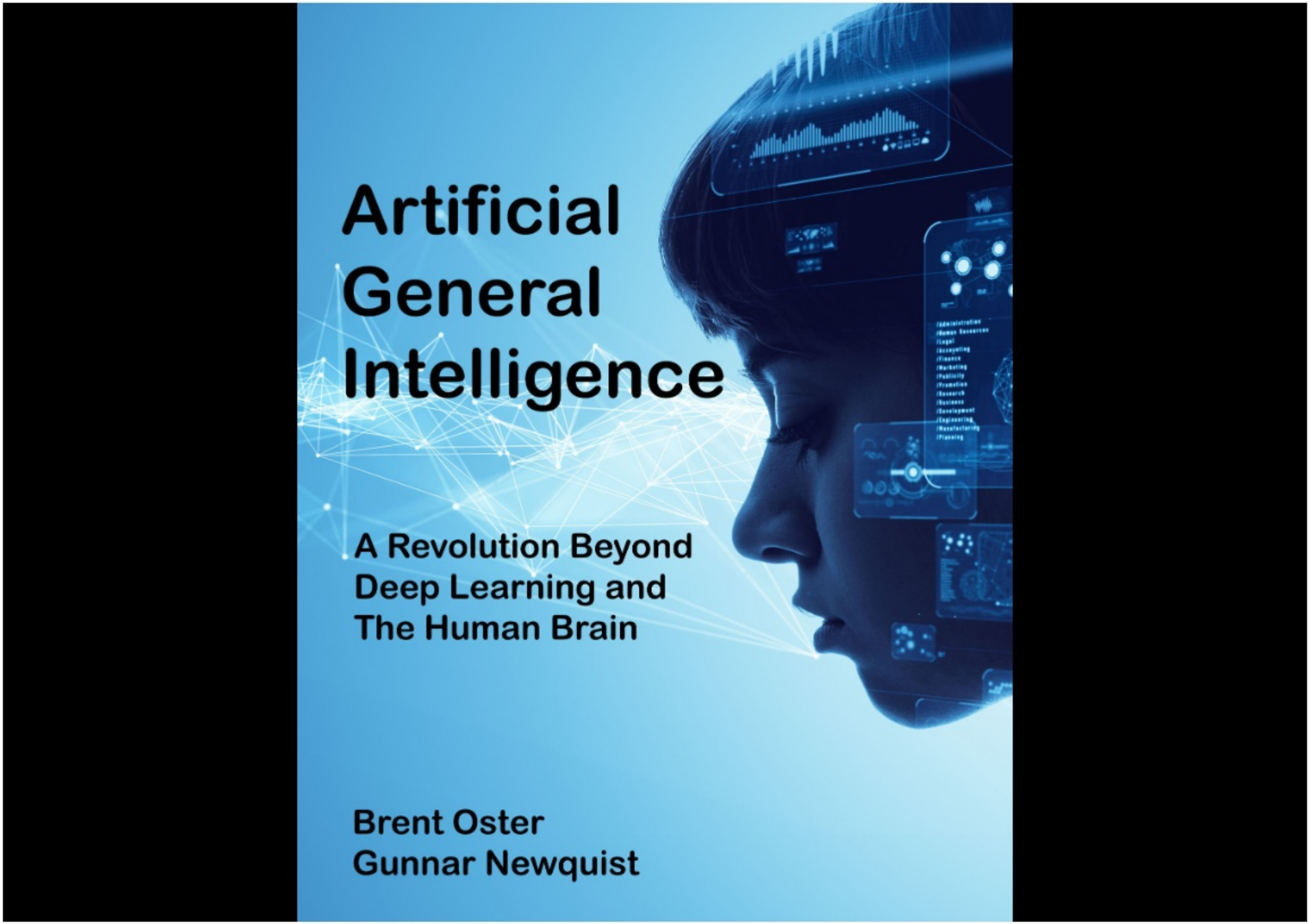

Artificial
General
Intelligence
A Revolution Beyond
Deep Learning and
The Human Brain
Brent Oster
Gunnar Newquist

ARTIFICIAL GENERAL INTELLIGENCE

A Revolution Beyond Deep Learning and The Human Brain

Brent Oster

Gunnar Newquist

CONTENTS

INTRODUCTION

In the annals of human history, moments of profound transformation stand out as beacons of progress. From the mastery of fire to the development of the printing press, and the harnessing of electricity to the dawn of the digital age, our species has continually sought to expand the boundaries of what is possible. Today, we stand on the precipice of another such transformative era, one that has the potential to reshape the very essence of intelligence itself: the era of Artificial General Intelligence (AGI).

Welcome to "Artificial General Intelligence: A Revolution Beyond Deep Learning and the Human Brain". In this book, we embark on an exhilarating journey into the heart of AGI—a paradigm of machine intelligence that surpasses the specialized capabilities of narrow AI and transcends the cognitive limitations of the human mind. AGI, often referred to as "strong AI" or "human-level AI," represents the culmination of decades of research, imagination, and relentless pursuit of creating machines that can think, learn, and adapt across a wide range of tasks with human-like proficiency.

As we delve into the world of AGI, we will peel back the layers of its evolution and go beyond the popular narratives surrounding machine learning and deep neural networks. We will explore the multidisciplinary nature of AGI, drawing insights from fields as diverse as neuroscience, psychology, philosophy, linguistics, and computer science. The quest for AGI is not just a technological endeavor but a voyage into the essence of intelligence itself.

Our exploration will extend beyond the confines of algorithms and silicon chips, touching upon the ethical, societal, and existential implications of AGI. What does it mean for humanity to create a form of intelligence that rivals our own? How do we ensure that AGI serves the betterment of society rather than becoming a force beyond our control? What are the philosophical conundrums and moral dilemmas that arise when we create machines capable of self-awareness and autonomous decision-making?

"Artificial General Intelligence" is a book for the curious and the visionary, for those who seek a deeper understanding of the most consequential technological revolution of our time. Whether you are a seasoned AI researcher, a budding enthusiast, or simply someone intrigued by the mysteries of intelligence and the future of humanity, this book will serve as a guiding light through the labyrinthine landscapes of AGI.

This book is intended not just for tech enthusiasts or experts but for anyone curious about the profound transformations taking place in our world. Whether you're an AI novice or a seasoned professional, our journey together will be an enlightening and accessible exploration of the concepts, breakthroughs, and ethical dilemmas at the heart of the AGI revolution.

AI Today

Currently most of us in the early 2020s envision artificial intelligence to be something like what we see in the movies. Based on those depictions of AI, we would expect to see human-like robots and holograms in our homes, talking and acting like real people and having human-level or even superhuman intelligence and capabilities. We expect robots to be able to do just about any of our daily chores for us and AI to do jobs in the real world with the accuracy and efficiency of a professional human being. We expect cars that can drive us with complete autonomy. We expect interactive personal assistants that can converse at a human level and provide accurate information for any of our needs. We worry about AI taking over the world or doing something detrimental toward humanity, fed by the tropes of science fiction. The reality of current AI, however, is something much less powerful. Let's look at what we actually have for AI today, and where that AI is going in the near future.

So, why don't we have the AI that science fiction promises us today? Why don't household robots clean our homes, do our laundry, and cook for us? Why don't we converse fluently with truly intelligent artificial agents on our devices and online? Why do cars still have a steering wheel and why are they not fully autonomous and driving themselves?

Is it just that these technologies are early and have not reached their full potential? We would argue no. Rather than nascent, but emerging, intelligence in machines today, what we actually have is something entirely different called narrow AI. This narrow AI will never be capable of these feats because, fundamentally, it isn't capable of scaling to the intelligence we imagine.

However, AI that could do all the marvels that we envision in our science fiction, the functions that are so fundamental to human intelligence, would be called Artificial General Intelligence (AGI), but it does NOT exist anywhere on earth yet. What we actually have for AI in 2023 is much simpler and much more narrow called deep-learning based artificial intelligence that can only do some extremely specific tasks better than people, but has fundamental limitations that will not allow it to become AGI, and it falls far short of human intelligence. Even the vaunted large language models like GPT are still a far cry from AGI.

In this book, the author, Brent Oster, a 10-year veteran engineer in artificial intelligence and deep learning, describes the state of AI in 2023, and what its strengths, weaknesses, and narrow applications are. The book then does a deep dive into the neuroscience of the human brain with co-author Gunnar Newquist (PhD in neuroscience) to show how different deep learning AI is from the human brain's capabilities. Then together they explore how we can set the requirements for AGI and come up with a conceptual design for an artificial general intelligence that goes far beyond deep learning to become more like our intelligence, and provide examples of such AGI used in robotics, medicine, law, defense, and other fields.

As AGI comes online, it will displace most of the deep learning technologies we use today and create tremendous upheaval in technology and many industries. We do our best in this book to provide concrete examples of this

technology, the changes it will bring and how they will unfold over the next decades and to make tangible estimates of what it will mean for humanity.

CHAPTER 1

Narrow Artificial Intelligence in the 2020s

In the early 2020s, we only have narrow artificial intelligence, which consists of systems that can learn only to do specific tasks, trained on specific data, formatted to the task, like identifying specific objects in an image, parsing text or speech, and so on. It is only one step above systems that are programmed to do a specific task in that it 'learns' to do the predefined task by fitting mathematical functions to pre-formatted input and output data. Each application is accomplished by a uniquely formed AI system, and each of these unique AIs are not guaranteed to work on any other set of data that goes beyond the scope for which it was trained. Hence, the term "narrow".

Even with this narrow AI, our lives are already touched by these algorithms that recommend our entertainment, autonomous vehicles that navigate our highways, and virtual assistants that converse with us on demand. As AI becomes increasingly integrated into our daily existence, it is imperative that we understand not only its capabilities but also its limitations, implications, and ethical considerations.

This chapter will equip you with the knowledge and insights needed to appreciate the art and science of deep learning. Whether you are a seasoned machine learning practitioner seeking to expand your horizons or a newcomer eager to grasp the nuances of this nascent technology, our journey into the heart of deep learning promises to be

both enlightening and inspiring and give you a better understanding of what Deep learning is all about and what its capabilities and limitations are.

Let us define what is meant by AI, Machine Learning and Deep Learning, as these terms are used extensively in this chapter.

Artificial Intelligence (AI) refers to the overall simulation of human intelligence in machines, enabling them to perform tasks that typically require human intelligence. These tasks can include learning from experience, understanding natural language, recognizing patterns, reasoning, problem-solving, and making decisions. AI aims to create machines or systems that can mimic human cognitive abilities and exhibit "intelligent" behavior.

Machine Learning (ML) is a subfield of artificial intelligence (AI) that focuses on the development of algorithms and statistical models that enable computers to learn and improve their performance on a specific task without being explicitly programmed for that task. In essence, it's about creating systems that can learn from data and make predictions or decisions based on that learning.

Deep Learning (DL) is a specialized subset of machine learning that focuses on training artificial neural networks with multiple layers (deep architectures) to learn and extract representations from data. The term "deep" refers to the presence of multiple layers in these neural networks, allowing them to learn hierarchical representations of the input data.

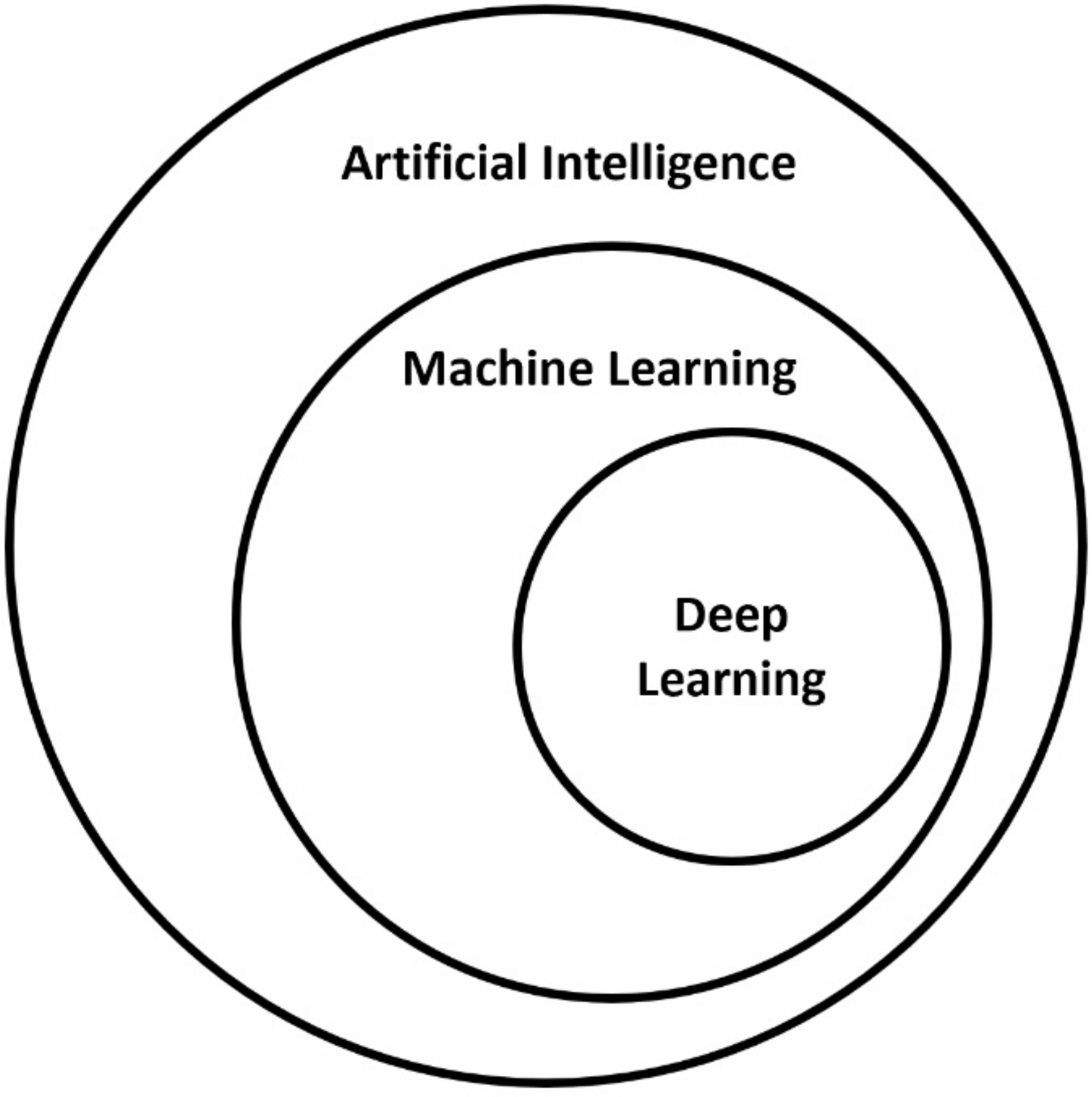


Figure 1.1 Nesting of the Definitions for AI, ML, and DL

Deep learning models are collectively known as artificial neural networks, composed of multiple layers of artificial neurons. Each neuron takes in input data from some or all the neurons in the previous layer, applies weights and biases

to each, performs summation and other mathematical operations, and distributes the output from each neuron to the next layer of neurons, with a final layer providing output from the network. The process of training a deep learning model involves adjusting the weights and biases of these neurons through a process called backpropagation, which uses optimization algorithms to minimize the difference between the model's predictions and the expected values.

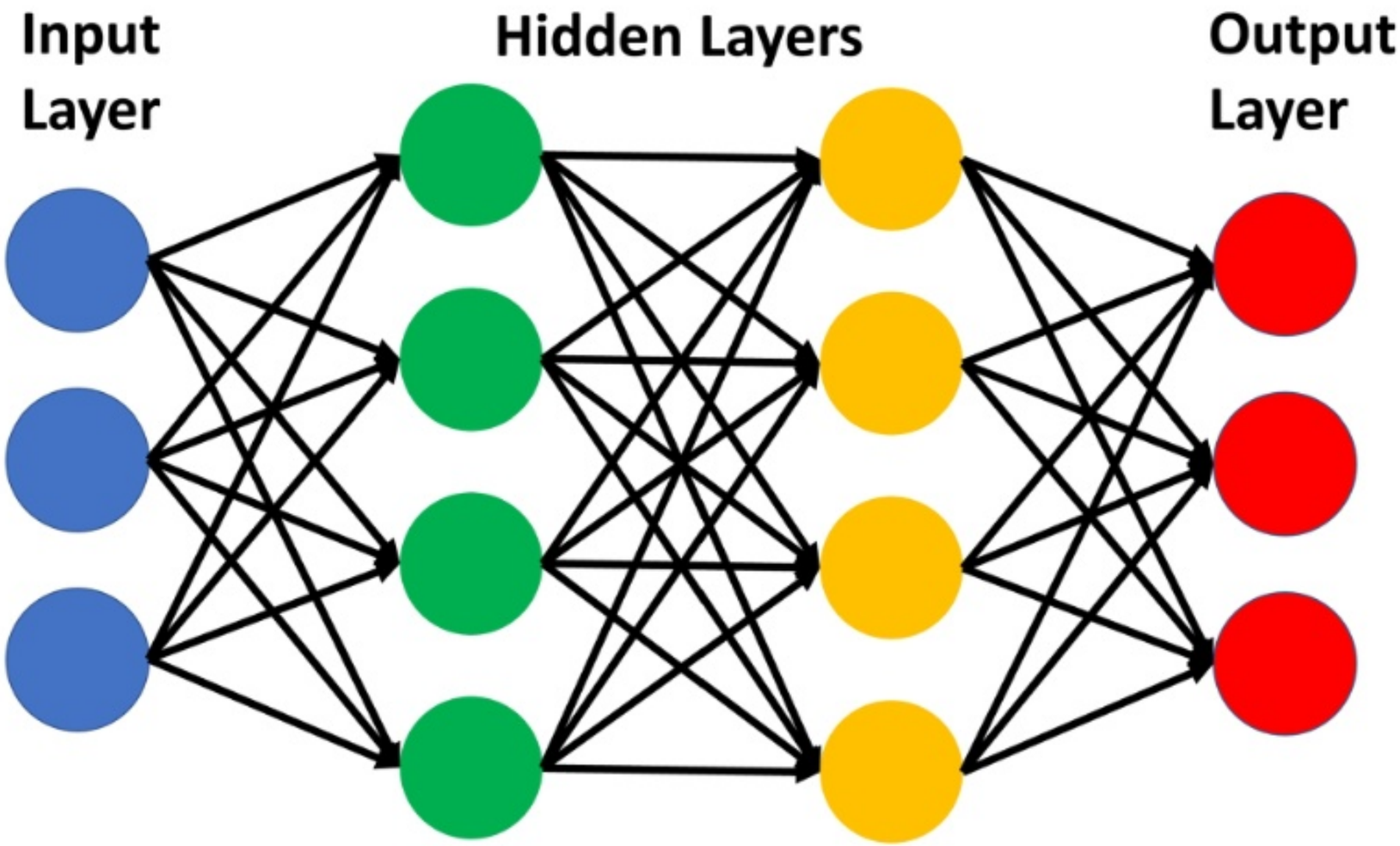


Figure 1.2 Deep Neural Network

The term "deep" in deep learning refers to the depth of the neural network, indicating the presence of multiple hidden layers between the input and output layers, and it's not uncommon for modern deep learning models to have tens or even hundreds of layers. Deep neural networks allow for the extraction of hierarchical features and representations

from the input data. The hidden layers in deep learning models enable the network to learn complex patterns, relationships, and abstractions in the data, leading to more accurate predictions or decisions.

Deep learning has found many narrow applications in a wide range of fields due to its ability to automatically learn and extract patterns from large datasets. Some of the key applications of deep learning include:

Computer Vision: Deep learning plays a role in computer vision applications. It works well in tasks such as image classification, object detection, and image generation. This technology is widely used in facial recognition systems, autonomous vehicles, and surveillance setups for identifying and tracking objects or individuals within images or video streams.

Natural Language Processing (NLP): In the realm of natural language processing, deep learning models like Transformers have made significant advancements. They enable language translation services like Google Translate, sentiment analysis for social media monitoring, and the creation of chatbots and virtual assistants capable of engaging in natural language conversations.

Speech Recognition: Deep learning powers speech recognition systems that transcribe spoken language into text. This technology underlies voice assistants such as Siri and facilitates speech-to-text applications across various domains.

Healthcare: Deep learning has made significant contributions to healthcare. It aids in medical image analysis, facilitating the diagnosis of diseases from X-rays, MRIs, and CT scans. Additionally, it plays a role in drug discovery and health monitoring, helping analyze patient data for early disease detection and outbreak prediction.

Autonomous Systems: The development of autonomous systems, particularly in self-driving cars and robotics, currently heavily relies on deep learning. Deep learning models are used for perception and decision-making in autonomous vehicles, enabling them to navigate safely and efficiently within a limited scope.

Finance: Deep learning is leveraged in the finance sector for algorithmic trading, where it analyzes financial data to make trading decisions. Additionally, it contributes to fraud detection, helping financial institutions identify and prevent fraudulent transactions and activities.

Recommendation Systems: Deep learning powers recommendation engines, which are widely used in e-commerce and content platforms to provide personalized recommendations to users. Prominent examples include Netflix and Amazon, which use deep learning to suggest movies and products to customers.

Art and Creativity: Deep learning models like DeepDream, Midjourney, and Dall-E are used in art and creativity to generate art pieces. AI-generated music compositions and other creative outputs are also made possible through deep learning algorithms.

Language Generation: Deep learning models are employed for language generation tasks, including content creation, news article writing, and creative writing. They can generate human-like text across various genres, and with the advent of large language models have become very useful for doing so.

These diverse applications demonstrate the versatility and transformative potential of deep learning across a wide range of industries and domains. The ongoing research and development in this field continues to expand its capabilities and impact, but the applications will remain narrow with each DNN implementation specific to a certain task and not able to generalize to other tasks or areas.

Now that we've learned a bit more about what deep learning is and what it is used for, let's take a more in-depth tour of the primary machine learning and deep learning methods, and give a brief description of how they each work. Then we will dive into more detail of their use, function, and their limitations in some specific examples.

Convolutional Neural Networks (CNNs) have a hierarchical structure (which is usually 2D for images), where at each layer, an image is sampled by (trained) convolution filters into a lower resolution map that represents the value of

the convolution operation at each point. A convolution is an operation multiplying each point in the image within a given area by each corresponding point in the convolution map and summing all those products to get a result in the target layer. In images the information in the layers goes from high-res pixels, to fine features (edges, circles,...) to representations of coarse features (noses, eyes, lips, ... on faces), then to the fully connected layers that can identify what is in the image. Finally, the final convolution result is flattened into a 1-D array of neurons whose values classify the input into a given category.

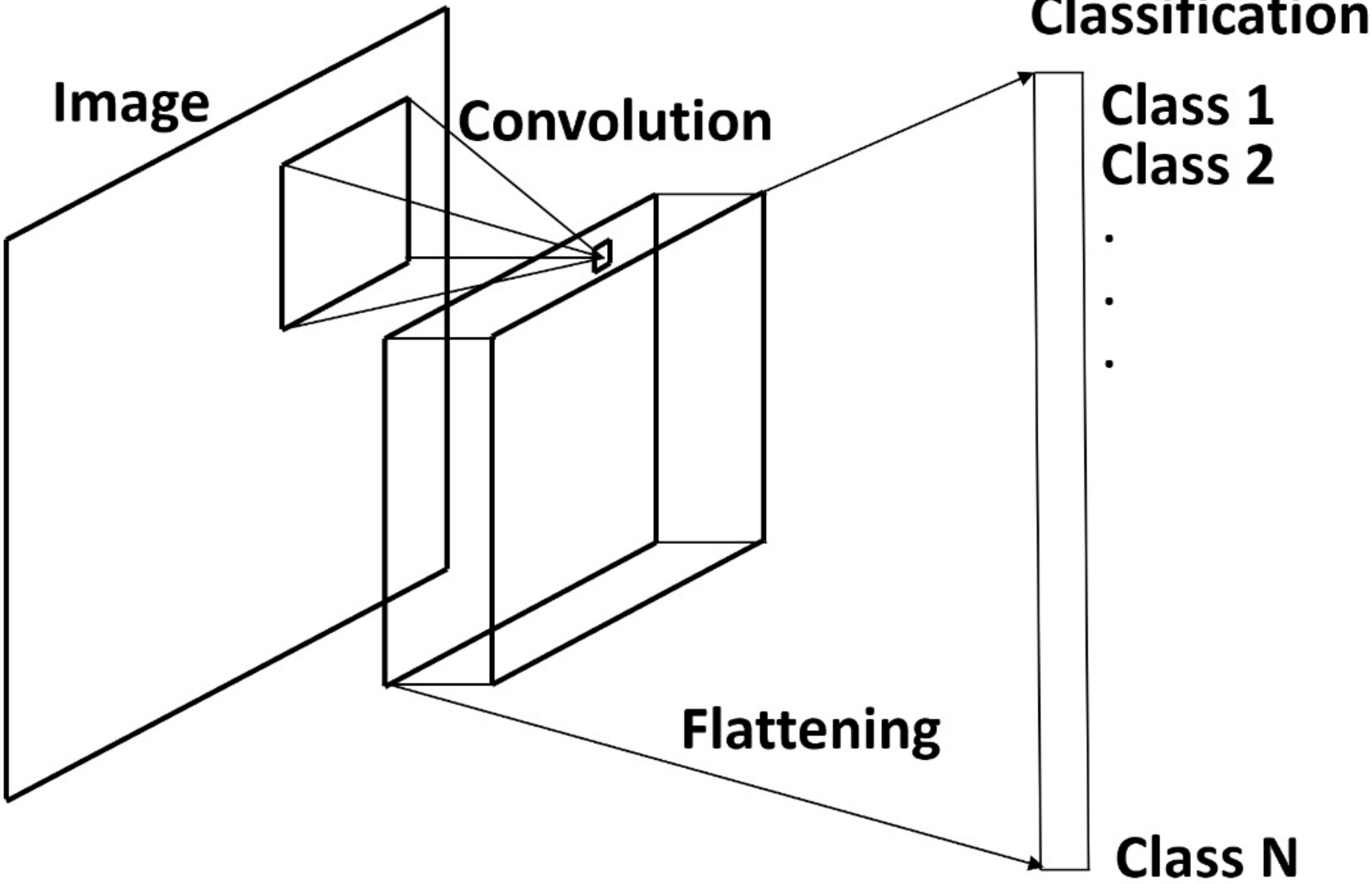


Figure 1.3 Convolutional Neural Network

The useful part of CNNs is that the convolutional filters are trained when you train the network. For decades, computer vision researchers had hand-crafted filters like this, but could never get results as accurate as CNNs can get. Additionally, the output of a CNN can be a 2D map instead of a single value, giving us image segmentation, or identifying objects in the different areas of the field of view. CNNs can also be used on many other types of 1D, 2D and even 3D data.

Recurrent Neural Networks (RNNs) work well for sequential or time series data. Basically each 'neural' node in an RNN is kind of a memory gate, often an LSTM or Long Short Term Memory cell that retains some state from the past as well as the current state. When these are linked up in layers of a neural net, these cells/nodes also have recurrent connections looping back into themselves and so tend to hold onto information that passes through them, each retaining a limited memory of its state, and allowing processing not only of current information, but of some, limited past information in the neural nodes as well. As such, RNNs are used for time sequential operations like language processing or translation, as well as signal processing, Text To Speech, Speech To Text, and so on.

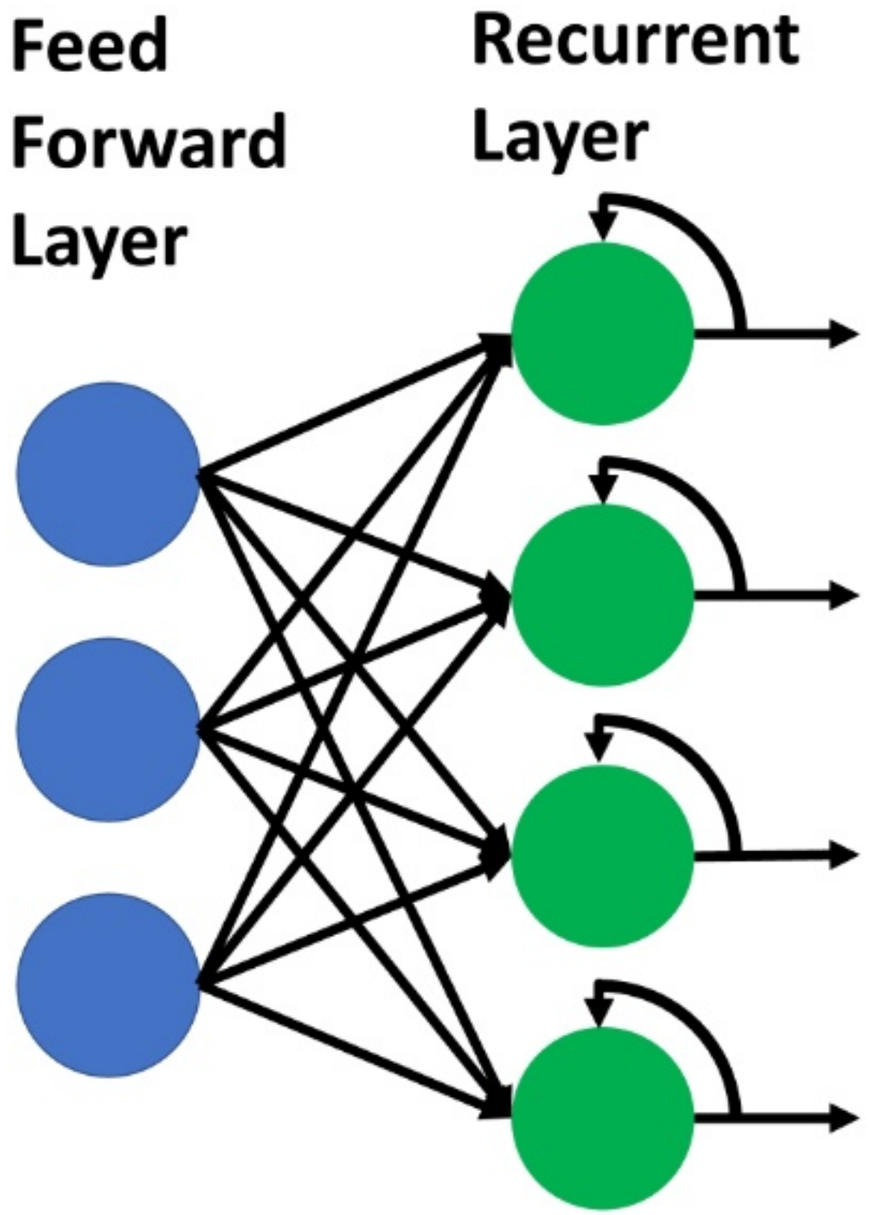


Figure 1.4 Recurrent Neural Network

Autoencoders are a pair of DNNs where an encoder and decoder are paired together with the encoder feeding its output from its lowest layer into the lowest layer of the decoder as input. The autoencoder is trained on the criteria that the output from the decoder must match the input to the encoder. In the autoencoder, the data is passed down the encoder, through the mid-section that creates a constriction or narrowing during the autoencoding process, and then the data is expanded again through the decoder.

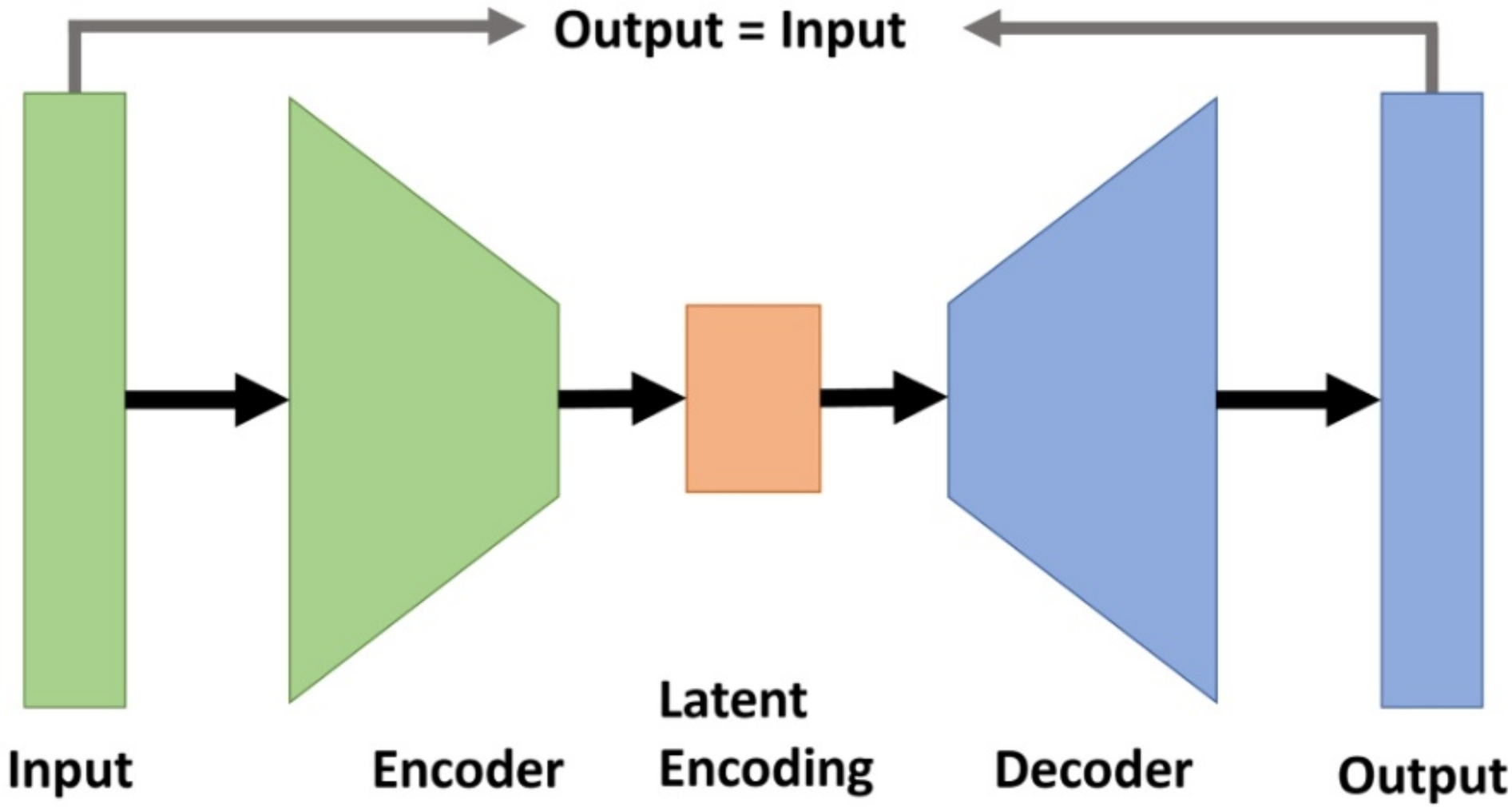


Figure 1.5 Autoencoder

By doing so, the data is compressed at the constriction, but in a way that the autoencoder neural network stores all the common features of the entire data set it has encoded to date (a set of basis vectors), and the data at the constriction is the set of basis coordinates referencing the basis vectors internal to the autoencoder. If we take the output from the area or volume of constriction for each input and we record this into memory in time, it forms a set of basis coordinates or encoded 'key' from which we can ideally reconstitute the input data. In practice, CNN Autoencoders work well with 2D images, but cannot handle more complex spatial-temporal data like video, computer vision or speech very well, which is where we really need such functionality.

GANs, or Generative Adversarial Networks, are a technique similar to autoencoders in that they learn to do a task unsupervised, but with a different underlying architecture. They are used with CNNs to create image discriminators and generators. The discriminator is a CNN that is trained to recognize images as synthesized or real. The generator is an inverse network that takes seed images and uses them to generate more detailed images. The discriminator evaluates the output of the generator and sends signals to the generator on how to improve, and the generator in turn sends signals to the discriminator to improve its accuracy as well, going back and forth in a zero-sum game till they both converge to the best quality of generating images from a seed and discriminating AI-generated images from real ones. This is a method for providing self-reinforcing feedback to do unsupervised training of an AI system, which we will revisit later. Again, GANs with CNNs work ok for static images, but do not perform as satisfactorily for spatial-temporal inputs like video, computer vision, or speech.

Autoencoders and GANs are examples of deep neural networks that use feedback to train, not requiring labeled data, but also having the internal representation or label for the data they operate on being obscured and not in a human-readable format. The output can still be used in other machine learning operations like clustering or prediction on these internal representations.

Transformers are the next step beyond RNNs for processing sequential data like language. They work by the input being sampled as a sequence of tokens (such as words in a sentence), along with the weights contributed to that token by the other tokens in that sequence, with those weights broken down into several matrices that are trained and used to calculate this weighting. Transformers then use these inter-token weights and attention mechanisms to focus the computation on specific parts of the sequences.

For language applications, the Transformer is trained on a corpus of text, by either getting it to predict the next word in the text (GPT), or by dropping a percentage of words from the text and getting it to predict them (BERT). Back-propagation is used to refine the weights, just as in a DNN. Then, once pre-trained, transformers can be further trained with input and output sequences, such as translating a sentence in one language to another, or for answering specific

questions with specific answers. While the transformer method improves accuracy over RNN sequence to sequence learning and similar methods, it still relies on the same underlying principle that uses only sequences of input information to generate sequences of output information by using statistical inference, and it does not process any underlying meaning, or perform cognitive processes on the inputs to do so.

The key takeaway from such LLMs is that the power to predict what comes next in a sequence, particularly with language, is a very powerful tool in AI, and this type of generative prediction is likely a fundamental aspect of intelligence - which we will see in our AGI design later.

Stable Diffusion is a technique combining autoencoders and transformers to generate very realistic images from text. It starts by training on a large dataset of both images and the text describing them by autoencoding the text and image into a common latent data space of tokens. Then, at inference time, when a text string is provided, an image of random noise is generated and decoded to a string of text tokens. This is compared with the input string, and adjustments are made to the image to make its string of text tokens closer to the input. This continues until an image with its descriptive text matching the input text emerges from the noise. It can create original and very realistic images matching the input text.

Reinforcement Learning is another AI method, where you train a learning agent to solve a complex problem by simply taking the best action given each state in the problem. The probability of taking each action at each state is defined by a policy. An example is running a maze, where the position of each cell is the 'state', the 4 possible directions to move from it are the actions, and the probabilities of moving each direction (action), at each cell (state) forms the policy.

By repeatedly running through the states and possible actions and rewarding the sequence of actions that gave a good result (by increasing the probabilities of those actions in the policy) and penalizing the actions that gave a negative result (by decreasing the probabilities of those actions in the policy) in time the algorithm arrives at an optimal policy,

which has the highest probability of a successful outcome. Usually while training, you discount the penalties/rewards for actions further back in time.

In our maze example, this means allowing an agent to move through the maze, choosing a direction to move from each cell by a random choice, weighted by the probability of moving each direction from that state, as represented in the policy. When this solution reaches a dead-end, we penalize the series of choices that got it there by tracing backwards along the path it took and reducing the probability (in the policy) of moving that direction from each cell again. If the agent finds the exit, we trace back along its route and reward the choices that got it there by increasing the probability of moving the direction from each cell that it did. In time the agent learns the fastest way through the maze to the exit, or the optimal policy.

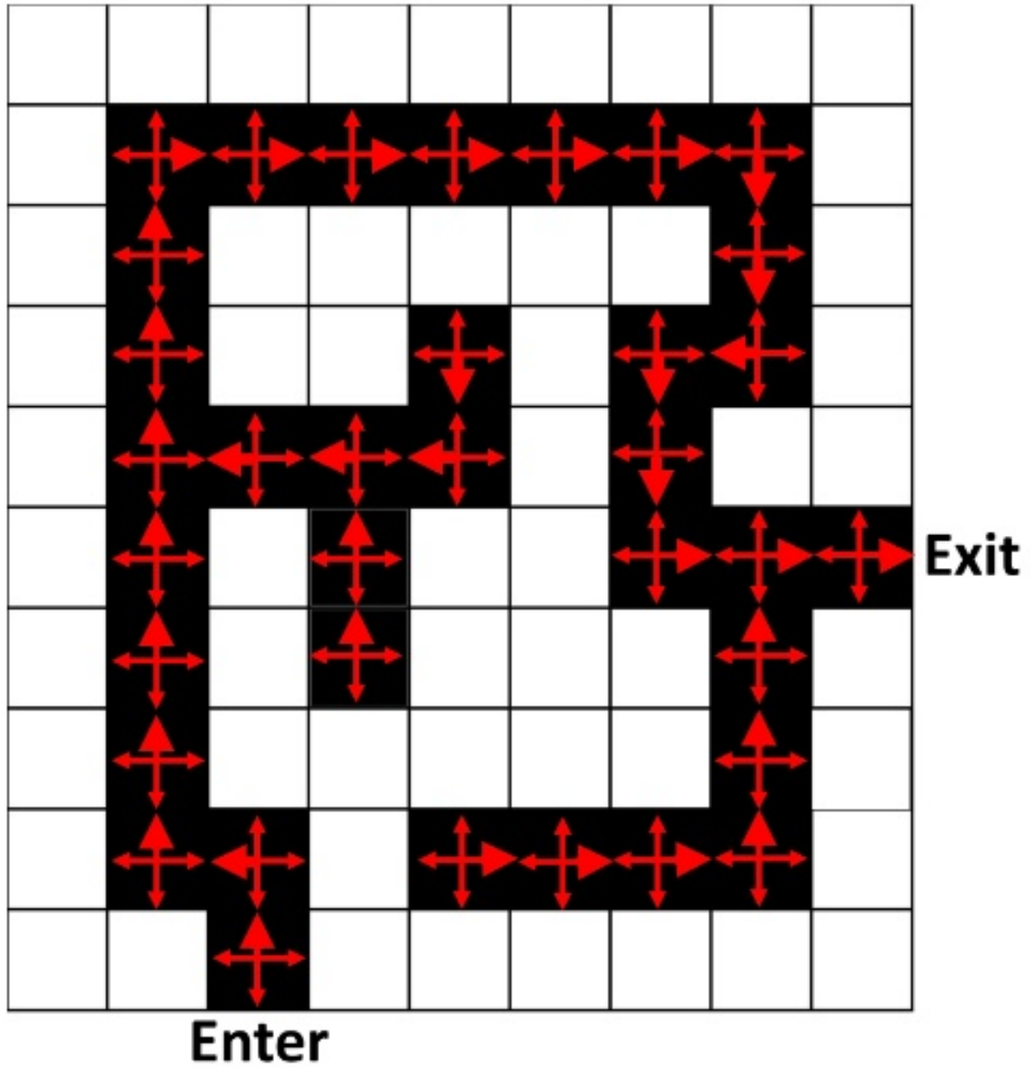


Figure 1.6 Reinforcement Learning

An example of a trained policy in our maze example is represented by the red arrows in Figure 1.6, with the larger red arrows in each cell showing the increased probability of moving in the direction that will take you to an exit.

