Chapter 20: Case Studies in NeuroAI

Chapter Goals

After completing this chapter, you will be able to:

- Analyze real-world applications where neuroscience has successfully informed AI development
- Evaluate the practical benefits of incorporating neuroscience principles into AI systems
- Identify common patterns and successful strategies across different NeuroAI projects
- Apply lessons from case studies to your own research or development projects
- Understand the specific challenges and solutions in translating neuroscience insights to AI implementations
- Recognize key success factors for interdisciplinary collaboration between neuroscience and AI

> **Note**
>
> This chapter features interactive examples to help you explore key concepts. Click the "launch binder" button at the top of the page or access the interactive notebook to experiment with:
>
> - Interactive PredNet visualization
> - Prioritized Experience Replay simulation
> - Vision Transformer attention mechanism
> - Interactive glossary with neural-AI connections
>
> For an enhanced learning experience, we've also integrated Jupyter AI assistance to help you generate code, get explanations, and create visualizations based on the case studies.

20.1 Introduction: From Theory to Practice

Throughout this handbook, we've explored the theoretical foundations of both neuroscience and artificial intelligence, examining how these fields inform and enrich each other. This chapter shifts our focus to real-world implementations, presenting detailed case studies that demonstrate how neuroscience principles have been successfully translated into practical AI systems.

These case studies represent the cutting edge of NeuroAI—where theory meets application, where biological insights drive technological innovation, and where interdisciplinary collaboration yields solutions that neither field could achieve alone. By examining these concrete examples, we gain valuable insights into the practical challenges and benefits of neuroscience-inspired AI approaches.

20.2 Case Study: Deep Predictive Coding Networks

20.2.1 Background and Motivation

Predictive coding is a neuroscience theory proposing that the brain constantly generates predictions about incoming sensory information and updates its internal models based on prediction errors. This first case study examines how predictive coding has been implemented in deep learning architectures to improve robustness and efficiency.

20.2.2 Implementation: PredNet Architecture

The PredNet architecture, developed by William Lotter, Gabriel Kreiman, and David Cox, implements hierarchical predictive coding in a deep learning framework:

```python
import numpy as np
import tensorflow as tf
from tensorflow.keras import layers, Model

class PredNetBlock(layers.Layer):
    """
    Implementation of a single layer of the PredNet architecture
    """

    def __init__(self, num_channels, **kwargs):
        """
        Initialize PredNet block

        Parameters:
        -----------
        num_channels : int
            Number of feature channels in this layer
        """
        super(PredNetBlock, self).__init__(**kwargs)
        self.num_channels = num_channels

        # Convolutional layers
        self.conv_pred = layers.Conv2D(num_channels, (3, 3), padding='same', acti
        self.conv_error_pos = layers.Conv2D(num_channels, (3, 3), padding='same',
        self.conv_error_neg = layers.Conv2D(num_channels, (3, 3), padding='same',
        self.conv_representation = layers.Conv2D(num_channels, (3, 3), padding='s

        # Pooling and upsampling
        self.pool = layers.MaxPooling2D((2, 2))

    def call(self, inputs, training=None):
        """
        Forward pass through the PredNet block

        Parameters:
        -----------
        inputs : tuple
            (current_input, representation_from_higher_layer)

        Returns:
        --------
        outputs : tuple
            (error, updated_representation, pooled_representation)
        """
        current_input, higher_representation = inputs

        # Generate prediction from higher layer representation
        if higher_representation is not None:
            prediction = self.conv_pred(higher_representation)
        else:
            # For the top layer, prediction is zeros
            prediction = tf.zeros_like(current_input)
```

```
        # Compute prediction error
        error = current_input - prediction

        # Split error into positive and negative components
        pos_error = tf.nn.relu(error)
        neg_error = tf.nn.relu(-error)

        # Process error
        error_processed_pos = self.conv_error_pos(pos_error)
        error_processed_neg = self.conv_error_neg(neg_error)

        # Combine processed errors
        combined_error = tf.concat([error_processed_pos, error_processed_neg], ax

        # Update representation based on combined error
        representation = self.conv_representation(combined_error)

        # Pool representation for the next higher layer
        pooled_representation = self.pool(representation)

        return error, representation, pooled_representation

class PredNet(Model):
    """
    Implementation of the PredNet architecture for predictive coding
    """

    def __init__(self, stack_sizes=(3, 16, 32, 64), **kwargs):
        """
        Initialize PredNet model

        Parameters:
        -----------
        stack_sizes : tuple of int
            Number of channels in each layer of the network,
            from input to highest layer
        """
        super(PredNet, self).__init__(**kwargs)
        self.stack_sizes = stack_sizes
        self.num_layers = len(stack_sizes)

        # Create PredNet blocks for each layer
        self.blocks = [PredNetBlock(stack_sizes[i]) for i in range(self.num_layer

        # Upsampling layers for top-down connections
        self.upsample_layers = [layers.UpSampling2D((2, 2)) for _ in range(self.n

    def call(self, inputs, training=None):
        """
        Forward pass through the PredNet

        Parameters:
        -----------
        inputs : tf.Tensor
```

```python
            Input image or sequence

        Returns:
        --------
        outputs : dict
            Dictionary containing predictions, errors, and representations
        """
        # Initialize lists to store layer-wise outputs
        errors = []
        representations = []

        # Bottom-up pass
        current_input = inputs
        higher_representations = [None] * self.num_layers

        for i in range(self.num_layers):
            # Process through current layer
            error, representation, pooled = self.blocks[i]([current_input, higher

            # Store results
            errors.append(error)
            representations.append(representation)

            # Set input for next layer
            if i < self.num_layers - 1:
                current_input = pooled

        # Top-down pass to update higher representations
        for i in reversed(range(self.num_layers - 1)):
            # Upsample representation from higher layer
            higher_rep = self.upsample_layers[i](representations[i + 1])
            higher_representations[i] = higher_rep

        # Return all outputs
        outputs = {
            'errors': errors,
            'representations': representations
        }

        return outputs

def build_prednet_model(input_shape=(128, 128, 3), sequence_length=10):
    """
    Build a PredNet model for sequence prediction

    Parameters:
    -----------
    input_shape : tuple
        Shape of input images (height, width, channels)
    sequence_length : int
        Number of frames in input sequences

    Returns:
    --------
```

```
    model : tf.keras.Model
        Complete PredNet model
    """
    # Create input layer for sequence
    inputs = layers.Input(shape=(sequence_length,) + input_shape)

    # Time-distributed PredNet to process sequences
    prednet = PredNet()
    td_prednet = layers.TimeDistributed(prednet)(inputs)

    # Combine outputs across time steps
    outputs = []
    for t in range(1, sequence_length):
        # Use previous frame's representation to predict next frame
        prev_representation = td_prednet[:, t-1]['representations'][-1]  # Top la
        prediction = layers.Conv2D(3, (3, 3), padding='same', activation='sigmoid
        outputs.append(prediction)

    # Stack predictions along time dimension
    predictions = layers.Lambda(lambda x: tf.stack(x, axis=1))(outputs)

    # Create model
    model = Model(inputs=inputs, outputs=predictions)

    return model
```

