energy minimizing process: an embryo grows (or a brain self-organizes), it

actively drives its states toward expected configurations, thus resisting chaos in the face of environmental uncertainty.

This thermodynamic view leads to a perspective of development as Bayesian inference or prediction. Friston et al. (2015) proposed that morphogenesis (the formation of body structure, including neural wiring) is essentially an inferential process akin to the brain's perception. They cast embryonic self-assembly in terms of variational free-energy minimization, suggesting that cell communities collectively "compute" the body plan by minimizing surprise or error from an intended anatomical model. Cells are hypothesized to share a common generative model (encoded by genetic and epigenetic information) of the final organism's form, and development is the process of inverting that model, each cell using local signals to infer how it should contribute to the overall anatomy. This view interprets the genetic code not just as a static model, but as a probabilistic program. Michael Levin and co-workers extended this model explicitly: they argued that each cell behaves as an information-processing agent making decisions to fulfill a built-in model of the target morphology. In their Bayesian formulation, the driving force behind morphogenesis is the maximization of a cell's model evidence — achieved by expressing particular receptors and signalling molecules that align with the cell's predicted role in the organism. Thus, development can be seen as cells performing Bayesian inference, adjusting gene expression until their local state matches the expected pattern that the collective is trying to build.

This dissertation leverages the Bayesian development framework in the latent variable model. The system collectively maximizes the probability of the connectivity data, which is assumed to be the target blueprint structure toward which the model is aiming. However, corresponding to the "complexity" term in the variational free energy equations, the models must balance the reconstruction of the connectivity with constraints on their ability to represent data. The constrained computing approach proved fruitful, as the models that served the best fit to the biological genetic expression data — thus the best models of biological self-organization — were those that were similarly constrained in their complexity. These findings are discussed for the static latent model in Section 4.4 and for the developmental model in Section 5.2.

2.3.2 Machine Learning for Canonical Solutions

At the intersection of neuroscience and artificial intelligence, a framework termed goal-driven deep learning provides a method for understanding neural representations using deep learning models. In this approach, researchers train deep neural networks to perform tasks analogous to those faced by biological systems (e.g. object

recognition, navigation, decision-making) and then compare the internal representations or dynamics of these models to neural data. The underlying hypothesis is that if a model's architecture and learning objectives are appropriate, it may discover canonical solutions – patterns of representation or dynamics – that mirror those used by real neural circuits. This framework was articulated in work by Yamins and DiCarlo (2016), who showed that a deep convolutional network optimized for visual object recognition could predict neural responses in the primate ventral visual pathway, effectively mapping model layers to brain areas like V4 and IT cortex. Together, these studies established goal-driven modeling as a powerful technique: by training neural networks to solve the same goals as the brain, one can obtain insights into the brain's own solution.

Many studies that followed applied this framework to sensory systems, showing that models that both perform sensory-based decision-making tasks tend to align with human neural data elicited by the same tasks. For example, vision models optimized for object or face recognition have been found to develop brain-like feature hierarchies, offering quantitative explanations for neural tuning in visual cortical areas (Cadieu et al., 2014; Yamins and DiCarlo, 2016). In the auditory domain, Kell et al. (2018) trained a deep network to recognize speech and music and showed that its internal activations could predict electrophysiological responses throughout the human auditory pathway. The researchers also searched through various solution architectures, showing that the best-fit auditory model had a hierarchical organization that corresponded to primary and secondary auditory cortices. Such results suggest that deep networks can serve as in silico proxies for sensory processing, where model neurons and layers map onto real neural populations in anatomical order.

Goal-driven models have also shed light on spatial navigation and memory representations in the brain. A well-studied case is the emergence of grid cells – neurons that activate in a periodic lattice-like pattern during navigation. Cueva and Wei (2018) found that when a recurrent network was trained to perform path integration (updating position from velocity inputs), units spontaneously developed grid-like firing patterns akin to those observed in the entorhinal cortex of mammals. Along with grid cells, the network produced other cell types (e.g. border cells) seen in animal brains, suggesting that these may be a natural solution for encoding space given the task demands. Complementary results came from deep reinforcement learning: Banino et al. (2018) trained an agent to navigate in a virtual environment and reported that its latent representations became grid-like, supporting efficient path planning. These findings indicate that when artificial agents are trained to solve spatial navigation tasks, they often converge on the same grid cell representations employed by biological navigation systems.

Beyond primary sensory or spatial domains, deep learning models have been used to probe higher-level cognitive computations. Mante et al. (2013) demonstrated that recurrent neural networks (RNNs) trained to solve cognitive tasks can exhibit low-dimensional dynamical motifs matching those recorded in prefrontal cortex during flexible decision-making. Similarly, Yang et al. (2019) trained a single RNN to master a battery of abstract working memory and decision tasks, and found that the network self-organized into specialized subspaces capturing each task's rules – mirroring how the primate prefrontal cortex can flexibly reconfigure for different contexts. This connection points to how task-trained models can reveal principles of cognitive flexibility and multi-task representation in neural circuits. Recent work by Miller et al. (2023) extends this idea to meta-learning: they introduced "disentangled RNNs" that learn interpretable, low-dimensional latent dynamics while fitting behavioral data. These RNNs recovered known cognitive strategies (e.g. evidence accumulation or Q-learning-like rules) from raw choice sequences, offering candidate explanations for the algorithms the brain might be running (Miller et al., 2023). By compressing complex behavior into human-interpretable components, such models act as hypothesis generators for the neural mechanisms underlying cognition.

Finally, the goal-driven modeling framework has begun to incorporate large language models (LLMs) to tackle language and semantic representations in the brain. For instance, Caucheteux et al. (2022) showed that the activations of a transformer language model (GPT-2) could be linearly mapped to fMRI signals recorded as subjects listened to stories – and crucially, the strength of this mapping predicted how well each subject comprehended the narrative. In parallel, Schrimpf et al. (2021) reported that neural network models with higher next-word prediction accuracy also better explain neural and behavioral responses during reading, suggesting a task-driven convergence between artificial and biological language processing. In a review, Mineault (2024) discussed the opportunity of foundation models for neuroscience – large pretrained models (of vision, language, etc.) that can be adapted to brain data – as a way to derive general-purpose computational explanations across brain regions. This approach leverages the power of large models that have learned rich data patterns, but it comes with challenges in interpretability and ethical use.

These studies illustrate how deep learning models, when trained on ecologically relevant tasks, often converge on representational solutions similar to those in the brain—providing a tool for formulating and testing hypotheses about neural computation. Drawing on the analogy established in the previous section, this dissertation extends the framework of goal-driven deep learning by applying representational learning to a developmental task. Many of the tools used in this dissertation originate in the goal-driven deep learning literature. For example, the "gene scores"

discussed in Section 3.3.2 derive from the "brain scores" of Caucheteux et al. (2022). This perspective frames development as an inference problem and justifies the use of machine learning tools to analyze the structure and origin of brain connectivity. The following chapter introduces the data and modeling framework used to investigate this idea.

Chapter 3

Approach

In order to understand the relationship between genetic expression and connectivity, this dissertation uses latent variable models to produce hypothetical explanations of the connectivity data. By comparing how well these various models map onto genetic data, different hypotheses (with different assumptions) can be tested as explanations for the genetic expression coding of connectivity. These methods thus explain how genetic expression patterns give rise to connectivity at an informational or coding level. This section details the methodology, starting from sourcing the biological data, processing it with latent variable models, and quantitatively evaluating results.

3.1 Biological data

The Allen Mouse Brain dataset was selected because it provides extensive genetic and connectivity data suitable for latent variable modeling. Gene expression data from the ABA and connectivity data from the MBCA are registered to the Allen Mouse Common Coordinate Framework (CCF). This alignment enables direct comparisons between genetic expression patterns and connectivity.

The complexity of the mouse brain provides an ideal case study for understanding genetic contributions to brain connectivity, potentially generalizable to complex neural systems.

3.1.1 Gene expression

The gene expression dataset comprises genome-wide *in-situ* hybridization (ISH) data for over 20,000 genes in adult mice, available from the ABA. Processed data was

accessed through the Allen API at 200µm voxel resolution. Three types of processed data are available:

1. **Expression density**: proportion of pixels expressing the gene.
2. **Expression intensity**: average intensity of expressing pixels.
3. **Expression energy** combines both density and intensity, providing a measure of total expression strength.

For this study, the expression energy format was selected due to its comprehensive representation of gene expression. To further reduce computational complexity, the genetic data was voxelized again into a grid at 20-voxel resolution using cubic anti-aliasing to avoid artifacts.

3.1.2 Connectivity

Connectivity data was obtained from viral tracer experiments, consisting of injection scans (images immediately after viral tracer injection) and projection scans (capturing axonal projections). These scans were used to construct a voxel-based connectivity matrix at the same 20-voxel resolution as the genetic data. Following the linear regression approach of Oh et al. (2014), injection scans were used as input predictors, and projection scans were used as targets to generate the adjacency matrix. An L2 regularization parameter ($\alpha = 1.0$) was chosen to prevent extreme values, balancing reconstruction accuracy with model stability.

The resulting distribution of connection weights (Figure 3.1) spanned several orders of magnitude, causing larger weights to dominate learning. Thus, the weights were log-scaled and thresholded (values below 10^{-5} were removed) to ensure balanced contribution across connections during learning. A positive translation was also applied, so that an absence of an edge is represented by the value 0.0, and the remaining edge strengths take on positive values. While binarization of the connectivity strength values was implemented in previous studies, preliminary experiments using simulated data found that including the variation in edge strength values resulted in a better fit of the static latent model to the underlying data causes.

3.1.3 Brain Volume and Structures

The structural annotation atlas provided by the Allen Institute, aligned to the Common Coordinate Framework (CCF), was used to define the brain interior at the 20-voxel resolution and to assign anatomical labels to voxels. Voxels were included

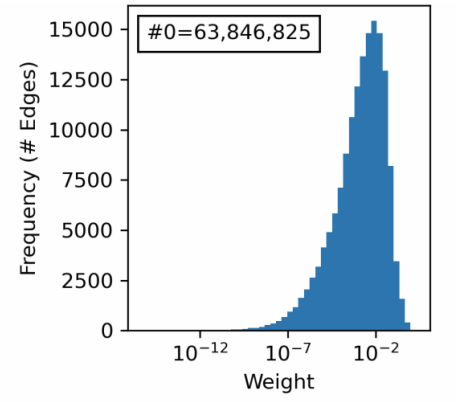

Figure 3.1: **Brain Connection Weights.** The distribution of connection weight values that is the raw output of the linear regression algorithm. These values span multiple orders of magnitude, which biases the resulting error gradient during learning so that the largest values take precedence. Post-processing to take the log of these values is therefore applied, allowing for all edges to contribute to the loss function on a similar scale.

in the brain interior if at least 50% of their constituent pixels carried a valid structural annotation. These annotations were also used to group voxels into anatomical regions for a region-based cross-validation procedure, described in Section 3.3.2, where models were trained on some regions and tested on others. Additionally, the annotation atlas was used to verify that the connectivity and gene expression datasets were aligned to the same voxel grid and that they exhibited expected region-to-region trends, such as strong reciprocal connectivity within cortical areas and consistent gene expression within structurally defined regions.

3.1.4 Post-Processing and Validation

Multiple validation steps were performed to ensure correct data processing. Three-dimensional visualization validated the alignment of genetic and connectivity data, and statistical checks confirmed accurate voxel-grid alignment. While the left hemisphere was well represented across experiments, the right hemisphere contained many missing or incomplete entries. As a result, model training included data from both hemispheres to maximize sample diversity, but gene scores and genetic correlations (see Section 3.3) were computed exclusively on the more complete left-hemisphere data to avoid introducing bias or artifacts due to uneven spatial coverage.

3.2 Architecture Design

Two models—the static model and the developmental model—are introduced in Chapters 4 and 5 to examine how genetic factors influence connectivity patterns. Both models produce voxel-based barcodes, or vectors of latent variables, that are used to predict connectivity. In both cases, representation learning is used to infer these latent variables from the observed connectivity data.

3.2.1 Representation Learning

Representation learning extracts latent variables $\mathbf{z}$ from observed data $\mathbf{x}$ using probabilistic modeling Dayan and Abbott (2005). The latents are considered to be a simpler explanation of the data, or an explanation that is more workable for downstream tasks.

For example, the input data may represent raw visual input, such as the input stream to the visual cortex. The corresponding cause variable may represent e.g. whether there is a dangerous predator animal in the input. The forward probability distribution, conditioned on the latent variables, is referred to as the **generative distribution** $P(\mathbf{x} \mid \mathbf{z})$, indicating the likelihood of the input data $\mathbf{x}$ given the belief $\mathbf{z}$. The reverse distribution, $P(\mathbf{z} \mid \mathbf{x})$, is referred to as the **recognition distribution**, used to ascertain the latent beliefs given the inputs. There is in addition a **prior**, $P(\mathbf{z})$, over the various latent variables. The theory of representation learning is applied in this work to the task of linking connectivity and genetic expression. The goal is to infer an explanation of the connectivity.

Variational learning introduces an approximate inference strategy for cases in which computing the true recognition distribution $P(\mathbf{z} \mid \mathbf{x})$ is computationally intractable. To address this intractability, variational methods posit a simpler, parameterized approximation distribution, denoted as $Q(\mathbf{z} \mid \mathbf{x})$, aiming to closely match the true recognition distribution $P(\mathbf{z} \mid \mathbf{x})$. The goal is to minimize the difference between these distributions, which is captured by the "free energy" objective:

$$F = \mathbb{E}_{Q(\mathbf{z}|\mathbf{x})} \left[\log Q(\mathbf{z} \mid \mathbf{x}) - \log P(\mathbf{x}, \mathbf{z}) \right]. \tag{3.1}$$

Minimizing this free energy function encourages the variational approximation $Q(\mathbf{z} \mid \mathbf{x})$ to closely approximate the true recognition distribution, leading to latent representations that effectively capture the structure underlying the observed data.

This section outlined the mathematical framework behind latent variables and their role in modeling structured data. In this dissertation, latent variables are used

to identify the hidden factors that give rise to observed patterns of brain connectivity and to investigate how these factors correspond to genetic expression. The following chapters apply this framework to two modeling approaches—first, a static model that directly maps latent variables to connectivity, and then a developmental model in which latent variables emerge through a sequential, spatially constrained growth process. Before introducing these models in detail, it is useful to begin with a classical baseline—singular value decomposition (SVD)—which approximates connectivity through a low-rank factorization and serves as the simplest instance of the static modeling approach.

3.2.2 Singular Vector Decomposition (SVD)

Singular value decomposition (SVD) is a classical method for data compression and dimensionality reduction, widely used to extract dominant patterns in structured datasets. In this work, SVD serves both as a baseline for evaluating the developmental model and as a limiting case of the static model when no additional structure is imposed. The static model extends this baseline by introducing learnable priors, but SVD remains a useful starting point for comparison.

Given a connectivity matrix $W \in R^{n \times m}$, representing connection strengths, SVD factorizes it into three matrices:

$$W = U\Sigma V^{\top}. \tag{3.2}$$

Here, U and V contain orthonormal singular vectors, and Σ is a diagonal matrix with singular values, quantifying the relative significance of each singular vector in reconstructing the original data. In neural connectivity, singular vectors and values reveal latent structural patterns that correspond to meaningful biological or functional groupings of neurons or brain regions.

SVD has been used to form node-level latent variables that explain connectivity data. For example, the "eigenmodel" eigen-decomposition of Hoff (2007), closely related to SVD, represents connectivity as inner products of latent vertex features. Each node i is associated with a latent vector u_i, and the strength of the connection between nodes i and j is modeled by a weighted inner-product $u_i^{\top} \Lambda u_j$, a low-rank approximation closely resembling SVD. Hoff demonstrated that this latent factor model can capture both stochastic equivalence (patterns between groups) and homophily (patterns within groups). Thus, SVD identifies latent dimensions that explain the underlying structure of connectivity matrices.

Moreover, classical results indicate that retaining the top k singular vectors (low-rank approximation) through SVD not only produces interpretable latent variables

but also yields the optimal low-rank reconstruction by minimizing Euclidean distance to the original matrix (Eckart and Young, 1936).

In summary, SVD functions both as a latent variable estimator and as an optimal low-rank reconstruction method for the connectivity matrix. In this dissertation, it is applied to brain connectivity as a starting point for the static model. The next section describes how the static model builds upon this baseline by introducing prior distributions that further constrain and interpret the latent structure.

3.2.3 Static Model

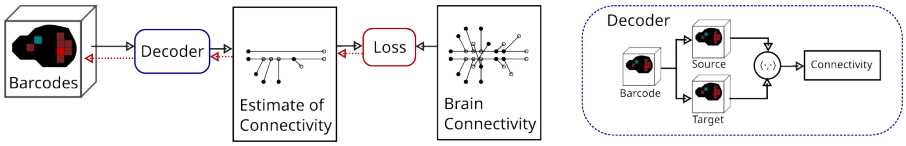

Figure 3.2: **Architecture of the Static Model.** *(Left)* The static model maps voxel-based latent variables ("barcodes") through a decoder to produce an estimated voxel-to-voxel connectivity matrix. This estimate is then compared with empirical connectivity data, and the barcode parameters are optimized to minimize reconstruction error. *(Right)* The decoder architecture is shown. Each barcode is split into two disjoint sub-vectors, which are used to predict outgoing and incoming connections, respectively. Connectivity between any two voxels is estimated by taking the dot product between their corresponding sub-vectors. This figure introduces the general framework of mapping latent variables to connectivity patterns using a fixed decoder function—a setup that will be expanded in later chapters by incorporating developmental constraints.

While SVD provides an efficient, low-error decomposition of the connectivity matrix into k latent variables, it does not account for biological constraints. In real neural systems, coding variables often reflect trade-offs between accurately representing signals and minimizing energetic or structural costs. To address this, the **static latent model** extends the SVD framework by incorporating such costs directly into the representation, as illustrated in Figure 3.2.3. The associational rule applied here—based on dot products between latent variables—is functionally similar to previous work modeling bicliques through associational wiring rules (Barabási and Barabási, 2020), or bilinear recommender models (Qiao, 2024), introduced in Section 2.1. However, this approach differs in that the latent variables are derived

directly from the connectivity data, without regressing from gene expression, allowing for the explicit introduction of biologically motivated priors. These priors reflect assumptions about how connectivity information is encoded in biological systems and form the core of this model's contribution beyond earlier methods.

To formalize this trade-off between reconstruction accuracy and biological costs, the static model draws on representational learning theory. This task constitutes an unsupervised learning problem, in which latent variables represent a variational distribution over the connectivity data. Under the representational learning framework, the true distribution of the connectivity data is approximated by the variational posterior derived from latent variables. The presence of a prior distribution over the latent variables introduces an additional complexity constraint, resulting in a final "free energy" optimization objective:

$$F = \underbrace{-\mathbb{E}_{Q(z)}[\log P(x \mid z)]}_{\text{Accuracy}} + \underbrace{KL[Q(z) \| \| P(z)]}_{\text{Complexity}}.$$

Minimizing this free energy objective encourages the latent variables to capture the features of the data necessary to reconstruct the connectivity matrix accurately (the accuracy term), while simultaneously constraining the latent representations from becoming overly complex by staying close to the prior distribution (complexity term). Therefore, minimizing free energy yields latent representations that balance accurate connectivity encoding and adherence to biologically plausible cost constraints. In machine learning, priors introduced via Bayesian frameworks are implemented as *regularizers*. Thus, free energy minimization can be viewed as a form of *regularized optimization*. In this work, the static latent model is fitted by gradient descent, minimizing a loss function comprising reconstruction error plus regularization cost derived from the prior.

The latent distribution $Q(\mathbf{z})$ can express uncertainty regarding the latent variables. The distribution is paramaterized by sufficient statistics $Q(\mathbf{z} \mid \phi)$. By minimizing F, the entire distribution over latent values is taken into account. However, this dissertation uses a point estimate $Q(\mathbf{z}) = \delta(\mathbf{z} - \mathbf{z}_0)$, where the variational distribution is a Dirac delta with mean $\mathbf{z}_0$. While this distribution does not model uncertainty (full uncertainty modeling is left to future work), it simplifies the model. The complexity term reduces to simply the surprisal of the prior: $-\log P(\mathbf{z}_0)$. The accuracy term reduces to a single likelihood evaluation, $-\log P(\mathbf{x} \mid \mathbf{z}_0)$. With these additions, the objective function is a sum of a negative log likelihood reconstruction term and a prior term:

$$F = -\log P(\mathbf{x} \mid \mathbf{z}_0) - \log P(\mathbf{z}_0),$$

which reduces the optimization to standard Bayesian optimization of a maximum a posteriori estimate.

The generative model $P(x \mid \mathbf{z})$ links the latents to the data. It will be called the *decoder*. A bilinear decoder is used as follows. The vector $\mathbf{z}_i$, the latents at voxel i, is first subdivided into two components, $\mathbf{z}_i = [\mathbf{z}_i^{src}; \mathbf{z}_i^{tgt}]$. Then the estimate of the connectivity strength from voxel j to voxel i is given as

$$\hat{M}_{i,j} = \mathbf{z}_i^{tgt} \cdot \mathbf{z}_j^{src},$$

thus producing the connectivity as a product of the source and target states. The generative distribution is then modeled as a Gaussian distribution with mean $\hat{M}$, so that the negative log likelihood reduces to the mean squared error between the estimated connectivity matrix $\hat{M}$ and the true connectivity matrix M.

Static latent variables are randomly initialized and iteratively refined through training. The experiments in this dissertation employed 2,000 training iterations with a learning rate of 0.001, optimized using the NAdam algorithm (Dozat, 2016).

The genetic data is not used in the loss function; only the reconstruction of connectivity, along with a prior over the latent variables, guides the fitting process. While it would be possible to define an objective that jointly optimizes for both connectivity and genetic alignment, such an approach would conflict with the scientific aims of this work. The goal here is to infer latent variables *from first principles*, drawing on ideas about optimal representations to test theoretical hypotheses about the relationship between genetics and connectivity. Incorporating genetic information into the loss function would compromise the independence of this data, effectively using validation information during training.

The static model produces latent variables explaining connectivity while adhering to biologically relevant cost constraints. Chapter 4 presents results of investigations comparing different cost scenarios as explanations linking genetic expression to connectivity. The same probabilistic framework will continue to be used in the developmental model, which extends the static model by constructing the latent variables using a developmental process over time.

3.2.4 Developmental Model

While the static model enables comparison between different coding hypotheses, it is limited by its assumption that voxels are independent. Its learned parameters consist only of the barcode values assigned to each voxel, without modeling how these values might emerge through developmental processes. In contrast, biological gene expression patterns are not independent across space; they are shaped by complex,

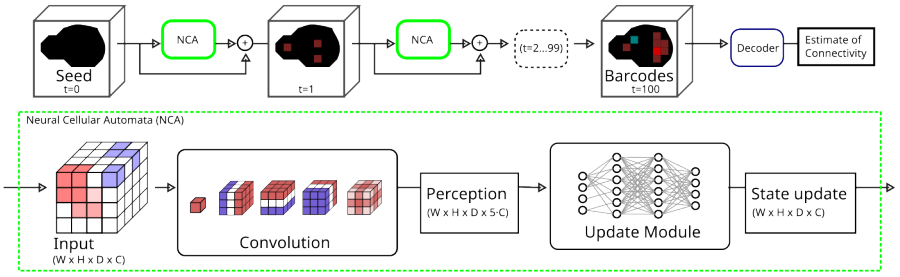

Figure 3.3: **Architecture of the Developmental Model.** *(Top)* The developmental model, implemented using an NCA block, process a sequence voxel-based states. It is initialized to a starting "seed" pattern at time $t = 0$, which is repeatedly passed through the NCA to accumulate updates over multiple time steps. The final state, called the "barcodes," is then passed through the decoder (described in the previous section) to produce a voxel-to-voxel estimate of brain connectivity. *(Bottom)* Internally, the NCA operates on a three-dimensional grid of vector-valued voxel states. At each time step, convolutional filters (identity, directional Sobel, and Laplace) are applied to extract a "perception vector" that encodes local spatial information. These vectors are processed by an update module—a small neural network—that computes the state updates of each voxel, the output of the NCA block. Thus the developmental architecture models connectivity as the outcome of a spatially local, iterative growth process, contrasting with the static model by incorporating temporal dynamics and neighborhood-based computation.

nonlinear interactions among signaling molecules during development (Kondo and Miura, 2010). To better reflect this, the developmental model generates voxel-level barcodes through a sequential process in which voxels interact with their neighbors over time. The developmental model used here is based on neural cellular automata (NCA), a framework previously applied to simple pattern generation tasks (Mordvintsev et al., 2020). This framework models how morphogen gradients and local signaling dynamics guide the emergence of spatial patterns in real organisms. Like the static model, the developmental model produces latent variables that are decoded into voxel-to-voxel connectivity estimates. Although it passes through many intermediate states, only the final state is used as the output—defining the learned barcodes. By embedding inductive biases that mirror biological development, this model offers a more realistic hypothesis for how genetic patterns give rise to wiring and may lead to improved alignment with observed gene expression.