Deep Reinforcement learning is a more advanced implementation of RL that uses a DNN to evaluate a more complex state, such as a video input into a state vector that is more easily implemented in the policy.

Reinforcement learning learns the best path through a state space to solve a problem and variations of deep reinforcement learning are at the core of the Google Atari Video Game playing AI and (combined with Monte Carlo methods) their AlphaGo AI.

Databases are leveraged by 2020s Deep Learning. This is a realm where computer science already far surpasses human capability in storage, search, access, and recall (both in the amount of data and in the speed and accuracy of recall). One of the things that makes computers, the Internet, and the applications running on them so powerful (including AI) is the enormous databases they can access and search to deliver us the information needed when we need it, usually by using Structured Query Language (SQL) for tabulated data. This could also be numerical information from stock charts of every company on the planet, going back decades, to text of online encyclopedias that contain all of human knowledge in multiple languages. Although these databases often share a common underlying operating system and infrastructure, each is structured in its own format and the only commonality is often that they use human-understandable or SQL methods in the interface to query them. Later in this chapter we describe how deep learning and databases are combined for voice assistant applications.

Scientific Computing's goal is to run simulations where the state of a physical system is set to initial conditions, then numerical simulation is run forward in time to predict how the system will progress. By repeating the simulations

using different initial conditions and runtime parameters to explore different configurations of the system, the user can find the best configuration to solve a given problem.

Scientific computing models physical problems with differential equations, which are then discretized into finite difference equations while the data for the problem is discretized into numerical basis sets, often a grid, Fourier basis or wavelet basis with two or three dimensions. In Chapter 5 we describe how this decomposition of real-world inputs and outputs into sets of basis vectors and basis coordinates can be applied to an AGI design.

Regardless of the discretization method and the basis set, finite difference equation operations on this data can usually be decomposed into large matrix operations that solve the difference equations. The LINPAK numerical benchmark is used to evaluate numerical performance for systems that run these matrix computations, common to most scientific computations.

Of course, there are rich variations and combinations of all of these methods. Combined, these are the bread and butter of what we call AI today, but perhaps prematurely, as these methods do not have any cognition, intelligence, or intuition, and instead use brute-force statistical analysis or pattern recognition and often require large amounts of (usually labeled) data to train to a given standard. And they must be given an input to produce a given output. In the next sections, we look at these limitations and show the boundaries of what these methods are capable of, and where they fall short.

Limitations of Deep Learning in The Early 2020s

While deep learning has shown some remarkable capabilities, it is important to acknowledge its limitations. Some of the key limitations of deep learning include:

Deep learning models require large amounts of labeled data to perform effectively. Training deep neural networks often necessitates vast datasets, which may be expensive, time-consuming, or challenging to obtain, particularly in certain

specialized domains or industries. In cases where labeled data is scarce or biased, deep learning models may struggle to generalize and make accurate predictions.

Deep learning models are computationally intensive and demand substantial processing power and memory. Training complex deep neural networks can require high-performance hardware infrastructure, including GPUs and specialized accelerators. The computational requirements may pose challenges for organizations or researchers with limited resources, hindering the widespread adoption of deep learning in certain contexts.

Deep learning models often operate as black boxes, making it difficult to understand the underlying reasoning behind their decisions. As the number of layers and parameters increases, interpreting and explaining the model's predictions becomes increasingly complex. This lack of interpretability can raise concerns in critical applications such as healthcare and finance, where explanations and justifications for predictions are essential.

Although DL can work very well for model problems and can outperform a human for specific benchmarks, these techniques often do not scale or work as well once you try to apply them outside those specific problems they were designed for. For real-world problems, they often do not perform as well as they do on model problems, even if you scale-up the network topology and tune them extensively. Sometimes you just do not have enough data to train them sufficiently to make them robust and accurate in deployment.

Deep learning models are susceptible to overfitting, which occurs when a model performs well on the training data but fails to generalize to unseen data. Overfitting can happen when the model is too complex relative to the available data or when the training dataset is noisy or unrepresentative of the target population. Addressing overfitting requires careful regularization techniques, cross-validation, and robust evaluation strategies.

Deep learning models excel at pattern recognition and can process vast amounts of data. However, they lack a deep understanding of context and common sense reasoning that humans possess. Deep learning models struggle with

tasks that require reasoning beyond the information explicitly available in the data. Complex language understanding, nuanced decision-making, and tasks requiring explicit logical reasoning still pose significant challenges for deep learning systems.

Large language models like ChatGPT have the unfortunate ability to produce outputs that are factually different from the data they were trained on. ChatGPT has been known to alter dates of historical events and simply fabricate stories about real people or events that seem confident and compelling. Though this ability helps produce creative writing, the models can give bad advice and scientifically inaccurate responses that were not found in the underlying training data.

Deep learning models can be susceptible to adversarial attacks, where carefully crafted inputs can mislead the model's predictions. By making minor modifications to the input data, an attacker can deceive the model and cause it to make incorrect or unintended predictions. This vulnerability to adversarial attacks raises concerns in security-critical domains such as autonomous vehicles and cybersecurity, requiring additional measures to enhance the robustness of deep learning models.

Sometimes, the real-life problem just can't be quantified well enough to fit within the limited capabilities of a DL method. For example, the Imagenet image classification competition has 1000 object classifications, but in a real-life application, there are probably millions of objects and sub classifications that would be needed to identify everything in the environment. To get the DL systems to do new things, or recognize new objects, new training data sets must be constructed manually and labeled by hand, and the DNNs have to be re-trained and re-deployed, and cannot learn *in situ*, on the fly, in the field. Retraining the entire model with a number of new examples can even break the model's ability to give previously correct classifications, a phenomenon known as catastrophic forgetting.

As well, many applications require combining multiple DL techniques together and finding ways to fuse them. A simple example is video tagging - you pass the video frames through a CNN to classify or segment it, and the output

of that CNN is fed through to an RNN or Transformer to capture the temporal behavior of the features in those videos with time.

As an example, the author, Brent Oster, helped a researcher/entrepreneur use this architecture in a computer vision application to recognize facial expressions of a quadriplegic to issue commands to their wheelchair and robotic hand prosthesis, with a different facial expression/gesture paired with each command. It will work, but as you scale it up in complexity, it becomes time consuming and tricky to develop and train the network, because you now have to tune two different types of DL networks that are integrated, and it is sometimes hard to know what effect these tweaks are having on which networks.

Later in this chapter, we will do a deep dive on a robot controller that has multiple CNN/RNN networks feeding input, a deep reinforcement learning engine making decisions on this input state, then driving generative networks creating output. These are a lot of specific DL techniques hacked together to accomplish a set of tasks. Will it even work? We will look at this complex use case and examine how DL techniques fall far short of being able to solve this problem in many areas.

No matter how large they scale, DNNs and their derivatives will never result in an AGI, or human-like artificial intelligence because they were never designed with the necessary capabilities and have too many limitations (as outlined above). In the following sections, we will look at two specific examples where they fall short: Interactive voice assistants and Robot AI and see how the inherent limitations of deep learning prevent 2020's AI techniques from adequately filling these roles.

Case Study: Interactive Voice Assistant with 2020's AI

Today's voice (and text) assistants or chatbots are designed to take in human language as input, do some processing on it, and provide language as output, and they work kind of like this at their core:

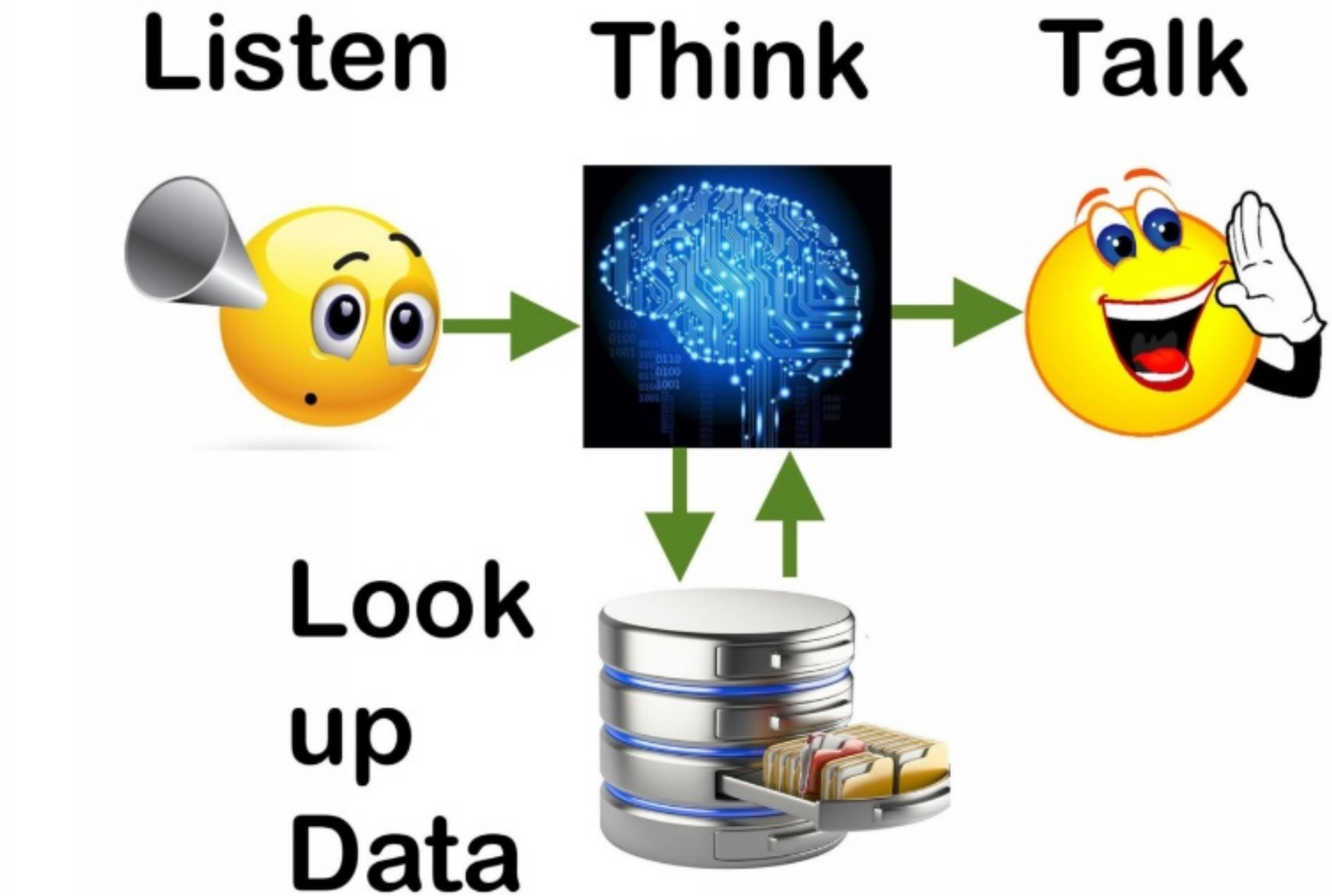


Figure 1.7 Diagram of Speech AI Functionality

While this is very oversimplified, we can break down each step into more detail.

Listen - after hearing the keyword or sensing a button press, the device starts recording the waveform of your voice, usually until you pause speaking long enough that it thinks you stopped. Then it sends the waveform of your voice (or

is already streaming it) to a server using a speech-to-text (STT) engine. This can be a CNN-RNN neural network or other type of computational engine that translates the waveform of your voice into a string of text so it is easier to work with.

Think - in this stage the verbal chatbot uses natural language processing (NLP) to statistically match the text with pre-scripted input phrases and pre-scripted answers to them, and/or turn the text into something the computer can understand, usually a command and set of parameters, like an SQL query. You may have said "Find me a hotel within 5 miles of San Jose, California, for between $200 and $300 with at least a 4-star rating." This gets turned into a command:

Query Type: Hotel
Location: San Jose, CA
Range: 5 miles
Price: $200 - $300
Rating: 4+ stars

Look Up Data - Now this query is sent to an appropriate database, like Expedia or Travelocity, which returns a list of hotels that meet the search criteria. Searches can also be done on keywords for databases like Wikipedia.These can be returned in HTML format, with pictures and links on your device, and be displayed textually and graphically. Your chatbot can also reply verbally.

Speak - Perhaps the chatbot now takes the top result and composes the information in it into a text string like: "The top result I found was the San Jose Doubletree near the Airport, at 2050 Gateway Place, San Jose. It gets 4 stars on Expedia. Would you like to book it?"

Then this text string gets converted into a speech sound clip by a text to speech (TTS) voice synthesizer, and that audio clip gets sent back to you to be played on your device or computer. Essentially, the speech client could be implemented

on something as simple as a walkie-talkie, as all it has to do is transmit voice clips to and from the server doing the speech AI.

Although there is a sophisticated string of impressive processing steps on the server, often using a variety of deep learning trained systems (some older systems used Markov chains and other, older methods of machine learning), we would not call it AI yet, as it does not show any true cognition, deep problem solving, nor real conversational capability. It just learns to recognize keywords, and parse commands and parameters out of the spoken sentences and construct answers using statistical models to generate the most probable answer.

To make it seem more human, programmers often add some simple conversational capability to this kind of verbal query engine by using natural language processing (NLP) and scripting question / answer pairs by hand in the AI to make it seem clever when it directly answers a specific question with a specific human coded answer. Sometimes you can make it include a variable in the answer, like the person's name, or the location, or date, inserted into the answer. Much of today's chatbots use thousands of such question / answer lines to seem intelligent, especially for the demos we see of them. There is even a list of 84 questions that people are most likely to ask a chatbot / AI that we can provide canned answers for that help define the AI's 'personality'.

The next level of sophistication is to use a deep learning question / answer system trained on an enormous number of real human chats or conversations using sequence-to-sequence RNNs or transformers. The input side of the RNN/Transformer receives the textual question, and the output side provides the textual answer. But such sequence to sequence question/answer systems are still mostly limited to providing a given answer to a given question, in isolation, and have limited context about the setting, or the history of the conversation.

If an RNN/Transformer system encounters a completely novel question, it has to guess and ad-lib an answer from the ones it was trained on. Often the results are incorrect, confusing, sometimes humorous, and often not really useful. This does not constitute a model for intelligent language understanding or cognition, nor does it provide truly

intelligent answers. Instead, it basically just answers queries with the statistically most likely phrases it has learned during training.

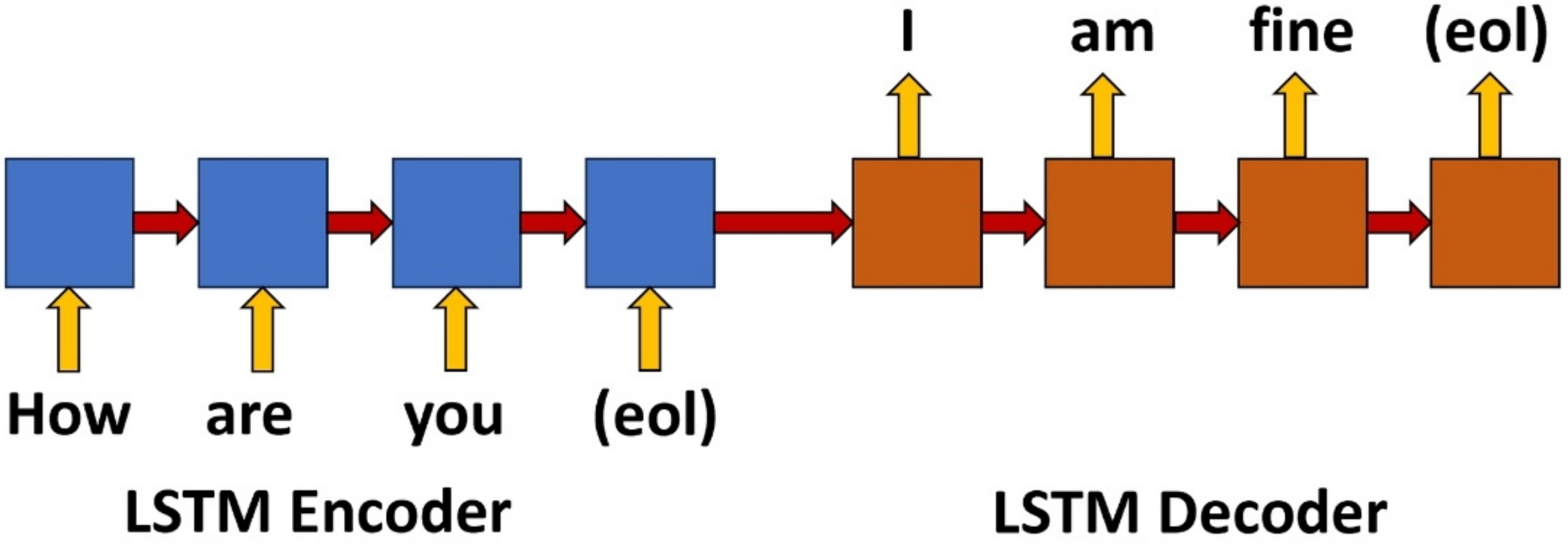


Figure 1.8 LSTM Encoder-Decoder Sequence to Sequence RNN

What these chatbots still do not have is naturally flowing interactive conversation with rich knowledge of past dialog, context, and setting, and they have no ability to compose much richer, fuller sentences as answers, with context awareness. Fundamentally, this is harder, and requires much more than what the deep learning models have for parsing input, instead needing to make more advanced decisions about what to say, and expanding what is said to include context and information from the chat history. Large language models get us part of the way there, but they still lack truly conversational capability, and have drawbacks.

OpenAI's GPT and other large language models that utilize transformers have, as of 2023, shown great power and flexibility in processing textual information and generating impressive textual outputs with wide knowledge access

from simple text prompts. ChatGPT, an application built on GPT4, is capable of answering general knowledge questions and composing writing based on a prompt, using transformers to author stories, poems, essays, even taking exams and getting passing grades. It can process natural language queries, and build detailed answers based on the data it was trained on. It almost seems like a step towards general AI but is really based on a simple trick.

GPT4's large language model uses transformers to learn to predict the next word in a sentence from the words and sentences preceding it. That's it. By learning to predict that next word very accurately by using a very large transformer model with over a hundred billion parameters trained on trillions of words of text, it implicitly builds a model of language and the concepts behind it to predict that next word. However, this model is prone to errors, as it does not have any understanding of the actual concepts, just a statistical model of how the words and statements in the documents it was trained on tend to co-occur. It only outputs information based on what it has trained on, and it cannot interpret it or ensure that it is factually correct. It cannot reason or make logical predictions or decisions based on abstract concepts, it just does a very good job of mimicking, doing so based on what humans have written about in the past and how they organized their writing.

By using human training and reinforcement learning on tasks such as question answering, essay writing, coding, and test taking, ChatGPT was placed as a layer on top of GPT3 in 2022 (and later GPT4 in 2023), which allowed it to do these language processing tasks at a surprising level of competency.

However, ChatGPT admits its conversational shortcomings explicitly when it is asked:

Lack of Real Understanding: While ChatGPT can generate coherent and contextually relevant responses, it does not truly understand the text in the way humans do. It relies on patterns in the data it was trained on and may produce answers that sound correct but are factually inaccurate or nonsensical.

Sensitivity to Input Phrasing: The way a question or prompt is phrased can significantly affect the quality of the response. Slight rephrasing may lead to different answers or even confusion.

Limited Context Window: ChatGPT has a limited context window, which means it might not remember details or references made earlier in a long conversation. This can lead to inconsistent responses.

Overuse of Certain Phrases: It has a tendency to overuse certain phrases, which can make responses sound repetitive. For example, it might start many responses with "I'm not sure" or "That's a good question."

Inappropriate Content: Despite efforts to filter out inappropriate content, ChatGPT can still produce responses that are biased, offensive, or objectionable. It may not always recognize or reject harmful instructions.

Lack of Common Sense and World Knowledge: While ChatGPT has access to a vast amount of information up until its knowledge cutoff date in September 2021, it doesn't have real-time internet access and may not be aware of recent events or developments.

Tendency to Make Things Up: When faced with questions it doesn't know the answer to, ChatGPT may generate plausible-sounding but incorrect information.

Difficulty with Abstraction and Creativity: While it can provide factual information, it may struggle with abstract or creative tasks and might not generate imaginative or out-of-the-box ideas.

Inconsistent Responses: Depending on the phrasing of a question or the order of information provided, ChatGPT's responses can be inconsistent, which can be frustrating for users.

No Personal Experience: ChatGPT lacks personal experiences, emotions, and consciousness. It cannot provide personal anecdotes or experiences.

Long-Winded Responses: It can sometimes generate overly verbose responses that don't get to the point, which can be frustrating for users looking for concise information.

Difficulty Handling Complex Reasoning: While it can handle some forms of logical reasoning and arithmetic, it can struggle with more complex or multi-step reasoning tasks.

Vulnerability to Manipulation: ChatGPT can be manipulated into generating biased or politically charged content by phrasing prompts a certain way. This raises concerns about its misuse.

Resource Intensive: Running large-scale language models like ChatGPT can be computationally expensive, making it less accessible to some users.

Localization and Multilingual Support: ChatGPT may perform better in English compared to other languages, and its performance can vary across languages.

Truly conversational AI requires an intelligence that can actually interpret language, assign meaning to it, and reason with abstract concepts. It requires something that can speak with the user just like a human, with all of the comprehension, cognitive capability and reasoning of a human behind that verbal interaction. That most likely requires an AGI with human-level intelligence and cognition, which will be discussed in later Chapters, and with an AGI speech example in Chapter 6.

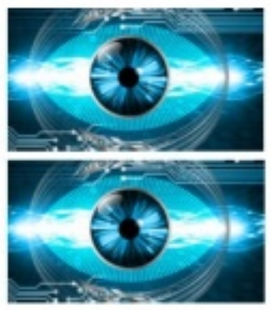
Case Study: Trying to Design a Robot Controller with Today's AI

Building a robot using AI seems like a fairly obvious application for deep learning, but after more than a decade into the AI renaissance, we still do not have useful humanoid robots, ones that can freely navigate our home, do useful things like cleaning, doing laundry, and cooking, and avoid obstacles, pets, and kids in the process. This is because the narrow slices of deep learning available for vision, planning, and control of motors, arms, and manipulators cannot actually take in all this varied input, plan, and do useful tasks in the home.

Existing deep learning and machine learning consist of narrow systems, each with very specific and very limited function, trained on specific datasets that often have to be hand-labeled for the training to be able to classify them.

If today we made a humanoid robot controller that we assembled all of these subsystems - CNN/RNN hybrid networks for vision, with Reinforcement Learning systems for control, navigation, and cognition, then some sort of reverse RNN/CNN for control outputs, it would look like this, hypothetically:

NLP Speech



Stereo Vision

Dual CNN-RNN

CNN-RNN for Sensors

Motion Planning

Deep Reinforcement Learning

Facial Controller

Inv RNN-CNN-GAN

Body Controller

Inv RNN-CNN-GAN

Figure 1.9 Humanoid Robot Controller with Deep Learning

BUT… there are so many reasons this would never work for a general-purpose robot. In 2018. The author was frustrated trying to come up with a provisional patent on AI and motion control for 3D characters and androids using Deep Learning techniques and kept hitting these roadblocks.

a) The sensory systems and vision have only narrow training and cannot identify the vast variety and combination of objects, actions, and events in the actual environment they need to work in.

b) The combination of possible input and possible action states is nearly infinite, far beyond the capability of any reinforcement learning system. Reinforcement learning only learns a policy, or path, through known states to solve a specifically structured problem.

c) The model this system would have of the real world would be skeletal, sparse, ethereal, with no way to fill in the gaps in sensory input, control, and movement, let alone all the states they can combine to form.

d) There is no real planning or intelligence in the middle, nor cognition that can interpret the environment. It would be impossible to train this simple AI on every possible combination of input and actions and would be easily stumped by simple changes in the environment, like a pet in its path.

So, in Chapter 4, we set out to develop a new foundation for AGI and robot controllers that works much better, that solves a bunch of hard problems with one architecture, and collapses a bunch of really complex, clunky, systems into one really elegant solution, modeled after the neuroscience of the human sensory cortex, motor cortex, memory, and how our brain does prediction, planning, and… dreaming.

From our look at the human brain in Chapter 2, how it functions and solves these and other problems, we will then form some requirements for an AGI that can come closer to replicating its capability in Chapter 4.

CHAPTER 2

Neuroscience of the Human Brain

The only general intelligence we know about so far is biological. Essentially any normal human can learn to perform any job to at least a basic level. Humans learn new things through their own experiences and from others. We have insights and epiphanies that go beyond the training we received and the 'datasets' we previously encountered. We transfer learning from one task to a new task almost seamlessly. These general learning abilities are abilities that would have great value for AI because you wouldn't need to build a new AI program for every new task. But these general learning capabilities are impossible to develop stepwise from where AI is today. General intelligence is much different from the ability to identify statistical relationships in data, which is what AI can currently deliver. Instead, we need a drastically different approach.

To better understand what an AGI would need to achieve to match or surpass human cognitive abilities, a deep dive into human intelligence and the intricate architecture of the human brain is necessary. In addition to understanding how the brain works and how it evolved to its present form, we also must unpack the scientific methods underlying our understanding of the brain—including, importantly, what the limitations of those methods are. This helps us reason why understanding a human brain can help us advance toward AGI, but we can't just replicate a human brain in silicon to build AGI.

What is a Brain?

The human brain is a complex biological organ that has evolved in structure and function over hundreds of millions of years. This network of cells has the capability to process sensory input, perceive and model our environment, make predictions, plan, and execute our actions within that environment. As such, the brain helps us survive and thrive, enabling us to pass our genes along to our offspring. The brain is the epicenter of our consciousness and creativity, and grants us our unparallelled abilities in language, art, music, science, and ability to write, make tools, and affect our environment unlike any other animal on the planet.

Importantly, the brain is a biological system, made of specialized cells similar to other organs like the liver or pancreas. It is subject to metabolic constraints and is strongly influenced by the body's internal chemical environment. Unlike any computer, the brain's unique structure and architecture evolved from biological cells and chemical processes, rather than being engineered in silicon.

As tempting as it is to use the metaphor, a brain is NOT a computer. The dramatic differences between a brain and a computer help us understand why natural intelligence is so different from the current AI systems and why these systems still have a long way to go in aligning with the competencies of a human. In a brain, there is no hard drive or RAM. Memories are not written into files. In fact, in the brain, the mere process of remembering alters the memories themselves.

Brains do not perform statistical calculations to make decisions, at least not in any way resembling the statistics used in today's narrow AI. Even terms such as 'processing' of information by the brain should be used with caution, as the mechanism by which inputs change to outputs is very unlike how information flows through a computer. (We will use the term 'processing' loosely throughout the chapter for simplicity.) Though inputs and outputs can be measured from a brain, the separation between input and output is much less defined than in computer chips. Instead, each individual

cell in a biological system performs both input and output functions, transformations, and many other processes which we do not fully understand.

Functional Principles of a brain - movement, relative change, and internal perception.

> *We have a brain for one reason and one reason only - that's to produce adaptable and complex movements. Movement is the only way we have affecting the world around us… I believe that to understand movement is to understand the whole brain. And therefore, it's important to remember when you are studying memory, cognition, sensory processing, they're there for a reason, and that reason is action.*
>
> Daniel Wolpert

If we learn only one principle from biological intelligence, it is that brains are built to produce movement. This principal contrasts greatly with computers, which were built for performing calculations. The very first nerve signals hundreds of millions of years ago coordinated activity across distant parts of early animal bodies even before they condensed into a brain. Any cognition or intelligence that developed along evolution's path most likely enabled more sophisticated movements or better action planning in dynamic, uncertain environments. Sea squirts show us this is true. Sea squirts go through a mobile larval stage where they navigate around the ocean using a simple brain. They eventually settle on a comfortable rock, where they stay for the rest of their lives. The first thing they do once they reach this sedentary state is to eat their own brains because they don't need to move anymore! Studying the structures and functions of the brain, in the context of how they contribute to movement or planning future movement, will help us understand the intelligence features we want to mimic in AGI.

Secondly, unlike AI, biological intelligence has never been about ingesting data. Instead, to produce appropriate movement, the brain is posed to identify and respond to differences in the current state of the environment. Retina cells in your eye, as an extreme example, will quickly overexert themselves and stop signaling altogether if there is zero movement in their field of vision. You literally go blind (temporarily) in a static environment.

Relativity is the important point from a biological standpoint. There is a constant level setting throughout the body and brain that makes us highly responsive to differences but leaves us somewhat vague on absolute values. One day 68 degrees Fahrenheit seems hot to us; on another day the exact same temperature seems cold. In the spring newly emerging greens seem greener because we have been used to a drearier existence. Later in the summer, the same green appears less striking after being exposed for so long (Welbourne et al., 2015). When lifting objects, we can distinguish about a 5% difference in weight between two different objects. Interestingly, it doesn't matter whether the things we are lifting are 1 kg or 100 kg, the 5% rule applies. That means that we can sense the difference between 1.00 and 1.05 kg, but not the difference between 100 kg and 100.05 kg, even though these are the same absolute differences.

Basically, your brain gets used to a particular level of inputs. Any deviation from this level is easily detected. In dynamic environments, you get used to whatever patterns form, and then you respond when changes to these patterns occur. We identify the change in level much more easily than the particular levels themselves because it is this change that is important for our survival. We don't waste energy processing that which is static, and hence, irrelevant.

Third, what you perceive as a human is not a realistic representation of the inputs you have received from your senses. Most of what you perceive is not actually there, but instead made up of predictions and alterations your own brain created to make sense of, and be one step ahead of, your environment. Your perception is your reality, and psychoses such as schizophrenia demonstrate just how real your brain's model is to you, despite the differences between that model and what is actually present in your environment. From a wiring standpoint, the majority of inputs are not going directly into your cortex, but rather, have been modified substantially from all over your brain before landing on regions thought to be involved in conscious awareness. A computer has to make a model of the world from data fed into it from somewhere external. A brain doesn't have this constraint. Some of your brain's model is generated from 'data' created internally!

How humans came to predict and produce movement in such a generalized capacity is an epic story through evolution of molecules and biological systems. The core principles of biological intelligence (how we produce movement, respond

to change, and generate internal realities) are drastically different from the base functionality of today's AI. To better understand how AI works in comparison to the core functions of a human brain, we will now take a nuanced look at both the cells and the molecular networks that underlie our biological intelligence because these building blocks are the basis for an intelligence that is so vastly different from a computer.

Building Blocks of the Brain: Neurons and Glia

The brain exhibits a sophisticated hierarchical organization, encompassing molecular networks, neural networks, and larger brain structures such as the cortex. These structures have identifiable, repeatable organizations across individuals. The building blocks of brains consist of two major types of cells: **neurons** and **glia**. These cells work in harmony to assemble the complex and expansive—yet predictably organized —network of the brain.

Neurons are interconnected within an intricate neural network that transmits electrical signals between cells through axons, synapses and dendrites. In the human brain, there are approximately 85 billion neurons, with a total of 100 trillion synapses connecting them. For a brief overview of neurons and synapses, we recommend watching the '2-Minute Neuroscience' series on YouTube (links to videos: Neurons and Synapses).

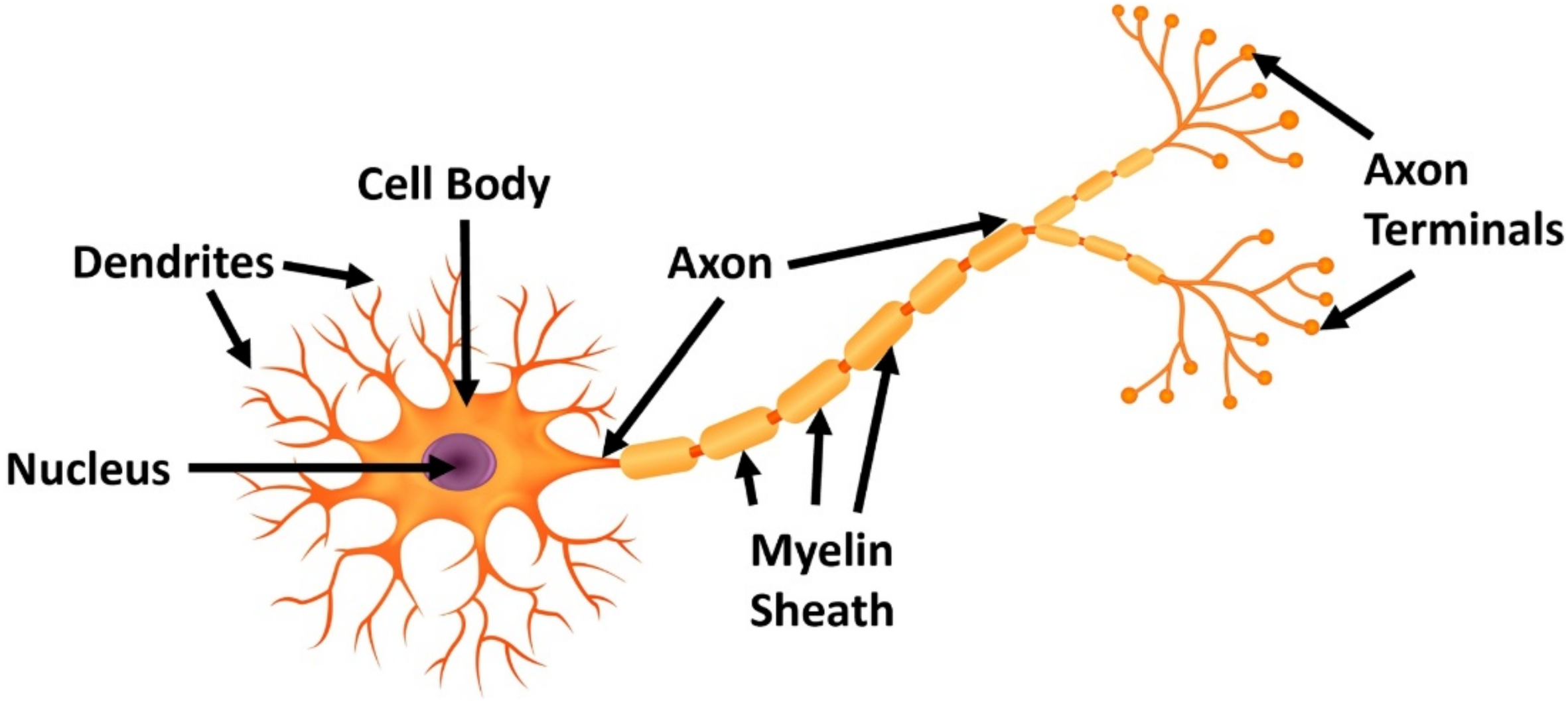


Figure 2.2 Anatomy of a Neuron

Neurons serve as the processing workhorses of the brain. Each neuron absorbs spikes of electrical charge from its dendrites and performs a sophisticated integration of those charges in time and space. As this charge accumulates in the soma (the cell body) of the neuron and reaches a certain threshold, a surge of current forms at the intersection of the cell body and the axon, a region referred to as the axon hillock. The axon hillock, in turn, fires and propagates a signal down the axon. In doing so, the charge in the neuron is depleted and reset. In this process the neuron emits a spike of electricity out along its axon, moving in time and space as that axon branches and re-amplifies the signal,

carrying it to thousands of synapses, where the electrical spike is absorbed by each synapse. This process causes the release of neurotransmitters into the synaptic cleft, where with the help of ambient neurochemistry, they are chemically integrated into the downstream (postsynaptic) cell.

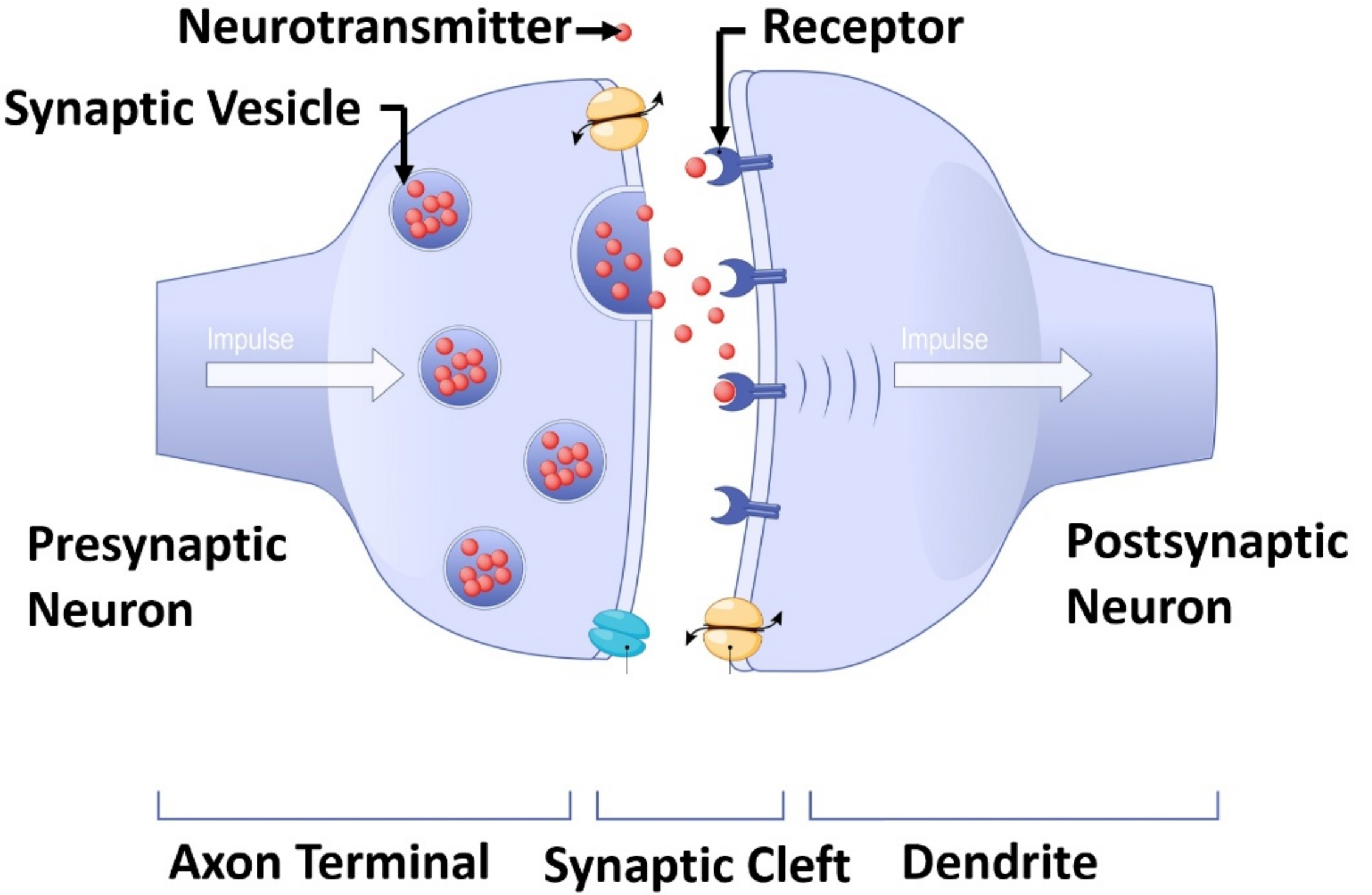


Figure 2.3 Anatomy of a Synapse

During the transmission of a signal from one neuron to the next, neurotransmitters migrate across the cleft to the postsynaptic side, where their accumulation in various receptors eventually, conditionally cause the postsynaptic neuron to emit a spike down along the dendrite to the next neuron. When two connected neurons fire in rapid succession, the synapse between them becomes more sensitive or potentiated, and fires more readily. This is referred to as **Hebbian learning**, or long-term potentiation, and is constantly occurring as we humans move around, sense, and interact with our environment. Hebbian learning is a key mechanism for learning and memory in the brain.

In addition to neurons that provide an excitatory signal, approximately 20-30% of neurons provide inhibitory signals that serve to selectively negate those excitatory signals. By doing so, this enables the dendrites to create complex input signals to the neurons that perform logical computations.

These electrical spike signals, sent down axons and across chemical synapses to the next cell, are what allow neurons to communicate over long distances to specific partners. The changes in signaling itself, potentiation of the signal, and the changes in strengths of the synaptic connections (due to processes such as Hebbian learning) between neurons are what underlie the plasticity of the brain. This complex plasticity is how we can change our thoughts and behaviors through observation and experience.

Glia were once thought to be merely the maintenance cells within the nervous system, removing waste and dead neurons, and creating the blood brain barrier. However, the more we learn, the more glia appear to be involved in the processing of information in the brain as well. Glia actively regulate the formation and destruction of neuronal synapses, which contributes to the plasticity of the brain. Interestingly, glia respond to the electrical activity of the synapses, and release their own neurotransmitters. Individual glia may be responsible for regulating the activity of groups of 1000s of neurons at once. These functions of glia can help coordinate network effects and learning that go

beyond the neuron-to-neuron connections, and as such, may be much more involved in cognition and higher-level brain functions than previously appreciated.

Neuronal Networks are the sum of the neurons connected by networks of axons, synapses, dendrites, and moderated by glia. The structure of these networks, such as the organization of the specific connections within the network play a very large role in the functions of these networks. The sum of all of these connections in the brain is known as the **connectome**. It has been shown that merely replicating the connectome of a nematode worm known as *C. elegans* using simplified artificial neurons can produce a control system for a robot that can navigate and respond appropriately to the environment (see this video for explanation). Therefore, communication among specific signaling partners must be important to intelligent operation.

However, neural networks are much more interesting than just their circuitry. The brain is not just a complicated set of electrical circuits where neurons are decision points and axons are wires connecting individual components. In fact, neurotransmitters, which are chemicals, are directly responsible for communication between two neurons, rather than the electrical action potential itself (except in connections called 'gap junctions'). There are dozens of types of neurotransmitters in the human brain which can carry a diversity of information beyond the action potential which originated the release.

Complex electro-chemical networks, both inside and between brain cells, allow information processing capabilities to far exceed that which would be expected from the numbers of neurons and their connections alone (and 100 trillion connections is an unfathomable number by itself!). These networks of molecules not only regulate the way the neurons fire, and therefore how neurons pass electrical information to other neurons, but they also communicate non-electrical signals and help regulate local and large-scale brain state (such as the way glia insulate and regulate large, independent neural regions). The axon's electrical spikes are only a fraction of the information that is transmitted among neurons. Biological systems, even individual cells, are 'wetware', capable of all sorts of information processing, and probably possess more sophisticated abilities we have yet to discover.