20.2.3 Results and Evaluation

The PredNet architecture was evaluated on several computer vision tasks:

1. **Video prediction**: PredNet demonstrated superior performance in predicting future frames in natural video sequences, particularly in handling object motion and occlusion.
2. **Sample efficiency**: Compared to standard CNNs, PredNet required significantly fewer training examples to achieve comparable performance on object recognition tasks.
3. **Error representation**: The explicit representation of prediction errors allowed the model to highlight unexpected or novel events in sequences.

20.2.4 Neuroscience Connection

This implementation connects to neuroscience in several ways:

- **Hierarchical processing**: The layer-wise organization mirrors the hierarchical structure of the visual cortex.

- **Prediction errors**: The explicit computation of prediction errors corresponds to theories about error-signaling neurons in the brain.
- **Bidirectional processing**: The combination of bottom-up and top-down signals aligns with bidirectional information flow in the visual system.

20.2.5 Limitations and Future Directions

While successful, the PredNet implementation faced several challenges:

1. **Computational efficiency**: The bidirectional processing increases computational demands compared to standard feed-forward networks.
2. **Hyperparameter sensitivity**: Performance is sensitive to the balance between bottom-up and top-down signals.
3. **Future work**: Ongoing research is exploring adaptive weighting of prediction errors and integration with reinforcement learning frameworks.

20.3 Case Study: Hippocampal Replay for Reinforcement Learning

20.3.1 Background and Motivation

The hippocampus plays a crucial role in memory consolidation, with “replay” events during sleep and rest periods helping to transfer experiences to long-term memory. This case study examines how hippocampal replay mechanisms have been incorporated into reinforcement learning systems to improve learning efficiency and generalization.