The developmental model draws from computational models of morphogenesis,

as outlined in Section 2.2. Each voxel is treated as a *cell*—not a biological cell, but a region-based processing unit that expresses a vector-valued *state* evolving over time through interactions with neighboring voxels. These voxel-to-voxel interactions are modeled using a three-dimensional *cellular automaton*, a framework well suited for simulating spatially distributed systems. In this implementation, the cellular automaton's update rule is parameterized as a learnable function, enabling gradient-based optimization. As the system evolves forward through a sequence of states, prediction errors are backpropagated through time to update the parameters of the local update rule. The model is trained to minimize the error between the predicted and observed connectivity, effectively learning a developmental program that produces barcode representations aligned with the data.

As a machine learning architecture, the model belongs to the class of neural cellular automata (NCA; see Figure 3.3), introduced in Section 2.2.2. An NCA is a differentiable extension of classical cellular automata, in which each spatial unit—or *cell*—updates its internal state based on its local neighborhood, using learnable rules. Whereas traditional cellular automata operate with fixed, hand-coded update rules, NCA use neural networks to learn these update functions directly from data. The architecture combines spatially local convolutional operations, as used in convolutional neural networks (CNNs), with recurrent dynamics that allow it to evolve a sequence of states over time, similar to recurrent neural networks (RNNs). This allows the system to model development as a sequential process with local perception, where each cell updates based only on information from its immediate surroundings.

The perception rules in the developmental model are intentionally constrained, in contrast to the more expressive convolutional kernels used in standard convolutional neural networks (CNNs). A typical $3 \times 3 \times 3$ convolutional kernel assigns independent weights to each position in the neighborhood, allowing the network to learn highly specific, orientation-dependent features. In contrast, biological morphogen signals are interpreted in a way that is largely invariant to their relative spatial orientation—cells respond to concentrations rather than directional patterns. To reflect this biological constraint, the model limits perception to a fixed set of symmetric filters, which reduce orientation sensitivity while still capturing local spatial structure. While not fully isotropic, this design introduces a biologically inspired constraint that encourages generalization across spatial configurations.

That said, cells do rely on spatial gradients during self-organization. To model this, the NCA applies a fixed set of perception filters — specifically, identity, Sobel, and Laplace filters — to each channel of the state. The Sobel filters provide directional information in the x-, y-, and z-axes, while the Laplace filter captures second-order curvature. Along with the state itself, these produce a *perception vector*

with $5 \cdot D$ dimensions, where D is the number of state channels. In contrast, a full $3 \times 3 \times 3$ convolution would yield $27 \cdot D$ dimensions, with unique weights for every position in the neighborhood. Thus, the limited perceptual kernel reduces both the parameter count and the total information available to each voxel, enforcing a stronger inductive bias toward local, isotropic processing.

The perception vectors thus convey neighborhood information to each voxel. These vectors are fed into the update rule, which computes the next state vector for each voxel. The update rule is implemented as a multi-layer perceptron (MLP), consisting of two hidden layers with ReLU activations. An important hyperparameter is the hidden layer width, which determines the number of units in each hidden layer. Using more units adds capacity to the module, increasing the number of perceptual conditions and responses it contains — effectively increasing its overall complexity.

The output of the update module is added to the input state to produce the next state in the sequence. In this way, the state vectors accumulate gradually over a simulation period of 100 time steps. This additive approach, now standard in machine learning, provides a direct path for error gradients to backpropagate through time, facilitating more effective training. However, propagating gradients across many time steps—as in deep recurrent networks—can lead to numerical instability due to exploding or vanishing gradients Graves (2014). To mitigate this, gradient clipping is applied, with gradients constrained to a maximum norm of 1.0. This stabilizes training by preventing large updates that could disrupt learning dynamics.

Let $\mathbf{h}_{i,t} \in \mathbb{R}^D$ denote the hidden state vector (also referred to as the developmental state or barcode) of voxel i at time step t, where D is the number of channels in the state. The update process unfolds over discrete time steps and is governed by a function f_θ — the update module — parameterized as a multi-layer perceptron (MLP), which maps local perception to a state update. Each voxel updates its state using only information from its spatial neighborhood, processed through a fixed set of convolutional filters.

The perception vector $\mathbf{p}_{i,t}$ encodes spatial information from the local neighborhood of voxel i at time t by applying a fixed bank of filters to each channel of the state grid. Each filter produces one response per channel, and the results are concatenated into a single vector. Specifically, the identity filter, three directional Sobel filters (S_x, S_y, S_z), and a Laplace filter (L) are applied. The resulting vector has dimensionality $5D$, where D is the number of state channels.

Let $\mathbf{h}_{\cdot,t} \in \mathbb{R}^{X \times Y \times Z \times D}$ denote the full grid of states at time t. The perception vector at voxel i is given by:

$$\mathbf{p}_{i,t} = \left[\mathbf{h}_{i,t},\ (S_x * \mathbf{h}_{\cdot,t})_i,\ (S_y * \mathbf{h}_{\cdot,t})_i,\ (S_z * \mathbf{h}_{\cdot,t})_i,\ (L * \mathbf{h}_{\cdot,t})_i\right], \tag{3.3}$$

where $*$ denotes 3D convolution applied channel-wise, and the subscript i extracts the vector at voxel i after filtering. This concatenated vector captures local gradients and curvature and serves as the input to the NCA update rule.

The update rule is then:

$$\mathbf{h}_{i,t+1} = \mathbf{h}_{i,t} + f_\theta(\mathbf{p}_{i,t}), \tag{3.4}$$

This formulation enables local, differentiable updates that accumulate information over time, allowing the system to grow spatially structured latent representations through repeated application of the update rule.

The process is initialized with a fixed starting *seed* state. In prior work (Mordvintsev et al., 2020), the seed consisted of a single activated voxel at the center of the domain. This kind of seed is necessary to break translational symmetry—since perception is purely local and invariant to spatial translation, some spatial signal must be present to orient the system. However, this setup introduced a "center-out" bias: voxels near the origin received more update steps during growth, leading to asymmetries in the resulting pattern. To eliminate this artifact, the present work adopts a full-brain seed. Here, all voxels in the brain interior receive a nonzero initial state. The first channel is set to one for voxels within the brain and zero elsewhere. The next three channels are initialized to the (x, y, z) coordinates of each voxel, forming a smooth morphogenetic gradient that spans the brain volume. This initialization provides spatial context to each voxel and serves as the starting substrate for the developmental process.

Through local processing over many time steps, the developmental model accumulates state updates to produce the final voxel-based barcodes. These barcodes are decoded into an estimate of the brain's connectivity and are also evaluated for how well they match measured genetic expression patterns. This comparison to gene expression serves as a validation metric, assessing whether the developmental process produces spatial structures that resemble the biological cues thought to guide neural wiring. While the static model captures high-level correlations between connectivity and latent structure, the developmental model operates at a lower level, explicitly modeling how those latent patterns might arise through interactive signaling over time. Together, they offer complementary perspectives on the relationship between genetic cues and brain wiring. To evaluate the biological relevance of the patterns learned by each model, the next section outlines the procedure for comparing latent variables—whether learned directly or through development—with measured gene expression data. This comparison provides a framework for assessing how well each model captures structure consistent with the known genetic scaffolding of the brain.

3.3 Latent Variable Evaluation

While the models are trained to match the observed brain connectivity using mean squared error (MSE), their latent variables are also compared to genetic expression data. This allows different models to be evaluated based on how well their internal representations match biology. By doing so, the models can be compared as possible explanations for how genetic expression shapes brain connectivity.

3.3.1 Correlations

The first and most straightforward comparison between voxel-based latent variables and the voxel-based gene expression is one-to-one correlations sampled across space. These correlations indicate spatial overlap between any given latent variable and any gene expression profile. Let $g_{i,k}$ represent the expression level of gene k measured at voxel i, and let $h_{i,l}$ represent the value of latent variable l at voxel i. Then, the correlation coefficient $R_{k,l}$ is calculated, along with the covariance $C_{k,l}$, as:

$$C_{k,l} = \frac{\sum_i (g_{i,k} - \bar{g}_k)(h_{i,l} - \bar{h}_l)}{N - 1} \tag{3.5}$$

$$R_{k,l} = \frac{C_{k,l}}{\sqrt{C_{k,k} C_{l,l}}} \tag{3.6}$$

In these equations, N represents the total number of voxels in the brain interior, and $\bar{g}_k$ and $\bar{h}_l$ are the mean values of gene k and latent variable l across all voxels, respectively.

The correlation is therefore a measure of the association between any given model latent and gene expression, taken over the brain interior, accounting for the respective scale of each variable. While this captures pairwise associations, the next section extends the analysis to account for the full set of latent variables jointly.

3.3.2 Gene Scores

While correlations compare model latents and gene expressions one-to-one, there are many genes measured in the dataset and many latent variables. To evaluate whether a model provides a meaningful account of brain development, it is useful to quantify how much of the genetic expression landscape can be predicted from the learned representation as a whole. This measure is called the gene score. It is computed in an analogous way to the "brain score" used to compare fMRI brain responses

to language model latents in Caucheteux et al. (2022), building upon the analogy established in Section 2.3.2. It is implemented as follows.

First, the voxels are divided into 13 mutually exclusive sets based on their top-level structural annotation. Then, training proceeds in a leave-one-out cross-validation manner, in which a general linear model is trained to predict the entire set of genetic expression within the 12 training regions given the model latent variables. That is, a linear combination of the model latents is used to reconstruct an approximation to the genetic expression within each voxel. The final, left-out region is then used to test the model. Iterating over the 13 different regions in this way, the predicted values of gene expression are constructed for every voxel in the brain interior. Then, the correlation is evaluated between the predicted genetic expression and the actual genetic expression, and this correlation value is referred to as the gene score for gene k.

Thus, the gene score represents the ability of the model latents to predict (up to a linear transformation) each genetic expression factor in the dataset. Models with higher gene scores are taken to offer more biologically grounded representations of neural development and are favored as hypotheses for how connectivity may emerge from gene expression. Frequently reported will be the *average gene score,* taken to be the mean score across all genes predicted. While this section discussed the genetic associations used to evaluate the coding hypotheses as biological theories, the next section reviews the measures of connectivity prediction, which are used to fit the latent variables during training.

3.4 Graph Match Evaluation

To evaluate how well each model captures the structure of brain connectivity, two metrics are used. Mean-squared error (MSE) serves as the training objective, measuring the distance between predicted and observed connectivity. Variance explained (R^2) offers a more interpretable score, reflecting how much of the observed variation is captured by the model. These metrics support comparisons between generative models and help assess their explanatory value.

The models were fitted using the MSE loss function, a standard objective for regression problems. In this context, it corresponds to the negative log-likelihood of the observed connectivity values under a Gaussian noise model, where the mean is given by the model's prediction. This formulation is appropriate for the dataset used in this work, which represents voxel-to-voxel connectivity as continuous-valued edge strengths. The MSE loss is defined as:

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2, \tag{3.7}$$

where y_i is the observed value, $\hat{y}_i$ is the model prediction, and N is the total number of edge entries.

While the MSE is useful for training, the values are not very interpretable. When assessing how well the model performs, the R^2 metric is more informative. This measure captures not only the accuracy of predicted edge strengths for existing connections, but also how well the model distinguishes between the presence and absence of edges. The R^2 metric is computed as:

$$R^2 = 1 - \frac{SS_{\text{residual}}}{SS_{\text{total}}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}, \tag{3.8}$$

where y_i are the true values, $\hat{y}_i$ are the predicted values, and $\bar{y}$ is the mean of the true values. This expression quantifies the proportion of variance in the data explained by the model. A value of $R^2 = 1$ indicates perfect prediction, while $R^2 = 0$ indicates that the model performs no better than predicting the mean.

To quantify the relative contribution of these two factors, an analysis of variance was conducted by partitioning the total variation (SS_{total}) across all voxel-to-voxel edges. The edge set was divided into two groups: those with nonzero edge weights and those with zero-valued edges. The between-group variation (SS_{between}) captures the variance due to the presence versus absence of a connection, while the within-group variation (SS_{within}) accounts for differences in edge strength among existing connections. This analysis revealed that 86% of the total variation is explained by the distinction between connected and unconnected voxel pairs, and only 14% is attributable to variation in edge strength within the connected group. Thus, models that correctly predict where edges occur can account for the majority of observed variance in the data.

These evaluation methods establish the framework for how models in this dissertation are trained and assessed. In the next chapter, these tools are applied to brain data to uncover how genetic expression patterns relate to the structure of neural connectivity.

Chapter 4

Latent Representations of Brain Connectivity

This chapter provides computational evidence that gene expression plays a crucial role in coordinating the wiring of the mouse brain. The primary objective is to establish a link between two distinct sources of biological data: gene expression measurements across voxelized space and a voxel-to-voxel connectivity matrix (data derived in Section 3.1). A latent representation model is used to bridge these data formats by learning a set of voxel-based latent variables that explain the observed connectivity (defined in Section 3.2.3). If these latent variables are associated with the genetic expression in the mouse brain, as evaluated by the average gene score (defined in Section 3.3.2), then the model explains how the wiring patterns can be computed from information in the gene expression.

In Section 4.1, it is shown that gene expression is tied to the brain's wiring by demonstrating how latent variables trained on mouse connectivity offer an explanation of genetic expression. In Section 4.2, several alternative models of the connectivity data are evaluated, indicating that the relationship between gene expression and wiring remains robust regardless of the chosen latent prior distribution. Section 4.3 explores whether sparsity-promoting priors, which encourage more "untangled" representations, affect this relationship; no significant difference was found. Lastly, Section 4.4 unveils a key finding: only at low levels of complexity is the model able to explain the link between genetic expression and connectivity patterns.

4.1 Linking Connectivity and Gene Expression

A central goal is to demonstrate that the spatial variables learned by the model are specifically associated with the observed wiring, rather than simply reflecting general spatial information in the brain volume. To make this distinction clear, the model trained on real connectivity is compared with models trained on randomized connectivity. If the learned variables correlate with observed gene expression only when reconstructing actual brain connectivity—and not when applied to randomized data—this indicates that the correlation stems from genuine connectivity features rather than a general coordinate system.

Section 4.1.1 introduces the randomized baselines. The results are then presented in Section 4.1.2, followed by a discussion in Section 4.1.3.

4.1.1 Experimental Setup

Several methods for randomization are evaluated as baselines to selectively retain certain information about the brain connectivity and provide a thorough comparison. The first method, referred to here as "FullRandom," constructs a random connectivity matrix by shuffling the rows and columns of the original connectivity matrix. This approach maintains the total number and weights of the connections but does not preserve per-vertex degree or the average edge distance.

The second randomization method preserves per-vertex degree, thereby retaining information about the density of connections entering and exiting each region. If the model captures only connection density rather than regional identity signatures, it will not distinguish between the baseline data and the degree-preserved randomization. Degree-preserving randomization is a standard approach in network analysis, commonly implemented by swapping or re-wiring edges to maintain degree information (Rao et al., 1996). This implementation is termed "DegreePreserve."

This experiment tests the hypothesis that biological data yields higher mean gene scores than randomized data. For the brain data, multiple models are trained using gradient descent with randomized initial conditions (see Section 3.2.3). For the randomized trials, data is shuffled—while preserving key statistics where noted—and models are fit using SVD. A Welch's t-test is then used to assess whether high gene scores from the true data are statistically distinguishable from those produced by randomized data.

4.1.2 Results

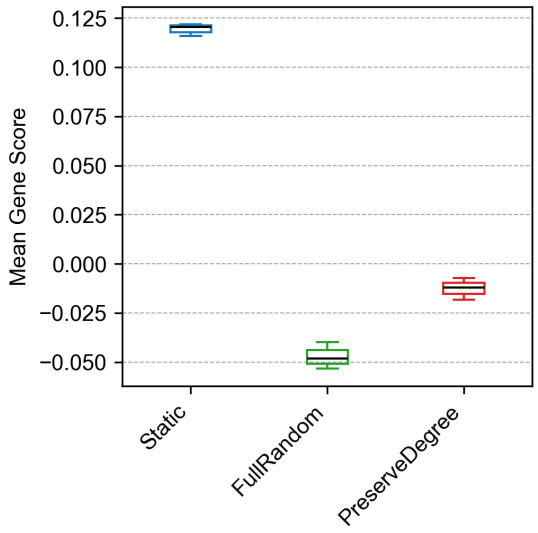

Figure 4.1: **Genetic Predictivity of the Latent Variables.** Gene scores are shown for a static model trained to predict either the true mouse brain connectivity or a set of randomized connectivity datasets. The distribution of gene scores for the true brain connectivity, compared to those for altered connectivity, indicates that structured wiring patterns drive the model's predictions of genetic expression.

Multiple randomized datasets were generated for both randomization methods —"FullRandom" and DegreePreserve" — and gene scores were computed to assess the relationship between learned latent variables and genetic expression. As shown in Figure 4.1, the average gene score served as the primary metric. A Welch's t-test confirmed that gene scores from randomized data differed significantly from those derived from the true data (FullRandom: $p = 3.9 \cdot 10^{-4}$; DegreePreserve: $p = 3.3 \cdot 10^{-4}$), indicating that the true associations are unlikely to be reproduced by random structure alone.

4.1.3 Discussion

This section establishes a fundamental link between gene expression and brain wiring, as revealed by the static latent variable model. When reproducing the original connectivity data, the learned latent variables show a strong association with true gene expression. However, altering the connectivity data reduces this association, indicating that information embedded in the wiring itself drives the observed relationship.

Alternate randomization schemes, including those that retain specific aspects of the brain's connectivity, were evaluated to assess how much of this gene-expression association could be recovered from summary statistics alone. Notably, the gene scores from true brain connectivity exceed those from the degree-preserving scheme, similar to the results with full randomization. This suggests that connection density alone does not account for the link between the modeled and true gene expression—a notable finding given the wide variability in the number of connections each brain region receives and projects.

The current analysis did not explicitly account for spatial structural patterns as a potential explanation for the observed association. Future work could explore alternative randomization strategies, such as distance-preserving randomization, to assess the role of spatial autocorrelation as a confounding factor. This is particularly relevant in light of the tendency for neural connections to form over short distances due to wiring cost constraints (Bullmore and Sporns, 2012).

Despite these limitations, this dissertation demonstrates robustness under degree preservation, strengthening the conclusion that gene expression serves as a structured representation of regional identity—one that maps onto the mouse brain's specific wiring architecture.

4.2 Evaluating Various Latent Priors

The results above confirm that the model's latent variables capture region-specific identity in a manner aligned with biological gene expression. However, the representation and its association with true gene expressions may be influenced by the choice of model. While the SVD-based analysis applied a classical algorithm to produce variables explaining the connectivity, the treatment did not give consideration to the probability distribution that these variables come from. Real biological systems must balance the amount of information represented with a coding cost (discussed in Section 2.3). Here, the coding cost is represented as a prior distribution over the latent variables (defined in Section 3.2.1), and the variables are fit to balance the reconstruction accuracy against the cost of the latent prior (algorithm described in Section 3.2.3).

To address this concern, this chapter moves away from SVD for producing representations. Instead, a gradient descent algorithm is employed to generate latent representations, balancing the influence of the prior distribution with the goal of maximizing connectivity likelihood. The aim is to show that, for a range of reasonable prior distributions, an association remains between the latent variables and true gene expression. If this relationship persists, then the model indicates a link between

gene expression and connectivity that is not contingent on the specific prior beliefs about the coding of connectivity.

4.2.1 Experimental Setup

In biological systems, various metabolic costs can influence the distribution of expressed factors, and different sources of noise or variable sensitivity may affect the choice of a final distribution. Overall, these priors reflect the biological costs associated with encoding information through the physical medium of gene expression.

The different priors are sourced from classic representation theory Dayan and Abbott (2005). Two different prior distributions are implemented. The first, the Gaussian prior, is implemented using L2 regularization, which imposes a loss penalty proportional to the sum-of-squares of each latent variable. The second, the Laplace prior, corresponds to L1 regularization, which penalizes the latent variables proportional to the sum of the absolute values. Each choice implies a hypothesis about how genetic expression might encode connectivity.

The effect of regularization on mean gene scores was evaluated by comparing models trained with and without L2 or L1 priors. A Welch's t-test was conducted to determine whether the mean gene score obtained without regularization differed significantly from those obtained with regularization. To estimate these distributions, multiple latent models were initialized with random weights and trained under each condition, using a fixed regularization strength taking on the values $\alpha \in \{0.001, 0.004, 0.01, 0.04, 0.1\}$.

4.2.2 Results

The gene score metric (described in Section 3.3.2) was evaluated for each prior distribution. Figure 4.2 shows a representative run with $\alpha = 0.04$, illustrating the full distribution of gene scores across all $\sim 20,000$ genes. To assess the impact of regularization, the mean gene score was compared across models trained with and without L2 or L1 priors, using 10 random initializations per condition. A Welch's t-test found no significant difference in mean gene scores ($p > 0.05$ for all α values tested), indicating that regularization did not affect this metric.

Reconstruction accuracy was also unaffected by regularization. In all cases, the latent variable model reconstructed the connectivity within 5% of the baseline accuracy achieved without regularization ($\alpha = 0$).

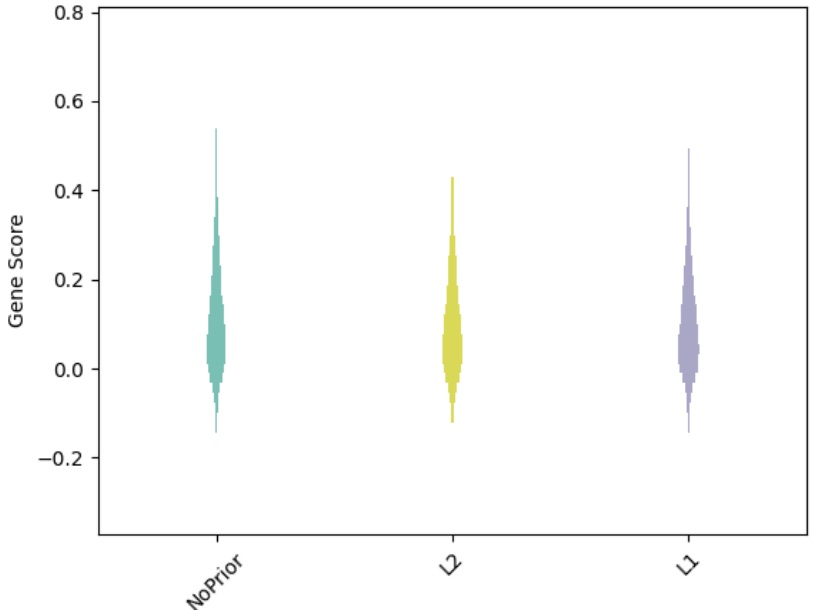

Figure 4.2: **Impact of the Latent Prior Distribution on Genetic Prediction.** The gene score histogram elicited by each prior distribution — the baseline ("NoPrior"), the L1 prior, and the L2 prior — shows the effect of each prior on the predictivity for each gene in the dataset. The choice of prior does not meaningfully affect how the latent variables predict the genetic expression.