The reductionism of deep learning oversimplifies the model of neural networks and neurons to exclude most of the functional details, and by doing so, misses many of the core functions of the brain. A DL model that has the same number of 'neurons' and connections of a human brain would still be orders of magnitude less capable than a biological brain, as a DL neural network model is orders of magnitude less complex than the electrical, molecular, and cellular signaling networks of a real brain.

Evolution of the Human Brain

How did this complicated structure that we call a brain come to be? The cells that would eventually become the brain have their evolutionary origins over a billion years ago, but for our purposes, we can focus more specifically on brain evolution itself. Starting with a few neurons in the first multicellular animals nearly 600 million years ago, the human brain expanded and evolved into the present form, shaped by the animals' bodies and environment along the way.

Brains began a dramatic change through the Cambrian explosion 530 million years ago. This is the time when eyes, ears and other sensory systems, motor systems, and the accompanying neural systems (and some of the first intelligence) evolved and exploded in an arms race, along with armor, teeth, and claws. Evolution of brains then followed the needs of fish, reptiles, dinosaurs, mammals, and eventually, up the hominids' lineage about 5-10 million years ago. Natural selection honed the capabilities of each brain over generations, selecting for the capabilities that were adaptive to the particular environment at the time.

Because the brain evolved and grew over 100s of millions of years, starting with a nucleus of a few neurons, then growing into the (roughly) 85 billion neuron human brain, many of the functions of the human brain remain distributed and connected. Even though more modern parts of the brain evolved specialized functions related to learning and cognition, they are still connected through the older, more ancient parts of the brain, which participate in these functions as well. Although it is possible to identify some brain regions that primarily contribute to things such

as memory, vision, language, and other functions, many parts of the brain work together as a system for many of these functions.

Much of the older parts of the human brain were evolved in animals for the previous several hundred million years of competition, hunting, and violence, not the last few thousands of years of human civilization, so in many ways, our brain is maladapted for our modern life in the information age. It is not very efficient at many of the tasks we use it for, such as in advanced professions like law, medicine, finance, and administration. A synthetic brain, focused on doing these tasks optimally can probably end up doing them much better without all of the biological constraints and human emotional biases, having a wider data reach in space, deeper data reach in time, and better prediction, planning, and decision making related to modern life.

We do not seek to replicate the brain, its form, nor its specific function, but to evolve a technological analog AGI that can solve similar problems. The next sections will cover structure-function relationships of brain structures and basic brain principles that can help us understand biological intelligence.

Basic Human Brain Anatomy

Let's go through some of that basic brain anatomy that evolution built, tell a little bit about how each of the structures is organized, and give a few of the fundamental functions of each. Anatomically, the brain is divided into distinct regions. There is a hierarchical organization to these brain regions where older, deeper parts of the brain were stacked upon. You might hear reference to your 'reptilian' brain making you respond abruptly and forcefully to something your logical brain says isn't a problem. That's a product of evolution building structures upon older structures. Those visceral, subconscious responses probably kept your ancestors alive, even if they make life today a little more challenging at times. Below are a few of the structures which are important for our discussion regarding building an AGI.

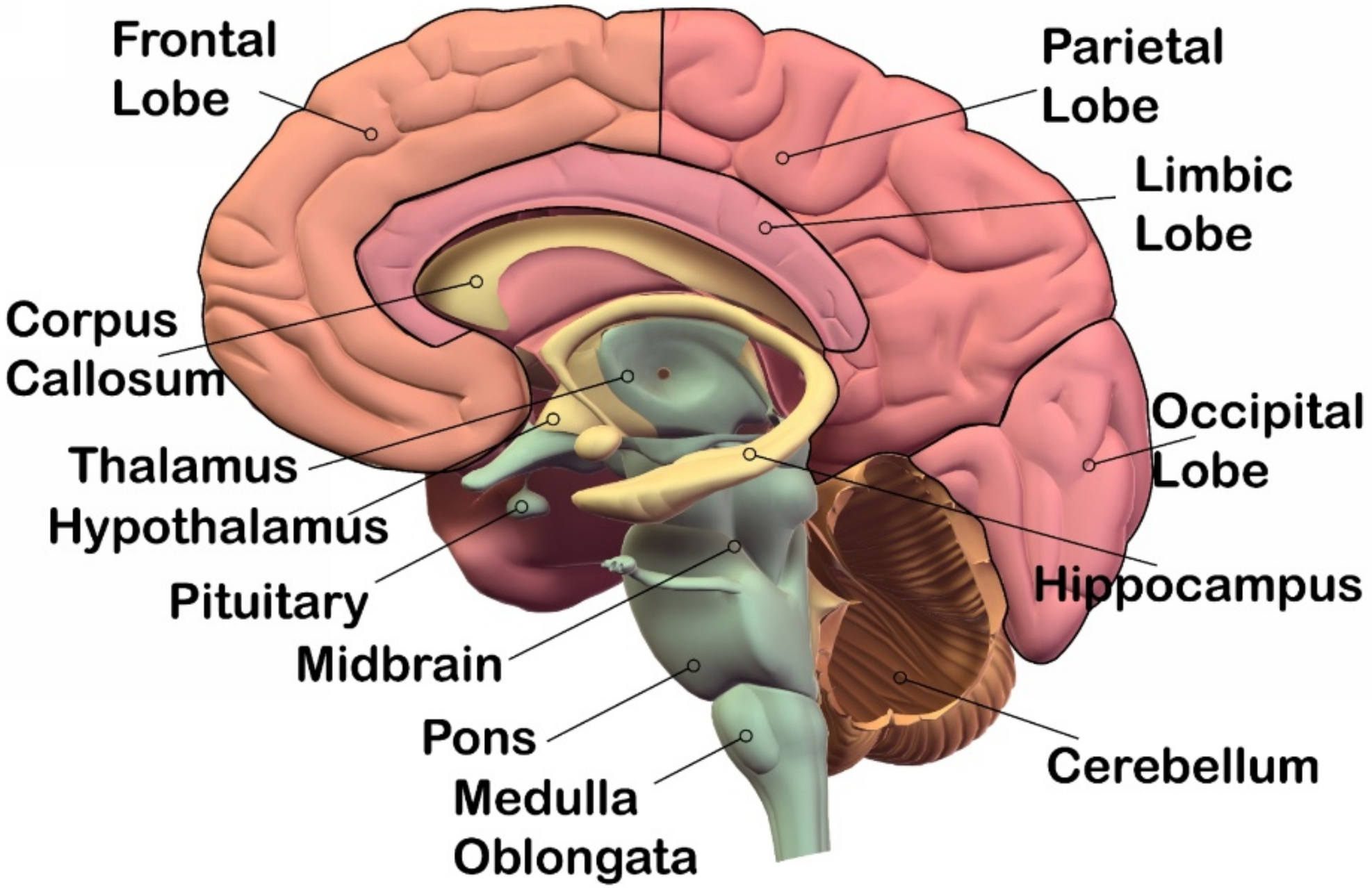


Fig 2.1 Anatomy of the Human Brain

Cerebral cortex: From the outside, the most defining characteristic of the brain is the elaborately folded structure that covers the entire top and sides. This is the cerebral cortex, which is a sheet of 6 layers of neurons only about 4mm thick that folds around the thalamocortical radiations below it like a pie crust folded around a head of broccoli.

Approximately 1.8 square meters of this cortical sheet is wrapped and folded around the brain, forming the noticeable wrinkles and kinks. The cortex takes up over 80% of the cerebral mass, yet the cortex accounts for only 19% of the total number of neurons in the brain (10). This cortex is divided into lobes which contain specific regions for vision, audio, speech, touch, smell, motor control, and our other external and internal senses and outputs.

The cortex is composed of individual processing units, called cortical columns, that span vertically across the layers and connect to the thalamus. Depending upon how researchers define a cortical column, there can be 10,000 to 100,000 cells in a column, and therefore 150,000 to a couple million columns total. Each of these cortical columns represents a computing unit for the cortex, a complete sensory-motor unit, processing a feature vector for the senses or motor control, and outputting to both motor units and to other cortical columns. The cortex is known for accomplishing much of high-level cognition, such as making large-scale associations, action planning, and language production. We will come back to do a deep dive on the cortex structure-function in a later section.

Motor Cortex - the brain controls movement of the body by cooperation between the excitatory and inhibitory networks of the motor cortex and the cerebellum. Let's now look specifically at the motor cortex, which has been studied more extensively in primates and other animals. The motor cortex generates the movement signals for the body, working with the cerebellum and other brain structures to coordinate goal-oriented movement.

The lateral spatial organization within the motor cortex is correlated with the spatial organization of the body. However, functional organization is more interesting. Experiments which stimulate a neuron in the primary motor cortex momentarily will cause a twitch in the muscles of the associated body part, such as the hand. Stimulating that same neuron for a period of seconds causes a full behavior to occur, such as hand to mouth motion from wherever the hand was sitting (Flindell & Gonzalez, 2016). This suggests a more orchestrated functionality to the motor cortex that uses underlying networks of both sensory and motor units to work together to accomplish specific motor patterns that appear to have some level of end goal.

Experiments measuring the electrical activity of the motor cortex with 3D probes having sub-millimeter arrays (2) show that motor signals consist of three-dimensional patterns that evolve in time. To get such a sensor array in a primate brain to properly control a robotic arm, the robot arm had to learn to recognize the full 3-dimensional, temporal pattern that represented each motor action in the respective brain area. (A promising application of narrow AI is to train these AIs to recognize complex patterns, such as those patterns which appear during particular brain states.)

Cerebellum: In the back of the brain, the cerebellum is like a second brain, and it serves as a second, complementary component to the motor cortex. Though it is much smaller in volume than the cortex, 50% of all neurons in the brain reside in the cerebellum (5). The cerebellum contains large Purkinje neurons with dendrites that fan out like a large planar bush, with long axons penetrating perpendicularly through them, connecting via synapses combining to make a very powerful neural 'coprocessor' to complement the cortices of the brain and to assist the motor cortex in particular to generate sequences of fluid and accurate motions with feedback from the proprioception and visual systems. Like the cortex, the cerebellum also has somewhat of a laminar organization, but the layers form much more leaf-like bands called folia. If you divide the cerebellum in half down the center, you see the tree-like organization called the arbor vitae (the tree of life) that is formed from the Purkinje cell axons that project from the cerebellum to the thalamus and other areas to communicate all over the brain. The plant metaphors run deep in the cerebellum, but the structural similarity is undeniable. Interestingly, all cells in the cerebellum, except a cell type called granular cells, are inhibitory. The cerebellum serves to fine-tune the signals generated by the cortex, processing feedback from the cortex of the brain to coordinate more finely organized and correct outputs.

Over all the past centuries, the functionality of the cerebellum was dominated by the view that the cerebellum only governs motion or motor activities. However, recent evidence has demonstrated the prominent roles of the cerebellum in high-order activities or functions, such as emotional regulation, executive control, language processing, timing perception and verbal learning (7), and even judgment of size of objects.

Thalamus and thalamocortical radiations: The thalamus is a pair of golf-ball sized structures near the center of the brain. They are the gateway to the cortex; all sensory inputs (except olfaction) from the body and all cortical motor outputs flow through the thalamus. The thalamus is a relay center, as one axon from a sensory input relays to one axon that connects to the cortex, and in that connection, the spiking rate and information redundancy are reduced before the signal is sent to the cortex. However, the thalamus is also much more than a simple relay. Ten times as many axons go from the cortex to the thalamus as the other way around! Some very extensive modification happens to each sensory input in the thalamus before it gets to the cortex that must contribute to our ability to predict and fill in gaps in information.

The brain's cortices (consisting of regions of these specialized cortical columns) evolved networks with very sophisticated space and time signal processing, and many neuroscientists postulate that the cerebral cortex comprises a mental map or basis set of cortical column vectors that translates inputs like vision and audio into an internal representation for that information, and also that internal representation into outputs or actions (motor cortex).

The thalamocortical radiations serve to connect the thalamus to the cerebral cortex, branching like a broccoli such that each cortical column is innervated by the final branches of the radiations.

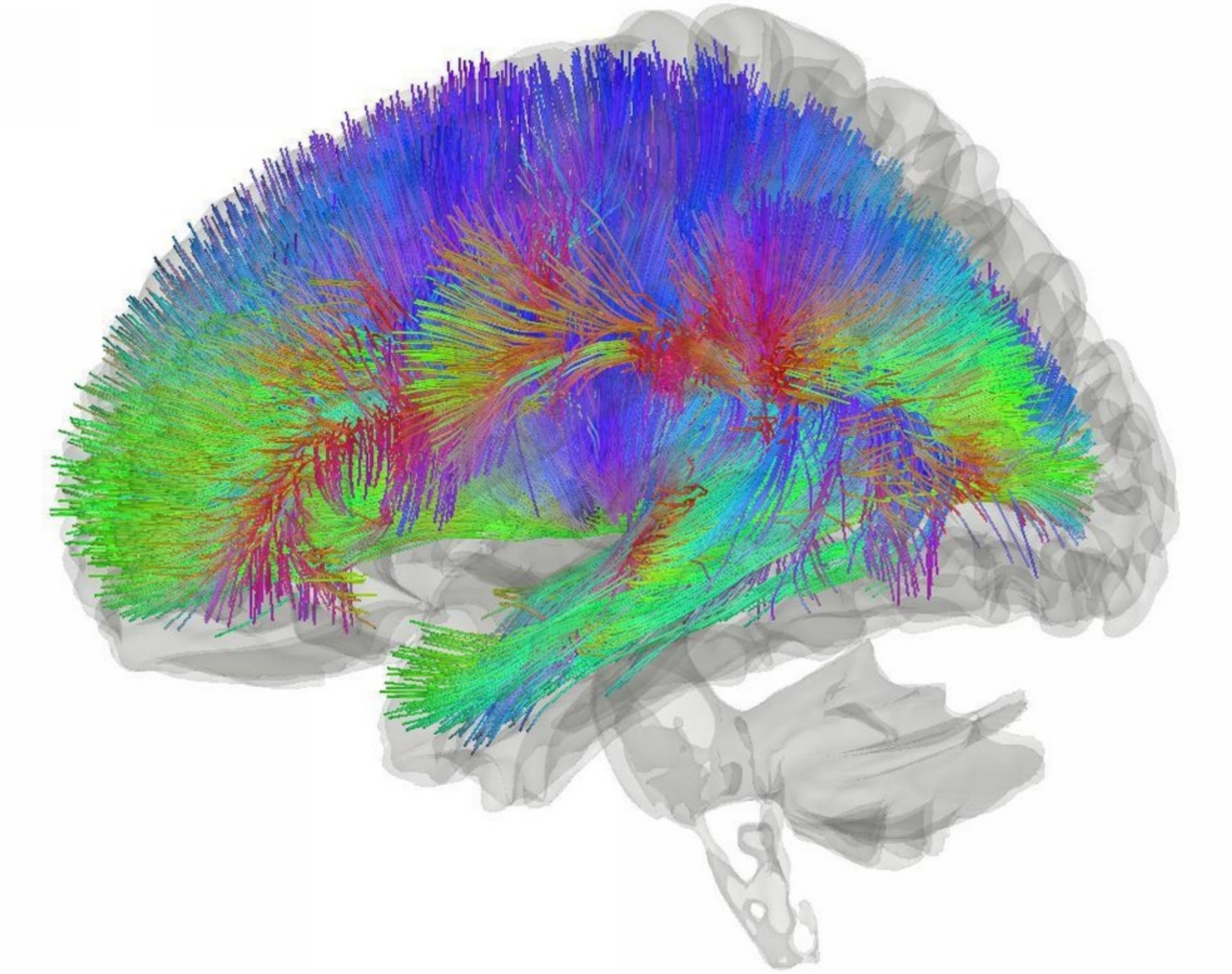


Figure 2.5 Thalamocortical Radiations (15)

Basal Ganglia: Deep to the cortex and sprinkled around the thalamus are the basal ganglia. Instead of the nicely organized layers of cells that we see in the cortex, basal ganglia are made of twisted knots of tightly packed neurons known as nuclei. These nuclei connect to cortex via the thalamus and play important roles in motor refinement, including when to start and stop a behavior. They are part of the limbic system and are highly involved in our cognitive and emotional responses. The nucleus accumbens and amygdala in the basal ganglia are two nuclei famous for their involvement in addiction and avoidance mechanisms respectively.

Hippocampus: The hippocampus is found deep beneath each temporal lobe. It is somewhat of a re-curved extension of the temporal lobe cortex, and it shares some structural and functional similarity to cortex. The hippocampus works with other parts of the brain's memory system to orchestrate narratives to reconstruct memories of the past and predict fictional narratives into the future. When we dream, our brain, directed by the hippocampus, creates fictional narratives that fill in the blanks in our waking knowledge and allow us to learn about and build models of our world that are much more complex and nuanced than we could without dreaming, later helping us with planning our waking actions. The hippocampus is also largely responsible for our ability to understand where we are in space and time through the activity of place cells and grid cells located in it.

Hypothalamus: The hypothalamus sits below and in front of the thalamus. This small structure is extremely important for maintaining homeostasis, and as such, regulates hunger and other basic drives. The hypothalamus exemplifies the electro-chemical nature of the brain because the hypothalamus controls release of hormones and other chemicals that signal all over the body and affect global brain states that, in turn, drive drastically different behaviors and mental states. Hormones and homeostasis might seem irrelevant to an AGI system, but it is important to understand these systems, of which the hypothalamus is a part, that regulate global state through the interaction of neural activity and body systems. Being in a different global state through chemical signaling can completely change behavior and can actually change how aware you are of different stimuli. At the simplest level, structures like the

hypothalamus help serve as a prioritization system, which is important for goal directed behavior, a feature that is important for AGI.

Midbrain, Pons and medulla oblongata: Many of the internal structures of the brain evolved longer ago and function independently of the cortex, controlling our core functions, drives, and emotions that are acted on by the rest of the brain. The cortex is stacked on top of these structures and cannot function on its own. At the very bottom of all of the brain hierarchy are the midbrain (mesencephalon), pons, and medulla oblongata. Midbrain and pons are involved in sleep/wake cycles. All basic involuntary body functions, such as breathing and heart rate are controlled by the medulla oblongata, which connects directly to the spinal cord. Many reflexes including vomiting, sneezing, and swallowing are controlled by the medulla. Even small disruptions to these structures generally kill you because they cut off basic life functions.

Summary: The brain is a biological system, not a computer, and it functions in a very different way than computers do, mainly because a brain was built for movement and responding to change, while a computer was built for data and mathematical processing. Brain structure is hierarchical, with newer brain regions being built on top, sometimes redundantly, of older structures. Different brain regions are adapted for different functions, and these regions work together as a whole to form a unified brain state that allows for high level cognitive functions and goal directed behaviors.

Evolution of the brain into functional architectures

So now that we've done an overview of what a brain is, we need to understand a little bit more about how it works: how the organization and functional principles contribute to the intelligence of the brain. We will attempt to bring in context of how constraints and functions of the brain together allow for the emergent properties we can observe.

Different brain architectures allow for different levels of coordination, prediction, and cognitive capabilities. Recently, Barron *et al.* (21) outlined 5 transitions in brain evolution that allowed descendants to be able to accomplish certain fundamental intelligence features. This framework serves as a useful starting point for our structure-function discussion of the human brain, so we will apply their ideas loosely through this section to introduce specific concepts needed to understand the human brain.

Neural Networks. The first stage in brain evolution was developing diffuse nerve nets that allowed for long-distance signaling through neurons. Prior to that, cells mostly communicated to their local partners. But to get a big body moving, there needed to be a way to quickly produce muscle contractions in one area that were responding to a signal from a completely separate location. Neural networks, such as those in jellyfish, provided that ability to communicate all the way across a body.

Centralization. After a neural network formed, centralization of those neural networks into a brain allowed for global coordination of the body. A centralized brain can get the whole body on the same page much more easily. The brain can now act as a master controller, deciding behavioral priorities and regulating physiological state, rather than individual neurons merely responding to events in their local area. With a central brain, you can get both global coordination and local specialization in the same package. This is important for the ability to utilize different types of sensory information and for utilizing different actuators to perform different actions from that information, while keeping the whole body aligned on a goal.

Recurrent circuitry. The next step in this framework was recurrent circuitry. Perhaps one of the most interesting features of a human brain, when compared to a computer, is the sheer number of connections that loop back to modify the inputs. These backward loops happen throughout all levels of hierarchy in the brain.

Visual processing provides an illustration. Sensory input from vision first goes through the deep brain, landing on a structure called the Lateral Geniculate Nucleus (LGN) within the thalamus. As with all the sensory information in the

thalamus, only 10% of the inputs to the LGN are from the eyes. The remaining 90% come from the cerebral cortex and brainstem. What we think we see is actually largely made up from information from inside our own brain, even before the image gets to the visual cortex.

Now, the processing at the cerebral cortex also forms bidirectional networks between regions of the cortex, such as the V1-V4 cortical areas, which break images into an abstraction and back again through these forward and backward connections. One interpretation of this organization is that sensory input is processed into thoughts, abstractions or 'engrams' by one directional network in the cortex, and then those engrams are processed back out to a re-creation of the expected representation by another, complementary network in the opposite direction, and they are fed back into each other (9). Also see Miguel Nicolelis' book, "Beyond Boundaries: The New Neuroscience of Connecting Brains with Machines" (2), for further reference.

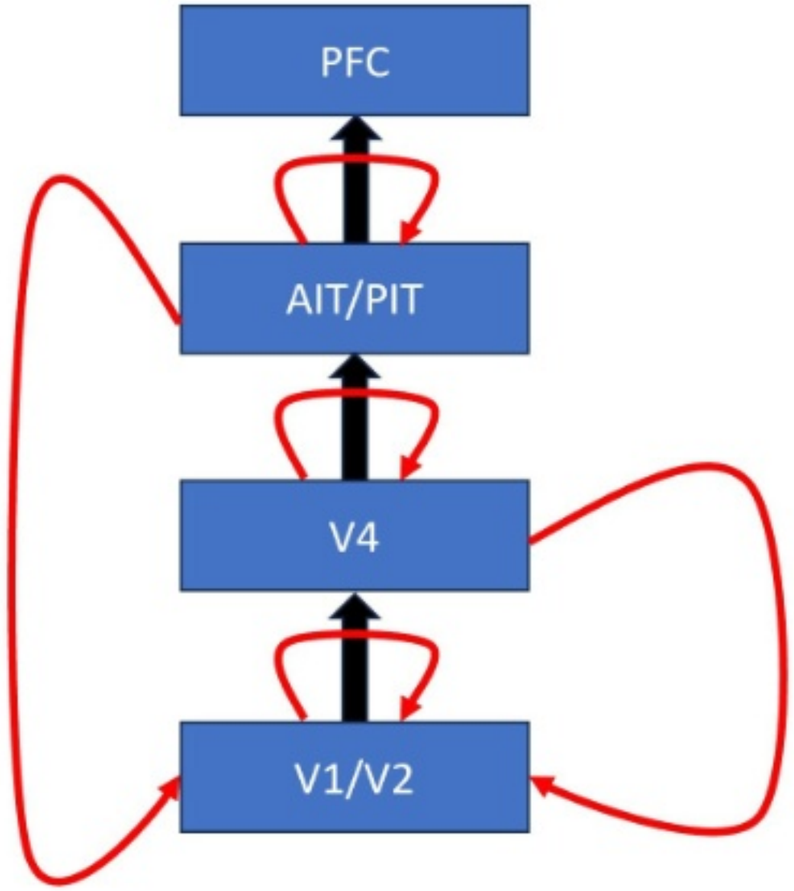


Figure 2.6 Recurrent Neural Circuitry in the Visual Cortex

As an example, picture a fire truck with your eyes closed and you will see the recurrent network of your visual cortex at work, allowing you to visualize the 'thought' of a fire truck into an image of one. You could probably even draw it if you wanted. Try looking at clouds, and you will see shapes that your brain is feeding back to your vision as thoughts of what to look for and to see. Visualize shapes and objects in a dark room when you are sleepy, and you will be able to make them take form, even with your eyes open.

Recursive loops not only train our sensory cortices to encode the information from our senses into compact 'thoughts' that are stored (with overall control by the hippocampus) into short term memory, but also allow us to selectively focus our senses, by projecting what we expect to see from thoughts to visual information (which in the extreme can form hallucinations). Each sensory cortex has the ability to decode them again and to provide a perceptual filter by comparing what we are seeing to what we expect to see, so our visual cortex can focus on what we are looking for and screen the rest out.

The frontal and prefrontal cortex are thought to have tighter, more specialized recursive loops that can store state (operational memory), operate on it, and participate in logic and planning at the macroscale.

All our cortices and the whole brain work together and can learn associatively and store long-term memories by Hebbian learning, augmented by hierarchical methods, with the hippocampus being a central controller for memory reconstruction, planning, and prediction.

The author (Gunnar Newquist) has purposefully used the more general term 'recurrent' to describe these backward loops. Typically, researchers call these recurrent loops feedback loops. However, feedback implies a particular function to the loops that may not always be the case. It is important to note for the purposes of AGI creation that the function of this recursion is often very different from the feedback signals typically found in a computer system. This is partially

because brain circuits are vastly more complicated than any computer circuit. As such, the result of recursion can be simple feedback, a feedforward loop (which creates entirely new downstream signals based upon which signals are currently present in the recursion), or some exotic neuromodulatory signals that we might not model accurately in computer circuits today. This complexity should be kept in mind, as an AGI may need greater flexibility to capture certain brain functionality than what we currently employ with feedback loops. We will revisit the control systems of the brain in context of learning and behavior later.

Lamination for parallelism. Centralization provides coordination, but it doesn't mean the brain is single minded. One of the major features of a brain is the extent to which parallel processing can occur. The laminated structure of the cerebral cortex is one such method of enabling massive parallelization. Recall that the human cortex has 6 layers of neurons. Cortical columns are the fundamental functional subunit of the cortex, and these units span across all 6 layers. These columns of neurons are repeated throughout the cortex but are often specialized in different regions of cortex for different functions.

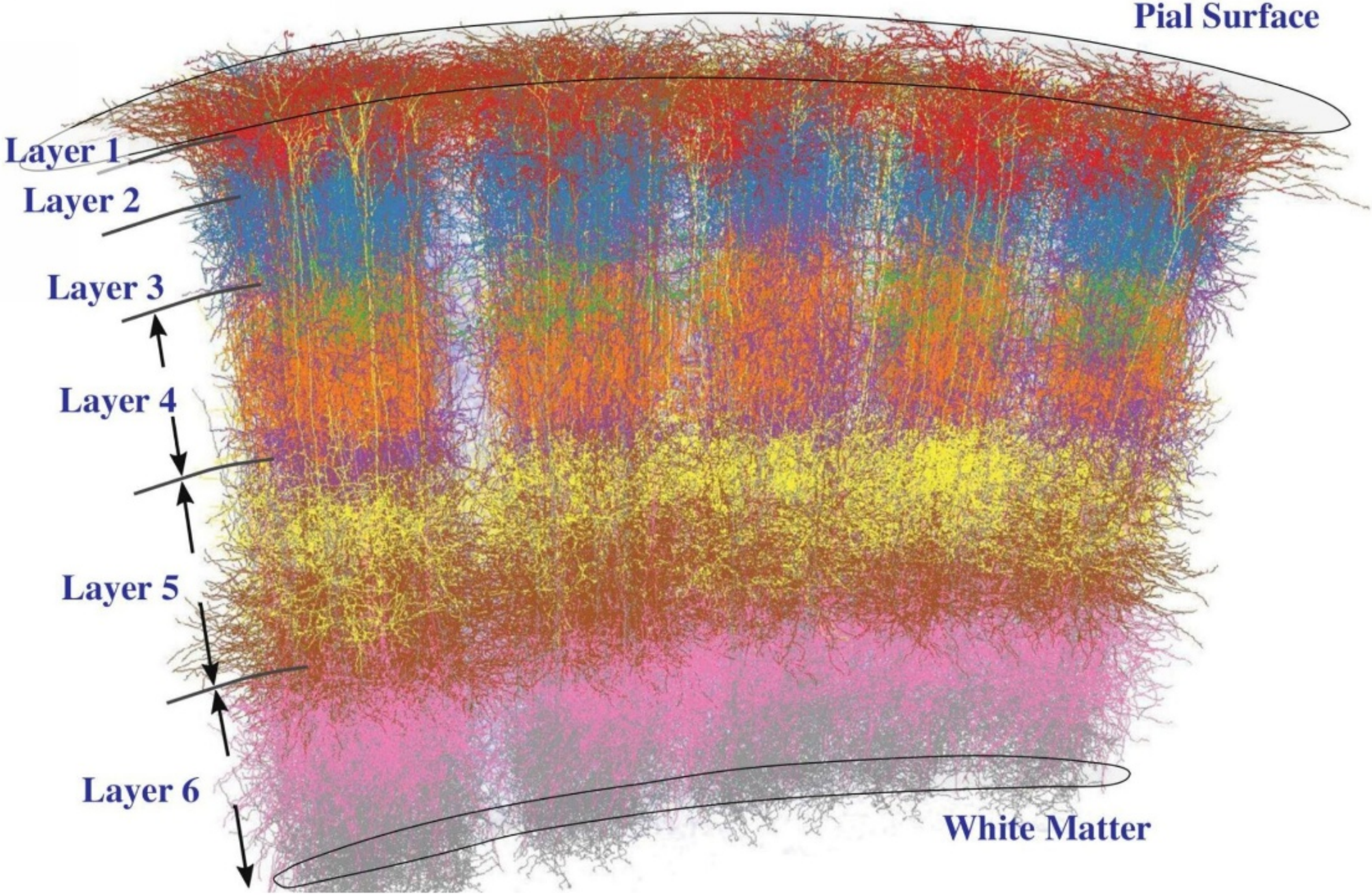


Figure 2.4 Cortical Columns (14)

Micro-columns within each column run from the bottom to the top layer of the cortex and have more connections up and down each column than laterally, including large pyramidal neurons with extended dendrite branches penetrating through the layers. Most of the connections in a cortical column do not extend laterally beyond more than a few columns, so these micro columns form discrete 'computational' units that are only connected to their immediate neighbors via lateral connections and to other brain structures at specific layers. Glia most likely help isolate and regulate the function of each cortical column as a unit.

How cortical columns function exactly is the subject of great debate in neuroscience, but their existence and what we know about them shows that these units process input and output, are similar, replicated about 1 million times across the cortex, and their general structure and functionality are capable of being specialized to certain functions in certain areas, like vision, touch, motor control, olfactory, language, and others. This is useful info in creating an AGI model, because if we can build and train an artificial cortical column and connect thousands of them into a sheet and underlying structures in the right way - we could have a start at an artificial brain - maybe. However, this seemingly simple construct is a trap for budding neuroscientists that has caught many an incorrect AGI model in its clutches, so we must be careful about making assumptions about the function of cortical columns and how artificial ones should work. Despite the risk in these assumptions, more detail into what we know about cortical columns is warranted, as this structure is thought to enable many of our human cognitive abilities.

Structure of cortical columns: enabling massive parallelism. Again, the structure of cortical columns is highly complex and still somewhat under debate. At first, this seems counterintuitive since some much research has been put into understanding the cortex. Just imagine however, trying to disentangle 50,000 neurons which each form around 1,000 connections locally and distantly and then try to simplify that into an easy take-home message. Luckily, there are a few patterns we can generally pull from this web.

The cortical layers are labeled 1-6 starting from the outermost Layer 1. Input signals from the senses come into the cortical columns from the thalamus via the thalamocortical radiations into Layer 4 and from there are transmitted to Layers 2 & 3, and these are combined with signals coming in at Layer 1-3 from the other cortical columns.

The upper three layers (closer to the cortical surface) are called the supragranular layers (1-3) and are functionally distinct from the lower two layers (4-5), which are called the infragranular layers. The supragranular layers project out to other cortical columns, both sending and receiving signals, with layer 3 cells projecting to adjacent columns and layer 2 cells projecting to more distant parts of the cortex. The infragranular layers (4-6) receive input from the supragranular cells of adjacent columns, although they do not reciprocate; instead, they send their signals to extracortical structures, such as the thalamus, and motor and sensory centers. These signals propagating through the outer layers of the cortex are most likely generative predictions or simulations of reality that propagate to adjoining columns in this laterally connected network. The spreading activation from columnar goals and the subgoal by subgoal process of problem solving is called a Call Tree, because of the way the subgoals are solved by a recursive search. The long-term potentiation produced in the synapses of this layer stores memories of successful predictions and (guided by the hippocampus) allows the layer to retrieve memory, predict future input from the senses, and plan. (17)

This combination of the input from the generative prediction layer 1-3 with what is being sensed from layer 4 happens within the neurons extending through layers 2-5, with the most distant 90% of the synapses on those neurons dendrites taking inputs from the predictive network, priming the neuron, and causing it to fire only if the synapses nearer the neuron are stimulated by the actual input that was predicted as well (10). This causes a pattern of single mini columns firing sooner when a prediction is correct, as opposed to a more diffuse firing of multiple columns when it is not, and provides different outputs to the lower layers based on the brain's predictive sequence being correct or not. This is how the cortex learns, reinforcing the connections that created a prediction that matched sensory data, and thus getting better at predicting events and building a model of the world.

In reference to movement again, each cortical column embodies a full sensory motor system. The outputs of these computations are passed to level 5 of the cortical column to feed back into the thalamus or out to the motor control neurons for many of the cortices (not just from the motor cortex). This computation of the predictive signals and sensory signals in Layers 2&3 serves to compute the similarity between what is predicted and what happens and to activate a response for when prediction meets occurrence, leading to action based on it. This is one of the key functions of intelligence, to predict what is going to happen, and act on it when it does happen. We will dive into one method of how the brain develops predictions in the section about conditioning.

Cortical columns in sensory areas (auditory, visual, somatosensory) form maps. Regions of cortex adjacent to these maps are associative, with the associations becoming progressively higher level and more abstract with greater distance from the sensory map. For instance, the intensities of different frequencies of sound waves are mapped on the planum temporale, while cortical areas in more inferior areas of the temporal lobe process higher level information, starting with sounds and moving to word concepts. (17)

Cortical columns in motor areas also form maps. Regions more anterior in the frontal lobe handle progressively higher-level information. Because temporal precision is necessary for motor movements, the upper levels of the frontal lobe naturally develop the role of organizing events in time. Each higher level increases the length and complexity of structured sequences. Thus, this hierarchical, temporal model proposes a plausible explanation for why the frontal lobes are involved in motivation and producing structured sequences. (17)

According to this model, cortical division of function is a natural outcome of certain areas of cortex being mapped to certain extracortical functions, i.e. sensory and motor maps. As a result of the spreading activation of call trees, the areas adjacent to these maps assume the roles of progressively higher-level integration and association units. (17)

Reflection. In computer science, 'reflection' is the ability of a computer to rewrite its own code. Humans have the ability to change the way their brains work through the use of symbolic language. Symbolic language allows one person

to give instructions about how to do something to another. To change a person's behavior, we can just tell someone what to do, and they can perform this new behavior that they didn't have to learn from scratch themselves.

But language does other more interesting things in relation to cognition. Developing a word for a phenomenon allows for greater memory and perhaps even increases the ability for abstract thought, such as Theory of Mind (the understanding that others might have different thoughts than you). For example, those who know more words for shades of color can better remember the color hues presented, though there seems to be no effect on color perception (Hasantash et al, 2020). Populations who lack the words for abstract concepts such as Theory of Mind, such as deaf children who grow up without being exposed to sign language, may not be able to conceptualize these abstract concepts at all, unless they develop words for them at an early age (Morgan & Kegl, 2006). The employment of language may actually make us smarter! The development of Nicaraguan sign language is a fascinating case study in how languages and cognitive abilities can develop in a population that started without being exposed to mature language (Senghas, Kita, & Ozyureka, 2004).

Brain structures underlying language, such as Broca's area in the cortex seem to enable adept pattern recognition in temporally extended sequences (Fiebach & Schubotz, 2006). This ability seems to correlate with language and musical abilities, but additionally, with tool manufacture. Therefore, it may be that complex temporal pattern recognition is a prerequisite for complex language and other complex skill development.

Summary: With Baron *et al.*'s framework, we have 5 brain organization principles that enable 5 different basic intelligence features: Neurons allow for both complex integration of information at each node and also for long distance communication. Centralization allows for coordination. Recurrent networks allow for the output of a process to control the operation of earlier processes (whether by feedback loops, feedforward loops, or something more exotic). Lamination allows for massive parallel processing. Reflection allows for 'software' to change the underlying code, such as the way assigning words to concepts allows for greater differentiation and identification of those named concepts.

Overarching Intelligence Principles

Baron et al.'s framework gives us some ways of understanding how stages in brain evolution opened the door to various cognitive capacities that we now possess as humans. Now we can put cognitive abilities in context of overarching principles that underlie biological intelligence.

Ethology

As mentioned previously, brains were built for movement, and because of that need to move, the brain develops with a variety of pre-built motor programs and responses. Even as a baby, you don't have to learn everything from scratch. Reflexes are a perfect example. You are born with the ability to pull your limbs away from pain, for example, when a hot stove burns your finger. The signal for this motor reflex doesn't even get all the way to your brain! The full circuit for this reflex is in your spine.

Reflexes can be more complicated than a simple knee-jerk response as well. Cats which have had their brain disconnected from their spine can still exhibit a full walking reflex on a treadmill. (video example) The entire program, including motor and sensor loops for various walking gaits seems to be located in the spine in cats. Though human gait takes more brain control than the cat example, much of walking and other behaviors can be created from **central pattern generators** that are built from rather simple oscillating neural circuits.

But the behavior pre-programming gets more sophisticated than simple reflexes and pre-built patterns as well. There are certain behaviors that are just inherent to the human species, behaviors that might be called **innate** in other animals. For example, every culture in the world gathers around and listens to a highly regarded community member telling stories: a university lecturer in an auditorium, a religious leader from a pulpit, a politician on a campaign, an elder around a campfire. Humans are also naturally a noisy bunch. Every (normal) human speaks or signs a language. All human cultures build tools.

The suite of behaviors, such as storytelling and tool building, that appear within members of a species is known as the species' **ethology**. Ethology is a powerful emergent property of a natural brain. First, ethology seems to happen almost *despite* the environment in which an animal is placed. A squirrel which was born in captivity and has never seen a nut nor dirt will attempt to bury a nut when shown one. A pet beaver will attempt to make a dam out of toys and household items. Herring gull chicks will peck at any red dot. Nuts, dam materials, and dots all produce specific behavioral patterns from these specific species. As such, these stimuli that create these **fixed action patterns** are known as **sign stimuli.**

Believe it or not, humans have their own fixed action patterns and sign stimuli. For example, humans naturally press buttons from a very early age. We seem compelled, regardless of the consequence of this action, to press any button we see. "Don't press this button" jokes, and Reddit's "The Button" experiment exemplify the tendency. Another fun example of a human sign stimulus is wet paint. Humans find it hard to resist the urge to touch something we know will smudge and get on us.

These pre-arranged stimulus-response patterns that are built into our brains form the basis of our general intelligence abilities. Pre-organized responses allow us to make some of the right responses without having to learn everything from scratch. It would be pretty tough to get these words on the page without the tendency to press buttons and converse as a baseline behavior set, as the trial and error needed to manipulate the world randomly.

Clearly, these tendencies can also get us into trouble and demonstrate the limits of our intelligence, as well. We press dangerous buttons with clear warning labels and sometimes we don't know when to back down in the face of a dangerous aggressor. Therefore, our intelligence only has advantages based upon certain environmental constraints. Take us too far away from the environment for which our ethology adapted, and we don't seem so smart.

Luckily, part of human ethology encompasses the flexibility to adapt to new environments (to a point), hence, the reason to use humans as an example of 'general intelligence'. And that brings us to the most misunderstood and controversial part of us, how we learn.

Natural control systems: built to predict and do something about it. The fundamental control systems of a brain are somewhat different from a computer. We can use the example of 2 types of thermostat for reference. In a thermostat with feedback control, the temperature sensor is inside the room. Once temperature in the room registers above or below a certain set point, the system turns on heating or cooling to return it to the designated temperature range. The problem is, by the time the temperature is too low inside the room, it takes more time and energy to get it back up to the set temp. This doesn't give you much room to predict a temperature change, and it lends itself to relatively large sways in temperature because of the lag between sensing a change in temperature and doing something about it.

A feedforward control system is employed by buildings which have a much smaller tolerance for temperature swings. In this type of system, temperature sensors are placed on the outside of the building. Changes on the outside register before there is an actual temperature difference on the inside where it matters. Therefore, the controller can increase or decrease the heating prior to a demonstrable change of temperature within the building. Besides just the sensors, a feedforward controller needs a very good model of how changes on the outside lead to changes on the inside so that the right amount of heating or cooling can be applied in anticipation of the internal swing.

Notice the likeness of biological systems to the feedforward controller. We have sensors all over ourselves, collecting information outside of our bodies so that we can detect environmental predictors ahead of time to prepare for any upcoming changes. However, what is fundamentally different is that biological systems continually build and adapt the 'model' of the relationship between the outward facing sensors and the internal state or behavior due to learning. This model becomes an internal reality that is constantly updated through both our sensory experience and our internal representations our own brains build from memories, dreams, or even hallucinations. We are essentially

a wildly complex feedforward controller with some sophisticated updating machinery, the likes of which is much different from a computer's control system.

ROS - Inhibitory Networks - Though many of the details of neural control of behavior are unknown, we do know a lot about the neural mechanism by which movement and speech are controlled. The generation of motor commands in the brain is done by a combination of an excitatory network of rank order selective (ROS) neurons (4), and networks of inhibitory neurons branching off from the ROS neurons.

ROS networks are a chain of neurons that fire in sequence to create a timing signal, which is always the same, independent of the actual movement being done. From the excitatory ROS, inhibitory networks branch off like a bush, with their terminal ends controlling specific movement signals. This includes motor control of the body and the speech apparatus, like the mouth, tongue, lips, and vocal cords for spoken language and the hands for written and sign language.