20.3.2 Implementation: Prioritized Experience Replay

Deep Q-Networks with Prioritized Experience Replay, developed by researchers at DeepMind, implement a biologically-inspired memory system:

```python
import numpy as np
import tensorflow as tf
import random
from tensorflow.keras import layers, Model, optimizers
from collections import deque

class SumTree:
    """
    A sum tree data structure for efficient sampling based on priorities
    """

    def __init__(self, capacity):
        """
        Initialize the sum tree

        Parameters:
        -----------
        capacity : int
            Maximum number of experiences to store
        """
        self.capacity = capacity
        self.tree = np.zeros(2 * capacity - 1)
        self.data = np.zeros(capacity, dtype=object)
        self.n_entries = 0
        self.write_index = 0

    def _propagate(self, idx, change):
        """
        Update the sum tree by propagating a value change up the tree
        """
        parent = (idx - 1) // 2

        self.tree[parent] += change

        if parent != 0:
            self._propagate(parent, change)

    def _retrieve(self, idx, s):
        """
        Find the index of the leaf node where s falls within its priority range
        """
        left = 2 * idx + 1
        right = left + 1

        if left >= len(self.tree):
            return idx

        if s <= self.tree[left]:
            return self._retrieve(left, s)
        else:
            return self._retrieve(right, s - self.tree[left])

    def total(self):
```

```python
        """
        Return the total priority sum
        """
        return self.tree[0]

    def add(self, p, data):
        """
        Add a new experience with priority p
        """
        idx = self.write_index + self.capacity - 1

        self.data[self.write_index] = data
        self.update(idx, p)

        self.write_index = (self.write_index + 1) % self.capacity

        if self.n_entries < self.capacity:
            self.n_entries += 1

    def update(self, idx, p):
        """
        Update the priority of an existing experience
        """
        change = p - self.tree[idx]

        self.tree[idx] = p
        self._propagate(idx, change)

    def get(self, s):
        """
        Get an experience using priority-based sampling
        """
        idx = self._retrieve(0, s)
        data_idx = idx - self.capacity + 1

        return idx, self.tree[idx], self.data[data_idx]

class PrioritizedReplayBuffer:
    """
    Prioritized experience replay buffer for efficient and effective learning
    """

    def __init__(self, capacity=10000, alpha=0.6, beta=0.4, beta_increment=0.001,
        """
        Initialize the prioritized replay buffer

        Parameters:
        -----------
        capacity : int
            Maximum number of experiences to store
        alpha : float
            Controls how much prioritization is used (0 = no prioritization, 1 =
        beta : float
            Controls importance sampling weights (0 = no correction, 1 = full cor
```

```python
        beta_increment : float
            Amount to increase beta over time
        epsilon : float
            Small value added to priorities to ensure non-zero probabilities
        """
        self.tree = SumTree(capacity)
        self.capacity = capacity
        self.alpha = alpha
        self.beta = beta
        self.beta_increment = beta_increment
        self.epsilon = epsilon
        self.max_priority = 1.0

    def add(self, experience):
        """
        Add an experience to the buffer with maximum priority
        """
        priority = self.max_priority ** self.alpha
        self.tree.add(priority, experience)

    def sample(self, batch_size):
        """
        Sample a batch of experiences based on their priorities
        """
        batch = []
        indices = []
        weights = np.zeros(batch_size, dtype=np.float32)
        priorities = np.zeros(batch_size, dtype=np.float32)

        # Calculate the priority segment
        total_priority = self.tree.total()
        segment = total_priority / batch_size

        # Increase beta each time we sample
        self.beta = min(1.0, self.beta + self.beta_increment)

        for i in range(batch_size):
            # Sample a value from the segment
            a = segment * i
            b = segment * (i + 1)
            s = random.uniform(a, b)

            # Retrieve the experience
            idx, priority, experience = self.tree.get(s)

            # Store the experience and its index
            batch.append(experience)
            indices.append(idx)
            priorities[i] = priority

        # Calculate importance sampling weights
        sampling_probabilities = priorities / total_priority
        weights = (self.capacity * sampling_probabilities) ** (-self.beta)
        weights /= weights.max()  # Normalize weights
```

```python
        return batch, indices, weights

    def update_priorities(self, indices, priorities):
        """
        Update the priorities of sampled experiences
        """
        for idx, priority in zip(indices, priorities):
            # Add a small value to ensure non-zero probabilities
            priority = (priority + self.epsilon) ** self.alpha
            self.max_priority = max(self.max_priority, priority)
            self.tree.update(idx, priority)

class DQNWithPER:
    """
    Deep Q-Network with Prioritized Experience Replay
    """

    def __init__(self, state_dim, action_dim,
                 learning_rate=0.001,
                 gamma=0.99,
                 per_alpha=0.6,
                 per_beta=0.4,
                 per_beta_increment=0.001,
                 replay_capacity=10000,
                 batch_size=64,
                 target_update_freq=100):
        """
        Initialize the DQN with PER agent

        Parameters:
        -----------