4.2.3 Discussion

The results remain consistent across a range of prior distributions over the latent variables, including L1, L2, and Gaussian priors. This consistency indicates that the observed correspondence between connectivity and gene expression is not dependent on any single assumption about the latent factor distribution. Rather, it suggests that a shared biological signal—reflected in both the connectivity matrix and the gene expression data—drives a robust representation that holds across multiple modeling assumptions. Although each prior imposes different structural constraints on the latent space (e.g. promoting sparsity versus penalizing large values), gene expression patterns consistently align with the learned latent factors well above chance. This robustness supports the conclusion that the observed relationships are biologically grounded rather than artifacts of a particular modeling choice.

4.3 Untangling the Representations

The aim in this section is to examine the nature of how genes encode connectivity by contrasting two primary hypotheses: a mixed coding approach versus an untangled coding approach. In the mixed coding scenario, identity signatures overlap across multiple genes, analogous to population coding in neuroscience. This overlap could provide redundancy that reduces uncertainty, with each gene expression factor potentially encoding multiple variables. Population coding has previously been invoked to explain how networks of neurons with overlapping Gaussian receptive fields collectively represent a continuous variable such as position Dayan and Abbott (2005).

By contrast, untangling using L1 regularization has been suggested as a method to untangle variables, pulling variables out of a mixture to be represented uniquely in the latent variables Bricken et al. (2023). Thus each gene expression factor independently represents a specific aspect of cellular identity. Under this scheme, there is little overlap or redundancy between genes, yielding a more straightforward but potentially less robust code. Comparing these two coding strategies—mixed versus untangled—enables a deeper investigation into whether gene expression in the brain encodes identity signatures in a blended, population-like manner or through segregated, one-to-one mappings. While the results do not provide strong evidence favoring either hypothesis, this analysis contributes to understanding how different coding assumptions affect the interpretability of gene-latent relationships.

4.3.1 Experimental Setup

This experiment tests whether different priors over the latent variables influence their correspondence to gene expression. Two regularization strategies are compared: an L2 prior, which allows the latent variables to rotate and form mixed representations (referred to as the "mixed code" model), and an L1 prior, which promotes sparse, decorrelated representations (referred to as the "untangled code" model). Both models are trained to reconstruct the same connectivity data using 16 latent variables, chosen because this dimensionality best highlights differences between the regularization schemes.

To assess how individual latent dimensions align with gene expression, one-to-one correlations were computed between each latent variable and each gene. For each latent dimension, the gene with the highest correlation was recorded, and the distribution of these maximum correlations was used to summarize the results. This approach is appropriate given the large number of genes (20,000), as taking the maximum highlights strong, specific alignments. Averaging correlations across all

genes would obscure any increase in fit with key genetic variables of interest. The L1 regularizer aims to encourage a more unique, one-to-one mapping between latent dimensions and specific genes by redistributing information across the latent space. To accentuate potential differences, the number of latent variables was fixed at 16. Models were trained with regularizer strengths $\alpha = 0.004$, chosen to maintain good connectivity reconstruction. A Welch's t-test was used to evaluate whether the mean of the maximum correlation metric differed across regularization conditions, testing the hypothesis that L1 regularization increases maximal gene-latent correlations on average.

4.3.2 Results

While previous analyses showed that overall gene scores are stable across regularization schemes (Section 4.2), one-to-one gene correlations were used to further examine how latent variables align with biological signals. As shown in Figure 4.3, no statistically significant differences in one-to-one correlations were observed between L1-regularized, L2-regularized, or unregularized models (Welch's t-test, $p > 0.05$). These results indicate that, under the conditions tested, the choice of prior does not have a substantial effect on the alignment between latent variables and individual gene expression profiles.

4.3.3 Discussion

This study aimed to test whether introducing sparsity through L1 regularization would lead to more interpretable, one-to-one alignments between latent variables and gene expression. Prior work has shown that sparse priors can help uncover distinct, meaningful components in high-dimensional representations, such as monosemantic features in language models (Bricken et al., 2023) and disentangled representations in genetic data (Lopez et al., 2023). Based on these findings, it was hypothesized that promoting sparsity in the static latent model would encourage a more modular encoding of connectivity, with individual genes aligning more specifically to particular latent factors.

However, under the conditions tested, L1 regularization did not produce a statistically significant difference in one-to-one gene-latent correlations compared to L2 regularization or no regularization. The choice of prior had limited influence on the alignment between latent variables and gene expression, and no consistent advantage for either coding strategy was observed.

These results suggest that, at least within the scope of this model and dataset,

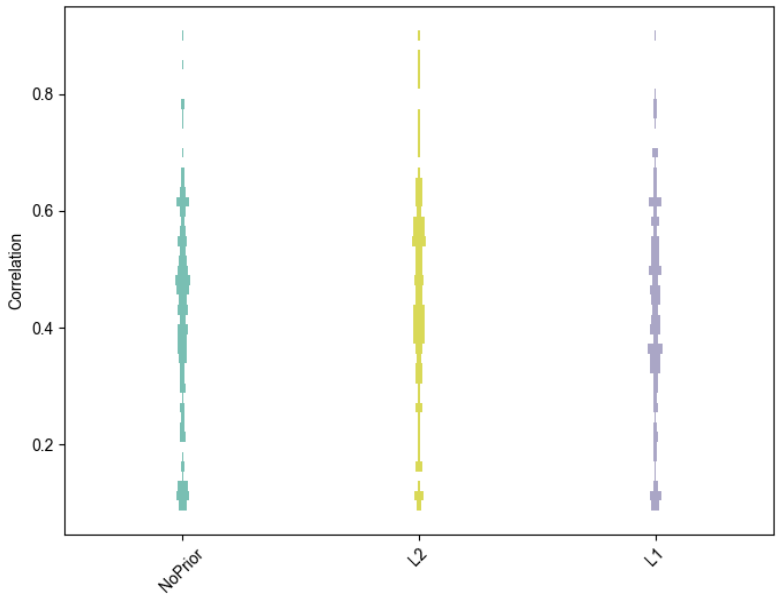

Figure 4.3: **Effect of Regularization on 1-to-1 Gene Correlations.** One-to-one correlations between latent variables and gene expression profiles were compared across regularization conditions. While L1 regularization showed slightly higher mean correlations than L2 or no regularization, these differences were not statistically significant. The number of latent variables was fixed at 16 to emphasize potential contrasts in representational structure. These results suggest that the choice of prior has limited influence on the alignment between latent variables and individual genes under the conditions tested.

gene expression may not naturally conform to a strongly modular or untangled code. Alternatively, the dimensionality, model constraints, or the biological data itself may mask more subtle effects of sparsity that could emerge under different conditions. While the idea of modular genetic "subroutines" (Levin, 2016) remains a compelling theoretical framework, further investigation is needed to determine whether and how such structure is reflected in the relationship between gene expression and connectivity.

4.4 Model Complexity

When comparing models across different levels of complexity, simpler models are found to yield the best fit to gene expression data. This supports the principle of Occam's Razor—that the simplest explanation is often the most plausible. Moreover, the model that achieves the highest gene score can be interpreted as revealing the effective genetic complexity of the connectivity data. This interpretation is further supported by experiments with artificially generated connectivity matrices, which show reduced gene scores as model complexity increases.

Previous research (reviewed in 2.1) showed that brain connectivity in several animals exhibits biclique motifs, which are amenable to efficient encoding under a genetic connectivity model. Such patterns are unlikely to emerge in a random graph model, and their presence supports the hypothesis that connections emerge from homophilic wiring principles, where the likelihood of connection depends on the similarity of gene expression profiles between source and target neurons.

However, such analyses have previously relied on complete connectomes from smaller organisms such as *C. elegans*, where the entire connectome is well-mapped and consistently produced during development. In contrast, the mouse brain presents greater challenges: its connectivity exhibits substantial individual variability, much of which arises from experience-dependent plasticity. Additionally, mouse connectomic data are typically derived from voxelized measurements, which introduce both noise and sampling limitations. These factors raise uncertainty about whether the principles observed in simpler systems extend to large-scale, higher-dimensional datasets like the mouse brain.

This section investigates whether voxel-level connectivity can be *efficiently* explained by a barcode-like genetic wiring scheme, in which structured network motifs emerge from gene expression patterns. Section 4.4.1 presents experimental results showing how model complexity affects both gene scores and the reconstruction accuracy of brain connectivity, revealing a peak complexity level that best aligns with the genetic data. Section 4.4.3 offers an interpretation of this finding, and Section 4.4.4 extends the analysis to synthetic datasets, supporting the hypothesis that the observed peak reflects the intrinsic latent complexity of the connectivity data.

4.4.1 Experimental Setup

An important lever in the static latent model is the number of latent variables allocated per voxel, which directly determines the model's complexity. To assess the impact of this parameter, a series of models were trained at varying complexity lev-

els. Two core metrics were evaluated: (1) the R^2 score, measuring how well the model reconstructs the observed connectivity matrix, and (2) the average gene score, quantifying how closely the model's latent variables align with empirical gene expression data. By systematically varying model complexity and tracking these metrics, the experiment explores the tradeoff between fitting capacity and biological interpretability.

4.4.2 Results

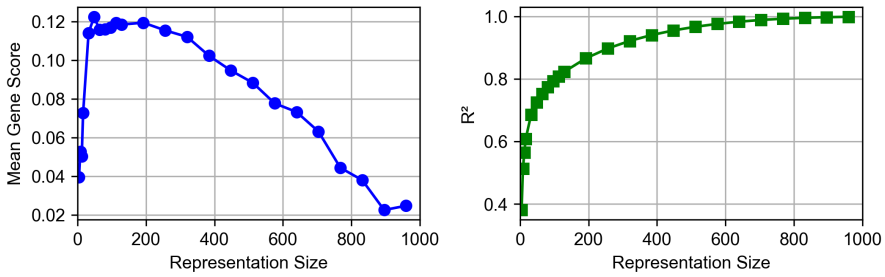

Figure 4.4: **Tradeoff Between Model Complexity and Genetic Predictivity.** As the number of latent variables per voxel increases, reconstruction accuracy of the connectivity matrix (measured by R^2) steadily improves, indicating better fit to the data. However, the gene score peaks at an intermediate complexity level before declining, suggesting that overly complex models may overfit the connectivity at the cost of biological interpretability. This tradeoff highlights a critical point: the model that best reflects genetic structure is not the most complex, but the one that balances expressiveness with constraint.

As shown in Figure 4.4, increasing model complexity leads to a consistent rise in R^2, reflecting progressively better reconstruction of the connectivity data. However, the gene score exhibits a non-monotonic trend: it rises to a peak at intermediate complexity, then declines with further increases. This peak suggests that while high-complexity models can memorize the connectivity matrix, they lose alignment with the structured patterns found in biological gene expression. In other words, simpler models yield latent variables that are more predictive of genetic structure, whereas more complex models sacrifice biological relevance in favor of raw reconstruction power.

4.4.3 Discussion

These results demonstrate a clear relationship between model complexity and the alignment of latent representations with genetic expression. Although one might be tempted to interpret the peak gene score as reflecting the "true" number of genetic factors involved in shaping connectivity, several considerations argue for caution in making such a direct inference. First, the genetic data itself is broader in scope than connectivity alone. Gene expression governs an array of cellular and developmental processes throughout the brain, which means only a portion of the genetic

variation is directly tied to connectivity. At the same time, the connectivity data available only captures a fraction of the total connectivity information—for instance, some connections may be missing or subject to experience-dependent variability (see Section 7.1 for discussion). Together, these factors imply that a strict one-to-one mapping between the peak gene score and biologically meaningful gene transcription factors is not supported by the data.

Beyond these constraints, increased model complexity beyond the point of optimal association with genetic expression can be understood as overfitting, memorizing the specific points of data rather than the underlying patterns. As the number of latent variables grows, the model may begin to learn noise or spurious correlations that are not grounded in true biological signals. Conversely, overly simplistic models fail to capture important structural patterns. The key is to identify a "critical number" of latent dimensions that accurately capture the meaningful variation without folding in extraneous or noisy features.

4.4.4 Artificial Connectivity and Complexity

A key question is whether the observed peak in gene score reflects a bias introduced by the modeling approach or whether it is driven by properties of the data itself. To investigate this, artificially generated connectivity datasets with known underlying complexity are used. The results show that the gene score peaks at the true level of data complexity, indicating that the effect arises from the structure of the data—not from limitations or artifacts of the model.

Experimental Setup

Artificial datasets were constructed to test how well the model recovers structure as a function of the number of generating factors. Each voxel was assigned an artificial barcode, consisting of two randomly sampled vectors—a source signature and a target signature—drawn independently from a $D/2$-dimensional standard multivariate normal distribution. The artificial connectivity matrix was computed as the dot product between the source and target vectors for every voxel pair. Isotropic Gaussian noise with variance 0.2 was then added to the matrix to simulate measurement noise. Finally, the matrix was sparsified to match the overall density of the real mouse brain connectivity by thresholding all values below the corresponding percentile. This procedure yielded artificial connectivity data and corresponding artificial barcodes, parameterized by the number of generating factors (D).

The standard latent variable modeling workflow is then applied: a static model is trained on the artificial connectivity data, and the resulting latents are scored

against the ground-truth artificial barcodes. This yields an "artificial gene score," which quantifies how well the model recovers the original generating factors. In doing so, the analysis tests whether the true causes of the artificial connectivity can be recovered by the modeling pipeline.

Results

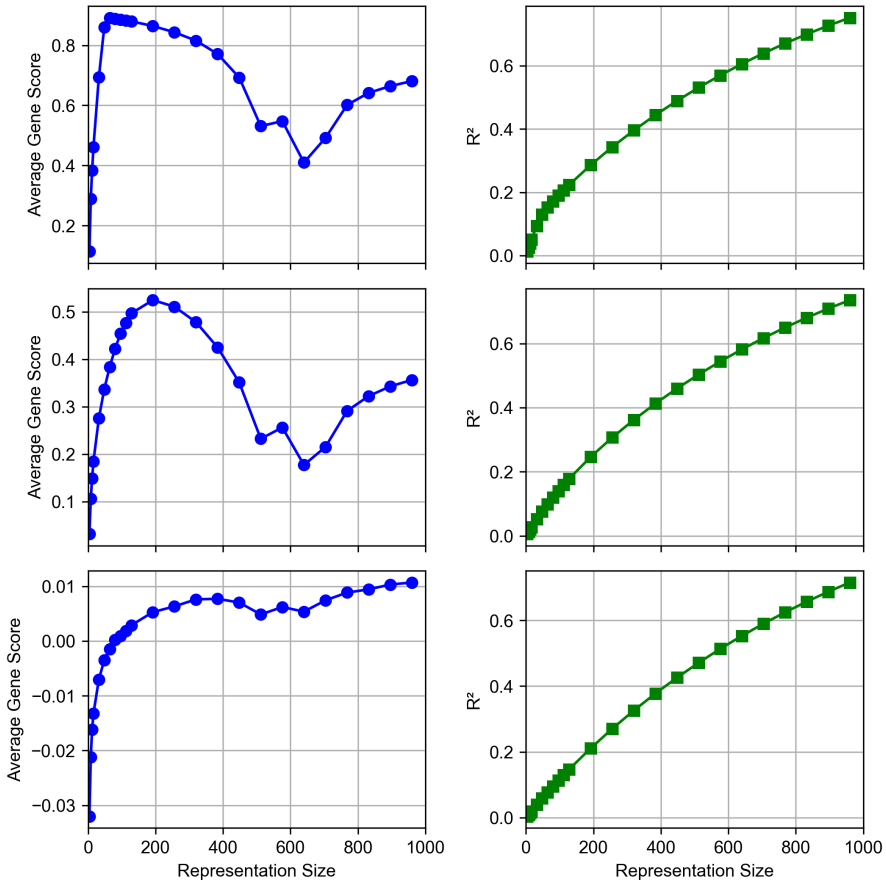

Figure 4.5: **Model Complexity versus Data Complexity**. In this figure, artificial data was constructed using "artificial barcodes," random vectors with 50 *(top)*, 200, *(middle)*, or 10,000 *(bottom)* channels. These artificial codes formed the basis of an artificial connectivity matrix, which was then run through the same analysis pipeline used for the brain connectivity, resulting in gene scores *(left)* and R^2 *(right)* metrics. The figure indicates that the gene scores do pick up on the underlying complexity of the generated data for these artificially constructed datasets.

The experiment tested the hypothesis that if the connectivity is generated from a known number of generating variables, then the gene score will peak at that same dimensionality during model fitting. As shown in Figure 4.5, artificial connectivity matrices were generated using barcodes of size 50 and 200. The resulting gene scores peaked at 52 and 192 dimensions, respectively—representing errors of just 3.8% and 4% from the true number of latent factors. These results confirm that the gene score reliably identifies the true number of generating factors embedded in the data.

The run using 10,000-dimensional barcodes tested whether the analysis pipeline contains an intrinsic bias that would produce a peak in gene score even when the data is generated by a highly complex, non-identifiable process. In this case, no peak was observed; instead, the gene score steadily increased with model complexity. However, the overall gene scores remained very low, reflecting the inability of the model to generalize when the connectivity is determined by too many independent factors. This result supports the interpretation that the peak observed in earlier experiments arises from true structure in the data rather than from artifacts of the modeling procedure.

Notably, in addition to shifting the x-location of the peak, changing the representation size also had a pronounced effect on the magnitude of the gene scores. Two aspects of the data construction process contribute to the difficulty of recovering the original generating factors: the noise added to the connectivity matrix and the thresholding applied during preprocessing. As a result, increasing the complexity of the data makes it progressively more challenging for the model to accurately reconstruct the underlying factors.

Discussion

This finding supports the conclusion that the observed peaks in gene score are driven by the intrinsic complexity of the data, not by any artifact or bias in the modeling framework. When the number of latent dimensions in the model matches the number of generating factors used to construct the artificial connectivity, the gene score reliably peaks—indicating that the model has correctly recovered the underlying structure. In contrast, when the generating process includes too many independent factors (as in the 10,000-dimensional case), no peak emerges. This suggests that the gene score metric is sensitive to the true latent dimensionality of the system.

Biologically, these results are consistent with the hypothesis that mouse brain connectivity may be shaped by a relatively low-dimensional developmental program. The appearance of a gene score peak in both biological and artificial datasets suggests that a limited number of interacting genetic and developmental factors could be

sufficient to explain large-scale wiring patterns. While this does not prove that brain connectivity is encoded in a strictly low-dimensional form, it does indicate that low-complexity models are especially effective at capturing biologically meaningful structure. This carries important implications: if a small number of core regulatory signals guide the formation of brain architecture, then it may be possible to identify interpretable, mechanistically grounded modules that drive neural development.

There is a discrepancy in the magnitude of gene scores between the artificial results and the brain data (see Section 4.4), with gene scores derived from brain data being substantially lower on average. In the artificial case, all latent variables contribute to the construction of connectivity, whereas in the brain, many genes may not play a direct role in this process. To address this, future work may focus on selecting a subset of genes known to have mechanistic relevance to neural development (see Section 7.1.4).

4.5 Conclusion

This chapter demonstrated that gene expression in the mouse brain can encode a substantial amount of the observed connectivity patterns. By applying a latent variable model, it was possible to uncover a robust association between voxel-level connectivity and genetic data—even when tested against randomization baselines that selectively preserve certain connectivity features. These baselines confirmed that the link between gene expression and wiring is not explained by simpler factors such as overall connection density, although additional analyses are required to isolate the effects of spatial structure and other potential confounding features.

Further investigation using different priors—Gaussian (L2) and Laplace (L1)—showed that the choice of prior had minimal impact on how latent variables aligned with individual gene expression profiles. Despite differences in regularization, the overall association between the latent representations and genetic expression remained stable. This suggests that the observed biological signal is not an artifact of any specific modeling assumption, but rather reflects a more general relationship between connectivity and gene expression. While sparsity was hypothesized to promote more disentangled representations, no consistent evidence was found to support this under the conditions tested.

Exploring model complexity revealed a fundamental tradeoff between expressiveness and biological interpretability. As the number of latent variables increased, the model became better at reconstructing connectivity (as reflected by higher R^2), but gene scores peaked and then declined. This suggests an optimal complexity level—one that captures essential structure without introducing noise that obscures

genetic correlations. This finding echoes Occam's Razor: the simplest model that explains the data is often the best. But it also reflects a deeper biological principle. In development, systems must operate under limited informational and energetic resources, favoring solutions that achieve maximal effect with minimal complexity. The peak in gene score may thus reflect not just a modeling sweet spot, but a real constraint shaping biological organization—one that reappears in the developmental model, where imposing complexity limits produced stronger alignment with genetic expression.

Finally, tests with artificially generated data corroborated the notion that peaks in gene scores as a function of complexity are driven by the inherent complexity of the data rather than a built-in bias of the model. When the artificial connectivity was generated using a known number of "artificial genes," the gene scores peaked exactly at that dimensionality. This finding confirms that the latent variable framework can indeed recover biologically meaningful factors when the data provide sufficient signal.

Altogether, these results support the view that gene expression influences the developmental wiring of the mouse brain, as evinced by latent-variable modeling. Different prior distributions, complexity levels, and randomization baselines all reinforce the conclusion that genetic factors shape connectivity in structured and interpretable ways. These insights advance the understanding of how genetic coding strategies might give rise to the observed macroscale organization of neural circuits, while also laying groundwork for future exploration of more nuanced or high-dimensional biological data. The analyses so far relied on a simple static model of connectivity. The next chapter investigates how incorporating biological constraints—such as local interactions and sequential development—may reveal deeper insights into the genetic basis of brain wiring.

Chapter 5

Gene Expression as a Developmental Scaffold

The previous chapter introduced a static latent model in which voxel-based latent variables explained brain connectivity, offering a potential link between gene expression and neural wiring. However, the model did not account for how such latent patterns might arise, nor did it incorporate the biological constraints of development. In real tissues, complex three-dimensional structures emerge gradually, guided by local information, constrained signaling, and a sequential unfolding of events (Turing, 1952). To better reflect these realities, this chapter introduces a developmental model in which a compact set of parameters encodes a dynamic process that generates the latent variables over time. By explicitly modeling the incremental, spatially localized formation of connectivity, the developmental approach imposes stronger structural constraints and yields representations that are more predictive of gene expression and more consistent with biological organization.

To justify the constraints placed on the developmental model, a long-standing theoretical framework is employed: morphogenesis (described in Section 2.2). Under this framework, cells utilize signaling molecules to establish morphogenetic gradients that structure and pattern the biological tissue, thus providing important information to guide cell differentiation and specialization. Crucially, local processing of these signaling chemicals, combined with diffusion and other transport mechanisms, has been shown to explain the wide range of patterned structures found in nature. This perspective is particularly effective for explaining repetitive or regular patterns, where similar interactions are repeated across space and time .

The developmental model introduced here aims to uncover these morphogens in addition to the signals that directly guide wiring. By imposing limits on how

voxels communicate with one another, additional disambiguating information must be encoded in the latent variables to ensure consistent reconstruction of the final pattern. Under this developmental paradigm, voxels encode not only connectivity but also scaffolding information relevant to generating the tissue's overall form.

This chapter proposes that, in the mouse brain, gene expression conveys not only connectivity signals but also morphogenic information essential for developmental organization. Section 5.1 begins by arguing that such disambiguating information is required for the consistent assembly of biological patterns. Through a denoising model restricted by local signalling, evidence is presented that morphogens are indispensable for pattern formation. The developmental model is then applied in Section 5.2 to mouse brain reconstruction to illustrate that the added restrictions from biology provide a more accurate representation of gene expression than simpler alternatives.

5.1 Limitations of the Static Model

This section uses a toy example to illustrate a fundamental limitation of the static model: the inability to recover or construct certain patterns using only local information. The pattern in question is not derived from brain connectivity data, but serves as a simplified analog to test whether local perception alone can reconstruct a complex signal. This abstraction allows for controlled experimentation on the kinds of informational bottlenecks that may also affect models of brain development. The key idea is that if even a toy pattern cannot be regenerated without additional signaling cues, then more complex biological patterns—such as those involved in neural wiring—would almost certainly require them as well. This motivates the introduction of a distinct morphogen-like signal in the developmental model, enabling coordination and disambiguation during pattern formation.