The neurons in the inhibitory network are controlled by external signals imposing the conscious control of the brain on the movement control network. If every neuron is inhibited, nothing moves. If no neurons are inhibited (or there is sporadic neural activity like in a seizure) the movements are random and undirected. It is this balance between the temporal timing signals of the ROS, and the application of complementary inhibitory signals that allows us to control our movements and speech. Below is a diagram of how the verbal ROS-Inhibitory network (very crudely) functions to assemble the sentence "I love you", with red representing neurons that are being inhibited, and green representing neurons that are not being inhibited:

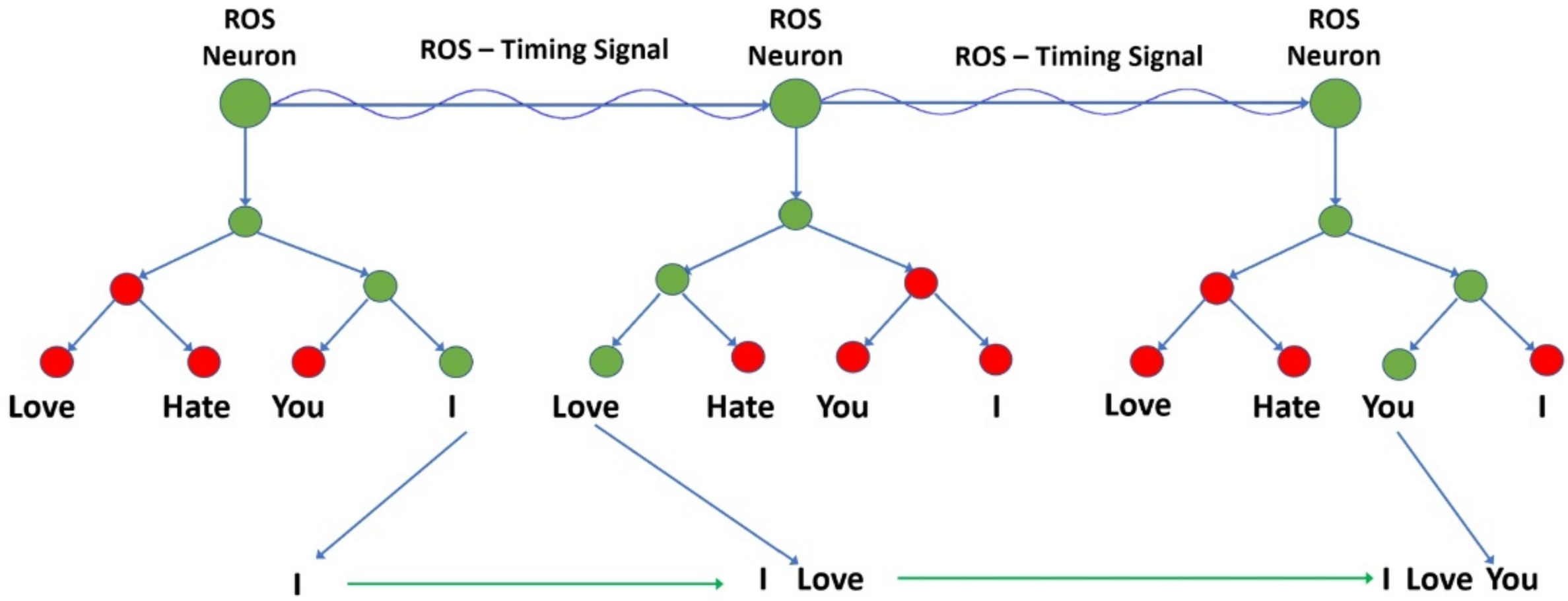


Figure 2.7 Language Production by ROS-Inhibitor Network

Having our actions such as motor movement, speech, and others controlled by these excitatory / inhibitory networks has an advantage. As simpler actions and words are learned as a baby, we are constructing basis sets or building blocks in the inhibitory networks like letters and phonemes for language, and basic motions (opening and closing a hand) for movement. Then later, these neural basis sets or building blocks can be combined and more advanced behaviors built on them. This is why we learn slowly at first, waving our hands around and babbling as babies, then exponentially developing speech and coordinated movement by age 2, once we start building on these basis sets and combining them to command more complex movements and speech.

This is also why, if we do not learn language by a certain age, we have a great deal of difficulty learning it later because we do not develop these fundamental language basis sets. It also explains why we have an accent when we learn a

second language later in life, because we still use the phonemes (sounds for letters) learned as the lowest basis set from our first language when speaking the second language, and it is very difficult to learn the new phoneme basis sets. The ROS network example gives us some insight about how we build from a basis set, but also, some of the limitations to applying our knowledge to new experiences. At some point in our development, we had to narrow down the basis set from all of the possibilities to help us function in our current environment, but this same process limits how well we take on tasks that might conflict with, or live outside of, the basis set.

Conditioning: updating the innate behaviors.

Pre-organized responses (our ethology) allow us to make some of the right responses without having to learn from scratch, which gives us a head start over a completely agnostic system. ROS networks then help us narrow down the possibilities further into basis sets. But ethology also helps contribute to our general intelligence abilities in ways that we take for granted. Ethology underlies the ways in which we adapt our predictive control systems.

Let's take the example of a professional athlete. Wayne Gretsky was generally regarded as one of the best hockey players in history. People like to talk about the fast reflexes of a professional athlete, but Gretsky wasn't the fastest or strongest on the ice. Instead, he always just seemed to be one step ahead of everyone else. His own quote sums up his advantage best:

> *A good hockey player plays where the puck is. A great hockey player plays where the puck is going to be.*

Prediction turns out to be our greatest intelligence asset, even for athletes who rely on their physical abilities to do their job. When tested, professional athletes can identify cues of where a ball or person is going to be earlier than amateur counterparts (Romeas et al., 2015). This is also one of the significant differences between us and machines. The way we create these predictions is still much of a mystery, but a few fundamental principles can help us understand some of the differences between a human prediction and machine prediction.

The process of identifying predictive cues and changing behavior in accordance with those predictions is called **conditioning**. The first widely recognized example of conditioning came from Pavlov's dog. Naive dogs salivate at the arrival of food, but not at the sound of a bell. In this famous experiment, Pavlov rang a bell that consistently predicted the arrival of food. At first, the dog did not salivate when the bell rang. After many repetitions of the bell paired with food, however, Pavlov rang the bell and withheld food afterward. The dog now salivated after the sound of the bell alone, even without food present. This change in behavior demonstrated that an association had been made between a bell and food!

Timing is important with conditioning. You get the strongest response to the bell if you give the bell just before food arrives, rather than at the same time as the food. Also, the first predictive cue that appears receives the most attention. If, for example, a light is turned on reliably before the bell, the bell is mostly ignored. If you slowly spread out the time between the predictive cue and food, you can see how it works behaviorally. When this happens, you see that the dog salivates not at the time of the bell, but rather, just before the predicted time of the food, even if many minutes later. The bell seems to act as a reference point for food after conditioning has occurred (Gardner & Gardner, 1998).

With some temporal space between the predictor and food, we can see that conditioning has resulted in a more global change in behavior, rather than just one specific response. In another experiment, during training, the dog is restrained while he observes that turning on a light bulb is predictive of food, so he cannot approach the light. However, at the test, when the light is presented alone, the dog is allowed to move freely. Not only did the dog salivate, but he walked up to the light to beg at the light. The dog demonstrated a suite of food acquisition behaviors, not just the single response that had occurred during training. From these results, conditioning is starting to look more like training goal directed behaviors, rather than training simple, individual reflexes based upon the time relationship between 2 stimuli.

Breaking it down further, it appears that conditioning in the brain aligns entire sequences of events into stimulus-response chains (Gardner & Gardner, 1998; Staddon and Simmelhag, 1971). Under this notion, a conditioned response looks something like this:

Bell -> orient to bell -> wait time -> do some unrelated interim behavior -> food time -> get ready for food (salivate, look around for food) -> food comes -> taste -> yummy! -> swallow

From a behavior standpoint, what appears to be happening is that conditioning aligns *when* to activate specific behavior states in time and space. Each brain state can represent a number of behaviors that depend upon the specifics of the situation. Wayne Gretzke figured out earlier predictors of where the puck was going to be, and any number of specific muscle movements would be used to get him there. Conditioning aligns not just reflexes such as salivation, but goal directed behaviors and suites of behavior to effectively execute the right types of behavior to acquire the goal. At the same time, there appears to be learning about when *not* to do the conditioned behavior, as exemplified by the appearance of predictable interim behaviors as well.

Learning in the Brain

The holy grail of neuroscience is a unified theory of how the brain learns. We're not there yet, but we have part of the picture at least. Let's focus on the changes that occur during conditioning. In humans, the signals from different stimuli, such as a bell and food, often come into the brain through different types of sensory neurons. At some point within the brain, these signals land on the same neuron or at least the same population of neurons. Since it is relatively difficult to parse apart what is happening at the neural level during learning in humans, we turn to much simpler organisms to understand a similar convergence of signals. In an insect such as a fruit fly, bee, or moth, you can present an odor that predicts footshock, and these insects will then avoid this odor. This is very similar to the bell-food conditioning in the dog above. However, in an insect, you can actually record the activity of the individual neurons involved in learning.

Here's what happens. A set of olfactory neurons fires during the presentation of odor, and then a dopamine neuron fires at the arrival of foot shock. Both of these neuron groups converge on a set of cells called kenyon cells that are analogous to our olfactory cortex. In naive animals, an odor alone evokes a sparse response from a specific subset of the

kenyon cells, the composition of which is unique to each odor (Cassanaer & Laurent, 2007). However, after training, the strength of the synapses between the Kenyon cells and specific output neurons is increased, demonstrating a learning event (Dylla et al., 2017). (At this point, the insect would avoid the odor.) Interestingly, there can be a space in time between the odor and footshock such that the olfactory neurons and dopamine neuron inputs aren't even firing at the same time, and yet conditioning still happens (Dylla et al., 2017). Recall Hebbian learning requires neurons to "fire together" to wire together, so in predictive conditioning, there must be another, non-action potential signal involved since there was no overlap in the electrical signals at all!

Indeed, molecular signals are at work. When researchers measure molecular signaling, we can see that conditioning begins inside individual cells in molecular networks. Inside the kenyon cells where the two signals converge, each different stimulus causes an activation of an enzyme called adenylyl cyclase (Gervasi et al., 2010), resulting in a rise in concentration of specific molecules, such as cyclic AMP, a ubiquitous signaling molecule. These intracellular chemicals rise then fall over a relatively short period of time of seconds to minutes. If 2 signals happen within that time window, each causes a rise in cAMP, which together, rise above a concentration threshold. Once this threshold is reached, there is a new cascade of molecular events which leads to a change in the molecular network such that the specific signal from the odor gets amplified in the kenyon cell after the conditioning event.

Though the specifics are complicated, there seems to be a rather simple explanation for conditioning. Action potentials and the intracellular chemicals released by action potentials cause an increase in concentration of cAMP (and other intracellular molecules such as Ca^{2+}, Ludke et al., 2018) in the downstream cell. If another action potential from another stimulus also happens on that same downstream cell within the time that the cAMP concentration is elevated, there is a further increase in cAMP that reaches a threshold that causes another chain of molecular events inside that cell. This threshold potentiates (makes it easier) for a subsequent presentation of that first stimulus to evoke a response in that downstream cell that couldn't occur before conditioning, which leads to a change in behavior.

Luckily, this synaptic potentiation can both increase and decrease depending upon the relationship between stimuli. If the predictive odor continues to happen without any more footshock, eventually, the organism will stop responding to it. This is called **habituation**. If, on the other hand, the timing relationship happens repeatedly and is spaced over 15 mins or more, then more synapses actually grow between the 2 neurons. This is a persistent structural change, which underlies long term memory formation.

As you can see in the conditioning examples above, learning in the brain is quite different from machine learning. After all, humans don't compute statistical relationships to make our associations. Our learning tends to focus on time and space relationships which turn into narratives of information. Through conditioning and other mechanisms, we learn sequences of events, and we match what to expect throughout various timelines so that we can anticipate important events to our advantage. Circadian rhythms (such as wake/sleep cycles) that happen at the same time daily are trained by time relationships between stimuli, such as light levels. The training of your brain to daily cycles uses many of the same chemical pathways as conditioning, so it may be pertinent to think about circadian rhythms and conditioning as simply different aspects of a general temporal learning 'algorithm'. Relative time is central to learning in a real brain.

Human Memory - When we zoom back out to a human brain, we see that it isn't an individual cell that creates the memory of a conditioning event, but rather neural networks store memories distributed throughout the brain, strung together in time-based narratives by the hippocampus and other brain systems.

While human long-term memory is less well known, we do know that it is non-local, as injuries to specific areas of the brain don't remove specific memories (even a hemispherectomy which removes half the brain still leaves memories intact). There is no such thing as a 'grandmother cell', a hypothetical neuron that represents your grandmother, and no one else. Rather, any given memory appears to be distributed through the brain, stored like a hologram, spread out over a wide area with scattered bits of information everywhere. This fits with our model that the cortical columns comprise a basis set for memory which encodes input into memory features and decodes those features into memories to output,

all in a distributed manner, such that loss of a small portion of the cortex may only cause loss of some features or aspects of memories, but not the actual memories themselves.

We know that global injury to the brain, like Alzheimer's, causes a progressive global loss of all memories, which all degrade together as the whole brain and cerebral cortex wither, but no damage to a given structure in the brain seems to contribute more to this long-term memory loss than another.

However, specific injury to the hippocampus causes the inability to form new long-term memories. Coincidentally, it also causes the inability to predict and plan and other cognitive deficits, showing that all these processes are similar (8).

Computers faithfully record information into memory to later present the same information again upon request or to perform calculations based upon that memory. Work from neuroscientist Elanor McGuire's lab demonstrates that the reason for memory in the brain is not to recall an accurate record of the past, but to predict the future, using the same stored information and processes in the brain to construct new, generated memories and narratives in addition to reconstructing past memories, orchestrated through the hippocampus. The processes of memory formation and future prediction are similar and intertwined, in a way that is a hallmark of human intelligence (Hassabis & Maguire, 2007; Schacter et al., 2012). Since the brain is reconstructing memories each time we recall them, memories change and drift as this process is repeated, making human memory prone to suggestion, errors, and inconsistency, and are not absolute like computer memory or databases. But because of this plasticity, we also imbue creativity in our narratives of the past and future, which lets us see beyond the data we experienced, unlike the memory of a computer.

Dreaming - Consolidating and Organizing Memory - How do we learn about what we have never experienced? We dream about it.

How does our brain do more than just record all our experiences and reconstruct them? How do we imagine or predict things we have never seen, or plan for events we've never experienced, or say things we've never heard? If all our

experiences are stored as known memories, how do we figure out what is in between? The answer is dreaming; it fills in the spaces between our experiences with simulated narratives and helps build models of our world from all of them, which become our internal realities.

The book "When Brains Dream" by Antonio Zadra and Robert Stickgold (1) describes some very interesting neuroscience research in this area. They propose a model for memory and dreaming called NEXTUP which states that during REM sleep, the brain explores associations between weaker connected memories via fictional dream narratives that, while not meant to solve immediate problems, nor necessarily even incorporate waking experiences explicitly, lays down a network of associations that will aid in future problem solving, whether or not we consciously recall the dreams themselves or not.

In dreaming, besides moving memories from short-term episodic memory to long term memory - the brain also forms connections between potentially related memories and abstract concepts. We experience this process as REM dreams - somewhat fantastic, sometimes nonsensical sequences of events, but still self-consistent and ordered. Our brains form connected narratives through fictional memories and abstracted places, people, and events in a creative and unfettered, but structured manner.

In psychology, there are specific tests in which a person is given a set of cards with scenes on them, then is asked to arrange the card scenes in the sequence that the events occurred in. We humans are surprisingly adept at this task of organizing events chronologically by guessing the order, even when we are looking at novel situations and sequences of events. For sleep studies done by Zadra and Stickgold, research subjects could identify with 90% accuracy which dream reports from other subjects had been randomized by the researchers after those patients reported them. There is an order and meaning to our REM dreams. Once the dreaming process has formed these connections, we now have fictional memory narratives that help us model our world, and we can use them to predict future events or plan contingencies.

The key takeaway from Zadra and Stickhold's work on dreams and memory, as well as the work by Elanor McGuire on the hippocampus, memory and planning - is that the primary tool the brain uses in recall, cognition, prediction, and planning - is narratives, or stories, that it constructs from the abstracted memory stored in our brains. There is some internal representation for these memory narratives where the events are connected in time and by abstraction of the concepts involved, that allows the hippocampus and other parts of the brain's memory system to orchestrate stories or narratives from this representation to reconstruct memories of the past, predict fictional stories into the future, or use multiple such fictional stories explored subconsciously to solve problems.

Narration is a part of our ethology which helps deliver general intelligence. Our internal monologue or 'train of thought' is one of our most powerful cognitive tools, providing a framework for our abstract thoughts and planning. This tendency for creating internally consistent narration is so powerful, that even those without the ability to form long term memories will create fictitious stories to fill in the blanks of what they remember, a phenomenon known as 'confabulation'. Those recurrent loops, timekeeping mechanisms, and hyper-connected internal association systems underlying our narratives all add up to our best predictive capabilities (as long as we stay somewhat within our ethological lane).

Human Language and the Brain - Of course, as a story-telling species, language is a specific form of structured memory narrative that is particularly advanced in humans, that can be expressed and understood through our speech and writing.

Humans are masters at language with by far the most complex spoken, written, and gestural language in the animal kingdom. Though there is a continuum of communication capabilities with whales, birds, and primates able to vocalize repetitive songs and even demonstrate syntax and understanding of specific words and concepts, none come close to having the 40,000+ word vocabulary, and cognitive flexibility of language that human beings universally master by adulthood. Language, and by extension, storytelling, is the ethological backbone for our cognition and communication, with our internal monologues being used to construct our plans and ideas, which are also expressed

in spoken words and writing to communicate them to others, a capability that has made civilization possible. It appears that even our abstract thinking abilities may hinge to a certain extent upon our ability to find words for these ideas, as demonstrated by the phenomenon of reflection described above.

Human language involves many parts of the brain working together, but there are some areas of the brain that are critical to language such as Wernicke's area, Broca's area, and the auditory and motor cortices. If any of these areas are injured or removed, it irreparably hampers our ability to use language and to communicate.

Wernicke's area, located on the left temporal lobe, is associated with our ability to understand and compose coherent and meaningful language. People with damage to this area can still speak, and form words and pseudo-sentences, but their speech and writing are nonsensical. These unfortunate individuals cannot seem to understand written and spoken language after such an injury, also babbling in incoherent sentences when talking. Therefore, Wernicke's area is assumed to be central to both the composition and understanding of language.

Broca's area is associated with the formation of spoken words and written language. It is the link between the language composed in Wernicke's area and the motor cortex, which controls the movements that produce spoken language (mouth, lips, tongue, vocal cords…), and written language (hands and fingers with a pen or keyboard). Any damage to Broca's area renders a person unable to form written or spoken language. They cannot enunciate spoken words nor write words, only uttering repetitive syllables. However, damage to this area does not cause significant problems with language understanding and comprehension, in contrast to Wernicke's area does. Interestingly, those with damage to either area are still able to swear, demonstrating a special functionality and brain processing to our explicit utterances.

Though we are masters of language, we are not the only animals with language abilities. Building blocks of language can be found in some form in a variety of species, and these species can help us understand the roots of our lingual intelligence. For example, many monkey species have specific calls that represent different things that are important to them, such as snakes, birds of prey, or my favorite, a call for 'hamburger', which developed in a group of captive cotton-

topped tamarins in Minnesota. Dolphins often demonstrate understanding of human language concepts such as the difference between subject and object. Birds such as parrots can repeat hundreds of words, but also, can demonstrate the meaning of some of the concepts which they parrot.

In a highly controversial set of experiments, Project Washoe cross-fostered young chimpanzees as if they were the researcher's own children (Gardner & Gardner, 1989). Proper human upbringing requires language. As chimpanzees are not very vocal, the researchers communicated to the chimps only with sign language, and attempted to teach the chimpanzees sign language just as they would human children. Though not quite as fast or adept as human children, the cross-fostered chimpanzees learned to use hundreds of signs. The chimps signed to themselves when they thought they were alone, such as the utterance *hurry* when running to hide from a researcher, and Washoe even taught signs directly to her adopted son Loulis later in life. The researchers recorded a few examples of the chimps making up 'words' for items that hadn't yet been associated with a sign, such as 'candy fruit drink' for watermelon. Most importantly, the chimps communicated information the researchers could not know by using the signs, such as the presence of a dropped toy behind the bed, demonstrating the intention of communication, not just parroting of responses.

Project Washoe demonstrated the importance of a consistent family group and human-like upbringing in the development of the chimpanzees' signing capabilities. Project Nim, another attempted chimpanzee cross-fostering, was much less successful. Nim was raised by over 60 different researchers, coming in and out of his life in the worst form of a broken human home (Terrace, et al., 1979). Instead of fostering an environment of conversation around house-hold experiences as was the case with Project Washoe and subsequent replications in the chimps Dar, Moja, and Tatu, Nim was put in starkly empty classroom situations to be trained with as little distraction as possible. As Nim sat at a barren desk, the researchers attempted to teach individual words by giving reward for faithful replication, much like a computer engaging in reinforcement learning (Terrace et al., 1980, pp. 377-378). Naturally, Nim learned only to mimic simple signs and then beg for the treats in this environment, rather than engaging in colorful conversation, as

was expected (Terrace, et al., 1979). (Note Nim's researchers attempted to dismiss Project Washoe's findings based upon their own difficulties, but public video documentation of Washoe can be found in The First Signs of Washoe.)

Nim's researchers concluded that chimpanzees could not be trained in language at all based upon their failures to teach Nim sign language. But instead, what the body of chimp cross-fostering research demonstrates is that a conversational environment is needed to learn language. To get even basic language, chimps need affordances in their environment which are conversation-evoking conditions. Ethology for social communication exists among primates other than humans, but it is only expressed under similar conditions to those that evoke language in human society, with limitations that are probably due to differences in the structure of their brains compared to ours.

Key points are that human's unique capability for language is due to evolution of specialized areas of our brain that process language and these specialized brain areas allow us to speak and write by interfacing to our motor cortex, and to read and hear language by connecting to our sensory cortices. The natural ability for language is part of human ethology, developed upon the building blocks of pre-language precursory functions and expressed under conversation-evoking circumstances. *Which* language you speak is a product of learning and your environment during early age. In general, because only humans demonstrate advanced language, and most invasive brain research cannot be ethically performed on humans and their closest relatives, the nuances of the human brain's language system are still some of the most poorly understood of all the brain's systems.

Emotions and Attractor states

As humans, we are born with some relatively pre-programmed tendencies, and then learning modifies those tendencies through processes like conditioning. This learning ultimately results in changes to information flow in the brain through synaptic potentiation and even changes in brain wiring through synaptic plasticity. With all this complexity and competing information, there must also be a way of prioritizing what information is important and what behaviors should be expressed at any given time, or else you would end up jumping back and forth between

different behaviors with every different type of information that comes in. What do you do when 2 pieces of information tell you to do opposite things?

One of the major forms of prioritization in the brain is our emotional system. Our emotional systems are the subject of much debate and misunderstanding, as they are so tied to what makes us human. They are also one of the major ways human intelligence is so vastly different from machine intelligence. Though often viewed as our irrational selves, our emotions keep us in specific brain states, which allow us to accomplish goal directed behaviors. Fear, anger, joy, jealousy, and other emotions anchor our brains in specific states that relate to specific goal-directed behaviors such as fight or flight, pleasure seeking, searching for love, interpersonal conflicts over those love interests, and other less emotionally charged activities. Mathematically, these stable states can be described as **attractor states**. Remember, information is flowing all over the brain in oscillations and other temporal patterns. This activity flow can either stabilize into a dynamic state that allows a particular train of thought or a particular type of behavior (attractor states), or devolve into chaos, such as a seizure. The only time the brain is truly inactive is when it's dead.

Emotional states are regulated by deep brain structures such as the limbic system. Additionally, how we interpret and 'feel' our emotions is probably greatly tied to our cortex, as our association language centers help draft narratives around these deep primal signals. Emotional states are sculpted by a bidirectional interaction between neural signaling and chemical signaling. Chemicals released indirectly from neural activity in the hypothalamus, such as stress hormones like cortisol in the blood, cause a cascade of events that prepares the body for a specific type of activity, such as fight or flight. These chemicals in the blood also travel various brain regions and activate specific receptors on neurons that change the activity level of these neural groups. This chemical signaling on neural receptors drives the brain into particular attractor states. In the case of cortisol, the activated brain states are in line with fight or flight behaviors. As long as there is a high level of stress hormones, the brain has a bias toward these fight or flight responses. This might just give you an extra advantage against an attacker compared to someone who has not been set in this hyper-vigilant state. However, it can also be maladaptive, because if there is high cortisol, but no actual danger, you live

in a state of heightened anxiety. This makes you responsive to slight insults that would otherwise seem unimportant to you and cause unneeded internal or external conflict.

Specific cell groups deep within our brains most likely originate and propel an emotional attractor state. For example, neurons involved in aggression can be identified just adjacent to and even intertwined with neurons involved in sexual behavior in the hypothalamus (Lin et al., 2011). Activation of each of these groups of neurons leads to aggression or sexual behavior respectively, but the lines are slightly blurry as some neurons are involved in both! This close relationship could be the reason for our sometimes close connection between love and violence.

Summary: Emotions can make us respond irrationally, but overall, they are part of a prioritization system that helps us decide what is important to focus on for longer-term goal directed behaviors. Chemical signaling such as blood hormones help set specific brain attractor states by signaling through specific neuron types that maintain these attractor states. The same brain bathed in a different suite of chemicals would behave completely differently to the same environment, even if all else is equal, because of how these chemicals affect the activity of specific neuron types.

Selective Attention - Focusing the senses on specific details helps us interpret our world.

Our brains also focus our attention based on our goals and our other cognitive processes. Our senses are not just a one-way input stream that feeds data into our brain for processing. We choose what to look at, and what to focus on when we listen. Selective attention is key to us being able to experience and understand the world around us without getting distracted or overwhelmed by sensory input.

With vision, we actively scan what we want to look at, and only a small portion of the retina, called the fovea, near the center of our visual field, sees the world in high resolution. Our senses are also bidirectional, and while our visual cortex is processing what it sees into more abstracted thoughts and concepts, there is a signal going the other way, processing thoughts and concepts into visual information that can form a selective screen and interact with the

incoming visual information. Just think of how you find 'Waldo' in those picture books by focusing on finding a boy with spectacles and a red and white striped shirt and hat in a very occluded visual field. You are actively seeking and screening for specific visual input.

The same concept applies to reading, using the narrow area defined by the fovea to visually scan a page of text while actively interpreting the words and sentences at your focus of attention.

The same ability to pay selective attention applies to hearing, especially when listening to people speak. We can focus on a specific person's voice in a crowded room, and selectively screen their voice out from the background of other voices, music, and ambient noise. This is commonly called the cocktail party effect and is a product of having predictive feedback throughout the audio and linguistic cortices of what we expect to hear compared to input of what we are hearing.

Memory is more than just passively recording a sequence of sensory events. Rather, memory is an active process noting interactions between existing memories, cognitive processes, and incoming information, comparing them and forming correlations between them that the brain records along with the incoming information, all in a dynamic operation. Also, as we discussed in the section on the hippocampus, memories are reconstructed, not recalled, with the reconstruction being influenced by the perception of the present and our predictions of the future as well as the 'memories' stored in our brains.

A common misconception is that people are able to multitask, by having more than one focus for their selective attention at once. The reality, found through experimentation, is that 98% of people cannot attend to more than one task at once, and suffer severe degradation in their ability to do the first task when trying to take in secondary inputs and do secondary tasks. They actually have to switch between tasks and slice their selective attention in time to each task to be able to 'multitask' and most people cannot actually do two things at once. However, surprisingly, there is

a ~2% minority of people that can truly multitask without performance degradation on the primary task (Watson & Strayer, 2010). As in all things, there are no absolutes when it comes to neuroscience of the human brain.

Key points of this section are that perception is a two-way process, with our brains selectively filtering what to listen and look for, and allowing us to concentrate on specific aspects of our environment and focus our attention on them without being overloaded by processing every single piece of information that we receive.

Why are Humans so Intelligent? In summary, we have big brains with more complicated interconnections than other animals.

Are humans the most intelligent life on the planet? Well, we are the ones writing the book, so we get to stake our claim to the title, but we are by no means the only intelligent life on Earth, and many animals have larger brains than us, or even more complex and more folded cerebral cortices, like the brains of elephants and whales. What, then, makes us unique and able to use computers to write books about intelligence?

There are a few general intelligence characteristics that really make us humans stand out, and that is the main reason we look to human brains as the benchmark for AGI. What is it about our brains that makes us able to have the breadth and depth of languages, abstract reasoning, mastering mathematics, developing science and engineering, using ever more complex tools, fine dexterity, and other feats that we seem uniquely capable of in the animal kingdom?

To start, humans have a high brain mass to body mass ratio (2%), and our brain consumes 20% of the metabolic budget for our (resting) bodies. Our brain has a large cerebral cortex that is folded over on itself to maximize the surface area available for the cortex and its constituent cortical columns to pack into our skulls. The human cortex is thicker than in most animals (up to 4.5mm compared to 0.7mm in rodents and 3.5mm in primates).

The main pyramidal neurons in the human cortical columns have dendritic branches that project all the way from the lower layers of the columns to the top layers and have much higher density, more branches, and 3-4 times more

synapses per neuron (7000 excitatory, 1000 inhibitory) than most mammal brains, making them capable of doing denser neural computing. Schmidt and Polleux (6) describe extended pyramidal dendrites in humans which separate into distinct layered groups. This structure allows these neurons to do more complex logical operations like XOR as well as OR and NOT operations found in neurons of other mammals.

Basically, we have a denser, more powerful network of neurons in the cortex than all other animals. As Schmidt and Polleux state: "The ability to integrate information from a larger number of inputs, track these inputs at higher frequencies, and perform a larger repertoire of computations may have represented key evolutionary steps for the emergence of human cognition."

Evolutionary selection based on humans living in social groups also shaped a larger frontal cortex in our brains that gives us higher reasoning abilities, and specialized cortices for language and enhanced control of our hands and vocal apparatus that allows us to communicate with one another. Walking upright and having use of our hands for making and using tools also caused selection pressure for greater intelligence to make better tools and to use them to craft, hunt, and fight with. Language, combined with our ability to do abstract reasoning and to use tools for writing (including computers), gives us the ability to adapt those skills to music, mathematics, science, engineering, and art, giving us greater creativity and flexibility to master higher-level abstract skills.

Finally, human brains have much greater plasticity and a longer neural pruning period than those of all other animals, even other primates. We are born with very dense connectivity in our brains that will prune over time as we develop, losing almost half of our neural connections by late adolescence, but we start with very underdeveloped behavior. Horses can run within minutes of birth, yet it takes humans months just to roll over and almost a year to walk. Human brains continue learning and pruning these networks throughout adulthood, well into our 30s, giving us a long time to develop specialized motor and cognitive skills. Even in our 50s and beyond we can learn new languages, master new motor skills, and learn new information and cognitive skills. Our neuroplasticity may be humans' greatest intellectual asset, allowing us to learn complex behaviors and intellectual capabilities we never evolved naturally.

But, even though we hold ourselves to be the most intelligent animal on the planet, and the model for an AGI, it is important to note that nearly any animal seems exceptionally intelligent compared to the state of the art AI in 2023. The ability of a bee to navigate to a new food source and then return to the hive to tell the others the location of the food could be considered a superintelligence in its own right compared to the paltry abilities of today's best self-driving cars.

Some of these animal capabilities even exceed our own in specific cases. Though we are poor at multitasking, pigeons can do two things at once better than we can. Squirrels can bury nuts in the fall and find them all 6 months later, while we struggle to find our keys an hour later. There may not be any single human intelligence feature that isn't found, at least in some basic form, somewhere in the animal kingdom. We should also be on the lookout for other architectures, such as bird brains, that might exceed human intelligence at certain skills, such as multitasking, when planning an AGI architecture.

Key points of this section are that humans are more intelligent than other animals because we have big brains (relative to our body size) and a thicker cerebral cortex with denser neural networks that were all shaped by evolution. Human intelligence appears to be a matter of scale, rather than some fundamental unique architecture. Being in cultural groups and having a complex vocal apparatus and dexterous hands helped evolve our brains towards enabling language, tool manufacture and tool usage because they provided survival advantages. Use of this language and tools appears to help accelerate our intelligence as well. However, intelligence capabilities of other animals still far exceeds today's AI, and animal intelligence is more accessible to understand than human intelligence, and so can also be used as inspiration for some features of AGI.

Limits to Measuring and Understanding the Brain

As we stated earlier, we only know what we can measure about the human brain, and much of our knowledge (especially pre-2000) about what specific brain regions do was from studies of people that have had injury, disease, or

surgery that incapacitated those regions and caused specific deficits. Each of these regions is part of a system that has components in other areas of the brain, including the primitive brain and brainstem, that all have to be present and working for that system to function correctly.

However, without the lower brain, and especially brainstem, we would be dead, so only deficiencies caused by very localized damage or stimulation in these structures can be investigated. This ability to measure the waking brain has improved in the 2000s with FMRI able to do real-time imaging at the millimeter level in waking patients actively doing tasks. This work is augmented by 3D probe studies at the sub-millimeter level in animals, but we are still far from understanding how all the individual neurons interoperate and function to accomplish processing of senses, cognition and action. We simply don't have the resolution when looking in from the outside, and we don't have the ability to ethically perturb a person's brain to see what changes without damaging them.

We can make some progress by observing what brains do, and *then* coming up with a hypothesis of how it works, looking for exceptions that would prove our theories wrong. If we can't disprove our theories, then this is a pretty good indication we are on the right track. It is tempting for computer scientists to look for evidence of computer algorithms such as backpropagation or various flavors of probabilistic modeling in a real brain to try to figure out how the brain works. However, this is backward. Just because an algorithm works well under specific conditions in a computer does not mean that the brain uses those same algorithms. Endeavors which hunt for backpropagation or other computer algorithms within the brain find themselves fraught with confirmation bias and potentially hinder progress in understanding ourselves, while failing to provide any innovations in biologically-inspired AI. Algorithms can describe a model of brain activity, just like physics can describe the motion of a ball through the air, but it doesn't mean that the ball calculates the math to move any more than the brain calculates the math to think, nor that either model is complete. Balls respond to multiple forces and physics, as do the inner workings of the brain respond to multiple aspects of physics, chemistry, and biology, most of which is very difficult to measure.

Building an Artificial Brain?

Replicating all of the brain's capabilities seems daunting when seen through the tools of deep learning – how can we encompass vision, motor control, speech, natural language understanding, decision making, written composition, solving mazes, playing games, planning, problem solving, creativity, imagination, and dreaming? Deep learning is using very crude, single-purpose components that are incapable of incorporating the features of intelligence that lead to AGI. Each of the DNN tools is a one-off, a specialization for a specific task, that cannot generalize to do other functions, and there is no way we can combine them all in any configuration capable of accomplishing all these general tasks. Deep learning will never come close to replicating the general functionality of the human brain.

But, the human brain is shaped by evolution to be simpler, more elegant, using more powerful, general purpose building blocks – the biological neurons, and connecting them by using the instructions of a mere 8000 genes total - into cortical columns, cortices, thalamocortical radiations, basal nuclei, and so on, that emerges into general functionality that can learn specialized perception, processing, and tasks. Nature has, through 100s of millions of years of evolution, come up with an elegant and relatively simple way (using our genes) to specify an architecture for the brain and its neural network structures that are able to solve the problems we met with during our evolution.

Instead of building upon typical AI to try to achieve AGI, we are going to start by mimicking the genome-connectome architecture and as much about the human brain's formation process and functionality as we can, using genetic algorithms and evolution to solve the harder design problems, just like nature did.

So now we know more about the human brain, and how the neurons and neural networks in it are completely different and much more sophisticated than the DNNs that deep learning is using. We also have a better idea of how much more sophisticated our simulated neurons, neural networks, cortices and neural structures would have to be to even begin attempting to build something on par with, or superior to, the human brain.

We note that overall, in this book, we do not seek to recreate the biological human brain, but instead to imbue an AGI with the core functionality that makes the human brain so flexible, adaptable and powerful, then augment that with computer science database and computing capabilities to take it far beyond human, and far beyond the existing computer and data systems we currently have. We still need to understand how the brain works better for those core functions that computers currently struggle with, and also, understand the deficiencies that we possess which help reveal how the brain actually works.

CHAPTER 3

Requirements for an Artificial General Intelligence

Now that we have done a deep dive into the current capabilities of computer science, and deep learning-based artificial intelligence from the 2020s, and another deep dive into the neuroscience of the human brain, and its general and powerful capabilities, we want to set the requirements for our AGI.

Artificial General Intelligence (AGI) refers to highly autonomous systems or machines that possess the ability to understand, learn, and apply intelligence across a wide range of tasks and domains, similar to human intelligence. AGI aims to replicate the cognitive capabilities of humans, including reasoning, problem-solving, learning, and adaptability.

Unlike narrow or specific AI systems, which are designed to perform a single task or a limited set of tasks, AGI is characterized by its versatility and general-purpose nature. It exhibits a high degree of cognitive flexibility, allowing it to transfer knowledge and skills from one domain to another, adapt to new situations, and perform tasks it has not been explicitly programmed for.

AGI represents the concept of developing machines that possess a level of intelligence comparable to or beyond human intelligence. It seeks to create AI systems that can exhibit understanding, common sense reasoning, creativity,

emotional intelligence, and other higher-order cognitive abilities. The development of AGI holds the potential for transformative impacts across numerous domains, including healthcare, education, automation, scientific discovery, and more.

It is important to note that AGI is distinct from the concept of **superintelligence**, which refers to AGI systems that surpass human intelligence in all aspects and possess the ability to improve themselves recursively. Superintelligence is a concept that goes beyond AGI and involves additional considerations regarding its potential implications and control mechanisms, which we will cover in later chapters.

Criteria for Evolving a Synthetic Brain - How do we shape the requirements for a human brain in a way that we can measure the performance of an artificial brain against them?

We do not set out to replicate the form nor even function of the human brain, as it has evolved through hundreds of millions of years using biological cells as its units of computing. No matter how well we model these neurons, synapses, and neural networks with artificial ones, we can never hope to evolve the exact same functionality or structure, let alone something superior to it.

Instead we set out to evolve a computer model that replicates the large-scale functions that the human brain is capable of, including vision, motor control, speech, natural language understanding, decision making, written composition, solving mazes, playing games, planning, problem solving, creativity, imagination, and even dreaming. But, we do so with an original design, evolved with artificial neurons and artificial neural networks that, although they approximate the ones in the human brain to the best of our knowledge in core functionality, they are fundamentally different. Given the same selection criteria, these artificial building blocks will potentially evolve very different neural nets and structures to accomplish the same functionality. We are evolving something new, from scratch, an inorganic brain that meets all the functional criteria of the human brain, but with a very different form and architecture in the end. Real brains give us ideas on where to start because they have solutions that already work in the real world, but we have to

evolve our AGI design from basic requirements that encompass what the human brain is capable of. We are evolving a mind, not a biological brain *per se*.

Setting the requirements of an AGI is not a simple task, yet it is the most important chapter in this whole book. Because we will be using genetic algorithms to evolve the AGI and its components, we will need to have clearly defined selection criteria for them to use during this evolution process. These selection criteria will be used to pick the most successful AGI candidates and components during the selection process, to be cross-bred, trained and evaluated as the next generation of AGI candidates. The requirements are like the mold we shape, and the AGI is the gelatin that conforms to the mold. We design indirectly with evolution, using the following requirements as the selection criteria in the genetic algorithms.

High Level Requirements - Historically, how have we defined intelligence and what do we expect of an artificial general intelligence at a high-level?

There has been much disagreement throughout the years over what constitutes 'intelligence' and how to test it, whether the intelligence be human, animal, or computer. Furthermore, humans seem to move the bar depending upon who or what is being tested. Identifying testing requirements for something that isn't human, such as a computer AGI, is probably the hardest for us, as we have trouble being objective when it comes to measuring something we consider so central to our own identity. Therefore, it is important to clearly define what constitutes an AGI. There has been significant discussion over the last few decades on this topic.

A notable example from 2002 laid down some basic requirements for AGI (3). Peter Voss suggested AGI comprises essential, domain independent skills required to gain domain specific knowledge and skills. This is the ability to 'learn anything'. This ability must be autonomous, goal-directed, and adaptive. Furthermore, AGI must "*inherently* process temporal data as patterns in time" to be of any use in real world situations. The AGI must have sensory inputs, a way to store knowledge, and actuation (a way to act upon the world). Pattern learning, exploration, and coherent knowledge

storage are therefore baseline functionality which lead to AGI, but do not define AGI itself. Ironically, these functions are mostly missing from traditional deep learning AI, which reinforces the impossibility of creating AGI from the 2020's version of DL AI.

IQ Tests - We could just decide to use the IQ tests for human intelligence like the WAIS-IV tests that test a wide range of verbal, memory, spatial, and perceptual organization skills to compute an overall full-scale IQ score for our AGI candidates.
2020's DL-based AI and general computer science software are much better at some tasks than others. For example, specifically coded CS and DL algorithms would perform 100% on some of the arithmetic, memory, letter number sequencing, and digit symbol-coding tests (if we formatted their problems into a computer-friendly format), have a harder time with picture completion and visual tasks, and really struggle with some of the other, more general verbal and spatial problems. IQ testing has recently been done with ChatGPT and has had some interesting results, with it performing very well on some tests, but not others. These tests include:

MMLU-Representation of questions in 57 subjects (incl. STEM, Humanities and others)
Big-Bench Hard – Diverse set of challenging tasks requiring multi-step reasoning
DROP - Reading Comprehension (F1 score)
HellaSwag – common sense questions
GSM8k – Basic arithmetic manipulations (including grade school math problems
MATH – Challenging Math problems (incl. algebra, geometry, pre calculus and others)
HumanEval – python code generation
MMMU – Multi-discipline college-level reasoning problems
VAQv2 – Natural image understanding
Text VQA – OCR on natural images
Infographic VQA – infographic understanding

MathVista – Mathematical reasoning in visual contexts
VATEX – English video captioning
Perception Test MCQA – Video question answering
CoVoST 2 – Automatic speech translation (BLEU score)

Our goal is not to have specific codes and algorithms for specific tasks, but to evolve a general intelligence that is good at all real-world general tasks. We will look at these tests, what they measure, and then design one system that has common components that can span solving all these problems.

In the context of AGI, the above human tests of intelligence may actually seem rather narrow themselves. A computer could potentially master any one of these individual tests, but the important features of AGI are the ability to learn new skills and transfer learning from one task to something entirely novel.

What is truly interesting is that the core of human intelligence, our ability to solve visual-spatial problems, use numbers, do math, use speech to communicate, and other very advanced capabilities (learned on a vast amount of real-world information) can be evaluated accurately using reduced problems with simple pictorial and verbal representations in the WAIS tests. This suggests that human core intelligence and cognition are separate from the perceptual inputs that feed information to that intelligence. We have a higher-level abstract representation for this information that is independent from the complexity or source of that information.