        state_dim : tuple
            Dimensions of the state space
        action_dim : int
            Dimension of the action space
        learning_rate : float
            Learning rate for the optimizer
        gamma : float
            Discount factor for future rewards
        per_alpha : float
            Controls how much prioritization is used
        per_beta : float
            Controls importance sampling weights
        per_beta_increment : float
            Amount to increase beta over time
        replay_capacity : int
            Capacity of the replay buffer
        batch_size : int
            Size of batches for training
        target_update_freq : int
            Frequency of target network updates
        """
        self.state_dim = state_dim
```

```
        self.action_dim = action_dim
        self.gamma = gamma
        self.batch_size = batch_size
        self.target_update_freq = target_update_freq
        self.step_counter = 0

        # Create replay buffer
        self.replay_buffer = PrioritizedReplayBuffer(
            capacity=replay_capacity,
            alpha=per_alpha,
            beta=per_beta,
            beta_increment=per_beta_increment
        )

        # Create Q networks
        self.q_network = self._build_q_network()
        self.target_q_network = self._build_q_network()

        # Use Mean Squared Error for loss
        self.optimizer = optimizers.Adam(learning_rate=learning_rate)

        # Initialize target network weights to match Q-network
        self._update_target_network()

    def _build_q_network(self):
        """
        Build the Q-network

        Returns:
        --------
        model : tf.keras.Model
            The Q-network model
        """
        inputs = layers.Input(shape=self.state_dim)

        x = layers.Conv2D(32, (8, 8), strides=(4, 4), activation='relu')(inputs)
        x = layers.Conv2D(64, (4, 4), strides=(2, 2), activation='relu')(x)
        x = layers.Conv2D(64, (3, 3), strides=(1, 1), activation='relu')(x)
        x = layers.Flatten()(x)
        x = layers.Dense(512, activation='relu')(x)
        outputs = layers.Dense(self.action_dim, activation='linear')(x)

        model = Model(inputs=inputs, outputs=outputs)
        return model

    def _update_target_network(self):
        """
        Update target network weights
        """
        self.target_q_network.set_weights(self.q_network.get_weights())

    def select_action(self, state, epsilon=0.1):
        """
        Select an action using epsilon-greedy policy
```

```
        Parameters:
        -----------
        state : np.ndarray
            Current state
        epsilon : float
            Exploration rate

        Returns:
        --------
        action : int
            Selected action
        """
        if random.random() < epsilon:
            # Explore: select a random action
            return random.randint(0, self.action_dim - 1)
        else:
            # Exploit: select the best action according to the Q-network
            state = np.expand_dims(state, axis=0)
            q_values = self.q_network.predict(state)[0]
            return np.argmax(q_values)

    def store_experience(self, state, action, reward, next_state, done):
        """
        Store an experience in the replay buffer

        Parameters:
        -----------
        state : np.ndarray
            Current state
        action : int
            Selected action
        reward : float
            Received reward
        next_state : np.ndarray
            Next state
        done : bool
            Whether the episode is done
        """
        experience = (state, action, reward, next_state, done)
        self.replay_buffer.add(experience)

    def train(self):
        """
        Train the Q-network

        Returns:
        --------
        loss : float
            Training loss
        """
        # Check if we have enough experiences
        if self.replay_buffer.tree.n_entries < self.batch_size:
            return 0
```

```
        # Sample a batch from the replay buffer
        batch, indices, weights = self.replay_buffer.sample(self.batch_size)

        # Unzip the batch
        states, actions, rewards, next_states, dones = zip(*batch)

        # Convert to numpy arrays
        states = np.array(states)
        next_states = np.array(next_states)
        actions = np.array(actions)
        rewards = np.array(rewards)
        dones = np.array(dones, dtype=np.float32)
        weights = np.array(weights)

        # Calculate target Q-values
        target_q_values = self.target_q_network.predict(next_states)
        max_target_q_values = np.max(target_q_values, axis=1)
        targets = rewards + (1 - dones) * self.gamma * max_target_q_values

        # Train the Q-network
        with tf.GradientTape() as tape:
            # Get the Q-values for the selected actions
            q_values = self.q_network(states, training=True)
            one_hot_actions = tf.one_hot(actions, self.action_dim)
            selected_q_values = tf.reduce_sum(q_values * one_hot_actions, axis=1)

            # Calculate TD errors for priority update
            td_errors = targets - selected_q_values

            # Calculate weighted loss
            losses = tf.square(td_errors) * weights
            loss = tf.reduce_mean(losses)

        # Get gradients and apply them
        grads = tape.gradient(loss, self.q_network.trainable_variables)
        self.optimizer.apply_gradients(zip(grads, self.q_network.trainable_variab

        # Update priorities in the replay buffer
        priorities = np.abs(td_errors.numpy())
        self.replay_buffer.update_priorities(indices, priorities)

        # Update target network periodically
        self.step_counter += 1
        if self.step_counter % self.target_update_freq == 0:
            self._update_target_network()

        return loss.numpy()
```