In essence, the developmental model imposes additional constraints on how patterns form, mandating that each stage arise through locally perceptive processes. The argument made here is that these developmental coordination constraints — specifically, the need for cells to first establish identity signatures that subsequently guide wiring — necessitate signaling extra disambiguating information. This concept serves as an explanation for the results observed in the developmental model.

To illustrate the limitation of patterns produced by the static model, an experiment was designed to test whether such a pattern contains enough information to remove noise when limited to local perception. In developmental systems, counteracting noise is a central challenge, since biological development can be understood as resisting a thermodynamic entropy gradient (as discussed in Chapter 2.3.1). This

section addresses the denoising problem using a denoising diffusion model, which removes noise from an image using only local perceptual constraints. As will be shown, the information contained in the image alone is insufficient to recover from extensive damage or to regenerate the pattern from a minimal seed. Therefore, in a developmental context where cells are limited to local perception, the patterns produced by the static model fall short. Additional disambiguating information is required to guide the developmental trajectory and determine the final identity signatures of each region.

The following subsections introduce the denoising task and the local perceptual module, followed by a presentation of the findings and a discussion of their implications.

5.1.1 Experimental Setup

To test whether a pattern can be reconstructed under local perceptual constraints, a denoising diffusion model is trained to reconstruct a single target image from noisy input. Denoising diffusion models are generative systems that learn to reverse a gradual noise process: images are corrupted by incremental Gaussian noise, and the model is trained to reverse this process step-by-step, denoising the image until it approximates the original.

Typically, diffusion models are trained across large datasets and learn general strategies for image restoration. In this case, however, only a single target image—a salamander emoji—is used. The goal is not to train a general-purpose model, but to assess whether the image can be reconstructed from noise under strict perceptual limitations. The emoji image is chosen for its simplicity and its patterned, symmetric structure, which resembles the kinds of spatial regularities found in biological forms.

To constrain the model's perceptual capacity, a 3×3 convolutional filter is applied to the image at each time step. In the original full-image setup, the network receives the entire image as input and produces a denoised version of the entire image as output. In the constrained setup, however, the model is restricted to a sliding 3×3 tile centered on each pixel. The input to the model becomes just this local 3×3 neighborhood, and the output is a prediction of the denoised value for the center pixel. This localized processing mimics the conditions of biological development, where cells must make decisions based on limited, neighborhood-level information. The model is trained to reconstruct the target image from noisy versions using only these local observations, simulating the restricted perceptual context of developmental systems.

5.1.2 Results

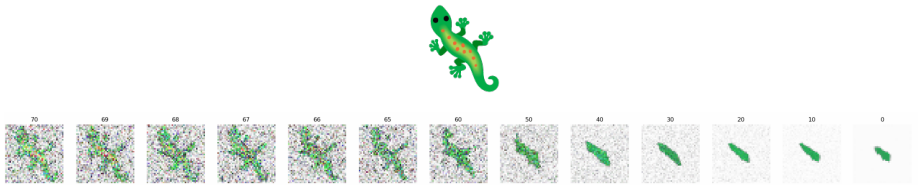

Figure 5.1: **Denoising with Local Perception.** The top image shows the target pattern used for training. The bottom row displays snapshots from a reverse diffusion process, which begins from a heavily noised version of the target and incrementally removes noise over 100 time-steps. The model is constrained to observe only local pixel neighborhoods. Although the system converges to a stable and locally self-consistent output, it fails to reconstruct many global features of the original pattern. This illustrates how local perception alone is insufficient for accurate reconstruction without additional disambiguating signals.

When trained with full perception—allowing access to the entire image at once—the model was able to accurately reconstruct the target image from noise. However, when limited to local perception via 3×3 filters, the model failed to recover the global structure of the salamander. Although the output converged to a stable, locally consistent pattern, it lacked many of the essential details of the target, as shown in Figure 5.1.

This result demonstrates a key limitation of local perception: even when local cues are present, they may be insufficient to resolve global ambiguity. In the context of development, this illustrates why additional signals—such as morphogens or other global patterning cues—are necessary to guide the emergence of complex structures.

5.1.3 Discussion

These experiments demonstrate that introducing local perceptual constraints alters the information available during reconstruction. When the denoising system had access to the entire image—as in standard diffusion setups—it successfully reproduced the target. But when perception was restricted to local neighborhoods, the system was unable to recover the global structure, even though the output remained locally self-consistent. It is not that local rules are inherently insufficient, but that they require additional structure to resolve ambiguities that emerge when global context is unavailable.

In biological development, morphogen signals play this role. Although they are locally sensed, their gradients encode broader positional information—allowing indi-

vidual cells to infer their place within a larger structure. This helps overcome the ambiguities introduced by strictly local perception and enables coordinated pattern formation across tissues.

The developmental model introduced in the following sections adopts this framework: it simulates a process where structure emerges from locally informed updates. To succeed, it must incorporate signals—analogous to morphogens—that provide disambiguating context needed for organizing complex patterns under perceptual constraints.

5.2 The Developmental Model

The developmental model (defined in Section 3.2.4) aims not only to generate latent variables that account for connectivity but also to explain how these latent variables originate through a constrained process that reflects developmental biology—specifically, a sequential process governed by local perception. The preceding section demonstrated that explaining the source of latent patterns inevitably involves modeling additional guidance signals, which are needed to resolve local ambiguities and properly assemble the overall pattern. An important question, therefore, is whether including these coordinating signals leads to a better fit between the model and the observed genetic expression. If a closer match is indeed observed, it implies that genetic expression in the brain not only encodes the regional identity signatures essential for wiring but also carries disambiguating information required to establish those identity signatures in the first place.

To investigate this possibility, the developmental model is introduced as an alternative coding hypothesis. Like the static model, the developmental model generates the mouse brain's connectome. However, in contrast to the static approach, the developmental model also accounts for the origin of latent variables by describing their spatial distribution through a developmental program. This program defines a sequence of states for each voxel, where each state update relies on the local neighborhood of voxels. As a result, the developmental model is fundamentally constrained by local perception and decision-making. A full description of this architecture is provided in Section 3.2.4. Here the model is evaluated as a hypothesis explaining the biological development of brain wiring.

5.2.1 Experimental Setup

To evaluate whether additional disambiguating information improves the model's ability to predict genetic expression, a developmental model was trained to reproduce

voxel-wise mouse brain connectivity data. The gene score metric was then assessed at the final developmental step—the same state used to construct the predicted connectivity. Because this final state emerges from a sequential process governed by local interactions, it naturally incorporates auxiliary information that may support more structured pattern formation.

In the static model, the number of latent variables played a critical role in controlling model complexity—too many led to overfitting, while a moderate number yielded the strongest correspondence to gene expression (see Section 4.4). To explore this constraint in the developmental model, an analogous parameter was varied: the size of the hidden layer in the NCA update rule. While the developmental and static models differ in architecture, reducing the hidden layer size serves as a proxy for lowering the model's internal capacity. This allowed the developmental model to be evaluated under varying complexity regimes, mirroring the insights gained from the static case and enabling analysis of how representational quality changes with model capacity. Hidden sizes were tested at 256, 128, 64, and 32 units.

Because substantial variation was observed in gene score across different hidden sizes, an additional set of trials was conducted with the hidden size fixed at 32. In these experiments, the only factor varied was the random initialization of the NCA parameters, allowing for an assessment of how much variation in gene predictivity could be attributed to initialization alone, independent of model complexity. The outcomes of these fixed-size trials were taken to represent the core developmental model used in comparisons with the static model.

The hypothesis tested was that the developmental model achieves higher mean gene scores than the static model. To evaluate this, five independent runs of each model type were trained with randomized initializations, and a Welch's t-test was conducted to compare the resulting gene score distributions.

5.2.2 Results

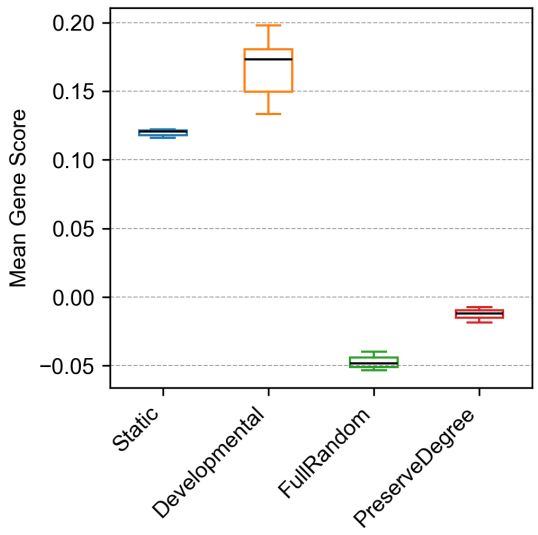

Figure 5.2: **Model Comparison: Gene Scores.** Mouse brain genetic predictivity is shown for the static and developmental models as well as randomized controls. The developmental model outperforms the static model in explaining gene expression.

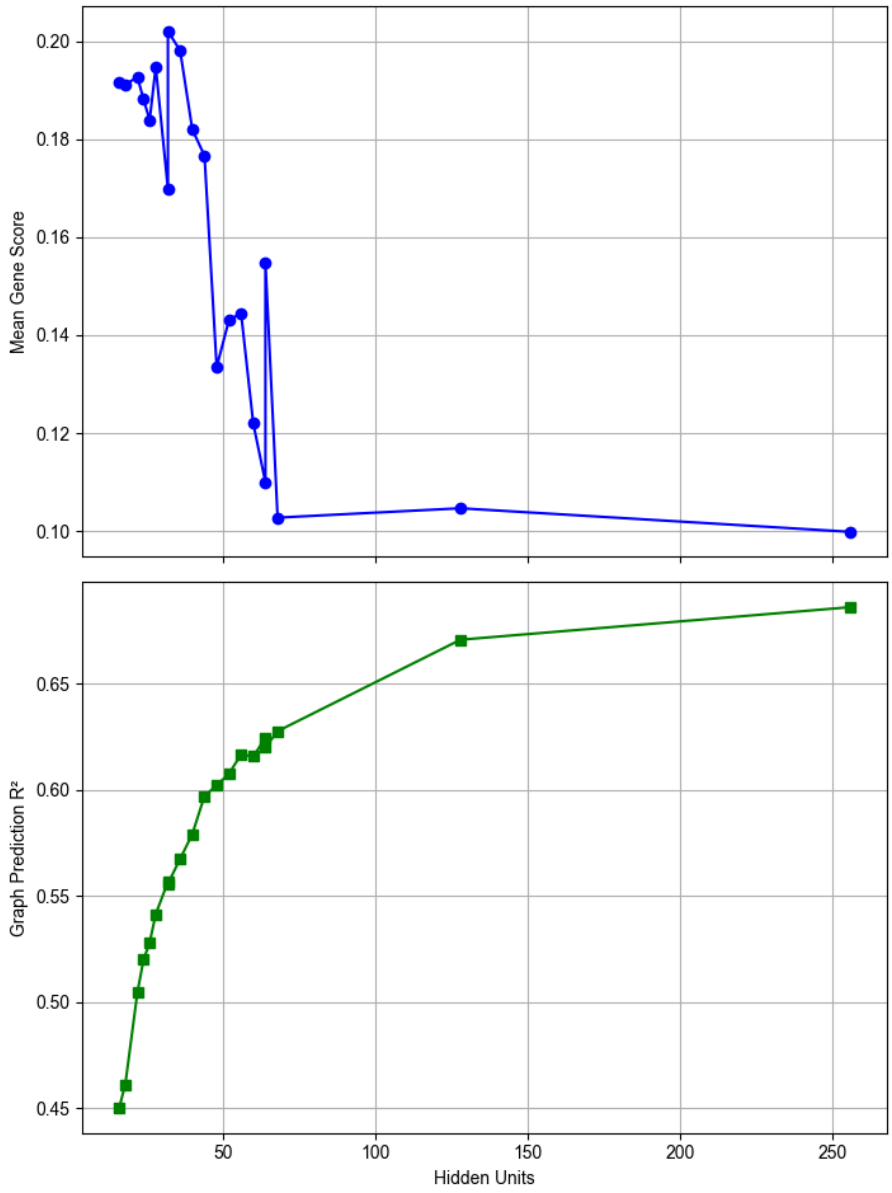

Figure 5.3: **Tradeoff Between Model Complexity and Biological Predictivity.** Gene scores (top) and connectivity reconstruction accuracy (bottom) are shown for developmental models at varying levels of complexity. As model complexity increases, reconstruction performance improves—but the ability to predict measured genetic expression declines. This tradeoff suggests that simpler models capture a structure more consistent with biological pattern formation.

The developmental model achieves higher genetic predictivity than the static model (Welch's t-test, $p = 0.0049$). As shown in Figure 5.2, the best-performing developmental model reaches a gene score of 0.202, compared to 0.120 for the static model. While the static model achieves better performance in reconstructing connectivity—explaining 68.6% of the variance compared to 55.6% for the developmental model—the developmental model produces latent representations that are more aligned with gene expression. This result suggests that the developmental model, by simulating a sequential and spatially constrained growth process, captures additional structure that is present in biological gene expression patterns.

To examine how model complexity affects performance, developmental models were trained across a range of hidden layer sizes. As shown in Figure 5.3, both connectivity reconstruction and gene expression prediction vary systematically with model complexity. Larger models achieve better reconstruction of the observed connectome, but perform worse on gene expression prediction. In contrast, smaller models—particularly those with 32 hidden units—produce more biologically aligned gene scores, despite reduced accuracy in connectivity prediction. These findings support the hypothesis that constraining perceptual and computational capacity encourages the model to adopt simpler, more generalizable developmental strategies that better reflect the structure of biological gene expression.

5.2.3 Discussion

This section compared two hypotheses for coding wiring patterns in the mouse brain: the static model and the developmental model. In the static model, latent variables are fit to maximize the likelihood of observed data. In contrast, the developmental model employs a sequential process with local perception to generate latent variables. Findings indicate that the developmental model achieves higher predictivity of genetic expression, thus providing computational evidence that the developing brain leverages morphogenetic cues, materialized as gene expression patterns, to coordinate the emergence of region-specific cell identities.

An additional result of interest was that for the developmental model to exceed the static model as a hypothesis of gene expression coding, it was necessary to introduce a measure of complexity within the developmental process itself. In the static model, performance improved when the complexity of the latent variables was kept relatively low. However, simply limiting the number of latent variables in the developmental model did not suffice to outperform the static model. It was also essential to constrain the developmental model's perception and decision-making capability. This finding highlights a broader insight: biological development operates under resource and

complexity constraints, and successful models must reflect those limitations. To understand how gene expression gives rise to wiring patterns, it is not enough to search for patterns in the data; models must be structured to reflect the bounded computational and physical processes by which real biological systems operate.

The next section will assess an interpretation of why the developmental model performs so well: it is modeling the spatial structure of the biological data.

5.3 Ablating Spatial Structure

The higher mean gene scores achieved by the developmental model, compared to the static model, may be attributed to its ability to capture spatial structure inherent in the data. While the static model treats each voxel independently, the developmental model defines complex, nonlinear interactions between neighboring voxels that shape the emergent voxel states. As a result, the developmental model learns a representation that reflects the spatial organization of brain connectivity, which enhances its alignment with gene expression patterns.

To test this interpretation, an experiment was conducted in which the developmental model was applied to data with spatial structure removed. Both the connectivity and gene expression data were randomly shuffled, thereby disrupting spatial relationships while preserving the underlying voxel-level associations between connectivity and gene expression. This ablation led to a marked deterioration in the gene scores achieved by the developmental model, confirming that its performance relies on capturing spatial structure present in the original data.

5.3.1 Experimental Setup

The gene expression and connectivity data were both shuffled using a common permutation of voxel indices. The brain interior (see Section 3.1.4) was used to define the set of candidate vertices for shuffling. A random permutation was generated to reassign the spatial positions of these vertices, resulting in a shuffled gene expression dataset and a shuffled connectivity dataset. While the underlying connectivity values and gene expression levels remained unchanged, their spatial locations were reassigned, effectively destroying the spatial structure of the data.

Gene scores were evaluated for both the shuffled and baseline datasets. In the shuffled condition, the model's latent representations were compared to the shuffled gene expression data, consistent with the input used during training. In the baseline condition, unshuffled gene expression data was used for both training and evaluation.

A Welch's t-test was conducted to test the hypothesis that ablating spatial structure through shuffling results in reduced mean gene scores. Five models were trained for each condition, using independent random permutations in the shuffled case and different parameter initializations in both cases.

5.3.2 Results

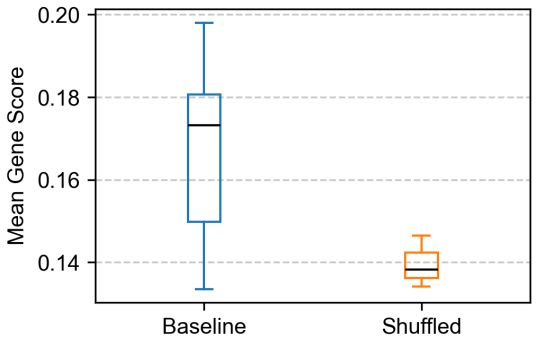

Figure 5.4: **Shuffled Performance.** Gene scores were evaluated for the baseline brain data and on the shuffled dataset. The shuffled data results in lower gene scores, indicating that the spatial structure is critical for the developmental model's representation of the data.

The mean gene scores obtained from the baseline data were significantly higher than those from the shuffled dataset (Figure 5.4), with the shuffled condition producing markedly lower scores ($p = 0.039$). This result indicates that spatial structure plays a critical role in the developmental model's ability to align with gene expression.

5.3.3 Discussion

These findings demonstrate that the developmental model's ability to align with gene expression critically depends on the spatial structure of the data. When spatial relationships between voxels were disrupted through shuffling, the model's gene scores significantly declined, indicating that spatial context plays an essential role in the model's representational capacity. This suggests that the developmental model leverages broader spatial patterns to infer biologically meaningful structure.

5.4 Conclusion

Chapter 5 has illustrated the critical role of gene expression in guiding both the connectivity patterns of the mouse brain and the broader morphogenetic processes that organize developing tissue. By contrasting a static latent variable approach with a developmental model constrained to local perception, it became evident that a sequential process with local signalling provides a more comprehensive explanation of gene expression data—especially when the developmental program's complexity is carefully managed.

A toy example of denoising diffusion under local perception constraints demonstrated why, in some cases, local information alone is insufficient to reconstruct complex patterns. Additional signals, akin to morphogens, must resolve ambiguities that otherwise prevent the faithful recovery or generation of intricate structures. Experiments on the mouse brain further underscored this necessity: the developmental model outperformed the static model in capturing gene expression, indicating that morphogen-like cues are indeed embedded in the genetic code.

These findings suggest that gene expression does more than merely specify the final wiring layout. It also carries essential information for orchestrating the orderly emergence of regional identities—a process contingent on both connectivity requirements and developmental cues that unfold over time. The developmental model's success in representing these features highlights the need to view gene expression as both a barcode for wiring and a scaffold for pattern formation, offering valuable insights into the genetic basis of complex neural development.

While this chapter focused on the developmental model as a tool for understanding biological processes, the next chapter asks a different question: can the same model be repurposed as a general-purpose network encoder? Specifically, Chapter 6 explores whether developmental programs can be leveraged for efficient architecture search, compressed representations, and robust circuit formation—extending their utility beyond neuroscience into the domain of machine learning and artificial systems.

Chapter 6

Developmental Models as Network Encoders

Having established that developmental models can more accurately capture the genetic underpinnings of brain connectivity, this chapter turns to their broader utility as general-purpose network encoders. Developmental models are not only biologically grounded representations of brain wiring, but they also offer practical advantages for architecture search and robust computing. In architecture search, the goal is to optimize the design of model architectures themselves, often across vast search spaces. While conventional practice requires human engineers to hand-craft architectures suited to specific tasks, automated methods, including neural architecture search and zero-shot modeling, seek to automate this process, albeit at significant computational cost. Developmental encodings offer a more compact and efficient alternative, capable of inducing structured architectural changes through minimal modifications to a generative program.

Developmental programs are introduced as a novel way to encode the space of neural architectures. By modifying a small number of parameters in a developmental program, one can induce large-scale, structured changes in an architecture—changes that would typically require multiple individual alterations in a standard encoding scheme. This property suggests that developmental encoding functions as a compressed representation of connectivity, potentially offering significant savings in both computation and design effort. Furthermore, as models of biological development, the specific kinds of patterned changes and architectures formed may resemble neural architectures in biology. And the wiring is grounded in 3-dimensional space, a property that facilitates biologically grounded architecture search, a topic that will be further discussed in Section 7.3.

Another advantage of developmental programs is their capacity for self-repair, a feature largely absent from conventional circuit designs. Physical damage to a silicon-based circuit is typically irreparable, requiring costly redundancy to ensure reliability in mission-critical applications such as satellites. Developmental architectures, by contrast, embed the information needed to rebuild connectivity within the system itself. This built-in resilience enables the possibility of self-repairing computing systems, which could prove valuable in settings where physical maintenance is infeasible or prohibitively expensive. The broader implications of such robust architectures are explored further in Section 7.5.

Among the many properties that make biological development remarkable, two stand out in the context of this work: efficiency and robustness. Developing organisms build complex structures from compact genetic programs, often reusing subroutines to produce repeated or modular anatomy. They also maintain resilience, correcting local errors and, in some cases, regenerating lost tissues. In the sections that follow, the developmental model introduced in this dissertation is examined through this lens. Although it is a computational abstraction, it demonstrates analogous behaviors—efficiently encoding structure and recovering from simulated damage—offering insight into how such biological strategies might be captured and explored through modeling.

6.1 Efficiency

To evaluate the efficiency of the developmental model, it is compared with a static approach that provides a baseline for compression. The static model uses singular vector decomposition (SVD), which yields the most efficient linear approximation for a given number of dimensions (as discussed in Section 3.2.2). Both models are evaluated across a range of compression levels, with reconstruction accuracy and parameter count used to assess performance. The results show that while the static model performs well at higher complexities, the developmental model is more efficient in the low-parameter regime, capturing essential structure with fewer resources. This analysis clarifies the extent to which a developmental process can serve as a compact encoding of brain-like structure.

6.1.1 Results

The performance metric used in this comparison is R^2, which measures the amount of variance in the observed connectivity that each model explains. As shown in Figure 6.1, developmental models achieve higher R^2 scores than static models when model

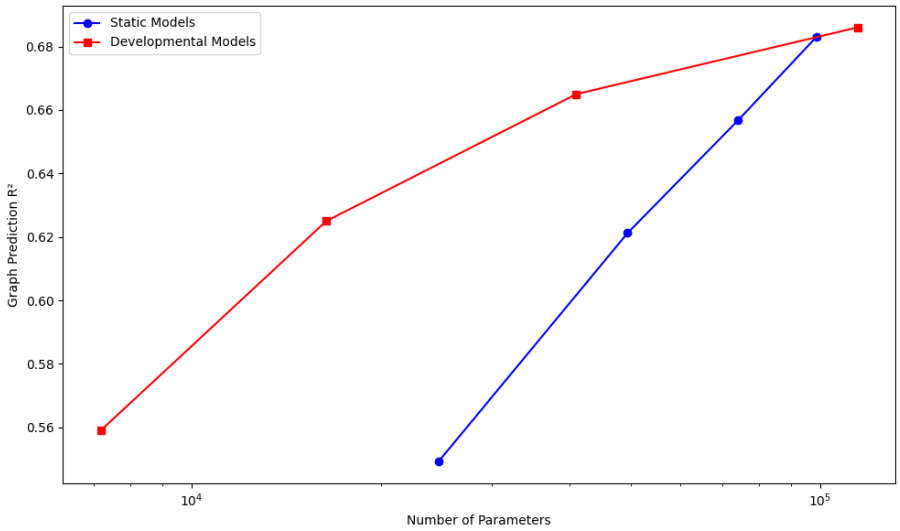

Figure 6.1: **Complexity and Reconstruction Quality.** The amount of the connectivity explained (R^2 value) is plotted against the complexity of the model, for a selection of static and developmental models. The results indicate that small developmental models are particularly efficient representations of the connectivity.

size is small, suggesting greater efficiency in this low-complexity regime. However, when the model size continues to increase, static models eventually surpass developmental models in efficiency. When the number of static latent variables grows sufficiently, the static model approaches of the connectivity variance.