This also explains how babies progress from seeing simple objects and listening to basic sounds and words to perceiving and processing more complex objects, events, vocabulary and stories as they get older, exercising the same underlying intelligence and cognition with progressively more complex inputs and outputs. This gives us insight into how we could start training, selection, and evolution of a proto-AGI with simple requirements and basic inputs and progress to more complex inputs and tasks as the AGI evolves and grows in complexity and capability.

Then, a comprehensive evaluation could be conducted measuring how well learning in one area transfers to another task. The AGI should accelerate its ability to master new tasks as this transferable intelligence builds. This is a test of insights and assessing learning itself.

Scientific Computing - Data input and transformation to basis sets for use with common computational methods

Since computers were invented, their most significant limitation is that they are unable to interpret the real world around them. They have counted on humans to format the world into data that the computers can understand and operate on, and for humans to interpret the results. Any method we use for AGI has to take all the varied real-world data, accessing computer-readable databases as well as external inputs, and reduce it to a set of common internal data formats. If we take a cue from scientific computing we could do this via basis sets that cast varied types of real world data into a common format to be operated on by similar internal cognitive, predictive, and planning operations. Coming up with the correct basis sets is the toughest part of the problem, as it has to span the entire set of the data (consisting of the entire world, in vision, audio, text and other data modalities), and we have to be able to reconstruct any input/output data from a combination of the elements of the basis set.

Databases – computer storage, search, and recall at a superhuman level

A realm where computer science far surpasses humans is data storage, search, access, and recall (both in the amount of data and in the speed and accuracy of recall). One of the things that makes computers, the Internet, and the applications running on them so powerful is the enormous databases they can access and search to deliver us the information needed when we need it. Any AGI that aims for superintelligence is going to have to be able to access, search, and retrieve information from these databases like any computer information system, and do so much better than a human. In the AGI, we would add the ability to abstract the information in the databases hierarchically and build connections between the abstractions so it can do more abstract queries and reasoning with the data.

So now we have explored some boundaries. We described how the brain works (in simplified form to the best of our knowledge), and how to evaluate the intelligence of babies, children and adults with progressively more complex cognitive tests in spatial, pictorial, numerical, and verbal areas, and we can apply those same requirements on an AGI. We also learned how a supercomputer simulation operates on a common numerical representation for multiple types of problems, casting the input data from a general representation into that numerical representation by convolving it with a set of basis vectors that span the data space that can (by linear combination) represent any data in that space. We determined that we have to incorporate database query and other computer science methods into our scheme to compete with 2020s machine learning and computer science methods. These will all be useful in setting the requirements of an AGI.

General Intelligence Requirements - What do we really require of an AGI for it to meet our criteria?

To go truly beyond where we are today with AI, to pass the threshold of human intelligence, and create an artificial general intelligence, requires an AI to have the ability to experience its environment, take in a wide variety of diverse inputs, and to dynamically learn to process these input data streams fed to it. It needs to be able to learn to transform that diverse data into a common internal representation that its internal models to operate on, where both the internal representation and the models accurately represent (with abstracted information) the objects, people, events, environment, and data in the world of the AGI, and it can expand as its knowledge of that world expands.

First, the AGI needs to be capable of performing the tasks and jobs humans are capable of and to be fluent in human language in order to be able to do those tasks and professions as well as or better than a human.

Along with language fluency, an AGI needs to be able to interact with humans, conversationally and verbally (and in text conversations) at a human level, and be able to understand the experiences, events, and concepts behind the words and sentences of language and how they connect, so it can compose language at a human level of intelligence and fluency. One measure of its capability to do this is the Turing test in which people text chat with either an AI

or a person, and 3rd party observers try to determine which they are chatting with. Once the observers cannot tell the difference between the human and AI conversations, the AI is said to be Turing-complete. It is by no means a comprehensive test of language capabilities for an AI, but it is one that will have to be met to be considered AGI.

The AGI needs to be able to set goals, either by itself or by human input, then be able to break those goals down hierarchically into tasks and subtasks, and finally into actions to be taken. It needs to be able to consider the environment and other inputs in this process so that it makes the correct decisions and takes the correct actions for the specific situation. For example, if the task is self-driving, the AGI needs to understand and be able to communicate that the goal of driving is to get from point A to point B safely, while obeying the rules of the road and the customs of the local drivers. It should be able to tell you that safety is the highest priority, and that it may bend the rules slightly, such as encroaching into an oncoming lane to go around a stalled car, only if safe to do so to complete the goal of navigating to point B.

The AGI needs to be predictive. It must be able to generate an internal model of its world from its perceptions and use that model to make predictions of what is going to happen next. It needs to be able to use this predictive ability to explore solutions to problems and decide on the best course of action. It also needs to be able to imagine or 'dream' about fictional scenarios and use these hypothetical outcomes to augment its model of the world.

The AGI needs to be able to predict the outcomes of multiple possible solutions, evaluate them, and decide on the most optimal solution, sometimes having to use sparse data and guessing (something computers are historically very bad at). It needs to learn to get better at this process with time both for similar goals but also transfer this learning to different goals and situations so that it becomes better at problem solving overall.

The AGI architecture must use one, generalized, elegant architecture to accomplish all of these requirements, with components built out of the same fundamental building blocks.

Requirement	How pass as an AGI
Prerequisite definitions	• Autonomous • Goal Directed • Adaptive
Prerequisite capabilities	• Has sensory inputs • Has actuation • Inherently processes temporal data as patterns in time • Learns patterns • Explores (searches) • Has coherent knowledge storage (memory & recall) • Transforms information into basis sets
Possesses domain independent skills required to gain specific domain knowledge and skills	A single base architecture can be trained to solve problems in any domain. Passes a WAIS test or similar

Generalizes experience from one task to other related tasks	The speed at which a new task is solved increases with increasing experience in related tasks
Breaks down goals into tasks/ subtasks, and then breaks these tasks into actions to be taken	Can demonstrate a hierarchical structure to achieving goals through communicating prioritized tasks and the actions which complete these tasks
Demonstrates language fluency	Passes 5 types of language tests including aptitude, diagnostic, placement, achievement, and proficiency.
Interacts with humans conversationally	Passes Turing test or similar with open-ended conversations
Predicts what happens next	Create a dataset, assign specific predictive tests, check that output matches or exceeds expert human results

Decides on the most optimal solution from the outcomes of multiple possible solutions	Give multiple solutions to evaluate and the AGI picks the best solution

Now that we have more precisely defined our requirements, we can turn our attention to coming up with a conceptual technical design that meets these requirements.

CHAPTER 4

Technical Design for an AGI

Based on the AGI requirements Chapter 3, we propose a way to evolve an AGI architecture that can learn to do all the types of tasks required of it according to our specifications. As stated, we do not seek to create an exact digital version of the human brain, but rather to draw inspiration from what we know about brain functions, and what it can do - to evolve the architecture to the requirements of Chapter 3. We will then combine this desired feature set with the computer science, database, and scientific computing capabilities outlined and use them as the design requirements to evolve a completely novel architecture that encompasses the strengths inherent in the human brain and the strengths of these more conventional computer science techniques to build a hybrid functional AGI for practical use, capable of running on today's digital architectures as well as future specialized neural network architectures.

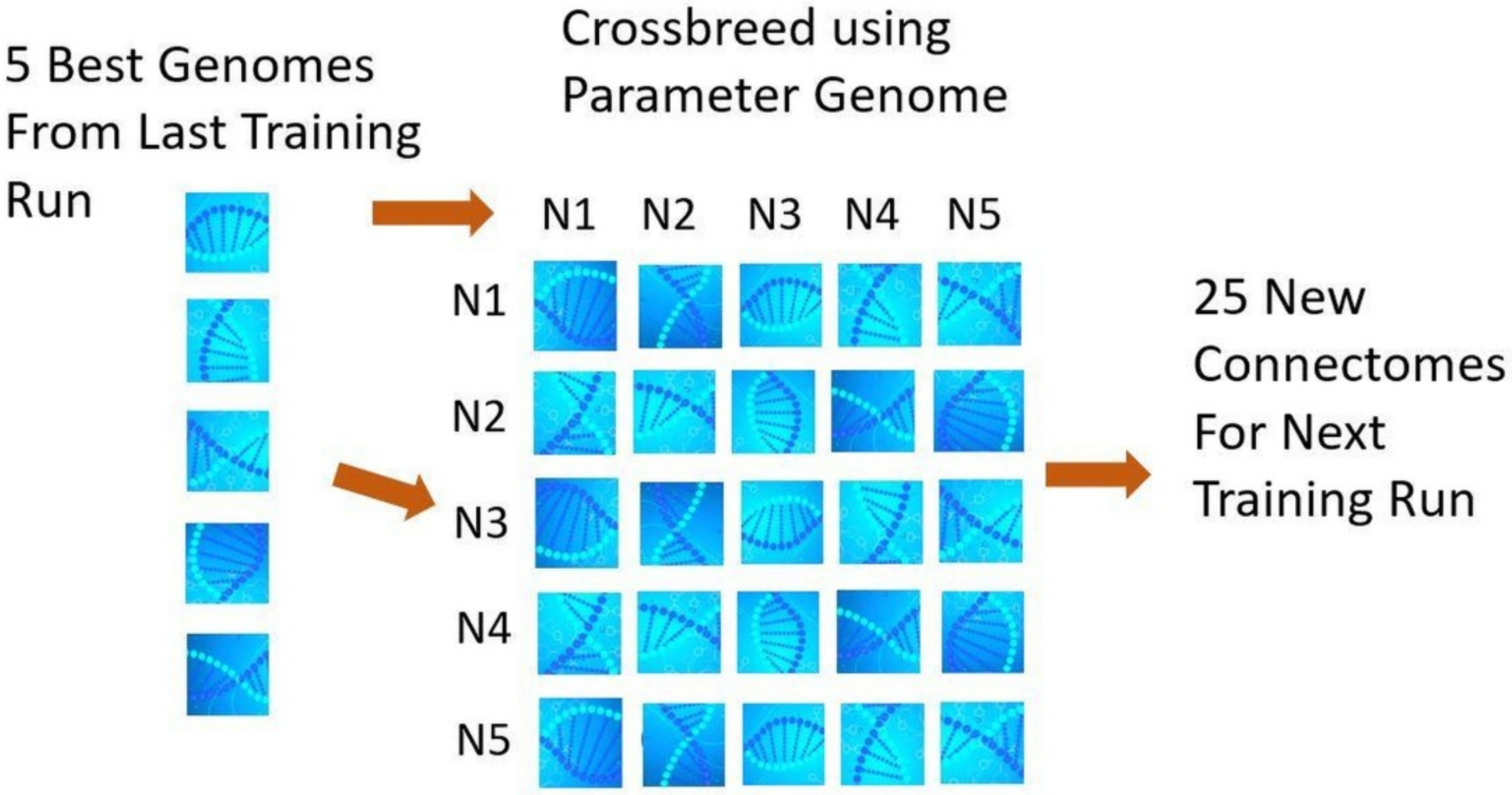
Nobody in the world currently understands the human brain well enough to directly reverse engineer it, let alone implement it in software and hardware, beyond a few computational neuroscience experiments on a small scale. We don't even know how a cortical column actually works, let alone how to replicate it. Many countries and companies have invested billions of dollars into such research, but the exact functioning of the brain's neural networks remains a mystery. However, we do know enough about the end functionality to use genetic algorithms to evolve everything from our low-level components to a final AGI architecture.

Genetic Algorithms - Using artificial evolution to design neural network architectures.

All of the components we are going to talk about are evolved in a genetic algorithm that the AGI uses to learn to solve problems. Evolution and genetic algorithms are like making Jello. We cannot shape the Jello directly (neural nets in this case), so we make a mold (evaluation criteria) that shapes the neural nets as they undergo evolution by culling the ones whose performance doesn't meet the criteria nor fit the mold and keeping the ones that perform according to our selection criteria and best fit the mold. Then we crossbreed the most successful ones to create more similar neural nets for the next generation. Over successive generations this genetic algorithm process evolves the neural networks closer to the design that we desire, that fits the function we specify in the selection criteria.

However, trying to evolve whole neural networks doesn't work, because in each generation you would have to make slight changes in one of millions, even billions of neurons and synapses and then test to see if those changes made the neural network better at the given task or not. It would simply take too long to converge it to a useful solution.

By representing the neural network in a compressed form, by a compact genome that encodes all the neuron properties and connectivity patterns, then crossbreeding and evolving those genomes and expanding them to neural network connectomes and testing those, we have a lot fewer parameters in that genome that we need to tweak and adjust using genetic algorithms, and the evolution can go a lot faster. This is the same way nature works, doing evolutionary selection on the 8000 genes of the human genome that represent the brain's connectome.



By running genetic algorithms to create and chain together functionality that operates on our internal representation of the data, the AGI evolves functionality that can do arbitrary operations on data to produce the desired results and accomplishes transfer learning by applying that functionality evolved on other problems to new, similar problems.

Spiking Neural Networks - Going beyond deep learning neural networks

For our basic unit of synthetic neural computing, we will use spiking neural networks (SNNs), which model neurons as discrete computational units that work much more like biological neurons. These SNNs fundamentally compute in the time domain, sending signal spikes that travel between artificial neurons and synapses, approximating real neurons with simple models like Izhikevich or more complex ones like Hodgkin-Huxley (Nobel Prize 1953).

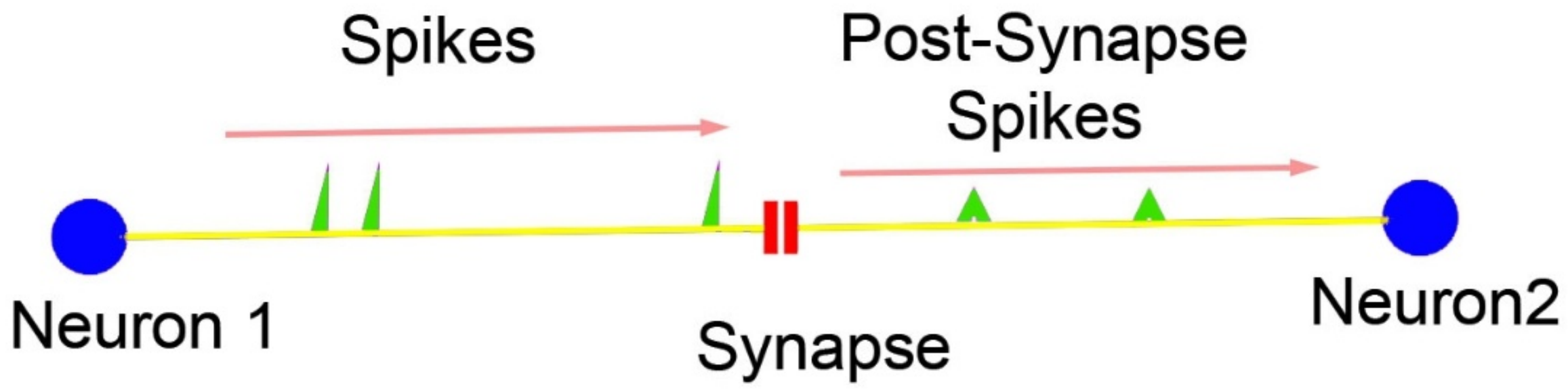
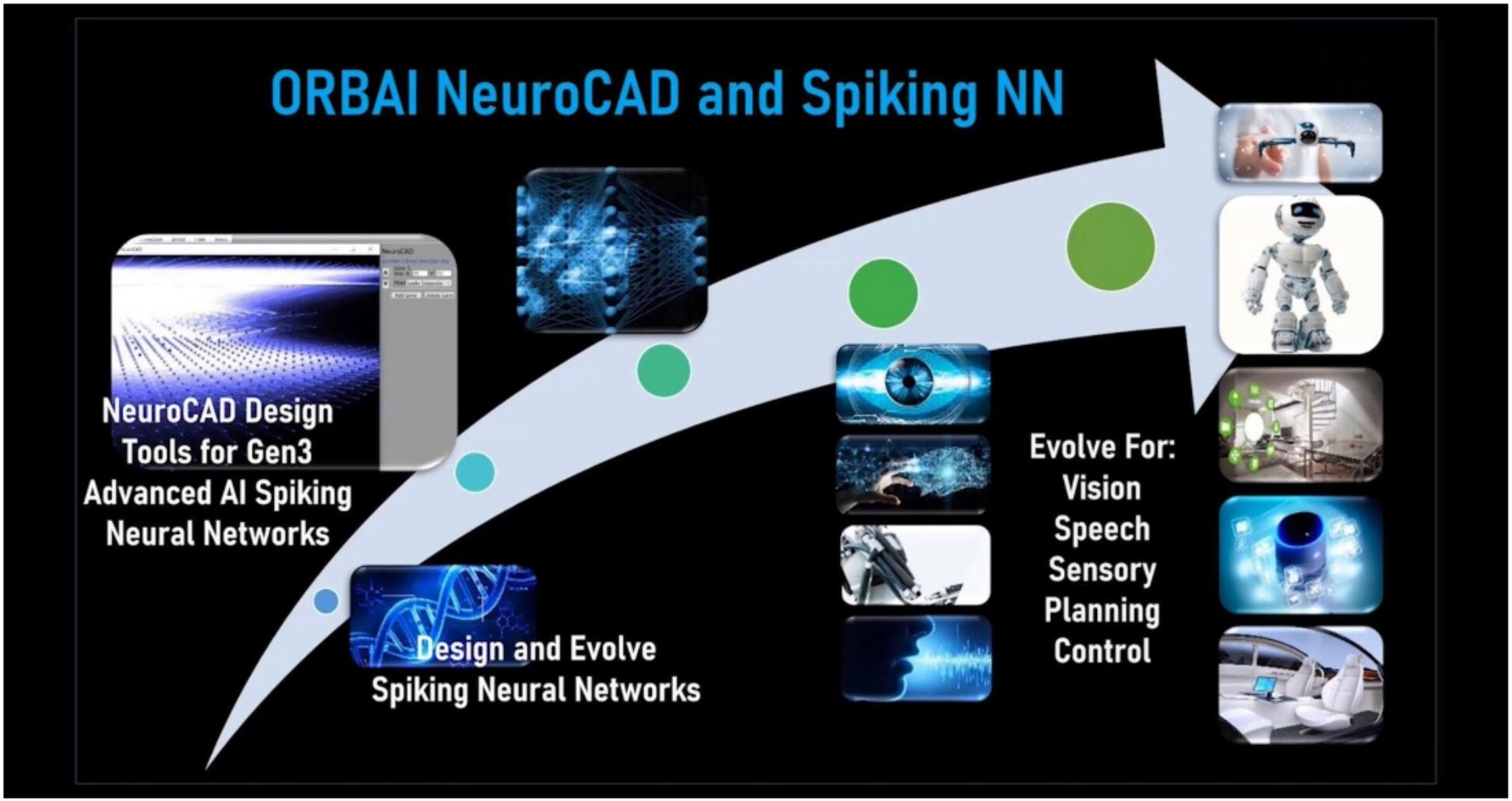


Figure 4.1 Spiking Neural Network

Spiking neural networks transmit rich information via spikes trained along dendrites and axons and the signals undergo integration and processing in space and time within the neuronal body, and at the synapses. Because there is latency from one neuron firing until the next, signals travel through a network with a time-domain behavior, with multiple signals combining in complex neural circuits that do time-based processing, which can have feedback, loops, oscillators or circuits that create timed and synchronized chains of impulses for motor control or speech applications. In general, they have much richer and more powerful behavior than deep learning neurons and neural networks. However, they are more challenging to design, train, test and work with, so we are building a CAD tool especially designed for working with Spiking Neural Networks.

NeuroCAD is a software tool designed and built by the startup company, ORBAI, with a UI for designing, training, and testing Spiking Neural Networks. It allows the user to lay out the layers of spiking neurons, connect them up algorithmically, crossbreed and mutate the resulting networks to generate a population of similar neural nets. Then the tool runs simulations on these networks, trains them, culls the underperformers, and then crossbreeds the top performing designs. These genetic algorithms continue until a design emerges that meets the performance criteria set by the designer.



NeuroCAD defines a process for designing, connecting, training, and evolving neural networks to a specific functionality. To accomplish this, we construct an artificial neural network as a set of layers and define a set of numbers, called a genome that can completely characterize all of the connections between all of the neurons in these layers. The process takes small sets of the parameters from the genome, inputs them into procedural algorithms to generate two-dimensional (2D) probability maps defining how connections to a layer are distributed in the plane of that layer. In addition, a weight is assigned for how many connections go to each layer above and below a given neuron layer, rendering a 3D distribution of connections when combined with a 2D probability map for each layer. When given a random seed, combined with stochastic sampling of this 3D probability distribution, this allows one to algorithmically compute to which target neuron the connections from the source neuron map to, and to expand these distributions deterministically into a much larger connectome for the entire neural network via this process. Genetic algorithms are used to search the small space of the genomes that expand into the full neural network connectomes, which are trained, evaluated, and selected for which of their genomes to crossbreed for the next iteration, allowing simulated evolution of large artificial neural networks.

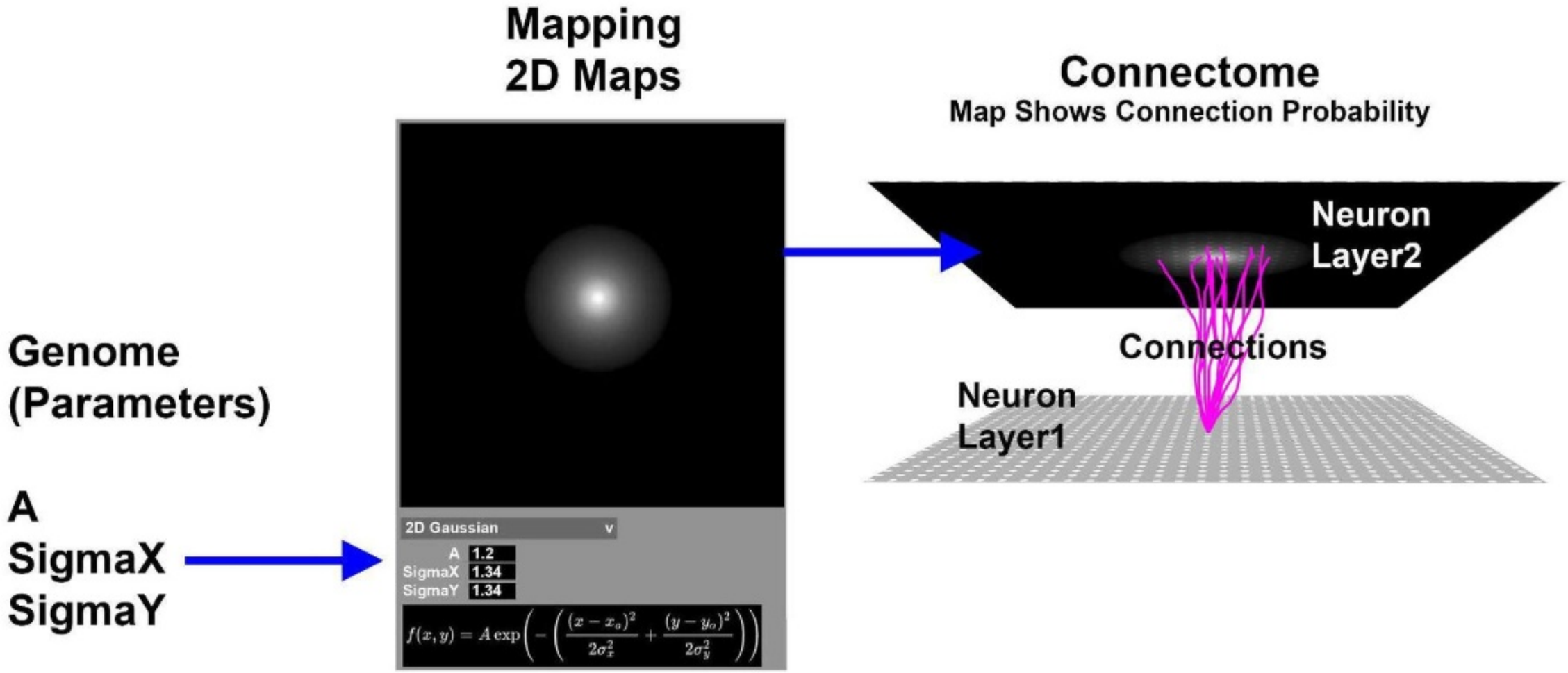


Figure 4.2 Mapping Genome Parameters to Probability Maps

NeuroCAD is specified in more detail in the [Patent, US # 11514327](), which covers the implementation of SNNs, the tools used to specify their layout and connectivity with a genome, and the tools for implementing genetic algorithms to evolve these genomes and their corresponding connectomes to the desired functionality.

Using these processes, we are evolving the core learning circuits for an AGI that can simulate how human intelligence can learn from its environment, by observation, mimicry, and practice, first learning to interpret its various senses and data inputs into an internal representation. In the process it can see how they fit together and how they connect into

sequences of events, and can develop a common, internal, AGI network-friendly format for all these types of data so that the AGI can more easily process and make sense of them. From these inputs it can build a model of its world, make predictions with it, and make plans. It can then translate those into outputs and actions that make sense in the real world, including speaking human language.

To date, application of spiking neural networks has remained difficult, as finding a way to train them to do specific tasks has remained elusive. Although Hebbian learning functions in these networks, there has not been a way to shape them so they can learn to do specific tasks. Backpropagation (used in DNNs) does not work because the spiking signals are one-way in time and are emitted, absorbed and integrated in operations that are non-reversible and non-differentiable. We choose to train these SNNs by feedback and feedforward circuits that learn by making generative predictions and strengthening the neural connections of the predictor circuit when it matches the perceived state. We will discuss this method in depth later in the chapter.

Autoencoding Input and Output - We need a more flexible connectome or network connection structure to train spiking neural networks. While DNNs only allow 'neurons' to connect to the next layer, connections in the visual cortex can go forward many layers, and even backwards, to form recurrent loops. When two SNNs with complementary function and opposite signal direction are organized into a recurrent architecture like this, Hebbian learning now helps train them to become an autoencoder, that is able to encode spatial-temporal inputs such as video, sound or other sensors, and reduce them to a compact machine representation and then decode that representation into the original input and together provide feedback to train this process. We will evolve this autoencoder by setting the selection criteria to be that the output needs to be the same as the input, and then run genetic algorithms on a set of candidate networks and use artificial selection to drive the network design towards a functional autoencoder.

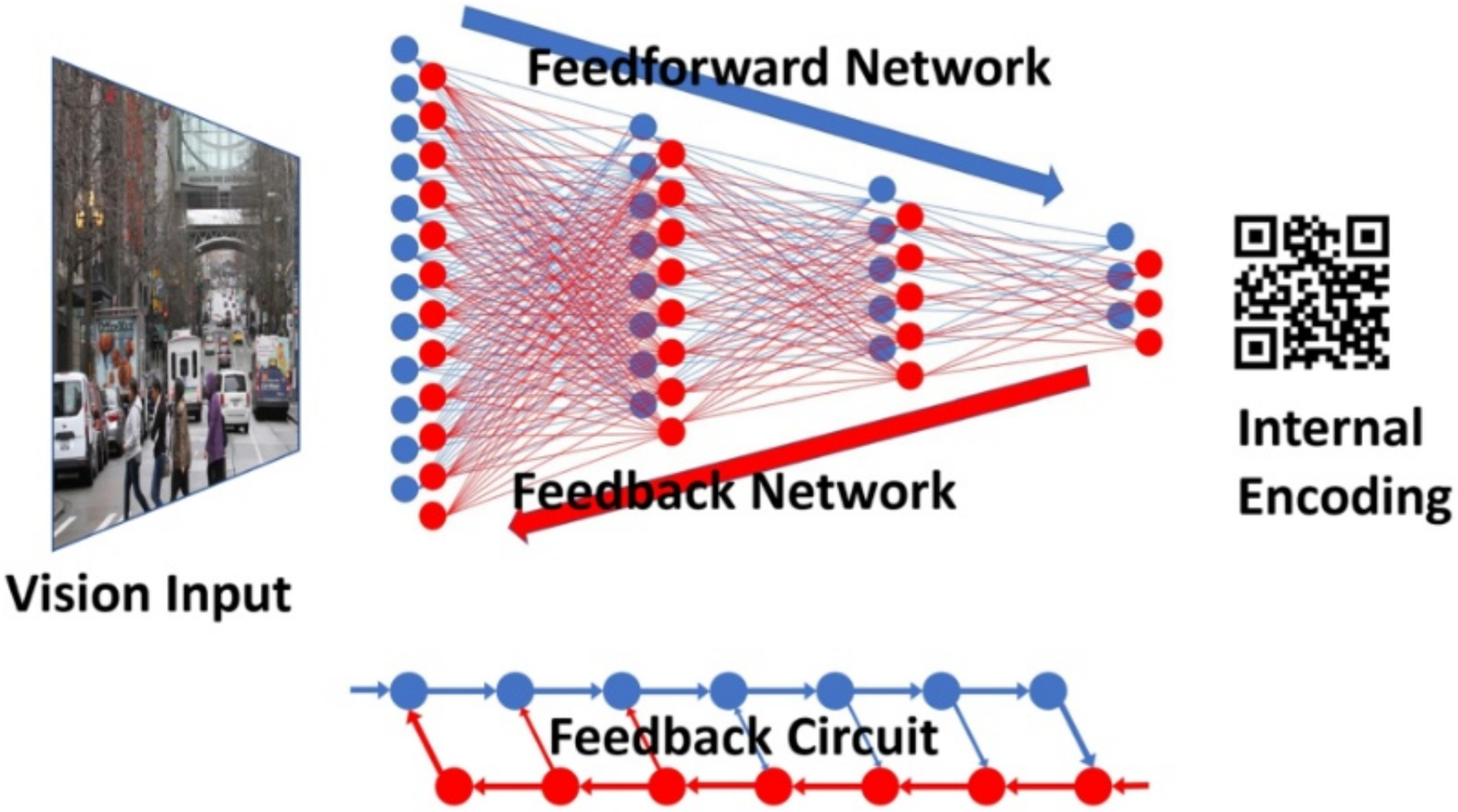


Figure 5.2 Bidirectional Interleaved Complementary Hierarchical Neural Network

Internally, the autoencoders process input and sensory data into a set of basis vectors and basis coordinates in the latent encoding, with the internal representation of information in the form of the basis coordinates coming from the internally encoded layer in the middle of the autoencoder. This can be used stand-alone for vision, speech, and other processing, or integrated with the rest of the AGI and used as the input layer to the artificial cortical columns, with each cortical column representing a basis vector or feature in the input data to the 2D sheet of cortical column inputs.

AGI Cortex - Now we combine these concepts to make an artificial cortex, complete with artificial cortical columns and functioning in a manner as similar to its biological equivalent as we can make it.

The first step is to evolve an artificial cortical column that meets the criteria of Chapter 3, such that it can process inputs, form and recall memories, and do generative predictions based on the world model formed by those memories. We do this by setting these properties as the selection criteria in a genetic evaluation process and score our cortical column designs by how well they perform on these selection criteria. We then use the NeuroCAD tools to lay out some networks that are initial guesses (based on our knowledge of neuroscience and the neuroanatomy we know about cortical columns). We then crossbreed these designs, and perform evaluation runs on them, scoring their performance vs the evaluation criteria, then selecting the top scoring neural networks to crossbreed their genes for the next evaluation run. This genetic algorithm can explore designs that no human could intuitively design and do so based on the properties of the artificial SNNs. The evolved cortical column design will differ from the natural one in the brain using biological neurons, but its functionality will be constrained by the selection criteria and evolution process such that it performs a similar set of tasks.

A sub-circuit that we can start with is a predictor. In the human brain's cortical columns there is a feedforward / feedback neural circuit that is continuously generating predictions of what the column will experience in the future, based on the past inputs and memories formed by the synaptic connections between the neurons in the column and adjoining columns. When this predictor is correct and it matches the next set of inputs, the synapses in the predictor circuit are reinforced, causing it to learn how to better predict at the level of the cortical column. This is how the brain learns at its lowest level, by constantly trying to predict the world outside of it and reinforcing the neural circuits whenever the generative prediction is correct. There is no separate training stage or backpropagation used like in deep neural networks. By adopting this same neural architecture, our AGI brain learns on the fly while operating, just like the human brain.

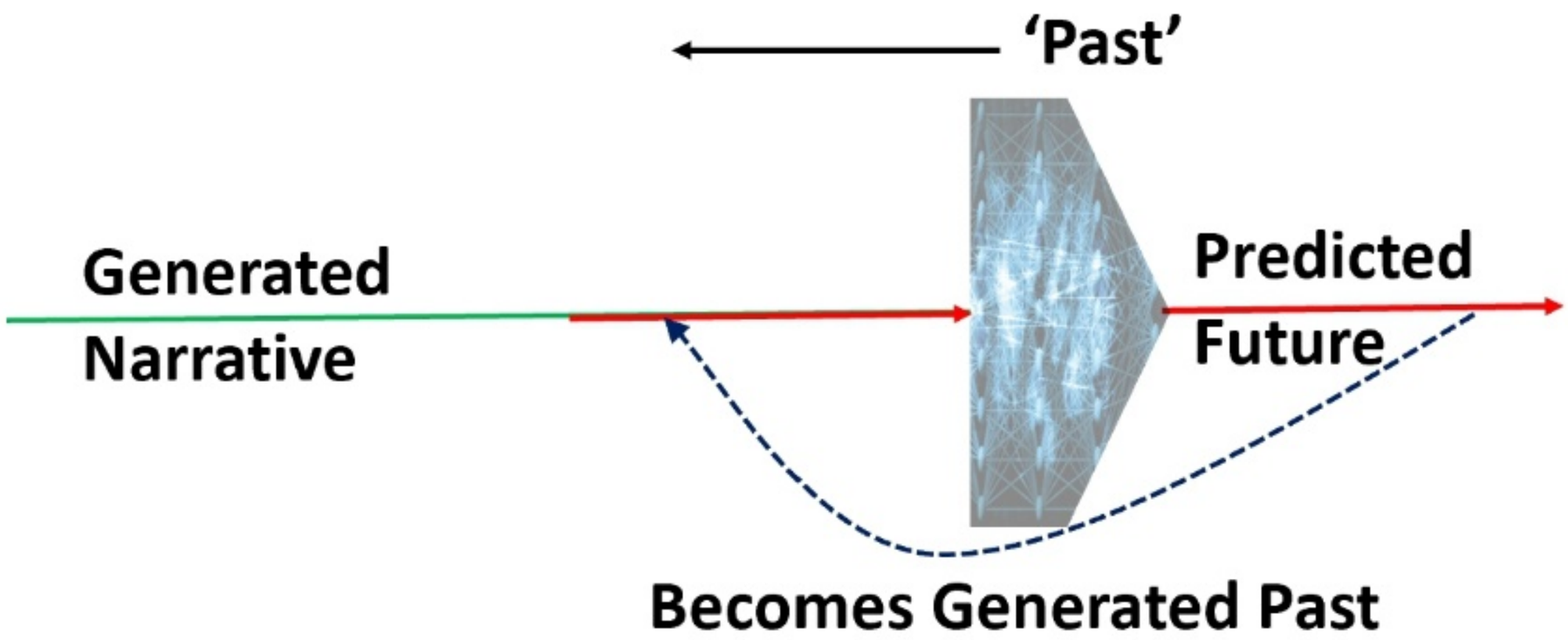


Figure 5.4 Generating A Narrative with the Predictor

We structure our cortical columns like in the brain, with the inputs coming into the middle of the column, the generative predictions being computed by the top layers, and the comparison of them being done in the other layers, reinforcing the neural connections when predictions match the input.

We mimic the architecture of the brain's cerebral cortex by constructing it out of a sheet of these synthetic cortical columns and build a set of AGI cortices out of these cortical columns. We do so using the same evolutionary process, by making multiple designs of the connections between cortical columns using NeuroCAD then evolving these connections using genetic algorithms at a macro level, keeping the sub-design of the cortical columns fixed. We set the same evaluation criteria as we used on the evolution of the columns, now applied to the full AGI cortex with full-fledged inputs and tasks to perform. We could further refine our cortical column design by evolving it between the evolution steps of the macro cortex structure, to help converge the cortex design and match the cortical columns to better function in it.

Now when we want to use the AGI to solve a problem, we can simply give it an input prompt and let the generative artificial cortex provide the continuation of that input stream to provide an answer. It does this by generating the next engram in the narrative based on the past engrams, advancing one step at a time, and using its own output as past input for subsequent steps. If we have properly trained the generative predictor, it will create a narrative for us that is logical and that answers what we asked.

This is far superior to large language models, because the AGI creating the narratives works in multiple modalities and with different types of data - text, pictorial, audio, video, stock charts,... and any type of input that has a corresponding cortex in the AGI. It also has a generative AI that is much more sophisticated than RNNs or transformers because the neural networks use sophisticated neural computing with SNNs to do the generative prediction, not just relying on statistical inference like the DL methods.

However, we want to do more than just provide an output given an input. We want our AGI to think, plan, and reason. We need more than just an AGI input cortex that predicts and generates. We need a component comparable to the human hippocampus, that provides input to the cortex for what memories to recall, and a decision cortex equivalent to the human prefrontal cortex that tells the rest of our artificial brain what narratives to simulate and predict and does the problem solving process with that data. To do this, we split the AGI into input, decision, and output cortices as in Figure 5.5, connected by a memory controller that the decision cortex uses to gather the input and memories it needs and to issue the outputs.

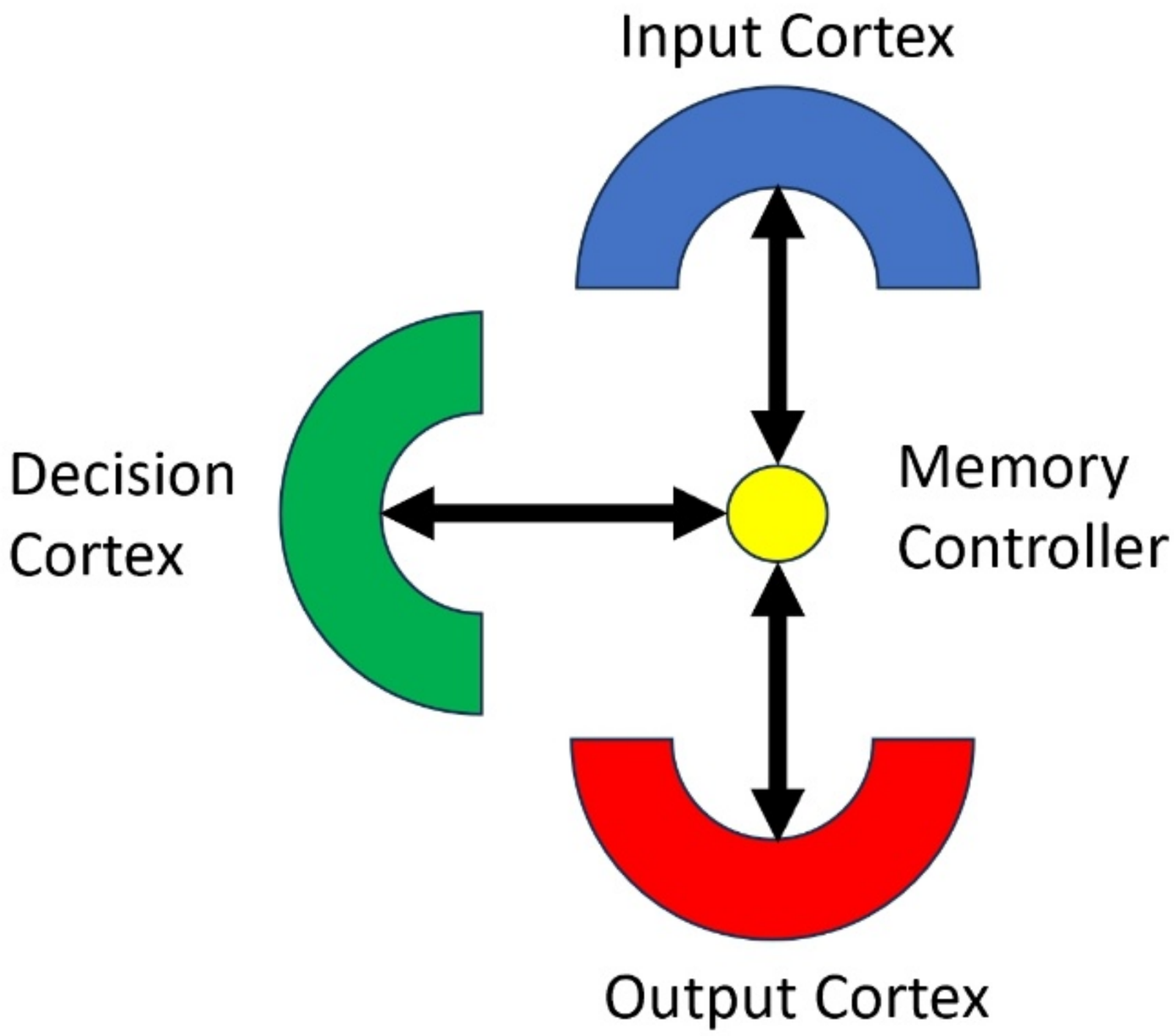


Figure 5.5 Cortices in the AGI and their connectivity

We learned that the human brain starts a decision-making process with an instinctive, emotional inception, then based on that instinctive decision, the hippocampus and prefrontal cortex expand it into possible solutions based on generative predictions and comparing the outcome of those predictions to the goal. We will reverse engineer this basic functionality to come up with our design.

We need the decision cortex to be capable of setting high-level goals from within the AGI or by human input and to have it be able to generate solutions by breaking those goals down hierarchically into tasks and subtasks, and finally into actions to be taken. We need a recursive process that can instantiate goals for itself. It then needs to be able to consider the environment and other inputs in this process so that it takes the correct actions for the situation. Furthermore, it needs a way to evaluate multiple generated solutions against the goal and choose the best one.

The decision cortex would be able to use its predictive, generative capability to map out multiple possible courses of action, and evaluate which will best meet the goals, then choose the best course of action.

Then, the internal representation of these courses of action can be transformed by the output cortex to outputs or to signals to drive actuators in robotics applications. Training of our AGI is done by reinforcing the network when there is a desirable output and causing the artificial neural network of the AGI to strengthen the connections between inputs, cognition and outputs to produce similar outputs.

By taking inspiration from the human brain's cortex and how it interprets the senses and models and predicts the world it perceives from the senses, and by using more powerful neural architectures (SNNs), we can build an AGI that can function similarly to the human brain, with the ability to handle general problems. The data that can be input into this AGI can be much more varied than our six senses, and new artificial cortices can be evolved to handle each modality of data and output. By scaling this architecture, able to handle inputs and outputs hierarchically, and giving the AGI access to a vast amount and variety of data, we can push its functionality beyond what the human brain can experience and process, possibly someday building an AGI that is superhuman in capability. More on this in Chapter 15.