20.3.3 Results and Evaluation

PER demonstrated significant improvements over standard experience replay in reinforcement learning tasks:

1. **Faster learning**: Systems with PER converged to optimal policies in 50% fewer training steps on Atari games.
2. **Better performance**: Final performance was improved by approximately 20% across a range of reinforcement learning benchmarks.
3. **Improved exploration**: The prioritization of surprising experiences led to more effective exploration of the state space.

20.3.4 Neuroscience Connection

The implementation connects to hippocampal replay in several ways:

- **Memory prioritization**: Just as the hippocampus preferentially replays behaviorally relevant experiences, PER revisits experiences with high learning value.
- **Surprise-based learning**: The prioritization based on TD error parallels the brain's tendency to strengthen memories associated with unexpected outcomes.
- **Interleaved learning**: Both biological replay and PER address the stability-plasticity dilemma by interleaving experiences.

20.3.5 Limitations and Future Directions

Key challenges and future directions include:

1. **Efficient implementation**: The tree-based sampling structure introduces additional computational overhead.
2. **Parameter sensitivity**: Performance depends on appropriate settings for alpha and beta parameters.
3. **Future work**: Ongoing research is exploring integrating episodic memory structures and context-dependent replay strategies.

20.4 Case Study: Attention Mechanisms in Vision Transformers

20.4.1 Background and Motivation

Visual attention in humans allows for selective processing of relevant information while filtering out distractions. This case study examines how principles from visual neuroscience informed the development of Vision Transformers (ViT), which revolutionized computer vision by applying attention mechanisms to visual data.