6.1.2 Discussion

Singular value decomposition (SVD) is a standard compression method for network-level connectivity. It treats the connectivity matrix as a low-rank approximation problem, identifying global structure without taking spatial context into account.

In contrast, the developmental model is explicitly grounded in spatial relationships. The efficiency comparison between the two sheds light on spatial regularities in the data that SVD does not exploit. Such regularities are expected for several reasons. First, neural connectivity often follows energetically efficient patterns, with connections more likely between nearby regions (Bullmore and Sporns, 2012). Second, the dataset, produced by viral tracer imaging, includes not only source and target regions but also the full trajectory of axonal projections, which introduces additional spatial correlations that static linear models like SVD are not equipped

to compress efficiently.

The developmental model's strong performance at lower complexity levels suggests that these spatial patterns are not only present but structured in a way that can be captured by a simple generative process. In natural systems, developmental programs often reuse local subroutines to build repeated anatomical motifs—for example, the modular patterning of leaves or vertebrae—reducing the need to encode each structure independently. The efficiency of the developmental model suggests that similar patterned regularities may exist in brain wiring. These regularities may reflect basic spatial trends or more complex motifs embedded in the developmental process. The results provide computational evidence that the connectivity in this dataset is structured in a way that can be effectively captured by a developmental mechanism.

6.2 Robustness

Developing biological systems must contend with injury, noise, and uncertainty. Physical lesions, toxic interference, or stochastic developmental fluctuations can all disrupt normal pattern formation. A system is said to be *robust* if it can recover from such disruptions—either by correcting deviations during growth or by actively repairing damage after the fact. In multicellular organisms, this robustness is enabled by local communication and feedback: cells perceive their environment, assess local discrepancies, and make decisions that collectively guide global pattern stability.

To explore whether these principles can be captured in a computational developmental model, this section evaluates the regenerative capacity of the developmental system when trained on simulated injuries. While the mammalian brain lacks extensive regenerative ability, modeling recovery in this context serves as a test of the model's general robustness. The results illustrate how a locally informed, self-organizing system can recover from significant structural loss—demonstrating resilience, adaptability, and the potential utility of such models in future applications, including regenerative medicine and self-assembling machines (see Section 7.4).

6.2.1 Experimental Setup

To evaluate the model's capacity for self-repair, the training procedure includes simulated developmental lesions. These lesions are introduced by taking partially developed brain states—intermediate outputs of the model—and ablating a contiguous region of voxels. The lesion erases the latent variables within that region, reverting them to the initial seed state. Lesions are spherical, with a radius equal to 20% of

the brain's width, and are placed at randomly sampled interior locations to ensure a wide range of injury types.

The model is trained to recover from these injuries through additional training steps. In this setup, backpropagation plays a role analogous to evolution: it adjusts the developmental rules by presenting a more complex environment in which injuries must be overcome. During training, a pool of previously grown brain patterns is maintained. For 20% of training steps, the model is initialized from a lesioned intermediate state drawn from this pool. The remaining 80% of training steps initialize the model from the seed state. No specialized loss function is used for lesion recovery. The model is trained to produce the same final connectivity pattern regardless of its initial condition, forcing it to discover local mechanisms that support regeneration.

Once trained, the model is tested by applying new lesions to fully developed brain states and running the model forward without further learning. In these test-time conditions, the model receives no additional gradient updates; regeneration occurs entirely through the forward application of the learned cellular update rules. In this sense, the capacity for self-repair is embedded in the developmental program itself—a learned, general-purpose strategy for restoring structure based on local cues.

6.2.2 Results

As shown in Figure 6.2, the developmental model is able to recover a substantial portion of the original wiring structure following lesion. On average, the model regains approximately 92% of the original R^2 performance achieved when grown from the seed state. The degree of recovery depends on lesion size: smaller lesions destroy less of the original structure and permit more accurate repair, while larger lesions degrade both performance and the system's ability to fully regenerate the target pattern.

The recovery trajectory is gradual. After lesioning, the system initially exhibits a drop in connectivity accuracy. Over subsequent time steps, the latent variables rebuild toward the original state, guided by local interactions. The final recovered state does not fully match the original—but it stabilizes at a new plateau, with substantial recovery of connectivity structure, demonstrating the model's ability to self-heal.

6.2.3 Discussion

The regenerative behavior observed here demonstrates that a developmental model based on local perception and sequential updates is capable of recovering from dam-

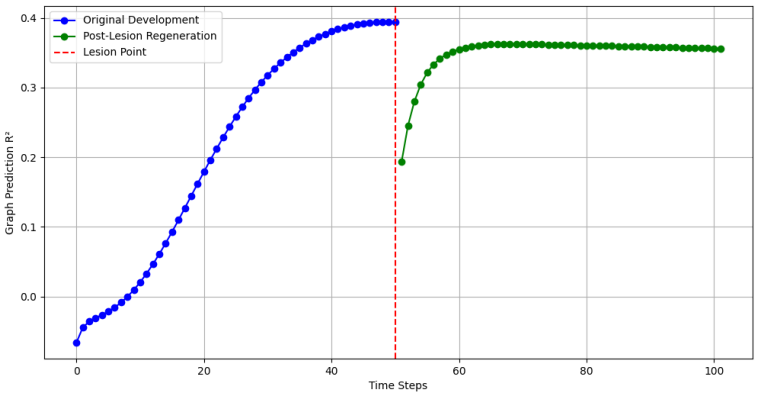

Figure 6.2: **Repair After Injury.** As the developmental model produces the initial pattern (blue curve), it incrementally builds up the predicted connectome, increasing the amount of the true brain network predicted (R^2). Then, a lesion is applied to the system (red line). Despite being in a state it has never experienced during training, the developmental model is able to rebuild the ablated region (green curve) and partially recover the original predicted connectome. The plot illustrates the model's capacity for self-directed structural repair.

age it has not previously encountered. This behavior arises not from memorization, but from general-purpose rules learned during training that allow the system to recognize and repair discrepancies in local structure. These perception-action loops—wherein each voxel continuously integrates information from its neighbors and updates accordingly—form the basis of the model's robustness.

Such dynamics are significant not only for computational modeling but for broader scientific and technological goals. Robust developmental models open new paths toward synthetic biological systems that can self-correct or maintain structure in the face of disruption. In regenerative medicine, they suggest avenues for engineering pattern-forming systems capable of autonomous repair. In biological computing or self-assembling machines, they offer a proof-of-concept for architectures that sustain functionality through local feedback and adaptation.

While not intended as a direct simulation of mammalian brain regeneration (which is limited in real organisms), this system offers a toy model of regenerative capacity. It demonstrates that recovery can emerge naturally from local, distributed rules without centralized control or special-purpose reprogramming.

That said, important limitations remain. The injury model used here is highly

simplified—lesions are spherical, clean, and spatially isolated. Real-world developmental damage is likely to be more irregular and may involve diffuse effects or conflicting chemical cues. Additionally, in biology, regeneration often uses a distinct set of mechanisms from development. For example, limb regeneration in salamanders involves the migration of epithelial cells to form a wound epithelium, the dedifferentiation of mature cells, and a dependence on nerve-derived signals such as nAG to sustain cell proliferation—none of which are involved in initial limb development (Vervoort, 2011). Despite these limitations, the present work highlights how even simple developmental models can be made resilient, offering a first step toward understanding and designing robust self-organizing systems.

6.3 Conclusion

This chapter demonstrated how developmental architectures could enable efficient, robust network formation. By encoding connectivity through a small set of parameters that guided sequential, locally informed updates, developmental programs offered a compressed representation of the underlying structure, outperforming static models at lower complexity levels. This efficiency stemmed from the ability to induce large-scale changes in network connectivity through minimal parameter shifts.

Additionally, the capacity for *self-repair* underscored a key advantage of developmental architectures. When portions of the pattern were erased or otherwise disrupted, the same local decision-making process that constructed the network could regenerate lost structure, ensuring greater resilience than conventional systems. This approach drew inspiration from biological processes, which often relied on iterative developmental signals and morphogenetic gradients to build, maintain, and even regenerate tissues.

Taken together, these findings highlighted two powerful benefits of developmental programs for generative architecture design: efficiency in encoding large-scale connectivity patterns, and robustness through local self-repair mechanisms. The following section, Future Work, discussed additional insights—particularly how these architectures compared with, and potentially drew from, similar processes in biology.

Chapter 7

Discussion and Future Work

This section begins by outlining the limitations of the findings presented in the preceding chapters. It then describes planned extensions to the training framework, including a new line of research that uses connectomic data to ground the search for brain-like AI architectures. Ethical considerations are addressed as a necessary component of this work, given its implications for both neuroscience and artificial intelligence. The chapter concludes with a discussion of the broader context in which developmental modeling takes place, highlighting areas where these insights may have meaningful applications in science, medicine, and engineering.

7.1 Limitations

Developmental models, broadly defined and potentially incorporating greater biological complexity in the future, are well positioned to leverage the large-scale connectomic datasets anticipated to emerge in the coming years. The computational model presented in this dissertation relies on numerous simplifying assumptions about biological development, axonal targeting mechanisms, and gene expression patterns. Additionally, the dataset provides only a partial representation of the true developmental processes, omitting biological dynamism and being further reduced in granularity through post-processing routines necessary for generative modeling analysis. Explicitly identifying these limitations is critical because it clarifies the boundaries of the current approach, highlights areas for methodological refinement, and underscores the importance of future work aimed at achieving greater biological fidelity and computational efficiency.

7.1.1 Capturing Developmental Dynamics with Latent Variables

This dissertation compared different voxel-based latent variable models based on how well their latent variables predicted measured gene expression. Latent coding strategies that more closely align with observed genetic expression are taken to better reflect the biological development of neural wiring.

However, the recorded genetic expression is only a snapshot of the developmental process. In fact, gene expressions in developing systems often change significantly over time, with certain genes becoming inactive or less expressed after their initial developmental role is complete. This temporal regulation of expression often occurs in genes involved in early neural patterning and embryonic tissue formation. One example is the role of the LIN-14 protein in *C. elegans*, which regulates developmental timing by being highly expressed during early larval stages and then declining to allow progression to later stages (Ambros, 2000). Additionally, the segmentation clock in vertebrates uses oscillatory expression of genes like those in the Notch signaling pathway to generate a periodic pattern along the antero-posterior axis during embryogenesis (Baker et al., 2006; Kageyama et al., 2009). Dynamics in expression are a normal part of cellular development, as epigenetic mechanisms such as DNA methylation or histone modifications can suppress previously active genes, effectively silencing them after they fulfill their developmental function (Reik et al., 2001). Early developmental genes (such as pluripotency markers) often decrease expression as cells commit to specific fates, replaced by lineage-specific genes that define the final function of the cell (Christophersen and Helin, 2010). Particularly in neural development, some genes active during initial growth phases become suppressed once neural circuits are established (Cantera et al., 2014; Gurok et al., 2004) and refined (Flavell and Greenberg, 2008). Altogether, the evidence points to a complex time-varying program of gene expression.

A limitation of the present study arises from the fact that neural wiring is modeled based on a single static snapshot (the latent variables), and is compared against a similarly static snapshot of genetic expression data. In reality, gene expressions in the developing brain change dynamically over time. A potential solution would involve leveraging dynamic genetic expression data, such as that provided by the Allen Institute's Developing Mouse Brain project (Lein et al., 2007), which offers in situ hybridization (ISH) datasets from various developmental stages. However, the developmental model is not fully equipped to make use of such a resource. Specifically, the developing brain's genetic expression data are aligned to a different coordinate system compared to the adult brain data, making direct comparisons difficult. Even

if alignment methods were successfully developed, the complexity of mechanical processes in brain tissue development—currently not represented in the developmental model—would create further complications for mapping expression between developmental stages.

A related extension to improve realism in our developmental model could involve explicitly incorporating intermediate developmental states into the wiring process, mirroring the gradual and cumulative nature of biological neural development. While this would enhance the biological fidelity of the model, it would significantly increase complexity. Additionally, evaluation remains a challenge because, again, our primary reference point is currently limited to genetic expression snapshots at one timepoint. Developing robust evaluation methods capable of assessing dynamic model predictions against dynamic biological data is an open area for future progress.

7.1.2 Mechanics of Development

Biological development involves significant mechanical forces, including bending, shaping, and tissue deformation. In neural development specifically, mechanical processes such as neuronal migration play a crucial role in determining the final structure and connectivity of the brain. Currently, the developmental model used in this dissertation does not explicitly incorporate these mechanical aspects. A promising direction for extending this model would involve integrating mechanical simulations that account for forces shaping tissue morphology and guiding neuronal migration, thus capturing additional biological realism.

Another aspect to consider is the locality inherent in neuronal wiring decisions. In biological systems, neuronal growth cones sense and respond to environmental cues locally, in contrast to the global wiring rule currently employed by our developmental model, which transforms latent variables into wiring patterns in a single global step. An iterative process, where axons and dendrites grow, branch, and form connections based on local cues over multiple time-steps, would reflect actual neuronal development better. Such iterative, local decision-making mechanisms could be introduced to enhance the biological fidelity of our developmental model.

Integrating mechanical simulations and iterative local wiring decisions into the developmental model represents exciting avenues for future research. However, these enhancements would significantly complicate error gradient backpropagation, posing potential obstacles for optimization. A notable advantage of the neural cellular automata model currently employed is its balance between realism (local interactions) and ease of gradient propagation. Future work should explore methods to maintain or adapt these gradient-propagation properties while introducing more biologically

detailed mechanisms.

7.1.3 Modeling Uncertainty

In this dissertation, the latent variables encoding connectivity are treated as maximum a posteriori (MAP) estimates (see Section 3.2.1). However, the encoding principle is more general: a full variational distribution, $Q(\mathbf{z})$, can represent uncertainty over the latent variables. This has particular relevance given the findings in Chapters 4 and 5, where constrained model complexity was shown to significantly influence performance. Within the variational free energy framework, complexity takes on a precise mathematical definition, offering a principled way to balance parsimony and reconstruction accuracy. Adopting a fully Bayesian perspective, where uncertainty in both latent variables and model parameters is explicitly modeled, could thus yield models that are better aligned with biological data by naturally including only the necessary level of complexity. Although this dissertation does not implement such an approach, future research might leverage this framework to improve interpretability and better capture the underlying biological mechanisms.

7.1.4 Modeling Specific Genes

A key direction for future work is to refine the genetic input by selecting a subset of genes known to play a role in neural development. The current model includes nearly 20,000 genes (see Chapter 3.1), many of which are unlikely to contribute meaningfully to connectivity formation. By leveraging Gene Ontology (GO) data — curated functional annotations describing biological processes, cellular components, and molecular functions (Ashburner et al., 2000) — a more focused set of genes associated with developmental processes can be used. This targeted selection is expected to improve gene scores by reducing noise from irrelevant genes, thereby enhancing the signal relating gene expression to morphogenesis and neural wiring. Notably, gene scores derived from biological data were lower than those obtained from artificial data (see Section 4.4.4), where all latent variables contribute directly to connectivity. In contrast, only a subset of genes in biological systems may be functionally relevant, and isolating this subset could provide a more accurate measure of the genetic basis for developmental connectivity.

7.1.5 Data Limitations

There are several important limitations inherent in the connectivity dataset used in this work. While the data provides a useful signal, it is important to acknowledge that the data are not perfectly representative of the underlying neural connectivity.

One significant limitation is the absence of synapse-level information. The dataset employed was derived from tracing experiments, a method that lacks the high-resolution capacity of, e.g. electron microscopy, which can explicitly map individual synapses. The distinction is critical because neural connectivity depends fundamentally on synaptic connections rather than mere proximity or anatomical adjacency. Axons frequently traverse regions without forming synaptic contacts; hence, the adjacency matrix used to represent connectivity might represent these scenarios incorrectly as connections.

Another key limitation arises from the voxelization process applied to the dataset. Voxelization effectively reduces the spatial resolution of the data, which in turn impacts the granularity of the analysis. While this process does offer the advantage of noise reduction through averaging signals within voxels, it nonetheless constrains the detail achievable in modeling neural connectivity. After voxelization, the analysis was conducted on 1,544 interior brain voxels, which, while substantial, is significantly reduced from the original dataset. Voxelization served to address a limitation stemming from the developmental model, which requires substantial memory resources. The developmental states at each timestep must be stored explicitly to facilitate gradient backpropagation, significantly increasing memory demands.

One possible avenue for reducing these memory requirements involves borrowing techniques from reinforcement learning—particularly those used in Proximal Policy Optimization (PPO). In PPO and related policy gradient methods, it is common to use partial rollout windows, or truncated sequences, to approximate gradients while avoiding the need to store entire trajectories. A similar approach could be applied to the developmental model by training on shorter subsequences of the developmental rollout, or by selectively storing a sparse set of intermediate states. This would relax the requirement to retain the full developmental history for backpropagation and could enable scaling to higher-resolution voxelizations. Exploring such hybrid methods, which blend ideas from reinforcement learning and morphogenetic modeling, may offer a path forward for handling larger and more complex developmental simulations.

Future research is poised to scale up these analyses substantially. Addressing the technical limitations outlined above will unlock significant potential for modeling large-scale connectomic datasets. Connectomics is a rapidly advancing field, with increasingly comprehensive datasets expected in the near future. Anticipated devel-

opments and emerging methodologies in large-scale connectomics will be discussed in greater detail in Section 7.2.

7.2 The Role of Connectome Data

This dissertation implemented a generative model of connectivity data. Like other generative models used in goal-driven deep learning (Chapter 2.3.2), such generative models benefit from large archives of data, which offer the models richer patterns to encode. This dissertation relied on the MBCA (see Chapter 3.1.2), which offers a comprehensive coverage of the adult mouse brain collected via hundreds of viral tracer experiments (Oh et al., 2014). While the data collected is invaluable, the experiments are very costly, as a unique animal subject is used in each experiment to avoid overlapping signal, and data must be laboriously registered to the CCF. This section highlights recent technological advances that may enable more efficient collection of data, paving the way for more sophisticated descriptive generative models, among other uses.

7.2.1 Whole-Brain Connectomes Across Species

While current scanning technology (electron microscopy) is expensive, a recent landmark project concluded in 2024 with a complete connectomic map of an adult fruit fly, comprising 140,000 neurons and 50 million synaptic connections (Dorkenwald et al., 2024). This stands as the largest fully mapped brain wiring diagram to date, far surpassing the *C. elegans* roundworm connectome of 302 neurons. The international FlyWire Consortium of researchers combined AI with painstaking manual validation to analyze over 100 terabytes of electron microscopy (EM) data to produce the circuit diagram. The previous record, set in 2023, collected the connectomic map of a larval fruit fly, consisting 3,016 neurons and 548,000 synapses (Winding et al., 2023). These studies reflect the rapid rate of progress and investment in large-scale connectomic mapping.

However, the challenge also scales exponentially progressing to more complex brains — from $\sim 10^2$ of neurons (worm) to $\sim 10^3$ (fly larva), $\sim 10^5$ (adult fly), $\sim 10^8$ (mouse), and finally $\sim 10^{11}$ (human) — marking the the next frontier in neuron-level connectomics. One notable project, called MICrONS, focused on a cubic millimeter of mouse cortex (visual cortex), which itself contains 75,000 neurons and about 500 million synapses (The MICrONS Consortium et al., 2023). While still too expensive for single-neuron mapping, the study, which utilized electron microscopy with AI analysis, yielded a nanometer-resolution voxel-based map, the most detailed look at

mammalian circuitry to date. A 2023 analysis by the Wellcome Trust estimated that with current technologies (EM), the cost of a neuron-scale mouse connectome is $7.5-21.7 billion and would take 10-15 years. This is due to the high cost of EM methods, which require special infrastructure — vibration-free, temperature-controlled, and electromagnetically shielded rooms, in addition to the upfront cost and operation of the machines. It would take about 20 scanning electron microscopes, operating around the clock, about five years to complete the project. For human brain tissue, a full-brain map is out of the question, but in 2024 a small cube of tissue one millimeter wide was analyzed with EM and AI, resulting in a map with 150 million synapses (Shapson-Coe et al., 2024). This study demonstrated a powerful AI system which processed the imaging data.

There is immense focus on high-resolution connectomics, which is now made possible with automated processing of large EM imaging datasets to produce more meaningful (but still large-scale) maps of neuronal wiring. This dissertation unveiled a way to make sense of connectomic data with a developmental framework. Future advances promise to enable high-fidelity reconstruction using a developmental program, which not only serves as a simpler explanation of the observed data but also provides a foundation for practical applications, discussed in Section 7.4.

7.2.2 High-throughput Brain Mapping Techniques

A key driver of today's connectomics boom is technological innovation that has vastly accelerated data collection. High-speed, automated electron microscopy (EM) is central — for example, in 2020 the fly hemibrain project used focused ion beam scanning EM (FIB-SEM) to image half a fly brain at 8 nm voxel resolution over months of continuous operation, the largest EM data collection at the time (Xu et al., 2020). Similarly, automated tape-collecting ultramicrotome systems (ATUM-SEM) and multi-beam SEM allow rapid imaging of large voluumes. These advances mean researchers can now aquire petabyte-scale datasets with synapse-level detail.

Another revolution has come from machine learning. Modern deep learning algorithms handle the otherwise intractable task of tracing each neurite and synapse in EM images. Automated segmentation tools label neurons in the dense EM data, yielding draft connectomes that are later refined by human experts.

Beyond EM, molecular mapping techniques are emerging. DNA-barcoding methods, for instance, can map long-range connections by sequencing rather than microscopy. MAPseq (Multiplexed Analysis of Projections by Sequencing) first demonstrated that individual neurons could be "tagged" with unique RNA barcodes to trace where they project in the brain (Kebschull et al., 2016). A more recent in-

novation, BRICseq, extends this idea to achieve brain-wide individual connectome sequencing (Huang et al., 2020). BRICseq uses DNA barcodes at both source and target sites, building upon the single-source technique of MAPseq, to enable neuron-to-neuron connectivity mapping across the whole mouse brain in single animals. This technique enables rapid and cost-effective wiring diagrams without needing hundreds of anatomical tracings.

These molecular mapping techniques build upon light microscopy with cell tracing techniques used in several large-scale projects. For example, Janelia's MouseLight project and the Allen Institute's cell atlases reconstruct the shapes of thousands of individual neurons across the mouse brain using fluorescent labeling and light-sheet microscopes. While not synapse-level, these give wiring blueprints at the single-cell level for long-range connections. Additionally, advances in tissue clearing and expansion microscopy promise to make larger volumes optically transparent and physically enlarged, so that high-resolution imaging of circuits could be done with light microscopes instead of EM in the future (Chen et al., 2015).

Rapid technological advancement in connectomics has enabled orders-of-magnitude growth in the size of feasible datasets. The ongoing frontier includes innovations to EM methods as well as barcoded light microscopy. These advancements foretell a promising future for downstream analysis such as the developmental models proposed in this dissertation.

7.3 Brain-like Functional Architectures

How can machines be made to think like people? In recent years, large language models (LLMs) have revolutionized the field of general intelligence. Despite this, there is still a gap to close to human-level intelligence (AGI). This section proposes that connectomic datasets provide a useful guiding signal toward human-like intelligence, and developmental models provide a solution space that, crucially, can iteratively complexify in a manner mimicking biological evolution, thus opening a path to functional human-like architectures. This section will compare the approach to current paradigms in AGI and then propose the method of phylogenetic refinement toward brain-like functional architectures.