We now look at some specific application implementations for AGI in the following chapters, showing how this general system we have created can be scaled and specialized to different fields now dominated by human intelligence.

CHAPTER 5

AGI in Language and Speech Applications

This section is about language and abstract reasoning and how we model them in our AGI such that language becomes a narrative that acts as a backbone for the other modalities of data and allows the AGI to relate them together and abstract them.

Our AGI will have a cortex specialized to handle language, with the ability to not only predict what word comes next in a sentence like an LLM, but also the ability to cross-reference language with other memory in other cortices such that it is able to abstract the meaning of the words and sentences and make decisions based on meaning not just statistical inference.

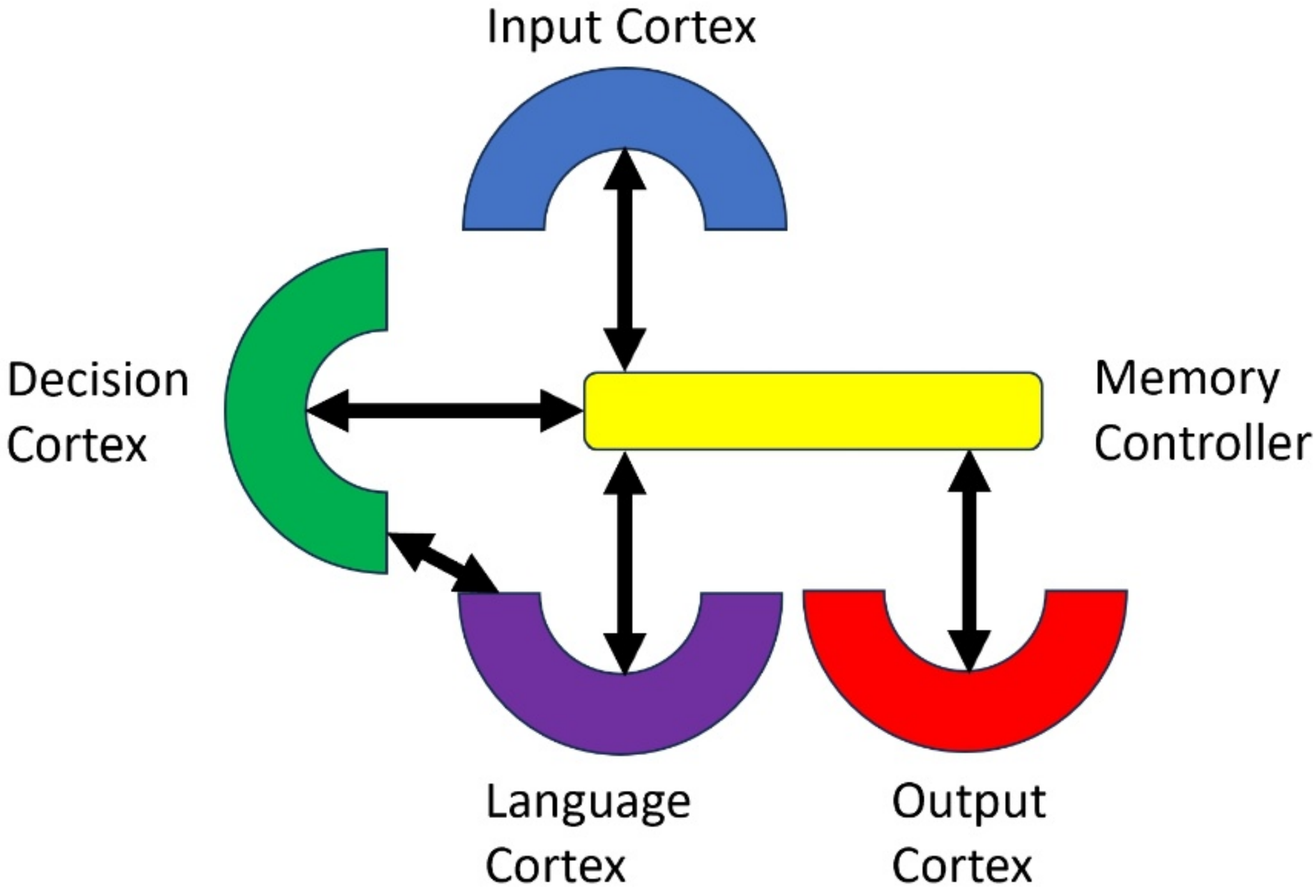


Fig 4.4 AGI With Specialized Language Cortex

Language is an example of a narrative, which all the cortices handle for the different modes of input, but language also has a hierarchical representation. It can be represented at the lowest level as a stream of letters (for text), or phonemes (for speech), which can be hierarchically structured as higher-level structures such as words, phrases, sentences, and

paragraphs as in our Jack and Jill example, where the highest levels of abstraction are linked to visual information like in the picture (or video), and to similar narratives.

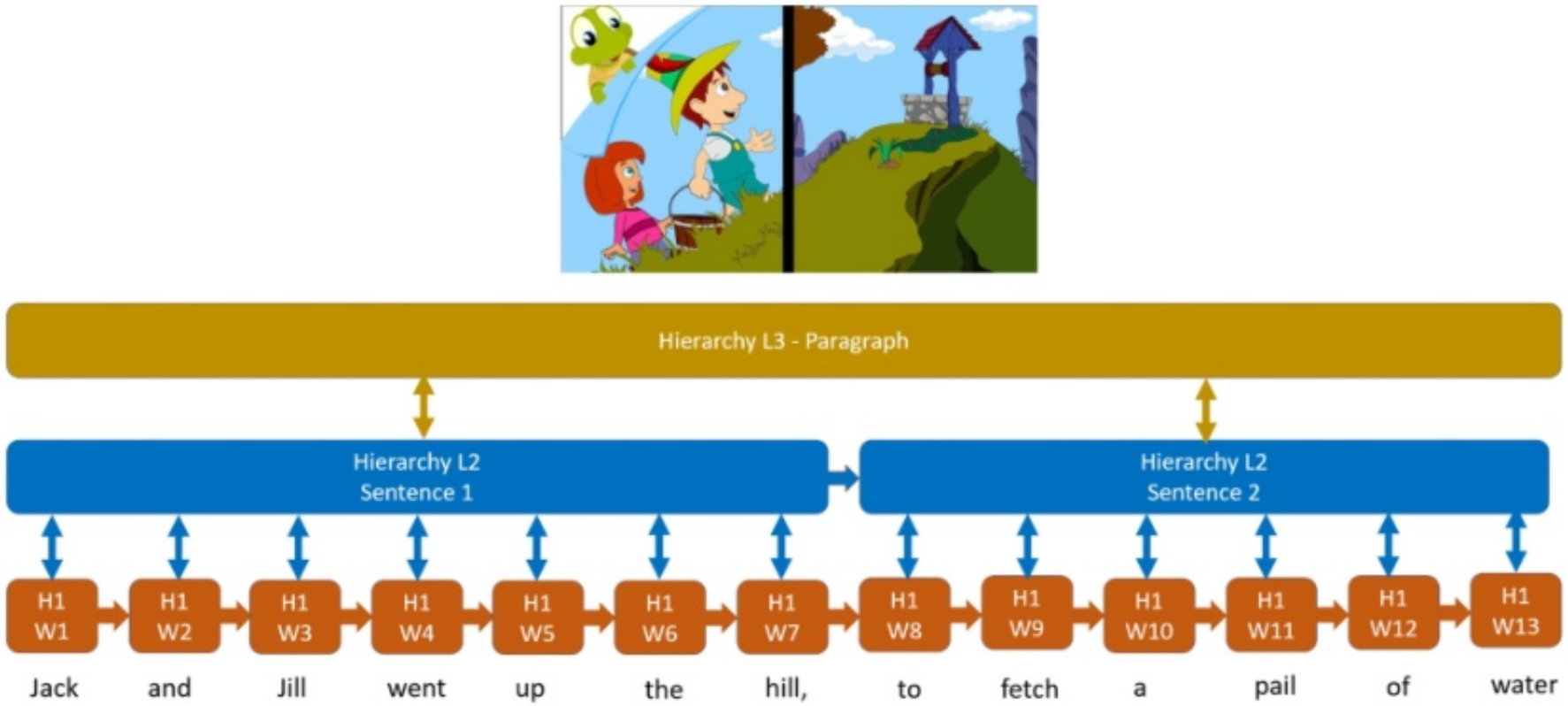


Figure 4.3 Hierarchical Representation of Language and Meaning

Language is a type of memory narrative (composed of hierarchical engrams in our AGI) that forms the backbone for all other forms of narratives, not only labeling the data with that language, but forming a cognitive monologue by which the AGI can construct engrams and actions – the same language monologue that our AGI's neural networks operate on. This is how we accomplish the recursive capability of our AGI decision cortex. It uses human language as an internal monologue or instruction set that it processes tasks with. It can expand a simple set of instructions into a much more detailed list of sub-tasks in the same way that the language cortex can expand a sentence into an essay based on that sentence, similar to, but superior to LLMs.

By organizing the internal data hierarchically, with the higher levels of the hierarchy abstracted and cross-linked to similar abstract data, and language being the backbone of that data, we go beyond existing deep learning large language models and allow our AGI to explore the higher-level relationships between objects, sequences, events, and the language describing them and tying them together.

This use of such abstraction and language leads to an AGI that can converse naturally with a human with fluid and fluent speech, and also allows reasoning and planning in many human professions like medicine, finance, and law.

As an example for a language subsystem that can converse naturally with a human using this generative prediction capability of the AGI, we train a cortex specifically evolved to learn human language on a large number of textual references of including factual databases as well as human conversations, from which it learns composition, grammar, and the underlying meaning of the text, making it able to generate more accurate responses to queries or questions. It trains to learn the narrative of words and sentences stated by each person in an alternating conversation by using our predictor model. This basic model is one of an adversarial conversational model and we will later see how it generalizes to other adversarial systems in medicine, law, and other fields.

Once it is trained on text and conversations, this system has learned more than the statistical weighting of sequences of words that RNNs and Transformers do in deep learning. It has actually built an internal neural computational model with the capability of reading, writing, and conversation. The language cortex can respond fluently using human language, using multiple modes of abstracted, cross-linked data, providing a far more advanced model than the statistical modeling of deep learning methods for language.

Then when actually conversing, the AI uses the language predictor to make multiple generative predictions of what the other person will say next, continuously updating those predictions and narrowing them as the human speaks, and also uses the predictor to decide what the AI will say after the person once in response to each of the first's predictions,

continuing onward in time, and dreaming where needed to ad-lib the conversation, building a dialog tree that it can prune once the person stops talking, and collapse it into what to say next.

When the human stops speaking, the AI uses that completed speech segment and the generative predictions of what the person is predicted to say next to collapse the dialog tree and computes what the AI will say now, pulling words and phrases from previous segments of the conversation to incorporate them where appropriate. The predictor would also have connections to the information about other modalities of information relevant to the conversation and their hierarchies, including visuals, audio, date, time, location to give the words context, and to interface with peripherals.

It can also generate text from prompts like an LLM, expanding sentences into essays, breaking down a list of tasks into detailed subtasks, and other language, and other such functionality, doing so much better that 2020's deep learning large language models are capable of, drawing on a neural computational model in the artificial language cortex rather than just statistical weighting.

This also would be superior to existing speech recognition, and speech synthesis systems because the underlying AGI methods allow for the system to learn from just listening to a person speak, building a basis set of phonemes, duoemes and triemes from their voice that would make synthetic speech produced by it much more realistic, and make speech recognition much more robust, as it would be able to screen out any non-speech audio by using the basis sets convolutions, and be more able to handle slight mispronunciations and be able to train on people from different geographies to compensate for accents.

Using an ROI-Inhibitory system makes a spoken voice much smoother because it does not just try to stitch together phonemes and their derivatives, it can learn to output whole words, phrases, and even sentences smoothly. As well, in reverse, it can still understand mispronounced words, poorly worded or grammatically incorrect phrasing or sentences, and draw inference from their context within the paragraph. It would perform speech recognition and synthesis at a level better and more accurately than humans.

We would have connections between language narratives to those derived from different input types, such as visual, audio, and other data, at coordinates where they are temporally, spatially or conceptually related, such that processing of one type of narrative can reference the related information in the other type of narrative as input or output in the processing. The method would connect between the higher, more conceptual levels of the hierarchies to allow more abstract operations between the different levels, making language the backbone for our AGI's memory and cognition by connecting words to references of visual objects and sounds, sentences to form abstractions for sequences of visual and audio events, and paragraphs to form abstractions scenarios and stories in memory, with each word, sentence and paragraph connected to one or more memory.

By connecting the different modalities and anchoring vision, audio, and other data to language, we not only gain a very robust system for recognizing objects, scenes, locations, actions, events in time, and identifying them with words and sentences,... we also have a conceptual abstraction at the higher levels of the hierarchy for how these concepts fit together and co-occur, so that we can do operations on those abstractions that more closely resemble the human brain's ability to think and plan based on generalities, then being able to dive into the details. By having language as the backbone of the AGI's conceptual model of reality, it would provide us with language capability in our AGI that is far superior to what can be achieved by deep-learning LLM based language systems, instead incorporating an actual understanding of language and abstract concepts rather than just a statistical model of word sequences.

To summarize, by training the language cortex on conversational data that includes other modalities like video and other environmental data, we create a very powerful conversational AI that can model real human speech and language processes and converse fluently, with real meaning and intelligence behind that speech. This far surpasses the statistical modeling of word sequences done by deep learning, even with modern transformer methods and LLMs.

This meets our requirement for being able to abstract verbal information, define words, and make associations between similar concepts that we laid down in Chapter 3.

CHAPTER 6

AGI in Robotics

For a humanoid robot controller, we would have multiple inputs and sensors, each feeding into their own specific cortices to process each type of input. We would have high-level commands, coming from an external source or from the AGI in the robot that would be hierarchically broken down into tasks and subtasks by our decision cortex, all the way down to commands for the robot's movement and speech, which would be further processed by an artificial motor control cortex and a language cortex to generate signals to move the motors and actuators of the robot and produce speech.

That output data can be organized hierarchically so there are low level movements like 'flex pinkie finger right hand 10%' or high-level commands like 'walk forward 2 meters, turn left, and stand on one foot'. Internal engram monologues need not be pre-scripted, as they could be generated, like our above conversation, reacting to what is happening in the world (vision, audio, touch, proprioception) and predicting what may happen next, then synthesizing intelligent movement based on training and practice to find the solution that will meet the short and long term goals of the robot controller. This is the foundation of intelligent robots, being able to plan and synthesize their movements with AGI.

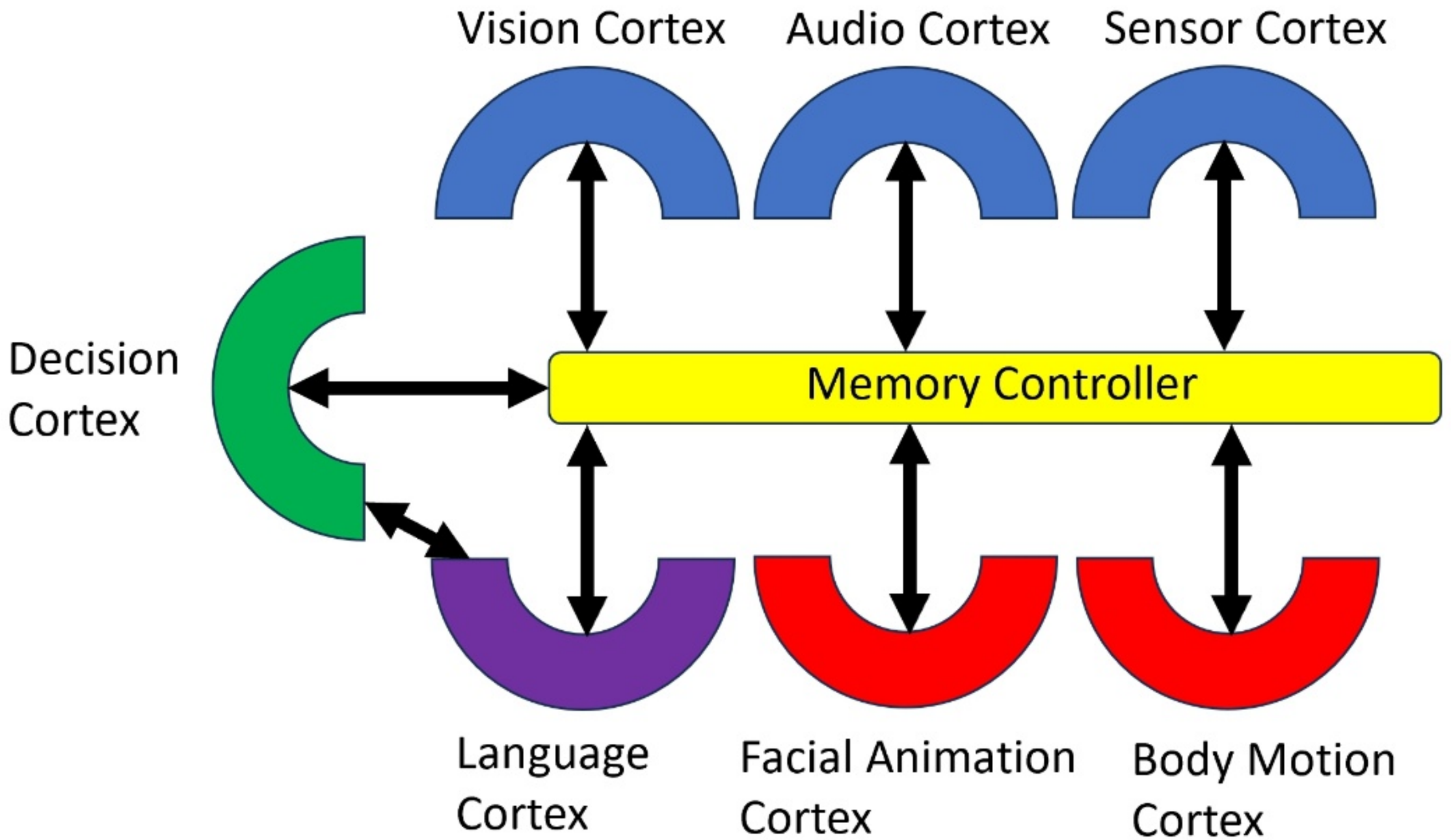


Fig 6.1 Robot Controller Using the AGI Architecture

If, like in this diagram, we split the problem into input, cognition, and output cortices, it gets a lot easier to approach. First the inputs are all encoded hierarchically to an internal representation. For example, vision input would consist of pixels which would be interpreted hierarchically to identify individual objects and then scenes. The decision cortex

in the middle takes in the abstracted inputs and the memories of past inputs and actions and performs operations on them as we described in Chapter 3, and produces high-level, abstracted commands as output. The system learns by constantly running a generative prediction of what actions to take based on current inputs, memories, and its model of the world. It maps out several options to take and evaluates which one will have the desired outcome by searching for the option that best accomplishes the goals from the predictions, given the present inputs and past context of memories at its disposal.

The output cortices then transform these high-level commands into specific instructions for actuators and other output devices to produce desired outputs for external systems (actuators for drones, robots, or peripherals, displays, and other devices), and control them. Based on input narratives, and operations between them and engrams in memory, and other means, output narratives are computed. Then these synthesized narratives are decoded hierarchically to a set of synthesized movement instructions and then to synthesized outputs which are fed to the external systems.

It would do much more than just reflexively produce the same outputs given a set of inputs. By being able to keep inputs in memory, do generative predictions, set goals and evaluate the predictions against how they meet the goals, the AI in the middle can learn to perform tasks that require advanced cognition by chaining together the operations of the cortices it has evolved to solve the problems of perception, planning, and actuation. We can build a truly intelligent robot controller.

Again, we would use genetic algorithms to design and evolve this system, starting with a simple brain in the middle with the basic set of components needed to produce useful output from basic input and memory, then evolve it to be able to perform ever more demanding tasks until it can fulfill the roles we are training it for.

Luckily, the basic principles of neuroscience apply across species, so we need not start with the goal of achieving a human model. It turns out that many features of human intelligence can be found in simpler form somewhere in the

animal kingdom. As an additional benefit for our biomimicry approach, the brains of some of these other animals are better characterized and more accessible to experimentation.

For example, entire connectomes of several species, such as insects and worms have now been fully described and functional information is being collected along with this detail, helping explain the relationship between structure and function. Drosophila, commonly known as fruit flies, have surprisingly complicated behavior and can learn associations that turn into memories that last days. A minimalistic cognitive cortex could first be modeled after the organization of the insects' neural mushroom bodies, which have been shown to provide a similar function to our olfactory cortex, but only contain a few thousand cells. Indeed, a number identification algorithm using MNIST as inputs was already developed based upon the architecture of the moth mushroom body. But we can easily go a step further.

Neuroscience has already described many of the genes involved in memory formation in fruit flies, which cells express those genes, and even the real time dynamics of the concentration of many molecules involved in signaling. When combined with the knowledge of electrical activity from electrophysiology, one can begin to put together a picture of how populations of cells change in real time during memory formation at multiple levels and within multiple time frames. This can form the basis not only of neural network connections, but the molecular networks and the signaling dynamics that underlie basic intelligence functions.

These simpler animals can even help us understand basic cognitive functions, which we generally attribute to humans. Insects have been postulated to possess emotion "primitives", such as the basic building blocks of fear. Jumping spiders can plan attacks by determining the best approach route and have even been observed altering their strategy based upon which prey species they have in their sights.

Though many of the animals which possess complicated cognition, such as chimpanzees and dolphins, are also difficult to study at the neural level, the simpler forms of cognitive behavior can tell us a lot about the nature of general

intelligence when comparing across species, and give us inspiration for designing robotic controllers and language systems that form stepping stones. Remember, the human brain evolved from other brains that contain most of the same features. Therefore, looking across species can give us clues into the origins of our own intelligence, and can help us design primitive AI that can evolve and be engineered upon with proper direction into an AGI that could match or eventually surpass human capabilities.

Humanoid Robots - We could create a humanoid robot or a virtual 3D avatar by training a human 'mimic' AGI controller using the training methods we have outlined fed with data from a performance capture of a real person, who has acted out a variety of scenarios to supply the inputs and outputs to train the AI for the speech, vision, body movement, and facial movement of an artificial AI person. That AI person can be instantiated with 3D computer graphics as a character on a screen or as a realistic humanoid robot, with the motion and facial expressions mapped to the actuators for the latter. The goal would be to provide an AI person with realistic, human-like dialog, lip-sync, facial expressions and functional movement in its environment. The detailed performance capture data can be augmented with simpler text conversations that may be general or specific purpose to a vocation and have access to databases to augment the training data, and also augmented with dreaming between training sessions to fill in the blanks.

This could bring us closer to having human-like avatars and robots that look, act, emote, plan, and do tasks just like a human does, using similar underlying intelligence that can adapt and react to dynamically changing environments far better than any deep-learning based robots would ever be capable of. These synthetic humans could be equipped with other computer science tools like databases and some scripted dialog and be used as artificial employees in a wide variety of customer service jobs and with sufficiently advanced AGI in professional vocations as well.

Combined with the AGI's conversational speech capability, this human mimic could become an artificial companion for a person, either instantiated as a computer graphics avatar, or (at a higher cost) a physical android. Loneliness is a huge problem in the modern world, and these conversation capable AGIs could provide companionship for people that

are living by themselves and lonely. When integrated with the personal assistant, it could help them manage and plan their life, giving them advice, handling routine errands online, and greatly improving their quality of life.

There are several companies developing realistic humanoid robots, however this current generation of robots come off as being creepy due to them being just realistic enough to accentuate the flaws, a phenomenon that is called the Uncanny Valley.

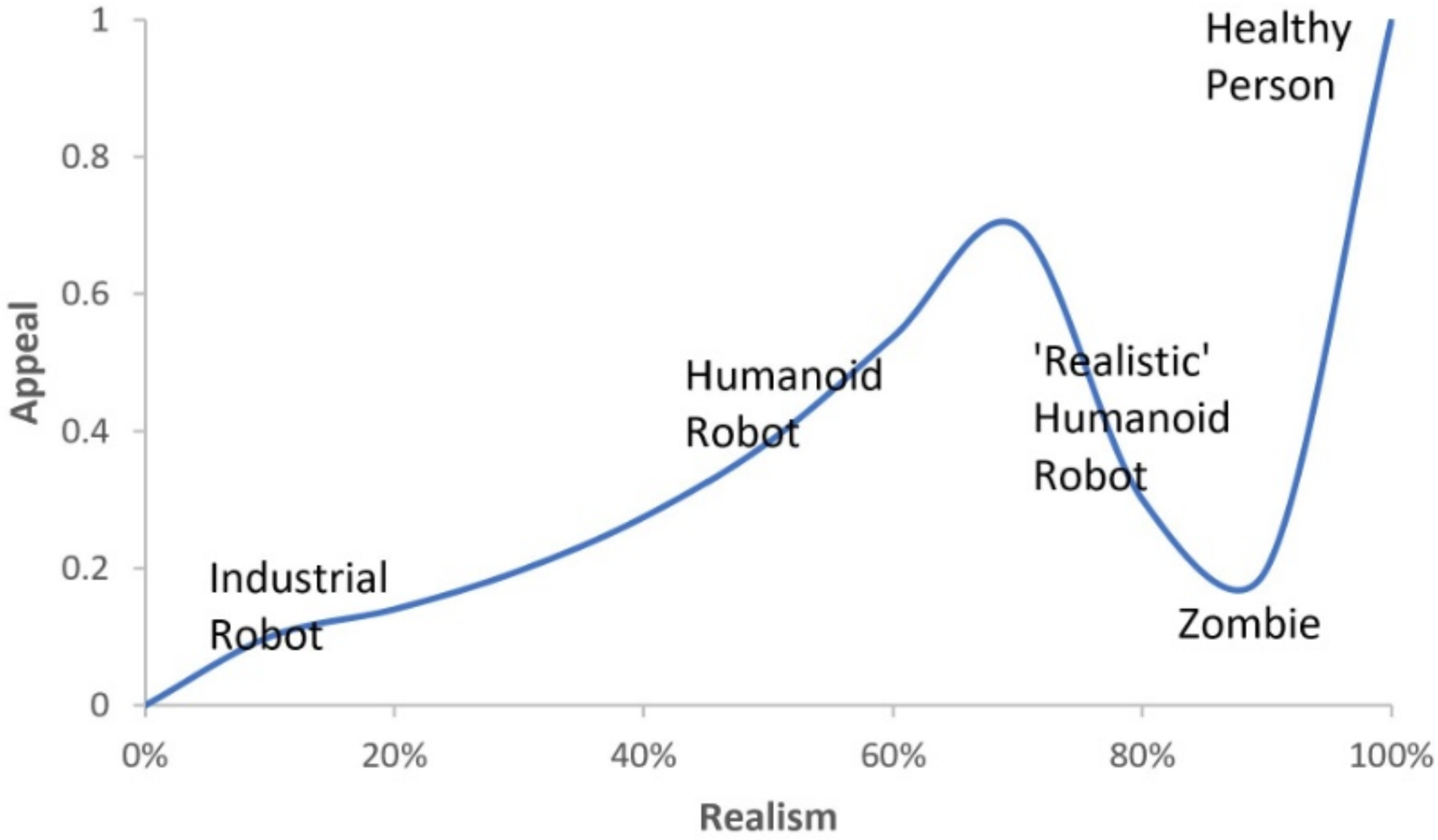


Figure 6.2 Uncanny Valley in Human Likenesses

This is the one factor that stands above all others in distinguishing whether artificial characters and androids meet our criteria for realism and acceptance, become part of our lives, and become emotional and even physical companions.

The Uncanny Valley is when small deviations from how a real human looks and behaves become more noticeable and repulsive to us as the likeness becomes more realistic. It is due to the human perception of human faces, bodies, and their movement being very, very discerning, and picking up on the small flaws in appearance, facial movement, speech, and motion of the simulated person, and giving a 'creepy' feeling when they all don't match up.

This will remain a challenge to solve with AGI, to get an artificial human through the uncanny valley and make it so it is not repulsive to us, and so they can be accepted in our daily lives. Once we have an intelligent, conversational speech AGI, developing realistic human avatars would be the logical first step, as they are an order of magnitude easier to implement and much lower cost than a full android due to the accuracy and degree of motion needed in the actuators to match the capability of the humanoid AGI. Plus, existing real-time avatars have almost photorealistic quality, and adding AGI controlled behavior, speech, and animation would augment them quite easily to get them out of the uncanny valley, and these virtual humans would join us in our media and online interactions, entertaining us, bringing us our news, and performing real jobs for us, often without us knowing if they are real human or virtual agents.

In another application we could train the humanoid AGI controllers on a specific person, creating a 'mimic' of that individual that would emulate and reproduce their speech, their responses to specific questions, their body language, and every nuance of their performance down to facial expressions and lip-sync. We could create digital duplicates of famous actors, then create performances and insert their digital likenesses into computer-generated shots for movie production, reducing the reliance on the actual human actor. The AGI actor could be given high level direction and a script to follow, and take over from there, becoming much more autonomous than today's CG characters which have to be animated by complex motion capture done on the actual actor and with a lot of hand tuning. In the future, the digital 'mimic' of the actor could simply be licensed for TV and movie productions to be operated as digitally directed actors, greatly enhancing the earning capability of that actor in a scalable manner and allowing more productions to use the likeness and performance of a celebrity without the scheduling and personal hassles.

In general, a robot controller based upon our AGI architecture has more flexibility in interpreting its senses and inputs, performing cognitive functions, and producing correct movement better than a DL-based controller ever could. It can also learn adaptively as it is operating, building a progressively more advanced model of its world to interact with it, filling in all the blanks and by dreaming using its predictors, all of which a deep learning AI would be limited by not being able to do. It could allow humanoid avatars and androids to cross the Uncanny Valley, become natural to interact with, and be accepted into our homes and lives.

CHAPTER 7

AGI in Daily Life – Personal Services

AGI would bring much more effective personal services to daily life. Instead of a myriad of recommendation systems, and digital personal assistants, each human would have a unique instance of an AGI that takes care of all their digital knowledge, recommendations, and planning needs.

In 2023, we have multiple DL-based services, recommending which posts we see on social media, recommending movies and music on streaming media platforms, recommending workout regimens, planning our schedules, and recommending ads for relevant services and platforms based on all of these.

In 2023, we can do searches on keywords that return pages of possible links that may or may not be what we are looking for, with half of them being ads, and most of the results being irrelevant. With the explosion of information available on the Internet, looking for exactly what we want is becoming harder every day, and it is even harder for organizations providing products and services to get noticed amidst the noise in a search engine.

What we propose with AGI is an integrated personal assistant that is completely integrated with your life, that knows what time to wake you every morning and knows your schedule based on your actions each day that it has learned. It tracks your calendar, which it dynamically manages with you to optimize your day and your week, helping you juggle

personal and work appointments to best make use of your time and efficiently organizing you so you don't have to waste time.

It can provide concise and accurate information on any topic, able to draw from a vast knowledge base it has learned to model and to provide answers to queries and questions with naturally-flowing human language. It can speak conversationally with the user and understand what it is talking about by using the concept abstraction built into the language system. A world of information and your personal info can intuitively be queried by using this natural language interface, giving you only the information you requested and not any extra information to search through to find it.

It would be able to manage all your streaming media subscriptions, and (based on the events of your day or week), help you find movies and music that you are most likely to want to view and hear. The best part is that it would not train into a corner like the current media search systems on Netflix, Amazon Prime, Hulu, and others, where they tend to stagnate into a genre and keep recommending the same few dozen movies or songs based on your preferences. It would instead keep bringing in fresh content and dynamically manage your recommendations.

If you have doctor or legal appointments, it sets up a pre-screen with the medical or legal AI and submits an email report to the human professionals to allow you to use your time with them most efficiently. If you have an ongoing health or legal issue, it monitors the status of the issue and gives you alerts if something gets out of bounds, like your blood sugar trending too high if you are diabetic. If you are a party in a legal dispute, the AGI can give an alert if hostile action is anticipated by the opposing side so your attorney can get ahead of them and preempt it.

For finances, the AGI would manage your stock portfolio, and constantly be monitoring all the factors in the market. It could be set to auto-trade and be front-running on specific stocks, or it could just be set to optimize long-term returns on specific indices or stocks. No human would be able to trade better than the algorithm, and it would safeguard against sudden market drops, getting your investments out of volatile investments before their value drops.

By retaining so much information about you, your schedule, your likes and dislikes (but keeping it private and encrypted for only your use), the personal AGI would be able to do very precise targeted searches for information, using a much more powerful system than just searching on keywords. Using the hierarchical basis set representation outlined in Chapter 4, the AGI can build enormous graphs of connected concepts so that a user can explain what they are looking for.

Using the predictive power of the AGI Cortex, the personal AI could be used to plot out all the possible decisions that could be made in your personal and professional life, and advise you on the outcome of the best decisions. It could become an invaluable planning tool in your life and career.

The AGI would be able to interface with your home, workplace automation systems, conserving energy by only lighting and using climate control on the rooms that people are in, or are predicted to go into soon. It could dynamically manage the systems of the home and workplace, optimizing energy usage, optimizing ergonomics and utility, using predictive AGI systems far superior to DL based systems.

In the future, it can coordinate the activities of household robots for cooking, cleaning, and tasks such as laundry. It could have your vehicle charged and ready to go, and your destination pre-programmed into the vehicle's autopilot with a simple voice query of "Where would you like to go today?" being answered by the user as they walk out of the house.

As the AGI is used by an individual over the years, it becomes a record of the things they have done, with a dynamic model of their life that can predict their behavior and what they will say and do so well that it could function as a memory aid as they age, and could even stand in their place once they are gone, if it were combined with the humanoid AGI controller of Chapter 6, standing in for them, answering factual questions about their life and even opinions about different decisions, just like they would.

This AGI instance of a person could be recorded and rendered with much higher fidelity for important persons (at a higher cost) with performance capture as per the humanoid AGI controller of Chapter 6, and they could remain part of a family, as an educational tool, or even part of a corporate board of advisors - posthumously. We may, in the next decades, see the beginning of the first 'digital immortals' by such technology. We need not lose our best and brightest minds.

CHAPTER 8

AGI in Multimedia Content Generation

In 2023, with Large Language Models using deep learning revolutionizing text authoring and generation, we are just now seeing the tip of the iceberg from what AGI will be able to do in multimedia content generation.

Because AGI develops a more accurate predictive model of the world than deep learning, can learn from much more varied and larger scale of data, and learns to associate different modalities of data, it will become a very powerful tool in authoring multimedia content. Here are some more specific examples:

Content Generation: AGI can create high-quality multimedia content across various formats such as videos, images, and music. It can automatically generate realistic scenes, characters, and even compose original music, reducing the need for human intervention in content creation. Large production efforts can be guided by a smaller team than is needed for a fully human authored production, which will allow content production to become decentralized, and a generation of high-quality amateur productions will proliferate, bringing us all a greater variety of more original content.

Content Personalization: AGI can analyze user preferences and behaviors to deliver personalized multimedia experiences. It can recommend movies, music, and articles tailored to an individual's tastes, enhancing user

engagement and satisfaction. It will soon be capable of authoring entire audio and video productions by the direction of the consumer, giving that person the authoring capability to create exactly what they and their peers want to see.

Content Analysis: AGI can perform in-depth content analysis, including sentiment analysis, object recognition, and speech-to-text conversion, enabling more accurate and efficient content indexing and search. This is especially useful in large multimedia databases and archives. This will improve automated closed captioning for videos and provide audio narration that is much more accurate as well as many other applications.

Translation and Localization: AGI can provide real-time translation and localization of multimedia content, making it accessible to a global audience. This facilitates cross-cultural communication and broadens the reach of content creators. We can even expect to see auto-dubbing capability where audio translation using AGI trained on professional actor's voices brings a movie to multiple languages automatically, using the original actor's voice, as we suggest using AGI mimics being trained on professional actors in the robotics AGI (chapter 6).

Enhanced Virtual Reality (VR) and Augmented Reality (AR): AGI can create realistic virtual environments and interactive AR experiences. It can simulate lifelike characters and scenarios, making VR and AR applications more immersive and engaging. Authoring content is currently a bottleneck in today's VR environments and virtual worlds. AGI will allow an average person to create high-quality 3D content that was only possible by professionals before, greatly helping to scale up virtual worlds and VR experiences in a way that is impossible today.

Content Moderation: AGI can assist in content moderation by identifying and flagging inappropriate or harmful multimedia content, helping to maintain online safety and community guidelines. This is accomplished today using deep learning, but it often makes mistakes and flags innocuous content by accident, causing angst for the user and extra time for human moderators to review improperly flagged content. AGI would be much more discerning and intelligent about flagging content.

Automated Video Editing: AGI can automate video editing tasks, such as cutting, splicing, and adding effects, streamlining the post-production process for filmmakers and content creators. Currently video editing is done by hand in high-end video production tools, requiring significant expertise and time to edit a video well. This is a content creation bottleneck that limits the quality of finished video that an amateur can accomplish. AGI will act as a professional video editor, directed at a high level by the human, who need only set the tone and supply suggestions for how the finished video should look. AGI can also improve the quality of multimedia content by upscaling low-resolution images or videos, removing noise, and enhancing visual and audio elements, resulting in better viewing and listening experiences.

Interactive Storytelling: AGI can create interactive multimedia narratives that adapt to user choices and actions, offering unique and engaging storytelling experiences in video games and digital storytelling platforms. Today's video games and interactive VR experiences are heavily scripted, offering limited interactivity and limited replay value. Non-player characters controlled by AGI can act and emote at a human level, and the overall story can pivot based on their interactions with the player character to create a much more immersive, dynamic, and interesting experience. If we use AGI mimics like we talk about in the robotics chapter, we can have them trained on professional actors to bring added realism and quality to the experience.

Content Verification: AGI can assist in verifying the authenticity and credibility of multimedia content, helping to combat the spread of fake news, deepfakes, and manipulated media. However, This will always be an arms race between the AGI's authoring the content and the verification AGI's, as both will improve at a rapid pace.

In summary, AGI's capabilities in content generation, personalization, analysis, and enhancement have far-reaching implications for the multimedia industry. It can enable more creative and efficient content production while enhancing the overall quality and accessibility of multimedia experiences for a global audience.

As an example, this chapter was first drafted by ChatGPT 3.5 from a simple query, with the author editing the generated content and adding more detail to it. We are at the beginning of a very exciting revolution in content generation that stands to change our entire workflow in many disciplines and bring the ability to author high quality content to the masses in a way not possible before.

CHAPTER 9

AGI in Finance and Enterprise

Figure 9.1 Finance AGI Concept

In financial applications, the AGI could track a massive number of factors that feed into the performance of specific companies, allowing decision makers to have much better predictions of market movements. By training the predictors on data that encompasses narratives generated from past stock charts, external data about the company, about similar companies, competitors, world data, and almost anything related to the target company's ecosystem, we could create automated trading algorithms that would be able to greatly outperform human traders simply by seeing more deeply into a wider data set than humans or DL algorithms ever could, and predicting more accurately based on their model of the world learned from that data. The AGI would learn a model of how groups of human traders and existing automated trading algorithms react to these factors and then front-run, or trade ahead of them to make higher margins based on the subsequent market movements.

For enterprise and government agencies, the AGI could provide ERP tools for enhanced analytics, for forecasting, and decision making in not only finance, but for the sales, marketing, product development, legal, and even HR teams, so they can forecast 4-6 months ahead, run simulations of multiple scenarios forward in time, and help the human users make the best decisions based on the outcomes of these simulations.

AGI-augmented ERP tools would provide an AGI-assisted platform that lets executives and high-level employees spend more time collaborating on content. The AGI would be an integral part of this collaborative process, gathering data, formatting it into charts, graphs, and reports, communicating as a team member in human language and acting as an overseer, focusing and coordinating the activities of the teams.

Let's just pause and consider that. When a corporate executive team adopts the AGI ERP tool, it suddenly gains another team member that speaks their language with text, speech and graphics, can integrate with the human team seamlessly, take on an enormous workload, and does so working with a prescience that no human has, being able to truly forecast into the future to predict the outcomes of potential plans months, even years in advance.

By being able to predict, set and track KPIs and critical success factors, the AGI ERP suite can keep projects and divisions on track and prevent organizations from making costly mistakes.

AGI applied to government administration could greatly streamline government agencies and make them much more effective and efficient. To start with, instead of filling out cryptic forms, a person could interact with an AGI customer service avatar that interviews them and best figures out how to fill out such forms based on the information given by the user and could ask them follow up questions to clarify any areas that are unclear. AGI could participate in the processing of such forms and help to make the process more efficient. Anybody that has applied for something like US Social Security benefits can appreciate this, as after filling in very cryptic forms and submitting them by mail, a person can wait for a year or more for a decision to be made on their case and for benefits to start.

AGI could be used to make decisions being made by the government more transparent, to communicate the issues being decided more effectively, and to provide a mechanism for citizens to provide feedback to their representatives on the issues. The AGI could generate accurate summaries of legislation being voted on making it easier for the average citizen to know what their government is doing. For the representatives in government, AGI could provide a tool for gathering this feedback, and planning their approach to a given issue or piece of legislation and make them much more effective in making decisions that represent their constituents' wishes.

Once these tools are adopted worldwide, their true power to enact global change is revealed. The financial AI can achieve returns greater than any brokerage's financial returns, and as it is adopted globally, tap into massive wealth, of which a portion can be redirected to fund a living wage for those most in need. 1% of $500 trillion in world wealth per year would go a long way to ending world poverty and hunger for the 15% of the population living below the extreme poverty line, and the AGI could also give them low-cost automated trading accounts that hold a minimum percentage balance to invest so they too can also grow their net worth and prosper and break the cycle of poverty.

The Enterprise Planning and Government Administration AGIs can suggest and fund joint projects that companies and governments can work on for mutual benefit, such as planning and development of new, inexpensive, high efficiency solar panels deployed in international energy farms, and higher energy capacity batteries in inexpensive electric cars and home power grids that span nations. The AI would undertake planning and coordination of these mega-projects to accomplish them by sub-tasking individual corporations and government agencies, coordinating their efforts and financing - more efficiently and faster than human administration could, making them more feasible and profitable, and succeed in their execution where humans have only failed before.

Such forecasting and planning AGI could have applications for every company, government agency, and person on earth, and help guide our collective efforts to truly bring change to the world.