20.4.2 Implementation: Vision Transformer

The Vision Transformer, developed by researchers at Google, applies the transformer architecture to image classification:

```python
import tensorflow as tf
from tensorflow.keras import layers, Model

class PatchExtractor(layers.Layer):
    """
    Extract patches from images
    """
    def __init__(self, patch_size):
        super(PatchExtractor, self).__init__()
        self.patch_size = patch_size

    def call(self, images):
        batch_size = tf.shape(images)[0]
        patches = tf.image.extract_patches(
            images=images,
            sizes=[1, self.patch_size, self.patch_size, 1],
            strides=[1, self.patch_size, self.patch_size, 1],
            rates=[1, 1, 1, 1],
            padding="VALID"
        )
        patch_dims = patches.shape[-1]
        patches = tf.reshape(patches, [batch_size, -1, patch_dims])
        return patches

class PositionalEmbedding(layers.Layer):
    """
    Add positional embeddings to patch embeddings
    """
    def __init__(self, num_patches, projection_dim):
        super(PositionalEmbedding, self).__init__()
        self.position_embedding = layers.Embedding(
            input_dim=num_patches + 1,  # +1 for the class token
            output_dim=projection_dim
        )

    def call(self, patch_embeddings, class_token):
        batch_size = tf.shape(patch_embeddings)[0]
        # Add class token to patch embeddings
        cls_tokens = tf.repeat(
            tf.expand_dims(class_token, 0), batch_size, axis=0
        )
        embeddings = tf.concat([cls_tokens, patch_embeddings], axis=1)

        # Add positional embeddings
        positions = tf.range(start=0, limit=embeddings.shape[1], delta=1)
        position_embeddings = self.position_embedding(positions)
        return embeddings + position_embeddings

class MultiHeadSelfAttention(layers.Layer):
    """
    Multi-head self-attention mechanism
    """
    def __init__(self, num_heads, projection_dim):
```

```python
        super(MultiHeadSelfAttention, self).__init__()
        self.num_heads = num_heads
        self.projection_dim = projection_dim
        self.head_dim = projection_dim // num_heads
        self.scale = self.head_dim ** -0.5

        self.query = layers.Dense(projection_dim)
        self.key = layers.Dense(projection_dim)
        self.value = layers.Dense(projection_dim)
        self.combine_heads = layers.Dense(projection_dim)

    def split_heads(self, x, batch_size):
        x = tf.reshape(
            x, (batch_size, -1, self.num_heads, self.head_dim)
        )
        return tf.transpose(x, perm=[0, 2, 1, 3])

    def call(self, inputs):
        batch_size = tf.shape(inputs)[0]

        # Linear projections
        query = self.query(inputs)
        key = self.key(inputs)
        value = self.value(inputs)

        # Split heads
        query = self.split_heads(query, batch_size)
        key = self.split_heads(key, batch_size)
        value = self.split_heads(value, batch_size)

        # Scaled dot-product attention
        attention_scores = tf.matmul(query, key, transpose_b=True) * self.scale
        attention_weights = tf.nn.softmax(attention_scores, axis=-1)

        # Apply attention to values
        context = tf.matmul(attention_weights, value)
        context = tf.transpose(context, perm=[0, 2, 1, 3])
        context = tf.reshape(context, [batch_size, -1, self.projection_dim])

        # Combine heads
        output = self.combine_heads(context)
        return output

class TransformerBlock(layers.Layer):
    """
    Transformer block with self-attention and MLP
    """
    def __init__(self, num_heads, projection_dim, mlp_dim, dropout=0.1):
        super(TransformerBlock, self).__init__()
        self.attention = MultiHeadSelfAttention(num_heads, projection_dim)
        self.mlp = tf.keras.Sequential([
            layers.Dense(mlp_dim, activation=tf.nn.gelu),
            layers.Dropout(dropout),
            layers.Dense(projection_dim),
```

```python
            layers.Dropout(dropout)
        ])
        self.layer_norm1 = layers.LayerNormalization(epsilon=1e-6)
        self.layer_norm2 = layers.LayerNormalization(epsilon=1e-6)
        self.dropout1 = layers.Dropout(dropout)
        self.dropout2 = layers.Dropout(dropout)

    def call(self, inputs, training):
        # Normalize and apply attention
        x = self.layer_norm1(inputs)
        attention_output = self.attention(x)
        attention_output = self.dropout1(attention_output, training=training)
        out1 = layers.add([inputs, attention_output])

        # Normalize and apply MLP
        x = self.layer_norm2(out1)
        mlp_output = self.mlp(x)
        mlp_output = self.dropout2(mlp_output, training=training)
        return layers.add([out1, mlp_output])

class VisionTransformer(Model):
    """
    Vision Transformer (ViT) model for image classification
    """
    def __init__(
        self,
        image_size=224,
        patch_size=16,
        num_layers=12,
        num_heads=12,
        projection_dim=768,
        mlp_dim=3072,
        num_classes=1000,
        dropout=0.1
    ):
        super(VisionTransformer, self).__init__()

        # Calculate number of patches
        num_patches = (image_size // patch_size) ** 2
        self.patch_size = patch_size

        # Patch extraction and projection
        self.patch_extractor = PatchExtractor(patch_size)
        self.projection = layers.Dense(projection_dim)

        # Class token
        self.class_token = tf.Variable(
            initial_value=tf.zeros([1, projection_dim]),
            trainable=True,
            name="class_token"
        )

        # Positional embedding
        self.position_embedding = PositionalEmbedding(
```

```
            num_patches, projection_dim
        )

        # Transformer blocks
        self.transformer_blocks = [
            TransformerBlock(num_heads, projection_dim, mlp_dim, dropout)
            for _ in range(num_layers)
        ]

        # Layer normalization and classifier
        self.layer_norm = layers.LayerNormalization(epsilon=1e-6)
        self.classifier = layers.Dense(num_classes)

    def call(self, inputs, training=False):
        # Extract patches from images
        patches = self.patch_extractor(inputs)

        # Project patches to embedding dimension
        patch_embeddings = self.projection(patches)

        # Add positional embeddings
        x = self.position_embedding(patch_embeddings, self.class_token)

        # Apply transformer blocks
        for transformer_block in self.transformer_blocks:
            x = transformer_block(x, training=training)

        # Layer normalization
        x = self.layer_norm(x)

        # Get class token output
        class_token_output = x[:, 0]

        # Classification
        return self.classifier(class_token_output)

def build_vit_model(
    image_size=224,
    patch_size=16,
    num_layers=12,
    num_heads=12,
    projection_dim=768,
    mlp_dim=3072,
    num_classes=1000
):
    """
    Build a Vision Transformer model

    Parameters:
    -----------
    image_size : int
        Size of input images (assuming square images)
    patch_size : int
        Size of image patches
```

```
    num_layers : int
        Number of transformer blocks
    num_heads : int
        Number of attention heads
    projection_dim : int
        Dimension of patch embeddings
    mlp_dim : int
        Hidden dimension in the MLP
    num_classes : int
        Number of output classes

    Returns:
    --------
    model : tf.keras.Model
        Vision Transformer model
    """
    inputs = layers.Input(shape=(image_size, image_size, 3))

    vit = VisionTransformer(
        image_size=image_size,
        patch_size=patch_size,
        num_layers=num_layers,
        num_heads=num_heads,
        projection_dim=projection_dim,
        mlp_dim=mlp_dim,
        num_classes=num_classes
    )

    outputs = vit(inputs)

    return Model(inputs=inputs, outputs=outputs)
```

20.4.3 Results and Evaluation

Vision Transformers have demonstrated impressive performance on computer vision tasks:

1. **Competitive accuracy**: When trained on sufficient data, ViT outperformed CNNs on image classification benchmarks like ImageNet.
2. **Data efficiency**: With pre-training on large datasets, ViTs showed better transfer learning efficiency than CNNs.
3. **Interpretability**: The attention maps provide visual explanations of which image regions contribute to decisions.