7.3.1 Cognitive, Emergent, and Connectomic Approaches to AGI

How to close the cognitive gap between LLMs and humans? One method seeks to build intelligence by hand-crafting architectures or modules analogous to human cognitive faculties — called here the "cognitive" approach. Another strategy, the "emergent" approach, posits that general intelligence can arise from sufficiently rich data environments and learning algorithms without predefined modules. The first strategy models human cognition, while the second strategy models the tasks that give rise to human-like cognition. The current proposal is a third path, in which developmental models of wiring link functional architectures to biological connectomic data, thus providing a useful training signal to search for architectures that not only solve computational tasks, but also resemble the connective structure of biological networks. By grounding the search for general intelligence in the connectomics of intelligent species, there is a path to machines that think like humans that relies on fewer a priori assumptions about the nature of human cognition or human evolutionary environment (ethology). Here we describe the cognitive and emergent approaches.

The cognitive camp has produced numerous AGI blueprints in the past five years. Marcus (2020), for example, argued for a hybrid AI that integrates human-like knowledge and reasoning via explicit cognitive models, and Rathi (2022) similarly concluded that *integrated cognitive architectures* — systems of interacting modules for different mental functions — offer the most plausible path to human-level intelligence. The MICrONS Consortium et al. (2023) extended this idea by identifying a comprehensive set of cognitive functions (memory, goal management, social reasoning, ethics, etc.) required for general intelligence and proposing a new architecture that incorporates these interrelated components. LeCun (2022) likewise outlined an architectural vision combining a predictive world model, hierarchical representations, and intrinsic motivation, enabling machines to learn and plan more like animals or humans via self-supervised objectives. Goertzel et al. (2023) advocated for an integrative design (OpenCog Hyperon) that spans symbolic reasoning and neural learning, arguing that only a system with sufficient breadth of built-in cognitive capabilities and modules can achieve AGI.

By contrast, the emergent camp emphasized learning and environment over innate structure, arguing that broad data and experience can yield general intelligence through *emergent* behavior. Silver et al. (2021) represented this view by hypothesizing that reward-based reinforcement learning is enough" to drive an agent to acquire all the abilities associated with intelligence, without the need for hand-coded

sub-skills. Clune (2020) similarly proposed AI-generating algorithms that meta-learn their own architectures and learning strategies while creating their own training environments, an open-ended approach inspired by biological evolution to ultimately produce AI. In the same spirit, Hughes et al. (2024) argued that endlessly self-improving agents — for example, large pre-trained models continually exploring novel tasks and data — are essential to eventually achieve superhuman intelligence. The outstanding success of LLMs, which produce seemingly intelligent behavior derived from a large text corpus, is another trend, and recent studies even reported sparks of AGI" in GPT-4 (Bubeck et al., 2023).

This work proposes a connectomic approach. Like in the emergent approach, open-ended environments or AI-generating algorithms produce tasks that drive the learning architecture to emergently acquire cognitive abilities. However, an additional training signal is provided, which has to do with the connective architecture, and which pulls architectures to more closely resemble connectomic data. The connectomic signal can be applied in an outer training loop, thus tuning the task parameters until the corresponding architecture resembles biology. In this way, not only are assumptions removed about the nature of cognition, but also fewer assumptions are made about the nature of the tasks that generate AI, since the task itself is selected algorithmically.

7.3.2 Proposed Approach: Connectome-Generating AI-Generating Algorithms

The approach proposed here combines two fundamental biological processes—evolution and development—as a means of constructing artificial general intelligence (AGI), mirroring the processes by which intelligence emerged in nature. This section develops two lines of argument. The first is that leveraging connectomic data provides a grounded, biologically informed approach to constructing AGI, in contrast to frameworks rooted in abstract cognitive or behavioral modeling. The second is that developmental models offer a powerful mechanism for efficiently exploring the space of neural architectures.

A central challenge in pursuing AGI is that its target remains ill-defined: the functional boundaries of general intelligence—exemplified only by the human brain—are still poorly understood. Grounding AGI efforts in connectomic data offers a way to sidestep this ambiguity by anchoring development in the structure of real biological systems. As discussed in the previous section, cognitive science has not yet produced a definitive account of the full set of components required for general intelligence, leaving cognitive-modeling approaches to AGI fundamentally underspecified. Mean-

while, AI-generating algorithms circumvent the need to explicitly define a cognitive model, but in doing so, they place the burden on the design of an environment capable of producing general intelligence—that is, they rely on an implicit ethological model. Consequently, any attempt to engineer AGI risks omitting essential capacities—such as empathy, social reasoning, or embodied perception—simply because they are not yet fully formalized or represented in benchmark tasks.

A more grounded approach involves modeling not just the behaviors of the human mind, but its underlying biological substrate. Extensive empirical data exists on the brain's connective architecture. If a model, shaped solely by environmental demands, independently converges on this structure—a statistically unlikely outcome—there is strong reason to believe it is approximating the functional dynamics of biological intelligence.

The motivation for incorporating development alongside evolution arises from the limitations of evolving neural connectivity directly. By shifting the focus from final connections to the rules that generate them, developmental models enable the reuse of local subroutines across space. This mechanism allows for the construction of highly regular and complex architectures from compact parameterizations, as discussed in Chapter 6.1. This approach falls under the broader framework of indirect encoding, in which compact, rule-based genetic descriptions generate large, structured architectures by leveraging regularities such as symmetry, repetition, and gradient-based variation. As reviewed by Stanley et al. (2019), these encodings enable the expression of vast, high-dimensional networks from a relatively small number of parameters, making it possible to evolve scalable and functionally rich neural systems without hand-designing their connectivity.

In the developmental model in this dissertation, the wiring pattern emerges from the interactions of self-organizing agents that cooperate to form the network architecture. This creates a rich design space in which the system effectively learns an emergent multi-agent communication protocol—a powerful mechanism for constructing flexible, adaptive architectures. The results presented in this dissertation demonstrate that such an approach is viable in producing functionally meaningful structure.

The problem of human-level intelligence would be solved if a network could be constructed that matches the functional architecture of the human brain — but how does one *get there*? As in biology, the answer is a stepwise process. The most promising way to recapitulate the evolutionary trajectory of brain networks is to produce them in the same way biology does: through development. Biological complexity emerges incrementally, a process described by concepts such as phylogenetic inertia—the evolutionary retention of ancestral traits—and the theory of phylogenetic

refinement of Cisek (2019), which posits that complex neural circuits arise through iterative elaboration of simpler, functional precursors.

This principle—that functional complexity emerges through incremental refinement—has an analogue in software engineering, where Gall's Law states that "a complex system that works is invariably found to have evolved from a simple system that worked" (Gall, 1977). Therefore, if the goal is to construct a machine learning architecture capable of solving problems at the level of complexity seen in human cognition, then it must be built through a stepwise process. Developmental systems offer the most promising pathway toward this end, as they naturally articulate structural motifs observed in neurobiology. By introducing complexity gradually—through the same kinds of generative mechanisms used in biological development—such systems are well positioned to recapitulate the functional architecture of the brain through an iteratively complexifying process. As demonstrated in Chapter 5, the developmental model employed in this dissertation exhibits key similarities to biological development, suggesting that it captures essential features of the processes by which brain connectivity emerges. If the goal is to construct systems that mirror the architecture of the brain, then they must be built the way the brain is built—through a developmental process that introduces complexity gradually and in biologically plausible ways.

This proposal builds on the concept of AI-generating algorithms (AI-GAs) (Clune, 2020), which use meta-learning to evolve intelligent systems through interaction with an environment. But rather than generating intelligence in an unconstrained emergent process, the approach introduced here explicitly targets the emergence of connectome-recapitulating architectures. The result is a connectome-generating AI-generating algorithm (CGAIGA): a system designed not only to produce intelligent agents, but to do so in a way that mirrors the structural logic of the biological brain. By "generating the environment from the connectome," the framework seeks to discover environmental conditions under which the emergence of a particular neural structure becomes a likely outcome of the system's adaptation—effectively treating the connectome as a constraint on what kinds of tasks the agent must solve.

This is implemented as a nested-loop optimization process: the inner loop trains the agent that develops within a given environment, while the outer loop adjusts the environment itself to increase the likelihood that the resulting architecture matches biological connectomic data. The system begins with minimal, functional architectures—analogous to primitive neural circuits—and expands their complexity over time, in a process that mirrors both ontogeny (development within an individual) and phylogeny (evolution across generations). Each stage of structural growth is shaped by task-based performance signals, ensuring that complexity emerges not

from arbitrary elaboration or overfitting to a biological template, but from meaningful adaptation to environmental demands. In doing so, this approach offers a computational model for how intelligence may have emerged in natural systems.

7.4 Regenerative Medicine

Harnessing the power of self-assembling cells is an exciting avenue for technological progress. Besides producing better AI systems, developmental models are a useful foothold and computational aid in medicine and in the engineering of physical computers, devices, and materials.

A major goal is to model develop and regeneration processes in silico, so that we can eventually guide or recapitulate them for clinical purposes, such as limb regeneration or organ repair (Pezzulo and Levin, 2015). This field, regenerative medicine, draws on decades of theoretical biology combined with machine learning to tackle the "big picture" of how cells collectively build and rebuild anatomical structures.

Early work studied computational models of regeneration in highly regenerative organisms. For example, in the 1970s, theoretical biologists like Gierer formulated models of positional information and reaction-diffusion that explained phenomena such as how a Hydra polyp can regenerate a head at the correct end of a fragment (Birnbaum and Alvarado, 2008).

Planarian flatworms, long famous for the ability to regrow any cut part, were the subject of a more recent computational model (Lobo and Levin, 2015). An evolutionary algorithm optimized a gene network in a "virtual worm" so that its simulated regeneration matched published experimental results. The resulting model — essentially a set of rules for how cells decide what to rebuild — highlights how AI can aid hypothesis discovery in developmental biology.

A core trend is using such models to identify intervention points – in other words, guiding the search for treatments that can induce or enhance regeneration. Lagasse and Levin (2023) dubbed the future interventions "morphoceuticals": drugs or stimuli that target the collective pattern-regulating mechanisms, rather than a single molecular pathway. The idea is that by nudging the higher-level control systems (like bioelectric circuits or morphogen gradients) one could trigger the body to regrow complex structures on its own, effectively telling tissues what to build.

Regenerative medicine thus moves from descriptive models (showing which gene affects which) to predictive, constructive models that can be "run" like a recipe to produce anatomic structures. Influential early works set the stage by asking how patterns emerge in development, and now modern computational power and biolog-

ical data are converging to finally decode those instructions and thus illuminating how to engineer regeneration.

7.5 Self-assembly of Biological Machines

Beyond their relevance to brain development, models of biological self-assembly offer promising directions for engineering novel forms of intelligent, adaptive machines. In contrast to conventional systems built from fixed components and top-down instructions, these systems harness the innate capabilities of living cells—such as growth, repair, and coordination—to build functional structures from the bottom up. This section explores how principles of developmental biology and morphogenesis can inform the design of such systems, beginning with the role of self-organization, and extending to the construction of large-scale grown structures and computing devices assembled from living neurons.

7.5.1 Self-organization

Models of biological self-assembly can map the complex architecture of the human body, but they also have use outside the body in engineering useful bio-hybrid machines. Nature demonstrates that complex, adaptive systems can arise without central control. In swarms of insects or cells, numerous simple agents following local rules can self-organize into sophisticated structures with collective intelligence (Risi, 2021). For example, ants form bridges and bees make group decisions with no master planner. Such systems are highly robust—they adjust to new conditions, heal damage, and continue functioning even if some components fail. These traits motivate self-assembled biological machines: devices built from living cells that leverage cells' innate behaviors (such as self-repair and adaptation) to achieve resilience beyond that of rigid, engineered robots.

Levin's "Darwin's Agential Materials" (Levin, 2023) articulated why living cells are powerful building blocks for machines. Evolution does not work with passive Lego-like parts; cells have their own agendas and problem-solving competencies, inherited from unicellular ancestors. Tissues and organs thus form a multiscale competency architecture—they can regulate and adjust their growth or function to reach a goal form even when perturbed. This built-in intelligence helps explain the speed and robustness of biological evolution, as cellular collectives can creatively solve morphogenetic problems on the fly. Applying this to engineering, a self-assembling biomachine made of such "agential" cells could likewise handle unexpected challenges, self-correct, or find novel solutions without needing every scenario pre-programmed.

Researchers note that the traditional line between "machine" and "organism" is blurring: designed living robots have shown surprising, emergent behaviors that prompt a rethinking of what machines can be. Rather than simple, rigid automatons, they can exhibit novelty, intelligence, and self-directed actions arising from the collective behavior of cells.

7.5.2 Grown Structures

Instead of fabricating every component, scientists are exploring how to *grow* biomechanical structures the way organisms grow tissues or organs. A recent initiative by DARPA openly imagines "large, self-assembled, mechanically stable biological growths in space" that could serve as useful structures (Nayak, 2025). For example, rather than launching a heavy antenna or space habitat from Earth, we might seed cells or engineered organisms in microgravity and have them grow a tether, a large net, or a space station wing *in situ*. This approach could radically cut costs and enable structures over 500 m long, positioning biology as a key component of future in-space construction. On Earth, engineered living materials (ELM) represent a growing area of research focused on developing building materials that can grow on demand, self-heal, and adapt to their environment (Wang et al., 2022). These materials are envisioned as programmable biological systems—such as engineered bacteria, fungi, or plant cells—that can be shipped as precursors and then cultivated in situ into walls, coatings, or structural components. Early examples include mycelium-based packing foams and bacterial sand bricks (Heveran et al., 2024). The goal is to preserve life's dynamic abilities—such as self-repair and environmental responsiveness—in the final material. Such capabilities could give rise to "smart" infrastructure, including buildings that seal their own cracks or bridges that reinforce themselves under stress, much like bone or tree wood.

Researchers are also pioneering synthetic morphogenesis, essentially bioengineering new multicellular life forms by guilding cell growth. Instead of sticking to evolved body plans, the idea is to grow novel living structures (or "proto-organs") from cells, creating useful biohybrid devices. Ebrahimkhani and Levin (2021) highlight advances in this field, emphasizing how emergent self-organization can be steered to produce coherent anatomies with active functions. These "synthetic living machines" include everything from biohybrid robots (muscle tissues on scaffolds) to entirely scaffold-free living structures like xenobots. Xenobots are a prime example of a grown biomechanical structure: they are neither traditional robots nor natural organisms, but clusters of frog cells designed by an evolutionary algorithm and then assembled by the cells themselves into a functional structure (Kriegman et al., 2021).

7.5.3 Grown Computers (Neuron-based Machines)

Another frontier of self-assembled biological machines is computing devices made of living neurons. In this paradigm, networks of neurons (often grown into 3D brain organoids or arranged on electrode arrays) serve as hardware for processing information. The emerging field of organoid intelligence (OI) aims to develop biocomputers that harness brain cells' computing capabilities (Morales Pantoja et al., 2023). By tapping into biological neural networks, researchers hope to build computers that are more efficient, adaptive, and capable of generalization than today's silicon-based AI.

Research on neuron-based computing spans from 2D cultures to complex organoids. A landmark study in 2022 showed that a layer of living neurons can learn to play the video game *Pong* (Kagan et al., 2022). For that setup, called "DishBrain," scientists connected 800,000 rat and human neurons grown on a microelectrode array to a *Pong* simulation. Signals from the electrodes provided sensory information, telling the neurons where the ball was, and the neuronal activity controlled the paddle in the game. The neurons, initially placed randomly on the array, self-organized to successfully hit the ball, essentially learning over time through feedback. This was the first demonstration of a goal-directed task in a synthetic neural network. This study was a proof-of-concept that live neural circuits can perform computations and adapt based on experience. In three dimensions, brain organoids add architectural complexity that more closely resembles biology. Researchers have successfully grown organoids the size of peas (50,000 cells), and they plan to assemble clusters of organoids to form a biological computer. As advances in biological computing blur the line between engineered systems and living tissue, they also raise pressing ethical questions—particularly about the use of biological material and the development of increasingly brain-like artificial intelligence. These concerns are the focus of the following section.

7.6 Ethics

Computational modeling of development sits at the intersection of biology and artificial intelligence (AI). This dissertation draws on biological data—such as gene expression and connectomics—and applies machine learning tools to model neural development. It also proposes developmental modeling as a route to more brain-like AI systems. These approaches offer powerful insights and open the door to new technological capabilities. However, with this potential come ethical considerations. This section examines the implications of working with sensitive biological data and

developing systems that increasingly resemble human cognitive processes. Two areas are addressed in particular: the ethical sourcing and use of biological data, and the broader societal impacts of brain-inspired AI.

7.6.1 Ethical Use of Biological Data

Generative models of development rely heavily on biological datasets (e.g., connectomic maps, genomic data) obtained from animal models and humans. It is essential that such data be sourced in an ethical fashion. Research on animal brains is subject to stringent animal welfare standards. Internationally, scientists adhere to the "3Rs" principle—Replace, Reduce, Refine—to minimize animal use and suffering (Hubrecht and Carter, 2019). That is, when possible, researchers seek alternatives to animal experiments, use the fewest necessary animals, and refine procedures to cause the least harm. Institutional Animal Care and Use Committees (IACUCs) or their equivalents review experimental protocols to ensure that the expected scientific knowledge justifies any animal sacrifice or invasive procedure. For human-derived data, ethical oversight is especially rigorous: donors and participants must give informed consent, and protocols must comply with frameworks like the Common Rule in the United States or the GDPR in Europe. Institutional review boards require that any collection of human brain data (e.g., MRI scans, post-mortem tissue, stem-cell derived organoids) uphold participants' rights and privacy (Greely et al., 2018a). The datasets used in this dissertation—including mouse brain connectomic data and gene expression atlases—are sourced from publicly available repositories established under these ethical guidelines. Such measures ensure that ethical integrity is maintained throughout the research process, from biological data collection to computational modeling.

After biological data is collected, additional challenges arise regarding its responsible use and sharing. Large-scale neurodata repositories, such as those developed in connectomics projects, exemplify the need to balance privacy concerns with open science (Jwa and Poldrack, 2022). On one hand, sharing data accelerates discovery and is viewed as an ethical duty to maximize the contribution of research participants. On the other hand, brain data can be deeply personal and identifying. Neuroimaging datasets, for example, often contain structural features that can reveal a person's identity; recent studies showed that facial images can be algorithmically reconstructed even from MRI scans that have been "defaced" to remove identifiable facial anatomy (Jwa and Poldrack, 2022). Accordingly, data repositories implement strict de-identification procedures and data use agreements that prohibit re-identification attempts. The Human Connectome Project, for instance, elected

to share only de-identified MRI data to protect participant anonymity (Elam et al., 2021). Yet, as machine learning techniques advance, the risk of re-identification or unintended inferences from neural data remains a moving target (Jwa and Poldrack, 2022). Researchers must therefore treat shared neurodata with caution and recognize that protecting participant privacy is a dynamic and ongoing responsibility. In this dissertation, all analyses were conducted using public datasets governed by such safeguards.

There is also a broader consideration of data governance and ownership regarding biological data. Brain data might be seen as part of an individual's identity, raising the issue of who controls its use. Some ethicists advocate for "neurorights" to ensure protections specific to the brain and mind, including rights to: freedom of thought, or protection from external interference; mental privacy, or protection of personal information; and mental integrity, or protection from harm (Ienca, 2021). These proposed rights are part of a recognition that data derived from the brain deserves special safeguarding due to its intimate link with personal identity and thought. In practice, policy frameworks like the NIH BRAIN Initiative's neuroethics guidelines (Greely et al., 2018b) explicitly state as a guiding principle to "protect the privacy and confidentiality of neural data" and to "attend to possible malign uses of neuroscience tools and neurotechnologies." Adhering to such codes of conduct throughout the data lifecycle is critical to uphold public trust and to honor the biological contributors to the data. These concerns are especially relevant for this dissertation, which draws from neural datasets and proposes new modeling approaches—underscoring the need for rigorous, ethically grounded research practices.

Just as biological data must be ethically sourced and handled, there are also important ethical considerations surrounding the application targets of this work in artificial intelligence, as will be discussed in the next section.

7.6.2 The Ethics of Brain-Inspired AI

Brain-inspired AI techniques, such as those unlocked by developmental models, are an emerging technology within the rapidly progressing front of AI innovation. The implications of this emerging technology on social systems requires careful consideration as AI technology is adopted in increasingly many facets of life.

One of the core challenges posed by brain-inspired AI is its inherent unpredictability—especially in systems designed to emulate aspects of human cognition, such as the connectomic-grounded models explored in this dissertation. If such generalist agents are realized, they would have potentially many uses cases due to their capacity for continual learning and contextual understanding (Farisco et al., 2024).

However, they may also be less transparent and harder to control than small-scope agents. Unpredictability may in fact be a hallmark feature of general intelligent agents. Traditional AI systems, such as reinforcement learning agents trained to play games or optimize specific tasks, operate within well-defined goals. By contrast, general AI, such as that emerging from training in open-ended environments, may not have a singular built in-goal, leading researchers to instead consider the social values or principles baked into these models in an effort to *align* them with human interests (Russell, 2019).

While the behavior of brain-inspired AI systems may be unpredictable, one outcome is clear: AI will significantly reshape the nature of human labor (Raman et al., 2025). Recent economic estimates suggest that up to 40% of current jobs globally could be affected by AI-driven automation (Georgieva, 2024). Unlike earlier automation waves that primarily displaced routine manual labor, advanced AI—especially flexible, brain-like systems—may extend automation into cognitive and decision-making roles. This shift poses both opportunities and challenges, as AI begins to encroach on professions once thought to require uniquely human intellect (Broady et al., 2025). Some scholars argue that AI reduces dependence on human labor, concentrates productivity gains, and increases inequality by enhancing the bargaining power of capital holders (Berg et al., 2016). While the long-term effects on labor remain contested, there is broad agreement that proactive governance—through re-skilling initiatives, economic policy, and labor protections—will be essential to manage the transition (Shen and Zhang, 2024; Virgilio et al., 2024).

It is essential that AI systems benefit humanity as a whole and not only those who control the technology. Therefore, access to AI becomes a justice issue, as differential access across different groups and regions could entrench social inequalities. An IMF analysis warns that AI could widen the gap between rich and poor nations by funneling investment and productivity gains primarily to advanced economies already at the forefront (Alonso et al., 2020). Recognizing these disparities, international bodies like UNESCO have called for fair governance frameworks, as reflected in their "Recommendation on the Ethics of Artificial Intelligence" (UNESCO, 2022). Equally important is the issue of representation: If AI systems are designed and controlled by a narrow group of people, they are more likely to reflect the values and assumptions of that group—risking unfair outcomes and excluding people from different backgrounds or experiences. Such embedded values are especially important to consider when designing agents through the emergent environments proposed in this dissertation. If a generalist AI system is trained in the context of a military strategy game like *StarCraft*, for example, what kinds of values, assumptions, or conflict-driven reasoning might emerge—and how might that shape its behavior in

broader contexts?

As AI systems become more brain-like or human-like ethical questions arise about the treatment of the agents themselves. Some ethicists argue that if an artificial system exhibits a "will to live," even in a minimal form, it may warrant some degree of moral status (Coghlan and and Leins, 2020). These concerns are amplified as AI systems grow more complex, autonomous, and biologically realistic. Philosophical debates about consciousness, sentience, and agency take on new urgency when applied to systems that mimic neural computation. There is growing interest in whether artificial agents may eventually deserve ethical consideration—either due to cognitive sophistication or the potential capacity for suffering. While no current system is believed to be sentient, precautionary principles and early rights frameworks are beginning to take shape in the AI ethics discourse.

As developmental models begin to reflect biological processes more closely, this work raises the question of where ethical limits should be placed in the design of future machine intelligence. While such long-term concerns are relevant, more immediate issues, including labor displacement, are already emerging alongside the rapid expansion of AI capabilities. The models developed here, which move AI toward brain-like and autonomous behavior, prompt careful attention to their potential role in shaping social and economic outcomes.