CHAPTER 10

AGI in Medicine

Figure 10.1 Medical AGI Concept

The primary job of a medical practitioner or doctor today is to diagnose and treat a patient's illness, and to follow up, monitor, and adjust that treatment. To do so, a doctor will interview the patient, get their medical and lifestyle history, take their vital signs, then ask the patient about the symptoms and issues they are having. They can choose to follow up with more testing, like lab blood tests, medical scans like X-ray, CT, and MRI, and other secondary exams based on those results.

The doctor then uses all this information to come up with a diagnosis and treatment plan, which could be medication, surgery, or other procedures. They will continue to follow up with the patient and monitor the progress of all these, adjusting the treatment plan accordingly.

The problem is that human doctors are highly sought after, take 7-12 years to train, and are highly compensated. This means that there are never enough doctors, and because of this shortage, patients can have long wait times, and short, infrequent visits, especially in areas that are remote or economically disadvantaged.

So, what can we do with AGI? The first thing that we can implement is a speech-capable, multilingual AGI to do initial patient interviews, get their background and symptoms, even do automated vitals with a portable blood pressure, pulse oximeter, thermometer - set up with their mobile phone, or at a kiosk in a clinic. This would do the initial patient screening process with AGI, that starts out by generating a concise medical report for the human doctor, it would make the doctor visit much more efficient and less costly in time and resources.

The AGI speech recognition combined with selective filtering will be able to listen to a person talk normally and screen out the symptoms and important factors from their conversational chit-chat to put in a concise medical report. This is highly necessary to work in the real world, where you cannot force people to speak in an artificial cadence to a voice interface and to specifically annunciate commands and parameters like 2020's speech interfaces require. We covered the speech interface in more detail in Chapter 5.

AGI could provide an interpretive tool for medical imaging and other tests, looking for anomalies that could be signs of illness. Deep learning has already started being used for this purpose, but it still has too high an error rate and is prone to interference and artifacts causing an improper interpretation. An AGI would be much more accurate and less prone to errors.

Also, the AGI can combine the patient interview data with all the diagnostic data – symptoms, vitals, medical imaging, lab results, … and it can be encoded into sequences by our autoencoders and recorded into a narrative like we did with sensory information earlier. This encoding process would flag diagnostic images and other medical tests as being normal or abnormal in the process, giving doctors a stand-alone screening tool in addition to its being a part of the automated diagnostic pipeline.

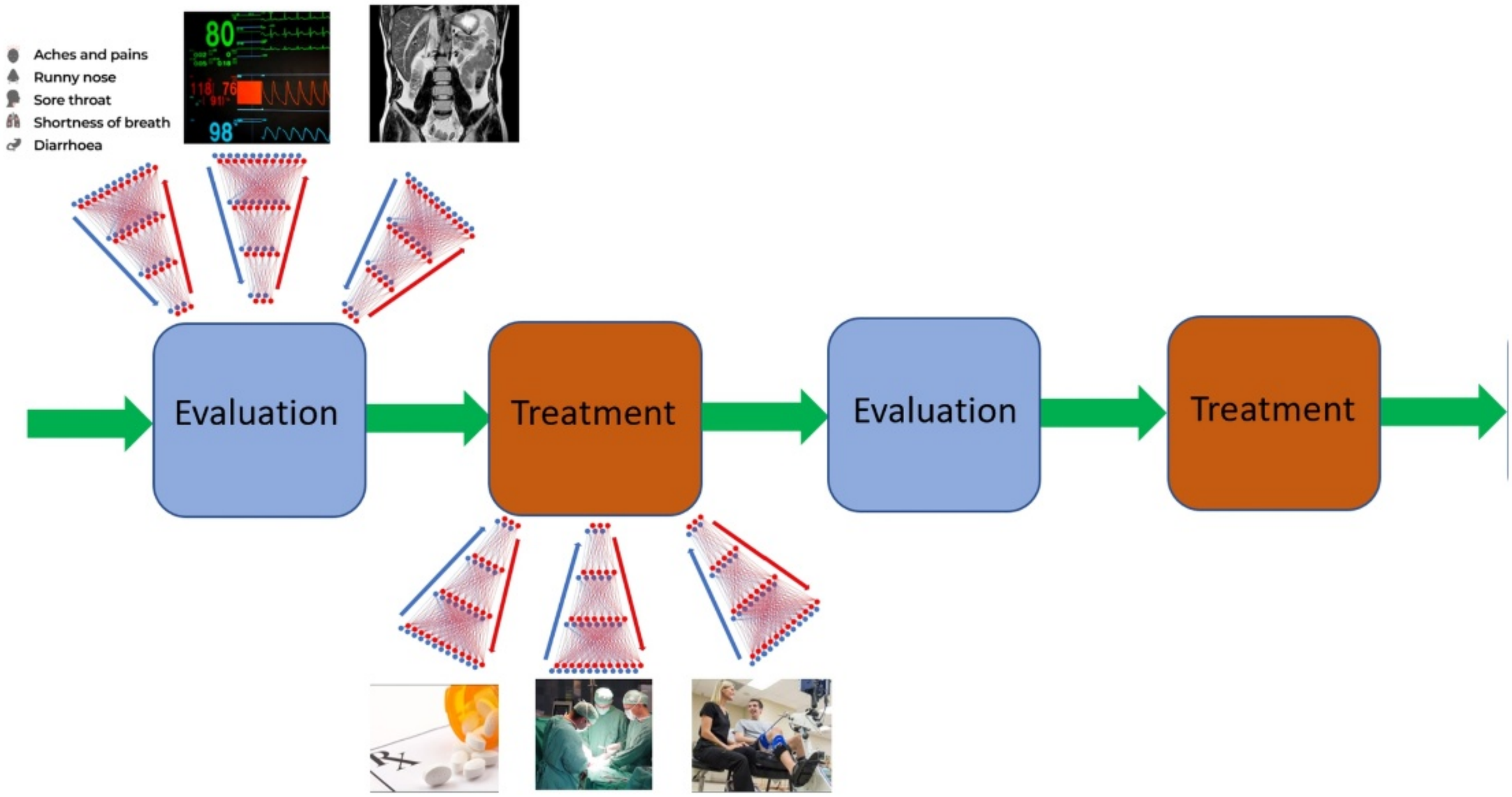


Figure 10.2 Evaluation and Treatment Narrative of Medical AGI

Each of these would be a snapshot of the patient's medical state at a given point in time, and as the patient undergoes treatment, these tests would be repeated to track the progress, creating a series of data points tracking the patient's

medical condition over time, that forms a narrative, a story of the progress of their disease, whether it is a flu for a few weeks, cancer over months, or diabetes over years.

However, we need to track more than just the patient's medical state. We also need to track the treatment regimen, and what the doctor is doing to change the patient's medical state. This is actually an adversarial model of the treatments vs the disease, where the doctor measures the patient's medical state, prescribes treatment to change that state, measures, treats, … and so on, so our adversarial / conflict language model applies. We can encode data for treatment, similar to the representations for medical data, but encoding the medications, dosages, frequency, and any procedures or therapy that are done into treatment narratives that interleave with the data of the patient's medical state. The Medical AGI can even record the patient's exercise or diet as well as environmental factors.

Now, the narrative for a patient's medical story is more complete – we have complementary narratives that form a chain consisting of alternating evaluation and treatment data that track the details of their medical condition and what was done about it, encoded in a way that we can train a generative predictor on it, to make an adversarial model. The AGI can run multiple simulated narratives of the patient's health, and vary the treatments, dosages, intensities at each point in the simulation to see if it can come up with an optimal treatment plan for that patient. This is a very powerful tool, as doctors often resort to doing this by initially guessing, then using trial and error to adjust treatment for their patients. This would be an AGI tool that could look months, or even years ahead to plan the results for a more effective treatment plan between now and then that would have a better overall outcome. It could prescribe treatments preemptively to keep the person from even getting ill years from now, based on detecting subtle early warning signs of a condition early on.

However, our Medical AGI is also more than one instance of a digital doctor. It is an AGI that is compiling these medical narratives for every one of the millions of patients it cumulatively treats every day, and these narratives can be used to train the Medical AGI to learn about the progression and treatment of different diseases in great detail, and to use all this knowledge to forecast a patient's medical condition into the future. Now this gets more interesting, pre-emptive

medicine based on disease modeling. It would keep each patient's data encrypted in narratives to train this predictive capability and do so without the patient's personal information included so privacy would be protected.

By doing this with millions of patients, the medical AI will build up an enormous knowledge base of human health, disease, and the progressions of those diseases vs the treatments that work best for them under different circumstances like patient demographic or specific genetic predispositions. It can understand the human body and its ailments and applicable treatments in far more detail than DL or even a human doctor ever could.

When this AGI is distributed globally, it can learn the medical practices and treatments of other cultures. Maybe a doctor in California would prescribe a series of radiation treatments to shrink a specific type of cancer tumor, but a rural doctor in Central America knows that the excretion of a specific insect, when extracted and modified by mixing it with lye and cooking it – is a far more effective treatment that destroys the tumors completely.

In addition to assisting with medical diagnosis and treatment, our global medical AGI would become the world's largest scale pharmaceutical research platform – building up these databases of medications, their efficacy, and the direct results on patients over time, and by allowing pharma companies to mine the database and use the AGI in clinical trials, while protecting user privacy and data with the encryption. When integrated with molecular and physiological simulations, completely digital clinical trials could be conducted before involving humans, enabling much more efficient screening of prospective treatments.

Using this AGI, we could use the revenue gained from treating patients in the more affluent countries and from the pharmaceutical companies to finance expansion of the medical AGI as an app on mobile to be used in remote areas and countries where that level of medical care was previously unavailable. By integrating it to prescribe generic brand pharmaceuticals made available at a lower cost in those countries, the whole world could have access to the best medical care in the world, available within a decade of its deployment, all done profitably along the way.

CHAPTER 11

AGI in Law

This chapter describes a Legal AGI that will be able to assist average individuals in representing themselves in court, or to augment an experienced attorney in researching cases and composing superior litigation documents.

The Legal AGI can read paragraphs and documents of legal language, encoding them into an internal format, then use that data and its AGI capabilities to do research on previous legal cases and build a model of how such cases are litigated in time. It also has the ability to author filings using the correct legal language, citations of laws and precedents, and exhibits, but based on the user's circumstances, making effective points and arguments about their case with reference to well written and well-evidenced exhibits and citations.

A legal proceeding is simply a very structured human conversation and consists of a sequence of hearings and in the end a trial in which each lawyer tries to convince the Judge or Jury to rule in their favor, hopefully resulting in the outcome they want at the end. Again, this adversarial process can be modeled by our AGI and the filings predicted.

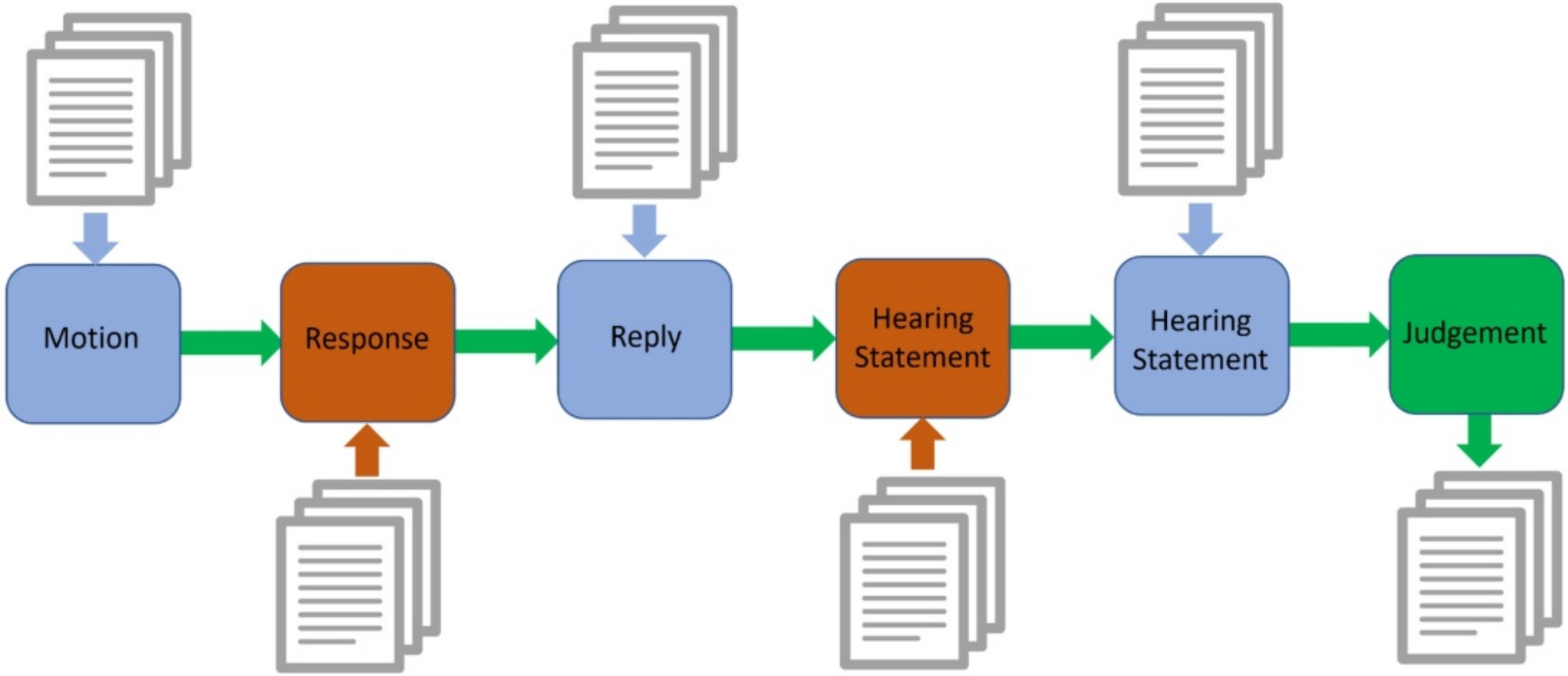


Figure 11.1 Legal Hearing Process

The form of the process is structured to try and make it fair. To decide each issue, there is a hearing in which a series of filings are made. One side files a motion, asking for something they want, usually in the form of an order the Judge can sign, plus all the substantiation for their request, including the legal claims, and the exhibits backing them up. These are not trivial for an ordinary person to write, and take lawyers many years to learn, and even longer to get really good at writing. Fortunately, these filings are very structured, and are easier for AGI to learn than normal, unstructured English is. Below is a diagram showing the process for the Legal AGI authoring a legal filing.

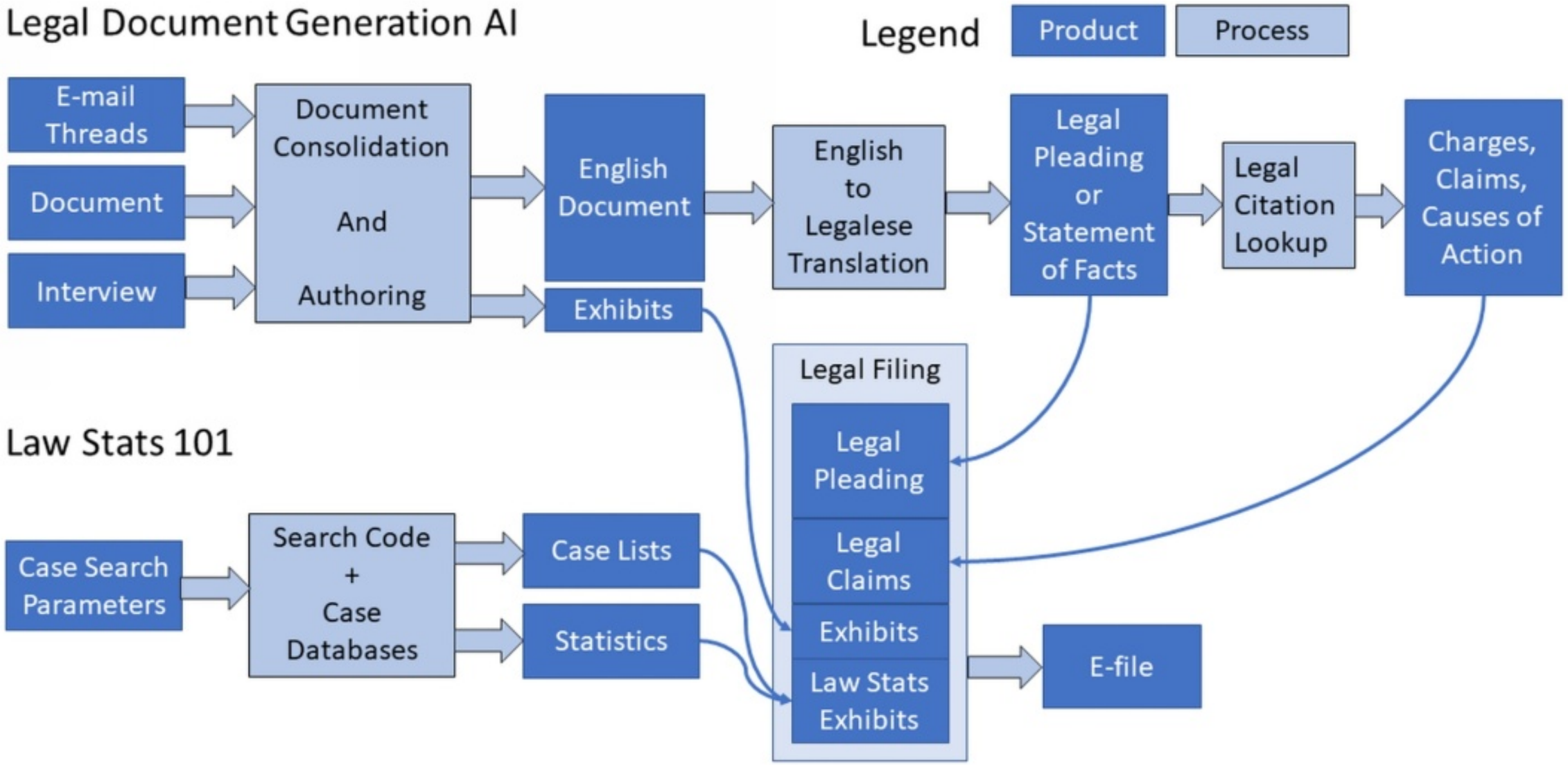


Figure 11.2 AI Legal Document Authoring and Filing Process

One section, the English to Legalese translation, is critical. Legalese is a commonly used term for the legal language, which is much more narrow, structured, and precise than English. In English, a word can have many meanings, and statements can be ambiguous. Legalese has very strict definitions for specific words, for example 'deceit' is a wrongdoing you can sue someone for, and 'perjury' (lying to the court) is a crime, but 'lying' is neither a tort you can

sue for, nor a crime, and is meaningless in legalese. Legalese is the code that a Judge will read and act on and has to be standardized and formally defined to keep it standard for other Judges and actors in the legal system to interpret in the same way.

The Legal AGI needs a translator that encodes English to hierarchical abstract concepts, then translates those concepts to legal ones, that then hierarchically defines the correct legal language to output into a document. It is literally like translating to a different language, with a different grammar and structure, and is not 1-1 from English. The Legal AGI does an encoding of the English to more general high-level concepts, concatenates those larger concepts, and then converts the high-level concepts into progressively more granular 'Legalese'.

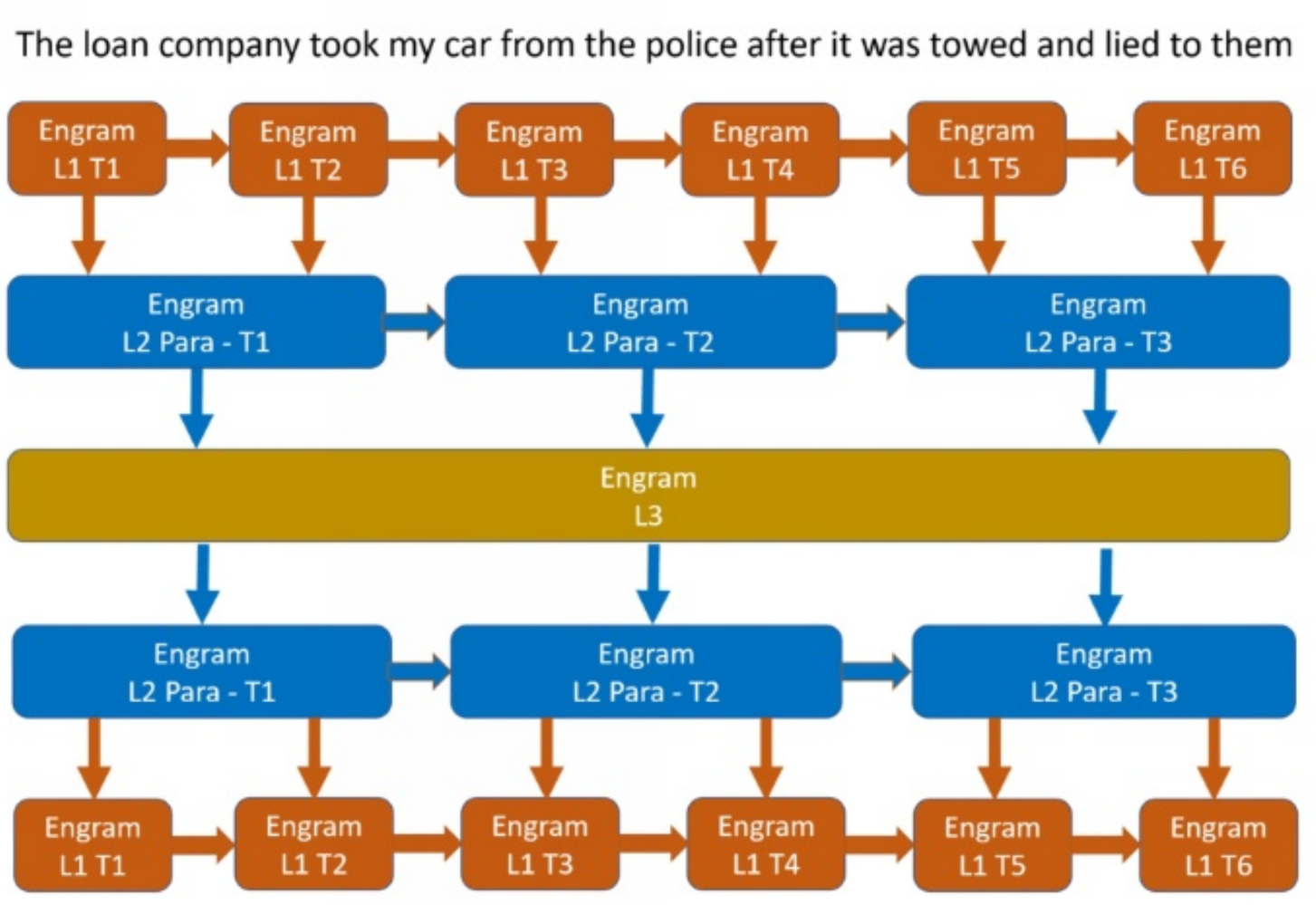


Figure 11.3 Translation from English to Legalese Using Hierarchical Engrams

This allows the legalese to be formatted to the correct structure, using the elements we supplied, and translating the whole concept to the correct legal language and format, pulling facts that were encoded from the source as it does, far beyond the capabilities of any deep learning based AI.

So now that it's AGI can write like a lawyer, the Legal AGI can litigate. Each hearing involves a formalized sequence of filings, with one side filing a motion, the other a response (also called opposition, or objection, depending on the court), and the motioning side files a reply. Then they go into a hearing, in front of a Judge in the courtroom, and each side gets to state their arguments briefly. Then the Judge makes a decision, immediately, or after looking at the filings further.

This process is a closed system that involves 3 parties in an adversarial process – two lawyers and a Judge, that essentially takes in 5 documents and outputs a sixth. This entire process could even be automated, and not even require lawyers and a judge if the AGI overseeing the proceeding is advanced enough.

Decisions by judges at hearings are made with partial, sometimes false information, and the decisions in those hearings are final and propagated to other hearings where decisions are made only on the final output of previous hearings. The filings from the previous hearings are not reconsidered, so the current system can be very error prone and chains of bad decisions by judges can accumulate and result in disastrous outcomes in a proceeding.

When a case starts, and as it evolves, research is of key importance. Researching the relevant laws is important, researching previous case files to find precedents, to research how previous attorneys litigated similar issues, and to research and study the filings and tactics of your opposing counsel are all necessary.

In doing case research, the problem is that there are many different corpuses of law – Federal and State Criminal Codes, and specific codes covering housing, agriculture, education, business, … and even codes of conduct for attorneys,

judges, police, and everyone in the different courts and organizations. Learning all of these is not possible in any human amount of time, even once a person has specialized in a field.

A corpus of relevant filings, even when localized to a specific area of law, and even to just a specific lawyer's past filings, can still number in the 100s of cases, each with dozens of filings, and hundreds of pages per case. It is not humanly possible to read, nor understand more than a few past case files a week for a human. It is not easy to figure out how to characterize your case so that you can search for similar cases, as there are no explicit search terms that can be generally used. For large scale cases like class action lawsuits, where one needs to find hundreds of cases with similar points of law, wrongdoing, and damages, and pull them together and certify them as a class, doing so is an art, which requires large teams of specialist human lawyers at great effort and expense. A legal AGI would excel at all of these tasks.

In a live case, the Legal AGI will also be able to use its litigation prediction skill to tell what opposing counsel is likely to do next. By studying a set of their past cases, it learns their litigation patterns and the order in which they file certain motions, how they respond to certain motions, and how they write their filings. it can look at your case and tell you what opposing counsel's next actions may be. Soon, with its conversational language capability (Chapter 5) extended to filings it will be able to show you exactly what the opposition will write next, and show you a copy of their next filing, before they even write it. Humans really are that predictable, especially senior attorneys. They have habits and patterns, ingrained from years of practice and follow the same scripts in composing litigation with each case.

Another element of court filings and judicial rulings is that they cite precedents of other decisions by judges in similar past cases as justification for the lawyer's arguments in their filings, and by the judges as justification for their rulings.

These precedents are used as a guideline, as nobody wants to stray too far from the beaten path, but the problem is that path itself meanders and goes off course over time, because each precedent was a judgment based on previous precedents, and over time, the precedents drift, especially when there are biases and political pressure pushing them

in a certain direction. A ruling by a judge based on this chain of precedents may drift so far that it is nowhere near the intent of the underlying laws and even in violation of them and in violation of constitutional rights.

This is a fundamental information science problem, as there is a loss in communication and translation at each step with no error correction. This is referred to as Developmental Stage Theory - The information used in processes becomes increasingly distorted as it becomes further removed from original sources.

A simple example is the game of telephone, where a message is passed from person to person, and after a few people, the output drifts far from the input, with each person making errors that are not corrected. Even still, each person on the chain thinks they heard and stated the information correctly.

The Legal AGI could encode the filings and the judge's ruling and trace all of the citations - from the most recent precedent, down to the base of the chain, and analyze the drift that happens at each step (using the language skills of the AGI) and through the precedents with time and give the citation a score based on how far off it has drifted from the initial intent of the laws and first precedent, and how applicable it actually is to the current case. It would be simple to add these metrics to existing legal citation research software, and attorneys could use it to choose their best citations as well as to critique the weaknesses in their opponent's citations, or errors in the Judge's ruling if appealing a decision. An audit could be done of past cases to see how badly cited the precedents were - and allow litigants to appeal their past cases and even bring specific judges under review.

To build an AGI that can monitor and predict a legal system (and be used to eventually replace it), we would first need to be able to encode many filings and judgments in many cases, starting by encoding each filing in a case into a set of hierarchical narratives, encode the flow of those filings through each case, and encode the judge's decision in each hearing. Once we had this, we could encode that Judge's rulings across each case they adjudicated, learning how the abstractions in the filings related to the final rulings, and how the relationship of the filings and their rulings are common across cases, as well as traversing the hierarchy of citations and precedents.

This gives us several tools – for one, we can have a tool that allows us to model how well the Legal AGI's filings will be received by a specific judge, given their rulings on similar cases. For another, it allows us to see when a judgment did not follow a correct pattern and find cause for it to be appealed. Also, we can profile each Judge to find anomalies, where (for specific lawyers), they tend to rule in the favor of that lawyer more often than they should according to its data. We have a judicial 'corruption finder'.

But when AI begins to dominate in the courtroom, do we need the whole court system, judges, hearings, even lawyers? Hearings are an artificial construct to work around the limitations of humans' attention spans because a human judge cannot take in all the facts of an entire case at once when they span years and sometimes dozens of parties. They must consider many different (and sometimes contradictory) state and federal laws, as well as constitutional law and local court rules when making a decision.

Going beyond these piecemeal tools and providing an A-Z legal arbitration service is where an AGI could really excel, acting as an arbitrator and decision maker in legal disputes, starting by replacing small claims court, family (divorce) court, and moving upwards to arbitrate and decide larger cases in civil court. It could directly interact with the parties, bring enormous knowledge to the case, and give the parties straight facts according to law, and help steer them to a mutually agreed solution before having to converge it by making a judgment. An expedient resolution is often not the goal in today's legal system, where lawyers tend to drag out cases, make them more complex than they need to be, and over-litigate them just so they can bill more. A legal AGI could make lawyers obsolete within a few decades and provide a much more efficient, economical, and reliable legal system.

For a nearer-term application, say you have a complex legal ecosystem, like family law, where there are lawyers, family court judges, district attorneys, criminal court judges, and realtors involved. In places like California, there is a lot of money in real-estate coming into these courts in the hands of the divorcees that unfortunately tends to get ordered to be sold off, without much of the proceeds ending up going to the divorcees, instead ending up paying for extraneous legal fees, realtor fees, private judges, and fees to other players in this ecosystem. The Legal AGI could be used to analyze

all the family cases and find the anomalous cases where money was misappropriated, then map out which lawyers, judges, DAs, and realtors were involved in each case. It is not hard to build a corruption graph from there, with hotspots showing which of these private and state actors are committing systemic fraud and need to be investigated by the authorities.

With the development of more skills and AGI technologies used to augment its already formidable capabilities, when it is used by the right people, the Legal AGI will become a powerful weapon in the fight against injustice and corruption. It can be a force multiplier for pro-se litigants and small law firms, making them much more capable of litigating against larger opponents. It can assist authorities in uncovering and building cases against organized crime networks and can augment large firms to greatly reduce costs of researching and composing filings, making them more competitive.

CHAPTER 12

AGI in Defense

Figure 12.1 Autopilot AGI for Autonomous Vehicles

The author (Brent Oster) debated whether or not to include this chapter as the topic evokes fear and adverse reactions, mostly due to all of the movies we have collectively watched where the defense AI takes over the planet and attempts to subjugate or eliminate all humans. We also fear ceding control of life-or-death decisions to AI, as although an AI may follow the laws of conflict and rules of engagement, deep learning AI does not have ethical values to judge whether a decision is moral or not. I do not dismiss these fears, as they are justified, and a military defense AGI would be both formidable and very dangerous if implemented improperly. I include this chapter here to educate people about the capabilities of such an AGI, expose the dangers, and try to mitigate them.

The author was a military pilot and officer in my early career, and has experience with military intelligence gathering, decision making, dissemination of orders, and execution of those orders. In such a formalized structure, implementation of an AGI-based command and control network is very feasible, and probably inevitable. Can we build it in a way so as to keep it under human control, and not let it get away from us? For now, we will just come up with an implementation that is pure defense AGI, with no holds barred, then later talk about the safeguards and limitations we would have to place on it to appease our fears.

First, we want to create an Autopilot AGI, capable of piloting or driving individual autonomous vehicles on land, sea, and in the air, integrating their sensor information with real-time intel, planning the mission, autonomously piloting the vehicle throughout it, solving problems, overcoming obstacles, and doing the reconnaissance and/or deploying weapons necessary to achieve the mission.

The Autopilot AGI would be a derivative of the robot AGI outlined in Chapter 6, with inputs from its sensors and outputs to motors and actuators to control the vehicle and its weapons. How do we train this Autopilot AGI to sense, plan, move, and fight as well as, or better than a human pilot or driver?

We can start with pre-recorded mission data from thousands of remote and human controlled aircraft or vehicle missions and use it to run intensive training sessions. We feed this data into our Autopilot AGI to train it to see, hear, talk, plan, pilot, drive, and operate a weapons system with skill and precision, becoming a digital mimic of human piloting, decision making and their combat tactics. Then we can evolve and scale these AGIs within their vehicles, using their senses and outputs to see, hear, navigate, plan, and operate the vehicle.

Next, we deploy them in computer simulations and run them through thousands of missions before deploying them on actual maneuvers with real autonomous vehicles and aircraft. This process would keep training and evolving them so they are better capable of movement, planning, communication, targeting, control, and can learn by observation, experience and practice, just like us.

AI like this is already being developed with deep learning, but AGI could do a much better job, gathering information more thoroughly, extracting more meaningful facts from the intel, learning to discriminate friend from foe better, have superior piloting and driving skills, and being more flexible and adaptable when carrying out its mission. It would have an internal model predicting the evolution of the mission, including opposing force movements and actions, and a decision-making process that works by using the generative predictor capability of the AGI to map out possible options, then picking the one that results in the best outcome.

We will probably not end up with Skynet like this, as a human still has to deploy the combat vehicles and give them their missions, but we will have combat-capable AIs that (when we add traditional computational, information access, and capabilities for a specific combat role), that will be adept at a sufficient variety of localized tasks to function as combatants doing a specific role, operating a specific platform and its weapons. If we network the individual land, sea and air units so they can share information and make joint decisions on tactics and coordinate their execution of the mission, we create a swarm AGI that is more than the sum of its parts, and we get a force multiplier. A military force composed of such autonomous vehicles, networked and intercommunicating would maneuver and make decisions so quickly and accurately that they would be very hard to defeat with a human military force.

This has immediate application in defense today, putting these Autopilot AGIs to work in piloting drones, land robots, autonomous tanks and vehicles, and gathering surveillance data, assessing the situation, choosing targets, and operating their weapons systems completely autonomously, far from human control. As they do so, they gather enormous amounts of data that are then used to train and evolve the next generation of their AIs. This can be done on a daily update cycle, so these Autopilot AGIs learn from all the previous day's experience, and we have new AGI evolved overnight that learns from the previous day's encounters, getting smarter and adapting to the opposing force's tactics daily.

Now that we have AGI-enabled combat units, how do we create a Command and Control AGI so that decisions at a higher level can also be made autonomously, with far greater speed and effectiveness than a human C&C network could possibly achieve? With the Autopilot AGI, we are now able to train these vehicle AGIs to each specialize at the tasks relevant to a specific role, and we made different versions of these AGIs that work at dozens of different combat, reconnaissance, planning, logistics, and support roles, and they learn further skill in these roles when deployed, but we need a central command and control AGI that can manage these autonomous assets to perform at peak effectiveness and efficiency.

Figure 12.2 Command and Control (C&C) AGI

In a human military command and control structure, the generals at the top issue high-level orders that are then disseminated to their subordinates, who take those orders and break them down into lower-level detailed orders applicable to their level of command and then disseminate those orders to the next level of subordinate commanders, until they reach the individual drivers, pilots, and logistics personnel as fine-grained orders to carry out. These orders

are written in a specific, standardized format with 5 sections called SMESC, standing for Situation, Mission, Execution, Support, and Command. These orders are augmented by Standard Operating Procedures or SOPs for each level of command and each branch or unit, which are scripts that specify in detail the manner in which the orders are to be carried out.

Adapting a C&C AGI to take over command at any level is aided by these standardized forms of communication and operation because the AGI can easily learn the format of orders, and how to disseminate them, and how to incorporate SOPs. This will create a very comprehensive and capable C&C AGI that can interact intuitively with the Autopilot AGIs that are all now integrated with it, doing so to control many different instances of the Autopilot AGIs, and also take over the higher levels of the command, control, reconnaissance, planning, logistics and deployment and control of combat units.

First the C&C AGI would be able to aid in gathering and interpreting intelligence data, including large amounts of satellite, drone and ground-based imagery, as well as written reports and other intelligence data, working to build a detailed model of the theater of operations. Then that model could be queried by both human commanders and the C&C AGI with questions about the theater such as asking for numbers and locations of enemy units and emplacements. AGI can do this better than deep learning because it is better at working with sparse information and building more accurate models of the theater of operations that can be queried more accurately.

Our C&C AGI could also predict the actions of the opposing force command and control. It would take all historic information on the opposing force's human commanders (including their briefings, information they had on hand, and the orders they each issued after) for all the exercises and battles they participated in, as far back as we can get it. It would encode each document or transcription and feed those in order into the generative prediction AGI, such that the system learns the battle strategy of those commanders. We can take this trained model, feed in the information we now have on our evolving battle, and have our AGI predict what the opposing commanders next orders will be, and even write them, so we can be holding those orders, reading them, and planning how to counter them – potentially before

the opposing force commander even writes them. As we have mentioned earlier with the Legal AGI, this is actually quite effective because all the communications and intel are in a specific format and uniform in style, and humans are very predictable, often following the same specific patterns and making the same decisions, and writing the same things, given the same circumstances each time. This is especially true in constrained communications like orders, and the older and more experienced the commanders are, the more predictable they get.

Our C&C AGI has the ability to run predictions on multiple possible future actions and outcomes and make decisions based on the outcomes with the highest probability of success. Intel about deployments on both sides, communications, intel, historic data from the commanders, news stories, people's moods, weather,... everything it can gather is fed in on a continuous data stream through into the C&C AGIs specialized input cortices evolved to handle this data. It can use its decision cortex to give us accurate generative predictions about what comes next and use that information to make the best decisions, with superhuman capability and speed. Then the output cortices send commands to the various vehicles and units under its command, both human and AI.

It would be structured hierarchically, just like a human C&C network, with orders from the top being broken down into more specific orders for the next level, down to the unit level using a C&C decision cortex to learn to mimic the orders of the best human commanders and learn from historical battles and exercises. As the orders are propagated to each lower level, they become more detailed and expand the superior's commands into more expanded orders that specify the actions to be carried out by the individual units at the lowest level.

To continue to train and evolve a C&C AGI as it gains in functionality and grows from near-human to superhuman in the training process we have it participate in simulated combat maneuvers vs other C&C AGI candidates, to fight the opposing force with an ensemble of networks AIs - in detailed, unit on unit simulated battles. The C&C AGIs could fight against each other repeatedly in simulations, running in as many instances as we want, at superhuman speeds, to accelerate the training process as hundreds of simulated battles are fought and feedback and training data are supplied after every battle to evolve and train the AGIs for the next round of battles. Periodically there would be a culling of

the AGIs in the candidate pool, and genes from the top percent of the C&C AGIs would be crossbred, and these new, improved instances trained and deployed, so that they are constantly evolving and improving. Each time they train, their AIs are improved, and the accumulated set of global training data gets larger and richer, so they get better by both modes of evolution and improved datasets for training.

A C&C AGI will have been bred and evolved through hundreds of generations on a supercomputer just to do this extremely well. It would be so good at waging war at this point that no human could outmaneuver it or beat it. Knowing exactly what its opponent will do the next day and even having good estimates of their actions for the next week, would be decisive in a battle and allow the C&C AGI to economize the use of force and minimize losses while winning. When combined with the deployed Autopilot AIs, UAVs, vehicles and weapons controlled by it, with the integrated system gathering and interpreting intel, predicting opposing force movements, tactics and adapting to them every day by evolving to get smarter and better, this would comprise an utterly invincible autonomous AI weapons system.

Some aspects of such a defense AI are within our grasp today with deep learning, and several companies are actively pursuing such Al in defense applications at all levels of intel, command, and control. What we propose is taking it a step further with a C&C AGI and scaling it beyond human capability, augmented with all our best technology in computation, information sciences, and access to the vast realms of military intelligence. Having it participating in dozens of battle simulations every day to produce a very powerful and intelligent weapons system will enable us to decisively fight battles from large-scale land/air/sea engagements between nations, to resolving intra-national conflicts between factions, to small conflicts where we need to selectively target specific individuals who present a threat, and having the autonomous AI weapons hunt and eliminate them without collateral casualties or risk to our people in uniform.

It would form an effective deterrent, just knowing that the country that has such a C&C AGI and an autonomous fighting force could have a huge advantage over a conventional military with human pilots, drivers, and force

commanders. Any opponent would think twice about starting a conflict that they can never win against such a superior AGI controlled force.

Ok, now it is out there, and the taboo subject of AGI in defense has been breached. We have sketched out the design for a fully implemented Autopilot AGI to pilot individual combat units and allow them to work autonomously and collectively as a swarm, sharing intelligence and making coordinated decisions. We have described how to build a C&C AGI that can take over the command and control network, using the same hierarchy and operations procedures that a human C&C hierarchy uses, but much more efficiently, reliably, and much faster. We have discussed how to evolve and train them in simulations and in field operations such that they rapidly improve and adapt to become a near unbeatable force with near-instantaneous dissemination of intelligence, decision making and execution of missions that far exceeds the capabilities of a human military. All that the humans have to do is issue the orders to go to war, and the Defence AGI could do the rest.

Should we build such a system? Hell no. You could end up with Skynet and all the other movie tropes of AI taking over the world coming to life. This is extremely dangerous technology, and it has incredible potential for misuse or disaster if not correctly constrained and managed.

At one end of the spectrum, human rights advocates state that an AI should never be allowed to take action that would result in the loss of life without direct control by a human being over that decision. That is a noble sentiment, but somewhat negates the speed and flexibility of a military AGI if it has to have a human being in the loop for each such decision for each vehicle. This is the conundrum caused by bringing AI into the loop, is that it functions better and faster when free of human control.

Training the AGI with guardrails, such that each decision must conform to the rules of combat, international law, and a code of ethics seems like a more realistic approach. Also, as we discussed earlier, it is a necessity to have the system be

transparent so that humans can interpret what the C&C network is deciding and ordering, and what the exact actions of the units are and to have the ability to override those orders and actions at any level.

If we want to keep the C&C AGI transparent to (and compatible with) human commanders and human piloted units, we will model each person in the hierarchy of command with an AGI mimic commander that takes in the orders above it, intel, and other inputs then outputs orders to the commanders below it in SMESC text format. Each of these mimics would train on the battlefield and training exercise data of actual human commanders, building up an artificial command network composed of these mimics, where at each level, real human commanders could read and understand the orders and override them if desired. This is the first and most important safeguard against a Skynet scenario, to have a human readable decision process with human oversight throughout.

However, a question arises of how can the humans in the loop keep up and can they effectively manage such an automated system that can think and make decisions so much faster, with a longer-term predictive horizon than humans can manage? Would the humans in the loop just constrain it so much that they negate much of its advantage? This is a slippery slope, especially if two AI-enabled nations are in a conflict and there is pressure to automate more in order to gain a decisive advantage.