20.4.4 Neuroscience Connection

The ViT architecture connects to visual neuroscience in several ways:

- **Parallel processing**: Like the visual system, ViT processes multiple parts of the visual field in parallel.
- **Hierarchical integration**: The transformer layers build increasingly abstract representations similar to the visual cortex.
- **Attention allocation**: The self-attention mechanism parallels how humans selectively attend to parts of a scene.
- **Context integration**: ViT's ability to relate distant parts of an image mirrors how the visual system integrates across the visual field.

20.4.5 Limitations and Future Directions

Key challenges and future directions include:

1. **Computational efficiency**: ViTs typically require more computation than CNNs for similar performance at small scales.
2. **Data requirements**: ViTs need more data to achieve good results without pre-training.
3. **Future work**: Ongoing research is exploring hybrid architectures combining CNNs and transformers, and biologically-inspired attention constraints.

20.5 Case Study: Neural Data Analysis with Latent Variable Models

20.5.1 Background and Motivation

Analyzing high-dimensional neural data requires methods that can identify underlying patterns and structures. This case study examines how latent variable models influenced by neuroscience principles have been used to extract meaningful representations from neural recordings.

20.5.2 Implementation: Latent Factor Analysis via Dynamical Systems (LFADS)

LFADS, developed by researchers at Stanford and Google, uses recurrent neural networks to model neural population dynamics:

```
import tensorflow as tf
from tensorflow.keras import layers, Model
import numpy as np

class Encoder(layers.Layer):
    """
    Bidirectional RNN encoder for LFADS
    """
    def __init__(self, hidden_size=100, **kwargs):
        super(Encoder, self).__init__(**kwargs)
        self.hidden_size = hidden_size

        # Forward and backward RNNs
        self.forward_rnn = layers.GRU(
            hidden_size, return_sequences=True, return_state=True,
            name="encoder_forward_rnn"
        )
        self.backward_rnn = layers.GRU(
            hidden_size, return_sequences=True, return_state=True, go_backwards=T
            name="encoder_backward_rnn"
        )

    def call(self, inputs):
        # Forward pass
        forward_outputs, forward_state = self.forward_rnn(inputs)

        # Backward pass
        backward_outputs, backward_state = self.backward_rnn(inputs)
        backward_outputs = tf.reverse(backward_outputs, axis=[1])

        # Combine states
        encoder_state = tf.concat([forward_state, backward_state], axis=-1)

        return encoder_state

class LatentDistribution(layers.Layer):
    """
    Variational distribution for latent variables
    """
    def __init__(self, latent_dim=50, **kwargs):
        super(LatentDistribution, self).__init__(**kwargs)
        self.latent_dim = latent_dim

        # Dense layers for mean and logvar
        self.mean_layer = layers.Dense(latent_dim, name="mean_layer")
        self.logvar_layer = layers.Dense(latent_dim, name="logvar_layer")

    def call(self, inputs, training=False):
        # Compute mean and logvar
        mean = self.mean_layer(inputs)
        logvar = self.logvar_layer(inputs)

        # If training, sample from the distribution
```

```python
        if training:
            epsilon = tf.random.normal(shape=tf.shape(mean))
            sample = mean + tf.exp(0.5 * logvar) * epsilon
        else:
            sample = mean

        return sample, mean, logvar

class Controller(layers.Layer):
    """
    Controller RNN for generating inputs to the generator
    """
    def __init__(self, hidden_size=100, **kwargs):
        super(Controller, self).__init__(**kwargs)
        self.hidden_size = hidden_size

        # Controller RNN
        self.rnn = layers.GRU(
            hidden_size, return_sequences=True, return_state=True,
            name="controller_rnn"
        )

        # Dense layer for controller outputs
        self.dense = layers.Dense(hidden_size, name="controller_output")

    def call(self, initial_state, sequence_length):
        # Create dummy input tensor
        batch_size = tf.shape(initial_state)[0]
        dummy_input = tf.zeros([batch_size, sequence_length, 1])

        # Initialize RNN state
        state = initial_state

        # Run RNN and get outputs
        outputs, _ = self.rnn(dummy_input, initial_state=state)

        # Apply dense layer to outputs
        controller_outputs = self.dense(outputs)

        return controller_outputs

class Generator(layers.Layer):
    """
    Generator RNN for modeling neural dynamics
    """
    def __init__(self, hidden_size=100, factors_dim=50, output_dim=100, **kwargs)
        super(Generator, self).__init__(**kwargs)
        self.hidden_size = hidden_size
        self.factors_dim = factors_dim
        self.output_dim = output_dim

        # Generator RNN
        self.rnn = layers.GRU(
            hidden_size, return_sequences=True, return_state=True,
```

```
            name="generator_rnn"
        )

        # Dense layers for outputs