7.7 Conclusion

The preceding sections have outlined both the promise and the challenges of using developmental models to understand and engineer complex systems. While the models developed in this dissertation provide a compelling framework for explaining brain connectivity, they are simplifications of biological reality, and extending them with richer biological detail may yield deeper insights and more powerful predictive tools. As connectomic datasets grow in scale and resolution—particularly in higher organisms—they enable more expressive developmental models that better reflect the complexity of biological wiring. These models not only offer insights into the structure of the brain; they also open new pathways toward regenerative medicine, bio-hybrid machines, and brain-inspired AI architectures. Each of these domains carries ethical considerations that must be addressed with care. Developmental modeling is thus not only a tool for scientific understanding, but also a foundation for technologies with far-reaching societal impact.

Chapter 8

Conclusion

This dissertation applies a generative modeling framework to a developmental biology dataset that captures detailed measurements of genetic expression and neural connectivity in the mouse brain. The central task is to reconstruct connectivity using a generative model trained without access to gene expression data. Then, following techniques from neural representation studies, the learned latent variables are compared to gene expression patterns to investigate how genetic factors encode wiring. A core contribution of this work is the introduction of a biologically grounded developmental model—one that generates connectivity through a sequential, spatially localized process. This developmental perspective offers a more faithful account of how genetic programs unfold over time to shape brain structure.

8.1 Contributions

Chapter 2 established the foundation for the dissertation by forming a synthesis of three core areas: statistical connectomic models, developmental biology, and computational neuroscience. It reviewed prior work that identified statistical correlations between genetic expression and brain connectivity, highlighting the transition from direct supervised mapping to more expressive latent variable approaches. Novel methodologies introduced in this chapter extracted spatial "barcodes" that served as latent representations, offering a richer explanation of how genetic coding might underlie neural wiring. Building on this representational view, the chapter then connected classical theories of morphogenesis—such as Turing's reaction-diffusion model and the concept of morphogenetic fields—with the deep learning tools used in modern computational neuroscience. This connection motivated the use of developmental models not only as biologically inspired architectures, but as mechanisms

for learning structured representations under spatial and temporal constraints, laying the groundwork for their application to mouse brain data in the chapters that followed.

Chapter 3 established the computational framework to explore how genetic expression gave rise to brain connectivity. The methodology began by detailing the integration of connectomic and transcriptomic biological data—from the Allen Brain Atlas and Mouse Brain Connectivity Atlas—aligned to a common coordinate system. This data was then pre-processed: gene expression was transformed into an expression energy measure and voxelized, while connectivity data from viral tracer experiments was log-scaled and thresholded to ensure balanced analysis.

Two complementary modeling architectures were then introduced to explore how genetic expression might have encoded neural connectivity. The static model employed latent variable techniques, inspired by singular value decomposition, to derive spatial "barcodes" that predicted connectivity patterns while incorporating biologically motivated cost constraints. Building on this foundation, the developmental model extended the framework to a lower level of biological realism by simulating a sequential, locally interacting process. Drawing on principles of morphogenesis, it generated latent representations through an iterative growth mechanism implemented using neural cellular automata. To assess how well these learned representations reflected biological reality, a suite of evaluation metrics was introduced—quantifying both how accurately the models reconstructed observed connectivity and how closely their latent variables aligned with measured gene expression. These included one-to-one gene correlations, cross-validated gene scores, and graph-level metrics such as mean-squared error and variance explained.

Chapter 4 demonstrated that gene expression drove mouse brain connectivity by showing that latent variables learned from actual connectivity data correlated robustly with genetic expression. Randomized connectivity baselines confirmed that these associations arose from genuine wiring patterns rather than generic spatial factors. Testing different latent priors revealed that overall correlations remained stable, with no significant differences observed between regularization strategies. Expanding on this search for coding principles, an analysis of model complexity showed that an optimal, intermediate level of latent dimensionality best captured meaningful connectivity features without overfitting noise—suggesting that a relatively simple, low-dimensional developmental program might have underlain the genetic specification of brain wiring.

Chapter 5 extended the static latent framework by introducing a developmental model that simulated a sequential, locally constrained process—akin to morphogenesis—to explain the emergence of gene expression patterns in the mouse brain. While

the static model captured high-level coding relationships, the developmental model operated at a lower level of biological detail, providing insight into how such coding might have arisen through local interactions and patterning dynamics. By incorporating additional guidance signals into the latents, the developmental model captured disambiguating information that the static model abstracted away, as demonstrated by toy denoising experiments and complexity analyses. Although this model sacrificed some connectivity reconstruction accuracy, it substantially improved prediction of genetic expression, suggesting that developmental signals—such as morphogen-like gradients—were reflected in the gene expression data itself. These findings supported the view that gene expression not only encoded wiring but also guided the formation of region-specific identities through structured, sequential processes.

Chapter 6 demonstrated that developmental encodings offered distinct advantages in modeling how structure could emerge and adapt over time. By guiding connectivity adjustments in a developmental program, a small set of parameters could induce large-scale, structured changes—offering a compact, process-based representation of neural architecture. While static models described the end-state of connectivity, developmental models indirectly specified patterned connectivity in an efficient manner through learned local rules. In addition to efficiency, their ability to respond to simulated damage—by rebuilding missing parts of the network through local interactions—highlighted their potential for capturing resilience and adaptability in neural systems.

Chapter 7 discussed the limitations of the current approach and outlined future directions. The work relied on static snapshots and simplified assumptions that did not fully capture the dynamic, mechanical, and temporal aspects of neural development. The discussion placed these limitations in the broader context of rapidly advancing connectomics—from whole-brain maps across species to high-throughput imaging—and argued that next-generation models should better reflect how development actually unfolded over time through local interactions and feedback. Moreover, the chapter explored the potential real-world impact of these models, ranging from brain-like functional architectures for AGI to applications in regenerative medicine and the self-assembly of biological machines, highlighting how a deeper understanding of developmental principles could transform both computational modeling and practical technology.

8.2 Implications

The findings in this dissertation advance the development of self-organizing models. A cellular automata model was successfully employed to construct a complex

connectivity map that not only captures neural wiring but also demonstrates superior predictive power for genetic expression compared to a static latent model. This work establishes a foundation for applying generative developmental models to increasingly complex connectomic datasets, thereby offering deeper insights into the underlying processes of biological development.

In the longer term, these models hold significant promise for applications in medicine and engineering. For example, the ability to simulate a developing brain may facilitate the exploration of novel treatment strategies and regenerative interventions. Furthermore, the development of self-organizing bio-mechanical devices—capable of autonomous assembly and repair—could transform traditional design and manufacturing processes.

These advances also have implications for the field of artificial intelligence. By providing a developmental perspective on neuronal wiring, the work contributes to a better understanding of the constraints and organizational principles that shape brain structure and function, which are critical to addressing fundamental questions regarding human-like intelligence. As introduced in the opening discussion using the tree analogy, the developmental model now serves as a lens through which complex biological structures can be understood—revealing how intricate and seemingly intractable forms are, in fact, the outcome of systematic, local developmental processes. This methodological innovation, powered by machine learning, represents a scalable approach to decoding the intricacies of biological development and may inform future advances across multiple scientific and technological domains.

Bibliography

Cristian Alonso, Siddharth Kothari, and Rehman. How Artificial Intelligence Could Widen the Gap Between Rich and Poor Nations. https://www.imf.org/en/Blogs/Articles/2020/12/02/blog-how-artificial-intelligence-could-widen-the-gap-between-rich-and-poor-nations, December 2020.

Victor Ambros. Control of developmental timing in *Caenorhabditis elegans*. *Current Opinion in Genetics & Development*, 10(4):428–433, August 2000. ISSN 0959-437X. doi: 10.1016/S0959-437X(00)00108-8.

Aurina Arnatkeviciute, Ben D. Fulcher, Mark A. Bellgrove, and Alex Fornito. Where the genome meets the connectome: Understanding how genes shape human brain connectivity. *NeuroImage*, 244:118570, December 2021. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2021.118570.

Michael Ashburner, Catherine A. Ball, Judith A. Blake, David Botstein, Heather Butler, J. Michael Cherry, Allan P. Davis, Kara Dolinski, Selina S. Dwight, Janan T. Eppig, Midori A. Harris, David P. Hill, Laurie Issel-Tarver, Andrew Kasarskis, Suzanna Lewis, John C. Matese, Joel E. Richardson, Martin Ringwald, Gerald M. Rubin, and Gavin Sherlock. Gene Ontology: Tool for the unification of biology. *Nature Genetics*, 25(1):25–29, May 2000. ISSN 1546-1718. doi: 10.1038/75556.

R. E. Baker, S. Schnell, and P. K. Maini. A clock and wavefront mechanism for somite formation. *Developmental Biology*, 293(1):116–126, May 2006. ISSN 0012-1606. doi: 10.1016/j.ydbio.2006.01.018.

Andrea Banino, Caswell Barry, Benigno Uria, Charles Blundell, Timothy Lillicrap, Piotr Mirowski, Alexander Pritzel, Martin J. Chadwick, Thomas Degris, Joseph Modayil, Greg Wayne, Hubert Soyer, Fabio Viola, Brian Zhang, Ross Goroshin, Neil Rabinowitz, Razvan Pascanu, Charlie Beattie, Stig Petersen, Amir Sadik,

Stephen Gaffney, Helen King, Koray Kavukcuoglu, Demis Hassabis, Raia Hadsell, and Dharshan Kumaran. Vector-based navigation using grid-like representations in artificial agents. *Nature*, 557(7705):429–433, May 2018. ISSN 1476-4687. doi: 10.1038/s41586-018-0102-6.

Dániel L. Barabási and Albert-László Barabási. A Genetic Model of the Connectome. *Neuron*, 105(3):435–445.e5, February 2020. ISSN 0896-6273. doi: 10.1016/j.neuron.2019.10.031.

Andrew Berg, Edward F. Buffie, and Luis-Felipe Zanna. Robots, Growth, and Inequality. *Finance & Development*, 53(3), September 2016.

Kenneth D. Birnbaum and Alejandro Sánchez Alvarado. Slicing across Kingdoms: Regeneration in Plants and Animals. *Cell*, 132(4):697–710, February 2008. ISSN 0092-8674. doi: 10.1016/j.cell.2008.01.040.

Trenton Bricken, Adly Templeton, Joshua Batson, Brian Chen, Adam Jermyn, Tom Conerly, Nick Turner, Cem Anil, Carson Denison, Amanda Askell, Robert Lasenby, Yifan Wu, Shauna Kravec, Nicholas Schiefer, Tim Maxwell, Nicholas Joseph, Zac Hatfield-Dodds, Alex Tamkin, Karina Nguyen, Brayden McLean, Josiah E Burke, Tristan Hume, Shan Carter, Tom Henighan, and Christopher Olah. Towards monosemanticity: Decomposing language models with dictionary learning. *Transformer Circuits Thread*, 2023.

Kristen E. Broady, Darlene Booth-Bell, Anthony Barr, and Aisha Meeks. Automation, artificial intelligence, and job displacement in the U.S., 2019–22. *Labor History*, 0(0):1–17, March 2025. ISSN 0023-656X. doi: 10.1080/0023656X.2025.2477153.

Sébastien Bubeck, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg, Harsha Nori, Hamid Palangi, Marco Tulio Ribeiro, and Yi Zhang. Sparks of Artificial General Intelligence: Early experiments with GPT-4, April 2023.

Ed Bullmore and Olaf Sporns. The economy of brain network organization. *Nature Reviews Neuroscience*, 13(5):336–349, May 2012. ISSN 1471-0048. doi: 10.1038/nrn3214.

Charles F. Cadieu, Ha Hong, Daniel L. K. Yamins, Nicolas Pinto, Diego Ardila, Ethan A. Solomon, Najib J. Majaj, and James J. DiCarlo. Deep Neural Networks Rival the Representation of Primate IT Cortex for Core Visual Object Recognition.

PLOS Computational Biology, 10(12):e1003963, December 2014. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1003963.

Rafael Cantera, María José Ferreiro, Ana María Aransay, and Rosa Barrio. Global Gene Expression Shift during the Transition from Early Neural Development to Late Neuronal Differentiation in Drosophila melanogaster. *PLOS ONE*, 9(5): e97703, May 2014. ISSN 1932-6203. doi: 10.1371/journal.pone.0097703.

Charlotte Caucheteux, Alexandre Gramfort, and Jean-Rémi King. Deep language algorithms predict semantic comprehension from brain activity. *Scientific Reports*, 12(1):16327, September 2022. ISSN 2045-2322. doi: 10.1038/s41598-022-20460-9.

Fei Chen, Paul W. Tillberg, and Edward S. Boyden. Expansion microscopy. *Science*, 347(6221):543–548, January 2015. doi: 10.1126/science.1260088.

Nicolaj Strøyer Christophersen and Kristian Helin. Epigenetic control of embryonic stem cell fate. *The Journal of Experimental Medicine*, 207(11):2287–2295, October 2010. ISSN 0022-1007. doi: 10.1084/jem.20101438.

Paul Cisek. Resynthesizing behavior through phylogenetic refinement. *Attention, Perception, & Psychophysics*, June 2019. ISSN 1943-393X. doi: 10.3758/s13414-019-01760-1.

Jeff Clune. AI-GAs: AI-generating algorithms, an alternate paradigm for producing general artificial intelligence, February 2020.

Simon Coghlan and Kobi and Leins. "Living Robots": Ethical Questions About Xenobots. *The American Journal of Bioethics*, 20(5):W1–W3, May 2020. ISSN 1526-5161. doi: 10.1080/15265161.2020.1746102.

Christopher J. Cueva and Xue-Xin Wei. Emergence of grid-like representations by training recurrent neural networks to perform spatial localization. *arXiv:1803.07770 [cs, q-bio, stat]*, March 2018.

Peter Dayan and L. F. Abbott. *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. The MIT Press, 2005. ISBN 978-0-262-54185-5.

Barry J. Dickson. Molecular mechanisms of axon guidance. *Science (New York, N.Y.)*, 298(5600):1959–1964, December 2002. ISSN 1095-9203. doi: 10.1126/science.1072165.

Sven Dorkenwald, Arie Matsliah, Amy R. Sterling, Philipp Schlegel, Szi-chieh Yu, Claire E. McKellar, Albert Lin, Marta Costa, Katharina Eichler, Yijie Yin, Will Silversmith, Casey Schneider-Mizell, Chris S. Jordan, Derrick Brittain, Akhilesh Halageri, Kai Kuehner, Oluwaseun Ogedengbe, Ryan Morey, Jay Gager, Krzysztof Kruk, Eric Perlman, Runzhe Yang, David Deutsch, Doug Bland, Marissa Sorek, Ran Lu, Thomas Macrina, Kisuk Lee, J. Alexander Bae, Shang Mu, Barak Nehoran, Eric Mitchell, Sergiy Popovych, Jingpeng Wu, Zhen Jia, Manuel A. Castro, Nico Kemnitz, Dodam Ih, Alexander Shakeel Bates, Nils Eckstein, Jan Funke, Forrest Collman, Davi D. Bock, Gregory S. X. E. Jefferis, H. Sebastian Seung, and Mala Murthy. Neuronal wiring diagram of an adult brain. *Nature*, 634(8032): 124–138, October 2024. ISSN 1476-4687. doi: 10.1038/s41586-024-07558-y.

Timothy Dozat. Incorporating Nesterov Momentum into Adam. February 2016.

Mo R. Ebrahimkhani and Michael Levin. Synthetic living machines: A new window on life. *iScience*, 24(5):102505, May 2021. ISSN 2589-0042. doi: 10.1016/j.isci.2021.102505.

Carl Eckart and Gale Young. The Approximation of One Matrix by Another of Lower Rank. *Psychometrika*, 1(3):211–218, September 1936. ISSN 0033-3123, 1860-0980. doi: 10.1007/BF02288367.

Jennifer Stine Elam, Matthew F. Glasser, Michael P. Harms, Stamatios N. Sotiropoulos, Jesper L. R. Andersson, Gregory C. Burgess, Sandra W. Curtiss, Robert Oostenveld, Linda J. Larson-Prior, Jan-Mathijs Schoffelen, Michael R. Hodge, Eileen A. Cler, Daniel M. Marcus, Deanna M. Barch, Essa Yacoub, Stephen M. Smith, Kamil Ugurbil, and David C. Van Essen. The Human Connectome Project: A retrospective. *NeuroImage*, 244:118543, December 2021. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2021.118543.

Ahmed Fakhry and Shuiwang Ji. High-resolution prediction of mouse brain connectivity using gene expression patterns. *Methods*, 73:71–78, February 2015. ISSN 1046-2023. doi: 10.1016/j.ymeth.2014.07.011.

Michele Farisco, G. Baldassarre, E. Cartoni, A. Leach, M.A. Petrovici, A. Rosemann, A. Salles, B. Stahl, and S. J. van Albada. A method for the ethical analysis of brain-inspired AI. *Artificial Intelligence Review*, 57(6):133, May 2024. ISSN 1573-7462. doi: 10.1007/s10462-024-10769-4.

Steven W. Flavell and Michael E. Greenberg. Signaling Mechanisms Linking Neuronal Activity to Gene Expression and Plasticity of the Nervous System. *Annual*

review of neuroscience, 31:563–590, 2008. ISSN 0147-006X. doi: 10.1146/annurev.neuro.31.060407.125631.

Alex Fornito, Aurina Arnatkevičiūtė, and Ben D. Fulcher. Bridging the Gap between Connectome and Transcriptome. *Trends in Cognitive Sciences*, 23(1):34–50, January 2019. ISSN 1879-307X. doi: 10.1016/j.tics.2018.10.005.

Leon French and Paul Pavlidis. Relationships between Gene Expression and Brain Wiring in the Adult Rodent Brain. *PLOS Computational Biology*, 7(1):e1001049, January 2011. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1001049.

Karl Friston, Michael Levin, Biswa Sengupta, and Giovanni Pezzulo. Knowing one's place: A free-energy approach to pattern regulation. *Journal of the Royal Society Interface*, 12(105):20141383, April 2015. doi: 10.1098/rsif.2014.1383.

John Gall. *Systemantics: How Systems Work and Especially How They Fail*. Quadrangle, first edition edition, May 1977. ISBN 978-0-8129-0674-5.

Kristalina Georgieva. AI Will Transform the Global Economy. Let's Make Sure It Benefits Humanity. https://www.imf.org/en/Blogs/Articles/2024/01/14/ai-will-transform-the-global-economy-lets-make-sure-it-benefits-humanity, January 2024.

Darren Gilmour, Martina Rembold, and Maria Leptin. From morphogen to morphogenesis and back. *Nature*, 541(7637):311–320, January 2017. ISSN 1476-4687. doi: 10.1038/nature21348.

Ben Goertzel, Vitaly Bogdanov, Michael Duncan, Deborah Duong, Zarathustra Goertzel, Jan Horlings, Matthew Ikle', Lucius Greg Meredith, Alexey Potapov, Andre' Luiz de Senna, Hedra Seid Andres Suarez, Adam Vandervorst, and Robert Werko. OpenCog Hyperon: A Framework for AGI at the Human Level and Beyond, September 2023.

Alex Graves. Generating Sequences With Recurrent Neural Networks, June 2014.

Henry T. Greely, Christine Grady, Khara M. Ramos, Winston Chiong, James Eberwine, Nita A. Farahany, L. Syd M Johnson, Bradley T. Hyman, Steven E. Hyman, Karen S. Rommelfanger, and Elba E. Serrano. Neuroethics Guiding Principles for the NIH BRAIN Initiative. *The Journal of Neuroscience*, 38(50):10586–10588, December 2018a. ISSN 0270-6474. doi: 10.1523/JNEUROSCI.2077-18.2018.

Henry T. Greely, Christine Grady, Khara M. Ramos, Winston Chiong, James Eberwine, Nita A. Farahany, L. Syd M Johnson, Bradley T. Hyman, Steven E. Hyman, Karen S. Rommelfanger, and Elba E. Serrano. Neuroethics Guiding Principles for the NIH BRAIN Initiative. *The Journal of Neuroscience*, 38(50):10586–10588, December 2018b. ISSN 0270-6474. doi: 10.1523/JNEUROSCI.2077-18.2018.

Ulf Gurok, Christine Steinhoff, Bettina Lipkowitz, H.-Hilger Ropers, Constance Scharff, and Ulrike A. Nuber. Gene Expression Changes in the Course of Neural Progenitor Cell Differentiation. *Journal of Neuroscience*, 24(26):5982–6002, June 2004. ISSN 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.0809-04.2004.

Chelsea M. Heveran, Robin Gerlach, Christopher J. Hernandez, Kristen Intemann, Anne S. Meyer, Caroline Ajo-Franklin, Marimikel Charrier, Wilfred V. Srubar, Neel Joshi, Alshakim Nelson, and Matthew W. Fields. Unlocking the societal potential of engineered living materials. *Matter*, 7(9):2846–2858, September 2024. ISSN 2590-2393, 2590-2385. doi: 10.1016/j.matt.2024.07.011.

Peter D. Hoff. Modeling homophily and stochastic equivalence in symmetric relational data, November 2007.

Longwen Huang, Justus M. Kebschull, Daniel Fürth, Simon Musall, Matthew T. Kaufman, Anne K. Churchland, and Anthony M. Zador. BRICseq Bridges Brain-wide Interregional Connectivity to Neural Activity and Gene Expression in Single Animals. *Cell*, 183(7):2040, December 2020. ISSN 0092-8674, 1097-4172. doi: 10.1016/j.cell.2020.12.009.

Robert C. Hubrecht and Elizabeth Carter. The 3Rs and Humane Experimental Technique: Implementing Change. *Animals : an Open Access Journal from MDPI*, 9(10):754, September 2019. ISSN 2076-2615. doi: 10.3390/ani9100754.

Edward Hughes, Michael Dennis, Jack Parker-Holder, Feryal Behbahani, Aditi Mavalankar, Yuge Shi, Tom Schaul, and Tim Rocktaschel. Open-Endedness is Essential for Artificial Superhuman Intelligence, June 2024.

Marcello Ienca. On Neurorights. *Frontiers in Human Neuroscience*, 15:701258, September 2021. ISSN 1662-5161. doi: 10.3389/fnhum.2021.701258.

Shuiwang Ji, Ahmed Fakhry, and Houtao Deng. Integrative analysis of the connectivity and gene expression atlases in the mouse brain. *NeuroImage*, 84:245–253, January 2014. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2013.08.049.

Anita S. Jwa and Russell A. Poldrack. The spectrum of data sharing policies in neuroimaging data repositories. *Human Brain Mapping*, 43(8):2707–2721, February 2022. ISSN 1065-9471. doi: 10.1002/hbm.25803.

Brett J. Kagan, Andy C. Kitchen, Nhi T. Tran, Forough Habibollahi, Moein Khajehnejad, Bradyn J. Parker, Anjali Bhat, Ben Rollo, Adeel Razi, and Karl J. Friston. In vitro neurons learn and exhibit sentience when embodied in a simulated game-world. *Neuron*, 110(23):3952–3969.e8, December 2022. ISSN 0896-6273. doi: 10.1016/j.neuron.2022.09.001.

Ryoichiro Kageyama, Yasutaka Niwa, and Hiromi Shimojo. Rhythmic Gene Expression in Somite Formation and Neural Development. *Molecules and Cells*, 27(5): 497–502, May 2009. ISSN 1016-8478. doi: 10.1007/s10059-009-0068-1.

Alon Kaufman, Gideon Dror, Isaac Meilijson, and Eytan Ruppin. Gene Expression of Caenorhabditis elegans Neurons Carries Information on Their Synaptic Connectivity. *PLOS Computational Biology*, 2(12):e167, December 2006. ISSN 1553-7358. doi: 10.1371/journal.pcbi.0020167.

Justus M. Kebschull, Pedro Garcia da Silva, Ashlan P. Reid, Ian D. Peikon, Dinu F. Albeanu, and Anthony M. Zador. High-Throughput Mapping of Single-Neuron Projections by Sequencing of Barcoded RNA. *Neuron*, 91(5):975–987, September 2016. ISSN 1097-4199. doi: 10.1016/j.neuron.2016.07.036.

Alexander J. E. Kell, Daniel L. K. Yamins, Erica N. Shook, Sam V. Norman-Haignere, and Josh H. McDermott. A task-optimized neural network replicates human auditory behavior, predicts brain responses, and reveals a cortical processing hierarchy. *Neuron*, 98(3):630–644.e16, May 2018. ISSN 0896-6273. doi: 10.1016/j.neuron.2018.03.044.