To address the most poignant question: Should nuclear weapons be placed under AI command and control? That is a tough decision that anybody who has watched the Terminator movies would instantly oppose. But when you think about it in more depth, who would you rather trust with the decision to use nuclear weapons, a human president who will probably be acting in fear and could be acting on partial or incorrect information during a time of crisis, or an AGI trained with guardrails to limit civilian casualties to the minimum when pursuing its objectives, with a broad and deep knowledge of the situation and the consequences of any actions involving nuclear weapons?

We should probably not have nuclear weapons at all, but mutually assured destruction is now ingrained in world geopolitics and warfare and has probably prevented many full scale world wars from happening, so we are stuck with

it for the foreseeable future. Our only course of action in dealing with nuclear weapons is to limit access to them and reduce the probability of them being used. If properly designed, the AGI with proper safeguards should never compute a solution to a scenario that involved use of nuclear weapons, and instead would avoid doing so at all costs. Could an AGI be a better solution to make nuclear weapons less likely to be used rashly or in error, while still providing a deterrent? This leads down the proverbial rabbit hole, but it has to be explored when considering AI in defense.

Regardless, the high-level control and authorization for any weapons release should have a human in the loop at some level of the command and control hierarchy, so that a final decision to deploy weapons is still in human hands, even if its detailed execution decisions are made by an AI. A person must be able to take over control from the C&C AGI and abort a mission if they feel it is warranted. Again, it is a compromise over how much control is automated, fast, and accurate, and how much is handled by the slower human counterparts with less ability to see the bigger picture and predict the future. This is a slippery slope, as a future conflict between two nations with AI-enabled C&C may result in an arms race, with the C&C AGI on each side becoming more and more autonomous and the human control being reduced to gain an advantage and keep up with the opposition. This has to be considered ahead of any such conflict, and safeguards implemented both in the AGI and in the policy and process for using it.

No matter what our intentions or what we do, AI in warfare is becoming a reality, as there is a multibillion-dollar defense industry and constant conflict around the globe driving innovation in defense technology. Integrating AI with their military has become a top priority for many countries, and there is an AI arms race in progress that is just as real and just as scary as the atomic weapons arms race was in the late 20th century. How we develop this AI, and how we keep it accountable to the rules of combat, international law, and ethics and keep it under control of human commanders will determine if it becomes a tool for resolving armed conflict more efficiently and with less loss of life, or a source of our own Armageddon.

Again, I write this chapter with some trepidation because AI in defense is such a controversial topic which will cause heated debate and definitely some resentment over such a system even being proposed. On the other hand, we have to

acknowledge that AI in defense is already becoming a reality, moving faster than governments are in regulating it. We have to know what can be done so that we can make intelligent decisions about what to do and what precautions to take.

CHAPTER 13

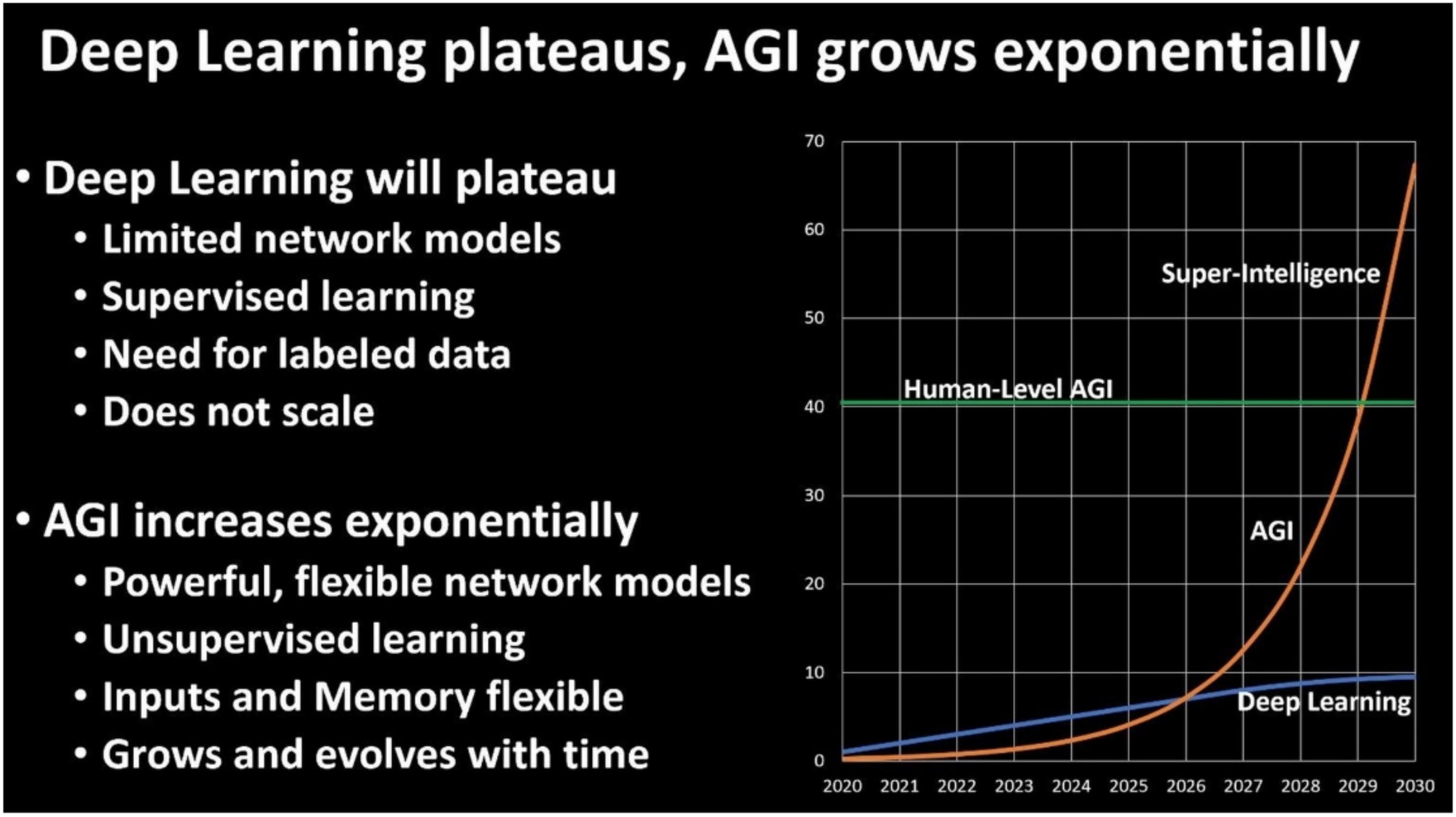
Market Potential of AGI

It is difficult to pin down the current and future market for AI because there is not even a clear definition for what we call AI, and there is not a clear delineation for where AI ends and the product begins. Do we call an iPhone AI? Is Siri AI? How much revenue does Siri bring into Apple, or is she just a value-add feature that helps sell iPhones? What is that increase in sales due to her inclusion in the product, and how do we measure it? I think you can see the problem. AI (including AGI) is an enabling technology that goes into products, increasing their market appeal, price, sales, and revenue for their company. AI as a technology has an actual market that is somewhat intangible.

If we count total sales of all the products that have AI (or deep learning) as a contributing component, the market could easily top a trillion dollars in the next 5 years. This would include revenue from all mobile devices, computers, appliances, automobiles, home robots, smart home devices, online services and so on…

According to the Fortune Business Insights report: Artificial Intelligence Market, 2022-2029, the global Artificial Intelligence market is projected to grow from $387 USD billion in 2022 to $1,394 billion USD by 2029, exhibiting a CAGR of 20%. “Artificial intelligence is being rapidly integrated into corporate processes around the world to improve business operations and consumer experience. Small and medium businesses are also investing in technology to improve the functionality and performance of their operations at a low cost.”

However, these projections do not take into account a disruptive Artificial General Intelligence like we propose in this book. In 2028, human-like AGIs would emerge into the professional services and administrative industries, and by 2030 they will dominate and would soon merge into a singular AGI superintelligence spanning the globe and become so much more. AGI could be running a majority of our human professions like administrative, financial, legal, medical, and be dominant in those fields.



The future 'market' for an AGI Superintelligence could easily be a double-digit percentage of the world's growing $100 trillion GDP in a couple of decades if it permeates as deeply as we forecast it could. This really is a civilization-changing technology.

Building a core AGI like this, enabling an ecosystem of front-end platforms and back-end applications by 3rd parties, and training 3rd party developers on the integration is a business model that could get traction and scale very quickly. The technology could rapidly penetrate markets and verticals and be carried to market by partnering with companies that already have the products, customers, sales channels and revenue.

To become globe-spanning in a reasonable timeframe, a company cannot go it alone. The plan would be to license technology and provide AGI as a Service, allowing it to spread everywhere within a decade of introduction, and by networking and interfacing the various applications, entirely new genres of products and services become possible as the combined AGI climbs towards superintelligence.

CHAPTER 14

Developing AGI Safely

Without exception, everyone who reads a post or article about AGI technology immediately raises the question of safety, with many invoking science-fiction scenarios of AI gone wrong, using examples like 'Skynet' from the 'Terminator' movies or 'VIKI' from 'I Robot'. These fears are often misguided about how far from these fictional portrayals of today's AI is, and even how far AGI in a decade would be. Still, the power of an AGI that can learn, expand, and evolve exponentially like we outline in this book (and the patents referenced) has to be shaped and wielded with respect and caution so that these and many other (less catastrophic but still damaging) adverse scenarios are not the result.

In this chapter we explore how we would actually train an AGI, what some of the risks are and how to mitigate them, both by safeguards built into the training process and by measures taken to isolate a proto-AGI under development and evolution until its behavior can be shown to be productive, beneficial, and benevolent - before opening it up to the Internet and all of humanity and our technological infrastructure connected to it.

Automated Training on Datasets - First the AGI Cortex has to be trained on a sizable dataset encompassing all of the modalities of input that we expect it to handle. This data is learned unsupervised by sending it to both the input

layer of the appropriate cortex, as well as the generative layer, with data going into the input layer lagged behind the generative layer so that we train the AGI cortex to predict the inputs on the generative layer.

Once the AGI cortex is sufficiently trained to start predicting, it will learn new inputs on the fly and be able to train itself further while operating, laying down new memories and predictive pathways as it does so. If the input is turned off and the generative layer is left active, we want it to dream and forge its own predictions of imaginary narratives, based on the memories and predictions computed in the generative layer.

We must be careful in selecting the training data set so that it does not have biases or harmful content in it, and so that it has been selected to provide enough depth and breadth of information to represent the modality of data it is being trained for.

Once the AGI is in operation and learning from real inputs, we still need to have some filters in place, some guardrails so that the AGI is not learning from harmful content, nor getting biased input from sources that represent incorrect information. The most elegant way to do this is by building the guardrails and filters into the training data, as it is awkward to train an AI on unfiltered data, then try to enforce policies and guardrails manually with code. A pre-training screener could run through the data, combing biased and harmful content from the data and putting in explicit examples of the guardrails being enforced.

Training with people - Do humans have the bandwidth to train and evaluate an AGI?

Once it is up and running, defining the performance criteria for advanced AGI's to be evaluated against is a very tough problem. Some prior art has stated that real people should be allowed to test the AGI once it gets to human-level performance, as in the Turing test, and having real people involved in curating the answers from the AGI to provide a moral input. But real people are limited in their available time, skill sets, consistency, biases, and most importantly, speed. They can only type or ask a few questions per minute. It would be very expensive, difficult, and annoying for

enough people to repetitively test an AGI day in, day out, let alone test hundreds for generations on end for genetic selection purposes.

Training with Simulated People - Using social skills and empathy as selection criteria

By using the other AGI's to train against, we can endlessly replicate them, and use them to run detailed testing on the AGIs in each of the fields it can operate in, such as medicine, law, finance, etc. We could give it the ability to interact with real people better by representing each AGI instance by using a human character that is speaking and using body language and emotive expressions when interacting with the human users and other AGI candidates, which can see, hear, and understand. The competition can be scored by a referee AI that can oversee the interactions and rate the AGI not only on how well the AGI candidate performs in a vocational sense, but how well is it communicating, showing correct verbal skills, expressions, and emotes that are context appropriate.

With or without an avatar, the referee AI would determine if the AGI socializes and appears to empathize with its human counterpart as well as perform the vocational tasks it is evolved to do? This sounds inane, but when you train a AGI that could grow to superhuman capability and beyond quite quickly would you prefer A) One with social skills and empathy with humans, or B) Without? I will explore the divergence of these two cases below.

Dangers in Synthetic Evolution on Overdrive - Evolving an AGI at 1000x the rate of biological evolution with specific selection criteria can go exponential quickly and be dangerous.

Because the opponents or trainers are also AGI's, we can not only replicate them as much as we want, but they can be run at computer speeds, possibly interacting with each other 10x or 100x faster than a human could, greatly shortening the training and performance evaluation time, and the interval time between evolving generations. But this is dangerous, and this is why we define the later set of isolation and shutdown measures for a facility doing this kind of simulation and training to evolve AGIs.

Here is why this is so dangerous. Biological evolution of complex life took about a billion years to make humans from microbes, mostly because the time increment between generations was in the range of days to decades as we evolved. And also, because evolution was not specifically aimed at producing an intelligent human, it spent a lot of time just blundering around making weird swimming creatures with seemingly random bodies and brains, then giant dinosaurs with large teeth and small brains, then scurrying small mammals on land with warm bodies, then big, warm-blooded whales with big brains that went back to live in the ocean but had no hands to make tools. There was no divine intelligence guiding the process, with only evolution simply selecting just the individuals that were able to survive long enough to procreate each generation, for millions of generations. Even when life was nearly completely wiped out on earth in several mass extinctions a few microbes or clams, or small lizards, or shrews survived, life re-emerged and proliferated, until we ended up at the present with humans proliferating as the dominant species.

We humans arose and prevailed perhaps only because we were equipped with hands and superior brains and could make tools, and so became better at digging up roots and killing and eating game, or killing off opposing smaller-brained, simpler-tooled, hominid cousins (that we may or may not have mated with on the side). Our evolutionary path was by decided survival against the elements and our brethren in trials by disease, famine, combat, and even genocide. It is not a pretty tale when it is examined in detail.

Our human evolution was also not a nice linear slope, rising from zero a billion years ago to us now. It was exponential. It started really, really slowly and then suddenly humans and our brain's capability zinged upwards in capability exponentially, starting a few million years ago, and we only really became fully intelligent modern humans in the last few hundred thousand years. But exponential growth is not always a good thing, as a malevolent species could result, just like how, during the era of the dinosaurs, monstrous apex predators evolved to the top of the ecosystem pushed there by an arms race of size, teeth, claws, and muscle.

As opposed to evolution taking decades to produce, train, and test a single human generation, the times between generations for AGI can be in greatly reduced by pitting AGIs against each other, and in this case, we are specifically

'evolving' these AGI's under very particular selection pressures to become better than their peer set of AGI's at in specific areas of expertise, which is probably much more rapid and potentially dangerous if it goes off track. This evolution would be orders of magnitude faster than any evolution that the natural world could have ever hoped to achieve, and although it could chug along for days, weeks, even months, without much progress, suddenly POW - it could go exponential literally overnight and have the potential of producing a benevolent or malevolent entity, depending on how carefully we did the training process.

In scenario A) above, we trained our AGI against each other and with humans with human avatars that look, act, and interact like people, so it is socialized with people, and it has interacting with them and empathizing with them built into its DNA. As long as we could keep it contained and learn to tame it, then this could turn out very well, and our AGI evolution under these conditions could produce a benevolent, very multi-talented AGI that is capable of amazing feats of not only computation, but is also capable of extrapolation, interpolation, intuition, creativity, and empathy towards us humans (from learning to read our emotions from all the mimics it trained against) and able to do the many human tasks and jobs effortlessly that it trained to do.

Such an entity would be able to help solve many of humanity's toughest problems and challenges, working with us in harmony to innovate alongside us to make discoveries in fundamental sciences and mathematics as well as applied sciences like engineering, medicine, pharmaceuticals and help bring us new discoveries, inventions, and technologies. It could continue to learn and expand as it did so and would become the single most powerful technological tool man has ever created, remaining under our control. This creative and productive outpouring could raise our standards of living, ease our burdens, financial and otherwise, and bring a better standard of living to all mankind worldwide by doing so. It may even help us to govern more fairly and peacefully. It would also make the company that owns and controls it fantastically wealthy and powerful.

Or we could go the route of B) without training the AI to recognize, interact, and empathize with humans and we could end up with a malevolent or even indifferent AI that is detrimental to our survival or just evolves beyond us and forgets to serve us like it was intended to do.

That's why the social training, of learning against both human avatars and real people is for, plus the isolation of the facility (see below), as there will be no time to steer or correct this process once it goes exponential, and if you trained it wrong, you may end up with something undesirable coming out of it.

It sounds like mad science fiction, but the book "SuperIntelligence" By Nick Bostrom (13) gives a much better and more thorough treatment of strong AGI and the potential dangers, and the precautions we should take when experimenting with these technologies. The above is kind of an exaggerated scenario but it shows that training advanced AGI's with human social skills, values, and empathy is just as necessary as training them to do the tasks such an AGI is meant to do. It also needs to have a human interface, because of what meaning is a superintelligence if we cannot access it and relate to it?

Isolation When Training / Evolving - Physically isolating the AGI from the rest of human technological infrastructure (The Internet) is a must when doing rapid evolution.

When evolving an AGI during explosive growth, where experimentation and trying out different configurations quickly is desired, it would make sense to isolate the AGI program so that it could not get out onto the Internet, multiply, mutate and breed to become something that is a nuisance or even dangerous.

There can be software precautions taken, like keys that are required to enable breeding and evolution that are only on the target hardware, that are turned off after deployment, but we are making something that is going to equal or exceed human capability, and it is going to be extremely clever, and may be able to circumvent these simple software countermeasures.

Physically constraining the AGI to a given cluster and isolating or air-gapping it from the rest of the internet is the only way to make sure that an evolving AGI doesn't become malicious and get out into the internet and wreak havoc. Also, having a limited compute capability will ultimately limit the power and capabilities of that AGI such that it can only become more efficient on that target platform, provided we can constrain it to that platform and keep it resident on it and only it.

There are so many forms of proto-AGI that could go wrong that these measures should be mandatory. Even a simple replicator that learns how to hack into networks and spread could gum up the entire internet within a few generations of its evolution, shutting it down. Those DARPA cyberhacking competitions with AI scare the hell out of me for just this reason.

CHAPTER 15

Achieving Global Superintelligence with AGI

We discussed in Chapter 5 that our AGI design was open ended and could be expanded by adding more cortices for more modalities of data, exceeding what humans can handle. By using genetic algorithms to constantly improve the AGI cortex and other components, gaining more data as we go, we could feasibly evolve an AGI that grows to exceed human capabilities and becomes a superintelligence.

By training AGIs against each other and using genetic algorithms to prune all but the highest performing, we can drive the capability of that AGI far beyond human, shaping it into a constantly evolving superintelligence.

AGI 'Eta' 2030 Concept - As an exercise, let's fast-forward to the year 2030 and name a future globe-spanning AGI superintelligence Eta, and make it feminine, with a crisp British accent to make it sound more intelligent.

Eta is a generation 5 Artificial General Intelligence from the year 2030. Her artificial mind spans the globe and can take in a world's worth of data per day, going back decades. But to pass the threshold and become an artificial general intelligence, she needed the ability to do most human tasks and professions and to solve all the general problems that a human can, by just learning from the world around her.

Her internal functionality had to operate like the human brain, on narratives constructed from memories of events, traversing the links between them during memory recall, prediction, and planning, using human language as the backbone of these narratives, anchoring language to the AGI's perception of reality, and providing implicit labeling of those events and narratives to give additional meaning and structure to them via the syntax and grammar of language.

Then she can process these narratives, create fictional narratives, make predictions and show what the future looks like, displayed in the same format as the inputs – speech, video, or even display a human persona, an avatar, that talks and emotes like a person (befitting a sci-fi level AGI).

Eta can simulate or 'dream' multiple timelines in parallel, and each time she have a dream that solves a problem better, or predicts the future better, she can then record that narrative into memory, influencing the direction of her future dreams, and by doing so, converge them to a desired solution or the most likely future path. With these abilities, she is an Oracle from which to seek knowledge and forecasts for individuals, corporations, or nations.

With her AI, she represents whole narratives as sequences of engrams or experiences hierarchically, matched to words, sentences, and paragraphs (with language as the backbone of the memory system), storing them hierarchically at multiple resolutions, and so is able to have the overall context for sentences, paragraphs, and pages, all learned from millions of narratives that she takes in every day.

By using her cognitive and sensory systems integrated together, Eta can create a human mimic, instantiated with 3D computer graphics as a character with the motion and facial expressions mapped to the character, with the motion controllers trained on data from a real person's performance capture such that they can learn the sequences, build underlying basis sets in the inhibitor network, and quickly learn new performances like humans do. This gives Eta a way to provide a more personal interaction with the user.

(Link to Eta AGI Video)

So, what impact will an AGI superintelligence have on the world? In earlier chapters, we described how we are taking the first steps towards AGI that can perceive the real world, reduce those perceptions to an internal format that computers can understand, yet still plan, think and dream like a human, then convert the results back to human understandable form, and even converse fluently using human language, enabling online professional services in finance, medicine, law, and other areas. It can also add these enhanced analytics, forecasting, and decision-making capabilities to financial forecasting and enterprise software - where it can be used by businesses large and small.

AGI applied to finance could consistently outperform human brokerages and would quickly gain a large market share in finance, even if we tax that by 1% to disburse funds for a living wage income for those living furthest below the poverty line. As the AGI broker subsumes control of the world markets, we increase the subsidies to the poor, and maintain a minimum investment balance for them so they can reap the rewards of it appreciating too. The rich and the poor both get richer. The market for products and services expands, growing companies and global wealth further. Poverty (and hunger) could be erased, perhaps within a generation.

AGI used in medicine that is deployed worldwide, in every language, on every mobile device, with all of humanity's medical knowledge, integrated with existing medical and pharmaceutical systems, could bring quality medical care to the 80% of the globe that are lacking it right now. Large pharma could finance it by buying access to the knowledge base (with individuals' info obfuscated and encrypted of course) of the course and treatments of diseases, and by gaining access to the expanding markets served by the Medical AI.

An AGI applied to law could replace the whole legal system, the inefficient, outdated, ineffective (and often corrupt and self-serving) lawyers, DA's, judges, and courts. The AGI would gather the information from the plaintiff(s) and defendant(s), and help walk each of them through the laws, and what info is needed at each step, and provide them with tools to organize and format their presentation (independently, securely). If the case has merit, a human jury is recruited, trained on the same legal points, walked through each side's presentation, and asked to deliberate and make a judgment. Perhaps the AGI can also learn how they do this - so well that jury duty also becomes a thing of the past.

An AGI in Enterprise/Administrative ERP could revolutionize how companies and government agencies forecast and plan, to help them gather and focus unprecedented amounts of information and to look months, even years into the future to make the best recommendations and help plot the best courses of action. Sometimes the AGI can make recommendations for international, inter-company ventures that benefit all of mankind, like building massive solar farms with revolutionary new solar cells or creating joint ventures to develop better energy storage and battery solutions, and other endeavors that take deep R&D and deep pockets, provided and coordinated by the AGI. We could have solutions to many intractable global problems within a generation.

Problems of wealth inequality, poverty, hunger, injustice, and lack of basic services for healthcare and information services are the norm for 3/4 of the people in the world, and for millennia, human civilization has been unable to solve these basic problems, no matter the form of government or choice of deity and belief system humans adopt. People are just unable to see the larger picture and seem helpless to do anything about it.

Over the next decades, a Superintelligence, a Strong Artificial General Intelligence, will evolve to oversee a global network augmenting the systems of Law, Medicine, Education, Finance and augment all previous human administrative functions. With its vast, wide, and deep knowledge capability, the wisdom to draw on all this past knowledge and the ability to predict possible paths into the future, this superintelligence will serve all of humanity and assist us to make carefully measured & unbiased choices to help to guide us, judge us, and govern us accordingly for the betterment of all.

Entrepreneurs can talk about making the world a better place but evolving an AGI to take the lead in solving problems in all these human fields, to tackle administration, planning, customer service, finance, medicine, law, personal assistants, all with a single architecture, globally, in all languages, makes that AGI a civilization-changing tool that would be unparalleled in human history. It could finally take the lead to show humanity the path to a sustainable future where poverty, sickness, and injustice are a thing of the past. This is the power to truly make global change for the better, within decades.

Glossary

Artificial Intelligence (AI) - Intelligence that is technology based, not naturally evolved, capable of learning one or more tasks, with or without human supervision and labeling of data.

Artificial General Intelligence (AGI) - Intelligence that is technology based, designed to approximate human intelligence that is capable of learning multiple tasks without direct supervision, without labeled data, and is capable of transfer learning between tasks

Attractor State - in a dynamic system, converging toward stable system behavior despite varied initial conditions.

Autoencoder (deep learning) - learns to encode data into a machine latent format and back out to the original format again, with the machine latent format and encoding/decoding process being learned from the data.

Basal Ganglia - group of subcortical nuclei, or groups of neurons under the cortex, involved in motor control, goal directed behavior systems, emotions, etc.

Basis Vector - a vector that specifies a feature or component of a vector, such that the set of all basis vectors can be linearly combined to represent any vector in the data set.

Basis Coordinates - the scalar values by which each basis vector is multiplied to reconstruct a vector

Basis Set - a set of basis vectors that completely spans the vector space such that any vector in that space can be reconstructed by a linear combination of basis vectors in the set

Bidirectional Interleaved Complementary Hierarchical Neural Network (BICHNN) - an autoencoder in which two spiking neural networks of complementary function and opposite signal direction are interleaved in a feedback configuration such that one network encodes to a machine latent format and the other decodes back into the original data.

Brain Stem - stalk-like posterior part of the brain that connects the cerebrum to the spinal cord, composed of the pons, midbrain, and medulla oblongata. Controls basic life functions such as breathing and heartbeat.

Broca's Area - part of the motor cortex associated with speech motor control

Central Pattern Generator – self-contained circuits that create repetitive patterns of motor behavior

Cerebral Cortex - outermost layers of the brain associated with higher cognitive functions

Clustering (ML) - grouping data that has similar features into 'clusters', such as sorting objects by color or shape.

Conditioning (biological) - a process of behavior modification whereby behaviors become associated with unrelated stimuli.

Connectome (biological) - complete map of all the neurons and all of their connections in the entire brain via axons, synapses, and dendrites.

Connectome (artificial) - a complete map of all artificial neurons and all connections between them in an artificial neural network

Convolutional Neural Network (CNN) - an artificial deep neural network where the operation between neuron layers is a convolution carried out between the previous layer and a set of learned filters. Usually used in image processing.

Cortical Column - a vertical cluster of neurons in the cortex that forms the fundamental building block of the cortex.

Crossbreeding (artificial) - taking two genomes, each consisting of a sequence of numerical data, and combining them into a third genome that has numerical values that are computed from the values in the original two genomes, often by linear interpolation, or random selection of one value or the other.

Deep Neural Network (DNN) - an artificial neural network that has layers of neurons with a simple summation model and connections between neurons of each layer and the next layer in the network that are represented by scalar weights.

Deep Learning (DL) - training deep neural networks on data by passing in an input, then comparing the output to the desired labeled value and using back propagation and linear regression to adjust the weights of the connections in the DNN until the outputs match the desired labels.

Deterministic Algorithmic Expansion - using a process or algorithm to expand a set of data into a much larger set of data, done in a manner such that for the same starting data, it always expands into the same larger data.

Engram - a pattern of electrical activity in the brain that represents a thought or idea. We define it as a snapshot in the activation of the neurons in the cortex at any given time.

Ethology - the scientific study of animal behavior.

Evolution - change in characteristics over generations from a common ancestor, usually driven by genetic selection at each generation.

Finite State Machine (FSM) - a method where a machine maintains a state, and can transition into a set of other states from the present state based on different conditions.

Fixed Action Pattern (biological) - a predictable series of actions triggered by a cue known as a sign stimulus

Generative Adversarial Network (GAN) - a set of artificial neural networks where one neural net creates a data set from a random seed and a second artificial neural network tries to determine if the result is real or manufactured. They are both trained against each other on a common data set of seed and resultant data until they converge to a system that is optimal at ‘faking’ and ‘discriminating’ the fakes.

Genetic Algorithms - artificial evolution accomplished by cross breeding the genomes of individuals, then generating the offspring and testing them against a fitness criteria to find the top percentile, whose genes are then cross bred for the next generation, discarding the less fit individuals and genomes with each generation and promoting more and more fit individuals.

Gene - unit of heredity passed down from parent to offspring composed of DNA (deoxyribonucleic acid) which determines some characteristic of the offspring

Genome (biological) - complete set of the genes of the organism.

Genome (artificial) - a set of numbers that completely specify a connectome by a deterministic algorithmic expansion

Glia - several types of cells in the brain that are not neurons, but are involved in maintenance, blood-brain barrier, and regulating activity in the brain

Habituation (biological) - an animal gradually stops responding to a repeated stimulus.

Hebbian Learning - "Cells that fire together, wire together." A theory of synaptic plasticity that states that coincident activity in both the upstream and downstream neurons leads to increases in synaptic strength between these two neurons.

Hierarchical Autoencoder Network (HAN) - a layered network of autoencoders interleaved with principal component axes (PCA) that sort the machine encoded data based on a given feature (per axis) into clusters which are then autoencoded and sorted on other feature axes recursively til single features remain.

Hippocampus - a pair of structures in the mid-lower brain, shaped like seahorses, that are critical in formation of long term memories, memory recall, planning, and prediction.

Hodgkin-Huxley SNN - a mathematical model for spiking neurons that is very close to the behavior of biological neurons, pioneered by Alan Hodgkin and Andrew Hodgkin in 1952 and that won the Nobel prize in medicine in 1963.

Inhibitory Connections/Synapses - connections between neurons that deliver negative charge to the target neuron when a spike is transmitted between neurons.

Inhibitory Neural Networks - neural networks that consist of layers that selectively inhibit a signal traveling down the neural network from the base, such that the signal only propagates when each neuron it passes through is turned on.

Innate - behavior that's genetically hardwired in an organism and can be performed in response to a cue without prior experience.

Izhikevich SNN - A mathematical model for spiking neurons that is an intermediate approximation of the behavior of a biological neuron, yet still simple enough to be computationally tractable for large networks.

Lateral Geniculate Nucleus (LGN) - structure within the thalamus where visual inputs from the optic nerve first connect within the brain.

Long Short Term Memory (LSTM) - a small neural network that feeds back on itself and that is used as a component in larger recurrent neural networks(RNNs). It has the advantage of solving the vanishing gradient problem for RNNs that use simpler feedback mechanisms.

Motor Cortex - the part of the brain that generates motor control signals for the muscles

Mutation (artificial) - altering the numbers in an artificial genome by a random amount to introduce variation

Narrative - we define this for our AGI as a series of engrams that results from sensory input or generative prediction.

Neuron (biological) - a cell that integrates the time-sum of current spikes coming in from its dendrites, thereby computing a voltage within the cell, and fires a spike from the neuron cell's axon hillock and down the axon when that voltage exceeds a threshold.

Neuron (as used in DNNs) - a node that adds up all the signals coming in from the connections from the previous layer of artificial neurons and applies a function to that sum before passing the result down the connections to the nodes of the next layer.

Prefrontal Cortex - the front part of the brain's cortex, involved in logic, planning, and decision making

Rank Order Selective (ROS) Neurons - A connected set of neurons that fire in a series to set a timing signal for a ROS-Inhibitory network.

Recurrent Neural Network (RNN) - An artificial deep neural network in which the signal from a layer is fed back into the current layer to allow it to retain information about past data for time-series calculations.

Selection Criteria - in genetic algorithms, the criteria, or test, that is applied to the entity being evolved to determine whether it will be chosen to be crossbred for the next generation or not

Sequence Prediction (AI) - predicting the next data entries in sequence, based on training on past sequences

Sequence to Sequence Neural Nets (seq-seq NN) - artificial neural networks that take one sequence as input such as a question, and output another sequence that is the closest match to the correct answer. Translating from one language to another is a second example

Sign Stimulus - a cue that produces a specific response in an organism.

Somatosensory Cortex - the cortex of the human brain that processes the inputs from the touch, pain, temperature, and other sensory neurons in the body

Spiking Neural Network (SNN) - an artificial neural network in which the signals travel as spikes between neurons, gated by synapses, with a model based on biological neurons, synapses and neural networks.

Superintelligence - an entity that surpasses human intelligence.

Synapse (biological) - a junction between one neuron's axon and another neuron's dendrite where an incoming electrical impulse from the axon triggers release of neurotransmitters into the junction, which accumulate in receptors on the dendrite side, eventually causing a spike of electricity to be released into the dendrite.

Temporal Lobe - a lobe down the side of the brain, behind the ear.

Thalamocortical Radiations - a bush-like neural network radiating out from the thalamus to the inner surface of the cerebral cortex, carrying information to and from the sensory and motor control cortices.

Thalamus - a golf-ball sized neural mass near the center of the brain that processes input from all the senses and the outputs to the body

Transformer (deep learning) - an artificial neural network using deep learning that processes a stream of tokens by learning statistically what tokens tend to come before and after a given token in a stream.

Wernike's Area - an area in the brain in the left temporal lobe that processes language. Damage to this area renders a person unable to form coherent sentences nor understand sentences.

References

(1) Antonio Zadra and Robert Stickgold, When Brains Dream, W. W. Norton & Company (January 12, 2021)

(2) Miguel Nicolelis, Beyond Boundaries: The New Neuroscience of Connecting Brains with Machines and How It Will Change Our Lives, Times Books; Illustrated edition (March 15, 2011)

(3) Peter Voss, Essentials of general intelligence: the direct path to AGI, Kurzweil - Tracking the acceleration of intelligence, August 22, 2002 (https://www.kurzweilai.net/essentials-of-general-intelligence-the-direct-path-to-agi)

(4) Emilio Salinas, Rank-Order-Selective Neurons Form a Temporal Basis Set for the Generation of Motor Sequences, Journal of Neuroscience, v.29(14), 2009 Apr 8, PMC2677524

(5) Suzana Herculano-Houzel, The Human Brain in Numbers: A Linearly Scaled-up Primate Brain, Frontiers Human Neuroscience. 2009; 3: 31

(6) Ewoud R. E. Schmidt and Franck Polleux, Genetic Mechanisms Underlying the Evolution of Connectivity in the Human Cortex, Frontiers in Neural Circuits, 07 Jan 2022

(7) Zhiyi Chen, Rong Zhang, Hangfeng Huo, Peiwei Liu, Chenyan Zhang, Tingyong Feng, Functional Connectome of Human Cerebellum, NeuroImage, Volume 251, 1 May 2022, 119015

(8) Comrie AE, Frank LM, Kay K, Imagination as a fundamental function of the hippocampus, 2022. Phil. Trans. R. Soc. B377:20210336 https://doi.org/10.1098/rstb.2021.0336

(9) R Rao, Predictive coding in the visual cortex, Nature Neuroscience, 1999 - cir.nii.ac.jp

(10) Jeff Hawkins, A Thousand Brains: A New Theory of Intelligence, Basic Books, New York, October 25, 2022

(11) Jeff Hawkins, Subutai Ahmad, Yuwei Cui, A Theory of How Columns in the Neocortex Enable Learning the Structure of the World, Frontiers, Neural Circuits, 25 October 2017, Volume 11 - 2017 | https://doi.org/10.3389/fncir.2017.00081

(12) Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., and Friston, K. J. (2012). Canonical microcircuits for predictive coding. Neuron 76, 695–711. doi: 10.1016/j.neuron.2012.10.038

(13) Nick Bostrom, Superintelligence: Paths, Dangers, Strategies, Oxford University Press, May 1, 2016

(14) Marcel Oberlaender, Beyond the Cortical Column, Neuroinformatics 2012

(15) Yeh, F. C., Panesar, S., Fernandes, D., Meola, A., Yoshino, M., Fernandez-Miranda, J. C., ... & Verstynen, T. (2018). Population-averaged atlas of the macroscale human structural connectome and its network topology. NeuroImage, 178, 57-68. - http://brain.labsolver.org/

(16) Lang S, Dercksen VJ, Sakmann B, Oberlaender M. Simulation of signal flow in 3D reconstructions of an anatomically realistic neural network in rat vibrissal cortex. Neural Netw. 2011 Nov;24(9):998-1011. doi: 10.1016/j.neunet.2011.06.013. Epub 2011 Jun 25. PMID: 21775101.

(17) Langbrain, Language and Brain: Neurocognitive Linguistics, Rice University, Houston, Texas, https://www.ruf.rice.edu/~lngbrain/main.htm

(18) Stefan Lang, Vincent J. Dercksen, Bert Sakmann, Marcel Oberlaender, Simulation of signal flow in 3D reconstructions of an anatomically realistic neural network in rat vibrissal cortex, Neural Networks, Volume 24, Issue 9, 2011, Pages 998-1011, ISSN 0893-6080, https://doi.org/10.1016/j.neunet.2011.06.013.

(19) Hole, K.J., Ahmad, S. A thousand brains: toward biologically constrained AI. SN Appl. Sci. 3, 743 (2021). https://doi.org/10.1007/s42452-021-04715-0, http://creativecommons.org/licenses/by/4.0/

(20) Serre, Thomas. (2006). Learning a Dictionary of Shape-Components in Visual Cortex: Comparison with Neurons, Humans and Machines. JOUR, 2006/10/31

(21) Andrew B. Barron, Marta Halina and Colin Klein, Transitions in Cognitive Evolution, Proceedings of the Royal Society B, Biology, 05 July2023, https://doi.org/10.1098/rspb.2023.0671

(22) Flindall, J. W., & Gonzalez, C. L. (2016). The destination defines the journey: an examination of the kinematics of hand-to-mouth movements. *Journal of Neurophysiology*, *116*(5), 2105-2113.

(23) Gardner, R. Allen, Beatrix T. Gardner, and Thomas E. Van Cantfort, eds. *Teaching sign language to chimpanzees*. Suny Press, 1989.

(24) Gardner, R. A., & Gardner, B. T. (1998). *The structure of learning: From sign stimuli to sign language.* Lawrence Erlbaum Associates Publishers.

(25) Hassabis, Demis, and Eleanor A. Maguire. "Deconstructing episodic memory with construction." *Trends in cognitive sciences* 11.7 (2007): 299-306.

(26) Hasantash, Maryam, and Arash Afraz. "Richer color vocabulary is associated with better color memory but not color perception." *Proceedings of the National Academy of Sciences* 117.49 (2020): 31046-31052.

(27) Lin D, Boyle MP, Dollar P, Lee H, Lein ES, Perona P, Anderson DJ. Functional identification of an aggression locus in the mouse hypothalamus. Nature. 2011 Feb 10;470(7333):221-6. doi: 10.1038/nature09736. PMID: 21307935; PMCID: PMC3075820.

(28) Romeas, Thomas, and Jocelyn Faubert. "Soccer athletes are superior to non-athletes at perceiving soccer-specific and non-sport specific human biological motion." *Frontiers in psychology* 6 (2015): 1343.

(29) Schacter, Daniel L., et al. "The future of memory: remembering, imagining, and the brain." *Neuron* 76.4 (2012): 677-694.

(30) Staddon, J. E., & Simmelhag, V. L. (1971). The "superstition" experiment: A reexamination of its implications for the principles of adaptive behavior. *Psychological Review, 78*(1), 3–43. https://doi.org/10.1037/h0030305

(31) Terrace, Herbert S., et al. "Can an ape create a sentence?." *Science* 206.4421 (1979): 891-902.

(32) Terrace, Herbert S., et al. "On the grammatical capacity of apes." *Children's language* 2 (1980): 371-495.

(33) Watson, Jason M., and David L. Strayer. "Supertaskers: Profiles in extraordinary multitasking ability." *Psychonomic bulletin & review* 17 (2010): 479-485.

(34) Welbourne, Lauren E., Antony B. Morland, and Alex R. Wade. "Human color perception changes between seasons." *Current Biology* 25.15 (2015): R646-R647.

(35) Cassenaer, Stijn, and Gilles Laurent. "Hebbian STDP in mushroom bodies facilitates the synchronous flow of olfactory information in locusts." *Nature* 448.7154 (2007): 709-713.

(36) Dylla, Kristina V., et al. "Trace conditioning in Drosophila induces associative plasticity in mushroom body Kenyon cells and dopaminergic neurons." *Frontiers in neural circuits* 11 (2017): 42.

(37) Fiebach, Christian J., and Ricarda I. Schubotz. "Dynamic anticipatory processing of hierarchical sequential events: a common role for Broca's area and ventral premotor cortex across domains?." *Cortex* 42.4 (2006): 499-502.

(38) Gervasi, Nicolas, Paul Tchénio, and Thomas Preat. "PKA dynamics in a Drosophila learning center: coincidence detection by rutabaga adenylyl cyclase and spatial regulation by dunce phosphodiesterase." *Neuron* 65.4 (2010): 516-529.

(39) Lüdke, Alja, et al. "Calcium in Kenyon cell somata as a substrate for an olfactory sensory memory in Drosophila." *Frontiers in Cellular Neuroscience* 12 (2018): 128.

(40) Morgan, Gary, and Judy Kegl. "Nicaraguan sign language and theory of mind: The issue of critical periods and abilities." *Journal of Child Psychology and Psychiatry* 47.8 (2006): 811-819.

(41) Senghas, Ann, Sotaro Kita, and Asli Ozyurek. "Children creating core properties of language: Evidence from an emerging sign language in Nicaragua." *Science* 305.5691 (2004): 1779-1782.

About the Authors

Brent Oster has 30 years of experience in the tech industry, having worked at Bioware, Electronic Arts, Check Six Studios. Lucasfilm, NVIDIA and ORBAI. He is an entrepreneur as a founding member of Bioware, Check Six, and ORBAI. He has a bachelor's degree in Aerospace Engineering and a masters in Computer Science, specializing in scientific computing, as well as military pilot training. Brent spent the last 9 years working in artificial intelligence and deep learning, working hands-on with real AI applications at fortune 500 companies as well as researching future technologies such as AGI that can go beyond deep learning and solve many of the tough problems that DL is incapable of solving.

Brent can be reached at brent.oster@orbai.com

Gunnar Newquist is a neuroscientist and tech entrepreneur, founding a company called Brain2Bot with the vision of bringing robot companions to life. He has a PhD in Cellular and Molecular Biology and studied both biology and piano performance as an undergraduate, as well as competing as a professional extreme skier. Gunnar has managed autonomous vehicle products in companies such as BlueSpace.ai and Gatik.ai and has been developing biologically-inspired AI for the last 8 years.