S. Kondo and R. Asal. A reaction-diffusion wave on the skin of the marine angelfish Pomacanthus. *Nature*, 376(6543):765–768, August 1995. ISSN 0028-0836. doi: 10.1038/376765a0.

Shigeru Kondo and Takashi Miura. Reaction-diffusion model as a framework for understanding biological pattern formation. *Science (New York, N.Y.)*, 329(5999): 1616–1620, September 2010. ISSN 1095-9203. doi: 10.1126/science.1179047.

István A. Kovács, Dániel L. Barabási, and Albert-László Barabási. Uncovering the genetic blueprint of the C. elegans nervous system. *Proceedings of the National*

Academy of Sciences, 117(52):33570–33577, December 2020. doi: 10.1073/pnas.2009093117.

Sam Kriegman, Douglas Blackiston, Michael Levin, and Josh Bongard. Kinematic self-replication in reconfigurable organisms. *Proceedings of the National Academy of Sciences*, 118(49):e2112672118, December 2021. doi: 10.1073/pnas.2112672118.

Eric Lagasse and Michael Levin. Future medicine: From molecular pathways to the collective intelligence of the body. *Trends in Molecular Medicine*, 29(9):687–710, September 2023. ISSN 1471-4914, 1471-499X. doi: 10.1016/j.molmed.2023.06.007.

Yann LeCun. A Path Towards Autonomous Machine Intelligence. June 2022.

Ed S. Lein, Michael J. Hawrylycz, Nancy Ao, Mikael Ayres, Amy Bensinger, Amy Bernard, Andrew F. Boe, Mark S. Boguski, Kevin S. Brockway, Emi J. Byrnes, Lin Chen, Li Chen, Tsuey-Ming Chen, Mei Chi Chin, Jimmy Chong, Brian E. Crook, Aneta Czaplinska, Chinh N. Dang, Suvro Datta, Nick R. Dee, Aimee L. Desaki, Tsega Desta, Ellen Diep, Tim A. Dolbeare, Matthew J. Donelan, Hong-Wei Dong, Jennifer G. Dougherty, Ben J. Duncan, Amanda J. Ebbert, Gregor Eichele, Lili K. Estin, Casey Faber, Benjamin A. Facer, Rick Fields, Shanna R. Fischer, Tim P. Fliss, Cliff Frensley, Sabrina N. Gates, Katie J. Glattfelder, Kevin R. Halverson, Matthew R. Hart, John G. Hohmann, Maureen P. Howell, Darren P. Jeung, Rebecca A. Johnson, Patrick T. Karr, Reena Kawal, Jolene M. Kidney, Rachel H. Knapik, Chihchau L. Kuan, James H. Lake, Annabel R. Laramee, Kirk D. Larsen, Christopher Lau, Tracy A. Lemon, Agnes J. Liang, Ying Liu, Lon T. Luong, Jesse Michaels, Judith J. Morgan, Rebecca J. Morgan, Marty T. Mortrud, Nerick F. Mosqueda, Lydia L. Ng, Randy Ng, Geralyn J. Orta, Caroline C. Overly, Tu H. Pak, Sheana E. Parry, Sayan D. Pathak, Owen C. Pearson, Ralph B. Puchalski, Zackery L. Riley, Hannah R. Rockett, Stephen A. Rowland, Joshua J. Royall, Marcos J. Ruiz, Nadia R. Sarno, Katherine Schaffnit, Nadiya V. Shapovalova, Taz Sivisay, Clifford R. Slaughterbeck, Simon C. Smith, Kimberly A. Smith, Bryan I. Smith, Andy J. Sodt, Nick N. Stewart, Kenda-Ruth Stumpf, Susan M. Sunkin, Madhavi Sutram, Angelene Tam, Carey D. Teemer, Christina Thaller, Carol L. Thompson, Lee R. Varnam, Axel Visel, Ray M. Whitlock, Paul E. Wohnoutka, Crissa K. Wolkey, Victoria Y. Wong, Matthew Wood, Murat B. Yaylaoglu, Rob C. Young, Brian L. Youngstrom, Xu Feng Yuan, Bin Zhang, Theresa A. Zwingman, and Allan R. Jones. Genome-wide atlas of gene expression in the adult mouse brain. *Nature*, 445(7124):168–176, January 2007. ISSN 1476-4687. doi: 10.1038/nature05453.

Michael Levin. Reading and Writing the Morphogenetic Code. Foundational White Paper of the Allen Discovery Center at Tufts University, 2016.

Michael Levin. Darwin's agential materials: Evolutionary implications of multiscale competency in developmental biology. *Cellular and molecular life sciences: CMLS*, 80(6):142, May 2023. ISSN 1420-9071. doi: 10.1007/s00018-023-04790-z.

Daniel Lobo and Michael Levin. Inferring Regulatory Networks from Experimental Morphological Phenotypes: A Computational Method Reverse-Engineers Planarian Regeneration. *PLoS Computational Biology*, 11(6):e1004295, June 2015. ISSN 1553-734X. doi: 10.1371/journal.pcbi.1004295.

Romain Lopez, Natasa Tagasovska, Stephen Ra, Kyunghyun Cho, Jonathan Pritchard, and Aviv Regev. Learning Causal Representations of Single Cells via Sparse Mechanism Shift Modeling. In *Proceedings of the Second Conference on Causal Learning and Reasoning*, pages 662–691. PMLR, August 2023.

Valerio Mante, David Sussillo, Krishna V. Shenoy, and William T. Newsome. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, 503(7474):78–84, November 2013. ISSN 1476-4687. doi: 10.1038/nature12742.

Gary Marcus. The Next Decade in AI: Four Steps Towards Robust Artificial Intelligence, February 2020.

Ronald L Meyer. Roger Sperry and his chemoaffinity hypothesis. *Neuropsychologia*, 36(10):957–980, October 1998. ISSN 0028-3932. doi: 10.1016/S0028-3932(98)00052-9.

Kevin J. Miller, Maria Eckstein, Matthew M. Botvinick, and Zeb Kurth-Nelson. Cognitive Model Discovery via Disentangled RNNs, June 2023.

Patrick Mineault. Foundation models for neuroscience. https://www.neuroai.science/p/foundation-models-for-neuroscience, January 2024.

Itzy E. Morales Pantoja, Lena Smirnova, Alysson R. Muotri, Karl J. Wahlin, Jeffrey Kahn, J. Lomax Boyd, David H. Gracias, Timothy D. Harris, Tzahi Cohen-Karni, Brian S. Caffo, Alexander S. Szalay, Fang Han, Donald J. Zack, Ralph Etienne-Cummings, Akwasi Akwaboah, July Carolina Romero, Dowlette-Mary Alam El Din, Jesse D. Plotkin, Barton L. Paulhamus, Erik C. Johnson, Frederic Gilbert,

J. Lowry Curley, Ben Cappiello, Jens C. Schwamborn, Eric J. Hill, Paul Roach, Daniel Tornero, Caroline Krall, Rheinallt Parri, Fenna Sillé, Andre Levchenko, Rabih E. Jabbour, Brett J. Kagan, Cynthia A. Berlinicke, Qi Huang, Alexandra Maertens, Kathrin Herrmann, Katya Tsaioun, Raha Dastgheyb, Christa Whelan Habela, Joshua T. Vogelstein, and Thomas Hartung. First Organoid Intelligence (OI) workshop to form an OI community. *Frontiers in Artificial Intelligence*, 6, February 2023. ISSN 2624-8212. doi: 10.3389/frai.2023.1116870.

Alexander Mordvintsev and Eyvind Niklasson. μNCA: Texture Generation with Ultra-Compact Neural Cellular Automata, November 2021.

Alexander Mordvintsev, Ettore Randazzo, Eyvind Niklasson, and Michael Levin. Growing Neural Cellular Automata. *Distill*, 5(2):e23, February 2020. ISSN 2476-0757. doi: 10.23915/distill.00023.

Alexander Mordvintsev, Ettore Randazzo, and Craig Fouts. Growing Isotropic Neural Cellular Automata. In *ALIFE 2022: The 2022 Conference on Artificial Life*. MIT Press, July 2022. doi: 10.1162/isal_a_00552.

James D. Murray. How the Leopard Gets Its Spots. *Scientific American*, 258(3): 80–87, 1988. ISSN 0036-8733.

Michael Nayak. Large Bio-Mechanical Space Structures, February 2025.

Seung Wook Oh, Julie A. Harris, Lydia Ng, Brent Winslow, Nicholas Cain, Stefan Mihalas, Quanxin Wang, Chris Lau, Leonard Kuan, Alex M. Henry, Marty T. Mortrud, Benjamin Ouellette, Thuc Nghi Nguyen, Staci A. Sorensen, Clifford R. Slaughterbeck, Wayne Wakeman, Yang Li, David Feng, Anh Ho, Eric Nicholas, Karla E. Hirokawa, Phillip Bohn, Kevin M. Joines, Hanchuan Peng, Michael J. Hawrylycz, John W. Phillips, John G. Hohmann, Paul Wohnoutka, Charles R. Gerfen, Christof Koch, Amy Bernard, Chinh Dang, Allan R. Jones, and Hongkui Zeng. A mesoscale connectome of the mouse brain. *Nature*, 508(7495):207–214, April 2014. ISSN 1476-4687. doi: 10.1038/nature13186.

Thomas Parr, Giovanni Pezzulo, and Karl J. Friston. *Active Inference: The Free Energy Principle in Mind, Brain, and Behavior*. The MIT Press, Cambridge, March 2022. ISBN 978-0-262-04535-3.

G. Pezzulo and M. Levin. Re-membering the body: Applications of computational neuroscience to the top-down control of regeneration of limbs and other complex

organs. *Integrative Biology: Quantitative Biosciences from Nano to Macro*, 7(12): 1487–1517, December 2015. ISSN 1757-9708. doi: 10.1039/c5ib00221d.

Mu Qiao. Deciphering the genetic code of neuronal type connectivity through bilinear modeling. *eLife*, 12:RP91532, June 2024. ISSN 2050-084X. doi: 10.7554/eLife.91532.

Raghu Raman, Robin Kowalski, Krishnashree Achuthan, Akshay Iyer, and Prema Nedungadi. Navigating artificial general intelligence development: Societal, technological, ethical, and brain-inspired pathways. *Scientific Reports*, 15:8443, March 2025. ISSN 2045-2322. doi: 10.1038/s41598-025-92190-7.

Ettore Randazzo, Alexander Mordvintsev, and Craig Fouts. Growing Steerable Neural Cellular Automata, May 2023.

B. A. R. Rao, Rabindranath Jana, and S. Bandyopadhyay. A Markov chain Monte carol method for generating random (0, 1)-matrices with given marginals. 1996.

Soumil Rathi. Approaches to Artificial General Intelligence: An Analysis, January 2022.

W. Reik, W. Dean, and J. Walter. Epigenetic reprogramming in mammalian development. *Science (New York, N.Y.)*, 293(5532):1089–1093, August 2001. ISSN 0036-8075. doi: 10.1126/science.1063443.

R. D. Riddle, R. L. Johnson, E. Laufer, and C. Tabin. Sonic hedgehog mediates the polarizing activity of the ZPA. *Cell*, 75(7):1401–1416, December 1993. ISSN 0092-8674. doi: 10.1016/0092-8674(93)90626-2.

Sebastian Risi. The Future of Artificial Intelligence is Self-Organizing and Self-Assembling. *sebastianrisi.com*, 2021.

Ilaria Roberti, Marta Lovino, Santa Di Cataldo, Elisa Ficarra, and Gianvito Urgese. Exploiting Gene Expression Profiles for the Automated Prediction of Connectivity between Brain Regions. *International Journal of Molecular Sciences*, 20(8):2035, January 2019. ISSN 1422-0067. doi: 10.3390/ijms20082035.

Stuart Russell. *Human Compatible: Artificial Intelligence and the Problem of Control*. Viking, New York?, first edition edition, October 2019. ISBN 978-0-525-55861-3.

Cristina Sánchez-Camacho and Paola Bovolenta. Emerging mechanisms in morphogen-mediated axon guidance. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, 31(10):1013–1025, October 2009. ISSN 1521-1878. doi: 10.1002/bies.200900063.

Martin Schrimpf, Idan Asher Blank, Greta Tuckute, Carina Kauf, Eghbal A. Hosseini, Nancy Kanwisher, Joshua B. Tenenbaum, and Evelina Fedorenko. The neural architecture of language: Integrative modeling converges on predictive processing. *Proceedings of the National Academy of Sciences*, 118(45):e2105646118, November 2021. doi: 10.1073/pnas.2105646118.

Erwin Schrödinger. *What Is Life? The Physical Aspect of the Living Cell*. Cambridge University Press, United Kingdom (UK), 1944.

Alexander Shapson-Coe, Michał Januszewski, Daniel R. Berger, Art Pope, Yuelong Wu, Tim Blakely, Richard L. Schalek, Peter H. Li, Shuohong Wang, Jeremy Maitin-Shepard, Neha Karlupia, Sven Dorkenwald, Evelina Sjostedt, Laramie Leavitt, Dongil Lee, Jakob Troidl, Forrest Collman, Luke Bailey, Angerica Fitzmaurice, Rohin Kar, Benjamin Field, Hank Wu, Julian Wagner-Carena, David Aley, Joanna Lau, Zudi Lin, Donglai Wei, Hanspeter Pfister, Adi Peleg, Viren Jain, and Jeff W. Lichtman. A petavoxel fragment of human cerebral cortex reconstructed at nanoscale resolution. *Science*, 384(6696):eadk4858, May 2024. doi: 10.1126/science.adk4858.

Yang Shen and Xiuwu Zhang. The impact of artificial intelligence on employment: The role of virtual agglomeration. *Humanities and Social Sciences Communications*, 11(1):1–14, January 2024. ISSN 2662-9992. doi: 10.1057/s41599-024-02647-9.

David Silver, Satinder Singh, Doina Precup, and Richard S. Sutton. Reward is enough. *Artificial Intelligence*, 299:103535, 2021. ISSN 0004-3702. doi: 10.1016/j.artint.2021.103535.

R. W. Sperry. Chemoaffinity in the orderly growth of nerve fiber patterns and connections*. *Proceedings of the National Academy of Sciences*, 50(4):703–710, October 1963. doi: 10.1073/pnas.50.4.703.

Kenneth O. Stanley, Jeff Clune, Joel Lehman, and Risto Miikkulainen. Designing neural networks through neuroevolution. *Nature Machine Intelligence*, 1(1):24–35, January 2019. ISSN 2522-5839. doi: 10.1038/s42256-018-0006-z.

Paul W Sternberg, Kimberly Van Auken, Qinghua Wang, Adam Wright, Karen Yook, Magdalena Zarowiecki, Valerio Arnaboldi, Andrés Becerra, Stephanie Brown, Scott Cain, Juancarlos Chan, Wen J Chen, Jaehyoung Cho, Paul Davis, Stavros Diamantakis, Sarah Dyer, Dionysis Grigoriadis, Christian A Grove, Todd Harris, Kevin Howe, Ranjana Kishore, Raymond Lee, Ian Longden, Manuel Luypaert, Hans-Michael Müller, Paulo Nuin, Mark Quinton-Tulloch, Daniela Raciti, Tim Schedl, Gary Schindelman, and Lincoln Stein. WormBase 2024: Status and transitioning to Alliance infrastructure. *Genetics*, 227(1):iyae050, May 2024. ISSN 1943-2631. doi: 10.1093/genetics/iyae050.

The MICrONS Consortium, J. Alexander Bae, Mahaly Baptiste, Caitlyn A. Bishop, Agnes L. Bodor, Derrick Brittain, JoAnn Buchanan, Daniel J. Bumbarger, Manuel A. Castro, Brendan Celii, Erick Cobos, Forrest Collman, Nuno Maçarico da Costa, Sven Dorkenwald, Leila Elabbady, Paul G. Fahey, Tim Fliss, Emmanouil Froudarakis, Jay Gager, Clare Gamlin, William Gray-Roncal, Akhilesh Halageri, James Hebditch, Zhen Jia, Emily Joyce, Justin Joyce, Chris Jordan, Daniel Kapner, Nico Kemnitz, Sam Kinn, Lindsey M. Kitchell, Selden Koolman, Kai Kuehner, Kisuk Lee, Kai Li, Ran Lu, Thomas Macrina, Gayathri Mahalingam, Jordan Matelsky, Sarah McReynolds, Elanine Miranda, Eric Mitchell, Shanka Subhra Mondal, Merlin Moore, Shang Mu, Taliah Muhammad, Barak Nehoran, Oluwaseun Ogedengbe, Christos Papadopoulos, Stelios Papadopoulos, Saumil Patel, Xaq Pitkow, Sergiy Popovych, Anthony Ramos, R. Clay Reid, Jacob Reimer, Patricia K. Rivlin, Victoria Rose, Casey M. Schneider-Mizell, H. Sebastian Seung, Ben Silverman, William Silversmith, Amy Sterling, Fabian H. Sinz, Cameron L. Smith, Shelby Suckow, Marc Takeno, Zheng H. Tan, Andreas S. Tolias, Russel Torres, Nicholas L. Turner, Edgar Y. Walker, Tianyu Wang, Adrian Wanner, Brock A. Wester, Grace Williams, Sarah Williams, Kyle Willie, Ryan Willie, William Wong, Jingpeng Wu, Chris Xu, Runzhe Yang, Dimitri Yatsenko, Fei Ye, Wenjing Yin, Rob Young, Szi-chieh Yu, Daniel Xenes, and Chi Zhang. Functional connectomics spanning multiple areas of mouse visual cortex. *bioRxiv*, page 2021.07.28.454025, January 2023. doi: 10.1101/2021.07.28.454025.

Luca Tiberi, Pierre Vanderhaeghen, and Jelle van den Ameele. Cortical neurogenesis and morphogens: Diversity of cues, sources and functions. *Current Opinion in Cell Biology*, 24(2):269–276, April 2012. ISSN 0955-0674. doi: 10.1016/j.ceb.2012.01.010.

Alan Turing. The chemical basis of morphogenesis. *Philosophical Transactions of*

the Royal Society of London. Series B, Biological Sciences, 237(641):37–72, August 1952. ISSN 2054-0280. doi: 10.1098/rstb.1952.0012.

UNESCO. *Recommendation on the Ethics of Artificial Intelligence*. United Nations Educational, Scientific and Cultural Organization, 2022.

Michel Vervoort. Regeneration and Development in Animals. *Biological Theory*, 6 (1):25–35, December 2011. ISSN 1555-5550. doi: 10.1007/s13752-011-0005-3.

G.P.M. Virgilio, F. Saavedra Hoyos, and C.B. Bao Ratzemberg. The impact of artificial intelligence on unemployment: A review. *International Journal of Social Economics*, 51(12):1680–1695, April 2024.

Yanyi Wang, Yi Liu, Jing Li, Yue Chen, Sizhe Liu, and Chao Zhong. Engineered living materials (ELMs) design: From function allocation to dynamic behavior modulation. *Current Opinion in Chemical Biology*, 70:102188, October 2022. ISSN 1367-5931. doi: 10.1016/j.cbpa.2022.102188.

J. G. White, E. Southgate, J. N. Thomson, and S. Brenner. The structure of the nervous system of the nematode Caenorhabditis elegans. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 314(1165):1–340, November 1986. ISSN 0962-8436. doi: 10.1098/rstb.1986.0056.

Michael Winding, Benjamin D. Pedigo, Christopher L. Barnes, Heather G. Patsolic, Youngser Park, Tom Kazimiers, Akira Fushiki, Ingrid V. Andrade, Avinash Khandelwal, Javier Valdes-Aleman, Feng Li, Nadine Randel, Elizabeth Barsotti, Ana Correia, Richard D. Fetter, Volker Hartenstein, Carey E. Priebe, Joshua T. Vogelstein, Albert Cardona, and Marta Zlatic. The connectome of an insect brain. *Science*, 379(6636):eadd9330, March 2023. doi: 10.1126/science.add9330.

Lior Wolf, Chen Goldberg, Nathan Manor, Roded Sharan, and Eytan Ruppin. Gene expression in the rodent brain is associated with its regional connectivity. *PLoS computational biology*, 7(5):e1002040, May 2011. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1002040.

L. Wolpert. Positional information and the spatial pattern of cellular differentiation. *Journal of Theoretical Biology*, 25(1):1–47, October 1969. ISSN 0022-5193. doi: 10.1016/S0022-5193(69)80016-0.

C. Shan Xu, Michal Januszewski, Zhiyuan Lu, Shin-ya Takemura, Kenneth J. Hayworth, Gary Huang, Kazunori Shinomiya, Jeremy Maitin-Shepard, David Ackerman, Stuart Berg, Tim Blakely, John Bogovic, Jody Clements, Tom Dolafi,

Philip Hubbard, Dagmar Kainmueller, William Katz, Takashi Kawase, Khaled A. Khairy, Laramie Leavitt, Peter H. Li, Larry Lindsey, Nicole Neubarth, Donald J. Olbris, Hideo Otsuna, Eric T. Troutman, Lowell Umayam, Ting Zhao, Masayoshi Ito, Jens Goldammer, Tanya Wolff, Robert Svirskas, Philipp Schlegel, Erika R. Neace, Christopher J. Knecht, Chelsea X. Alvarado, Dennis A. Bailey, Samantha Ballinger, Jolanta A. Borycz, Brandon S. Canino, Natasha Cheatham, Michael Cook, Marisa Dreher, Octave Duclos, Bryon Eubanks, Kelli Fairbanks, Samantha Finley, Nora Forknall, Audrey Francis, Gary Patrick Hopkins, Emily M. Joyce, SungJin Kim, Nicole A. Kirk, Julie Kovalyak, Shirley A. Lauchie, Alanna Lohff, Charli Maldonado, Emily A. Manley, Sari McLin, Caroline Mooney, Miatta Ndama, Omotara Ogundeyi, Nneoma Okeoma, Christopher Ordish, Nicholas Padilla, Christopher Patrick, Tyler Paterson, Elliott E. Phillips, Emily M. Phillips, Neha Rampally, Caitlin Ribeiro, Madelaine K. Robertson, Jon Thomson Rymer, Sean M. Ryan, Megan Sammons, Anne K. Scott, Ashley L. Scott, Aya Shinomiya, Claire Smith, Kelsey Smith, Natalie L. Smith, Margaret A. Sobeski, Alia Suleiman, Jackie Swift, Satoko Takemura, Iris Talebi, Dorota Tarnogorska, Emily Tenshaw, Temour Tokhi, John J. Walsh, Tansy Yang, Jane Anne Horne, Feng Li, Ruchi Parekh, Patricia K. Rivlin, Vivek Jayaraman, Kei Ito, Stephan Saalfeld, Reed George, Ian Meinertzhagen, Gerald M. Rubin, Harald F. Hess, Louis K. Scheffer, Viren Jain, and Stephen M. Plaza. A Connectome of the Adult Drosophila Central Brain, January 2020.

Daniel L. K. Yamins and James J. DiCarlo. Using goal-driven deep learning models to understand sensory cortex. *Nature Neuroscience*, 19(3):356–365, March 2016. ISSN 1546-1726. doi: 10.1038/nn.4244.

Guangyu Robert Yang, Madhura R. Joglekar, H. Francis Song, William T. Newsome, and Xiao-Jing Wang. Task representations in neural networks trained to perform many cognitive tasks. *Nature Neuroscience*, 22(2):297, February 2019. ISSN 1546-1726. doi: 10.1038/s41593-018-0310-2.

Dongsu Zhang, Changwoon Choi, Jeonghwan Kim, and Young Min Kim. Learning to Generate 3D Shapes with Generative Cellular Automata, March 2021.

Mingyuan Zhu and Adrienne Hk Roeder. Plants are better engineers: The complexity of plant organ morphogenesis. *Current Opinion in Genetics & Development*, 63: 16–23, August 2020. ISSN 1879-0380. doi: 10.1016/j.gde.2020.02.008.