        self.factors_layer = layers.Dense(factors_dim, name="factors_layer")
        self.rates_layer = layers.Dense(output_dim, activation=tf.nn.softplus, na

    def call(self, initial_state, controller_outputs):
        # Run RNN with controller outputs as inputs
        outputs, _ = self.rnn(controller_outputs, initial_state=initial_state)

        # Generate factors (latent neural dynamics)
        factors = self.factors_layer(outputs)

        # Generate rates (expected neural firing rates)
        rates = self.rates_layer(factors)

        return rates, factors

class LFADS(Model):
    """
    Latent Factor Analysis via Dynamical Systems (LFADS) model
    """
    def __init__(
        self,
        encoder_dim=100,
        latent_dim=50,
        controller_dim=100,
        generator_dim=100,
        factors_dim=50,
        **kwargs
    ):
        super(LFADS, self).__init__(**kwargs)

        # Model components
        self.encoder = Encoder(hidden_size=encoder_dim)
        self.latent_distribution = LatentDistribution(latent_dim=latent_dim)
        self.controller = Controller(hidden_size=controller_dim)
        self.generator_initial_dense = layers.Dense(generator_dim, activation="ta

    def build(self, input_shape):
        # Build the generator once we know the output dimension
        self.generator = Generator(
            hidden_size=self.generator_initial_dense.units,
            factors_dim=50,
            output_dim=input_shape[-1]
        )

        super(LFADS, self).build(input_shape)

    def call(self, inputs, training=False):
        # Get sequence length
        sequence_length = tf.shape(inputs)[1]
```

```python
        # Encode input
        encoder_state = self.encoder(inputs)

        # Sample from latent distribution
        latent_sample, mean, logvar = self.latent_distribution(encoder_state, tra

        # Generate controller outputs
        controller_outputs = self.controller(latent_sample, sequence_length)

        # Generate initial state for generator
        generator_initial_state = self.generator_initial_dense(latent_sample)

        # Generate neural rates and factors
        rates, factors = self.generator(generator_initial_state, controller_outpu

        # Create model outputs dictionary
        outputs = {
            "rates": rates,
            "factors": factors,
            "latent_mean": mean,
            "latent_logvar": logvar
        }

        return outputs

    def compute_loss(self, x, training=False):
        """
        Compute LFADS loss function

        Parameters:
        -----------
        x : tf.Tensor
            Input spike data
        training : bool
            Whether model is in training mode

        Returns:
        --------
        total_loss : tf.Tensor
            Combined loss
        reconstruction_loss : tf.Tensor
            Poisson reconstruction loss
        kl_loss : tf.Tensor
            KL divergence loss
        """
        # Get model outputs
        outputs = self(x, training=training)
        rates = outputs["rates"]
        latent_mean = outputs["latent_mean"]
        latent_logvar = outputs["latent_logvar"]

        # Compute Poisson reconstruction loss
        # log p(x|z) = sum_t sum_i (x_i,t * log(r_i,t) - r_i,t - log(x_i,t!))
        # We drop the factorial term as it's constant with respect to the paramet
```

```python
        reconstruction_loss = tf.reduce_sum(
            rates - x * tf.math.log(rates + 1e-8),
            axis=[1, 2]
        )
        reconstruction_loss = tf.reduce_mean(reconstruction_loss)

        # Compute KL divergence loss
        # KL(q(z|x) || p(z)) = 0.5 * sum_j (1 + log(sigma_j^2) - mu_j^2 - sigma_j
        kl_loss = -0.5 * tf.reduce_sum(
            1 + latent_logvar - tf.square(latent_mean) - tf.exp(latent_logvar),
            axis=1
        )
        kl_loss = tf.reduce_mean(kl_loss)

        # Combine losses
        # Optionally add weight for KL term (beta-VAE style)
        kl_weight = 1.0
        total_loss = reconstruction_loss + kl_weight * kl_loss

        return total_loss, reconstruction_loss, kl_loss

def build_lfads_model(
    input_shape,
    encoder_dim=100,
    latent_dim=50,
    controller_dim=100,
    generator_dim=100,
    factors_dim=50
):
    """
    Build an LFADS model

    Parameters:
    -----------
    input_shape : tuple
        Shape of input data (sequence_length, num_neurons)
    encoder_dim : int
        Hidden dimension of encoder RNN
    latent_dim : int
        Dimension of latent variables
    controller_dim : int
        Hidden dimension of controller RNN
    generator_dim : int
        Hidden dimension of generator RNN
    factors_dim : int
        Dimension of latent factors

    Returns:
    --------
    model : LFADS
        LFADS model
    """
    # Create LFADS model
    model = LFADS(
```

```
        encoder_dim=encoder_dim,
        latent_dim=latent_dim,
        controller_dim=controller_dim,
        generator_dim=generator_dim,
        factors_dim=factors_dim
    )

    # Build the model with sample input
    sample_input = tf.zeros((1,) + input_shape)
    model(sample_input)

    return model
```

20.5.3 Results and Evaluation

LFADS has demonstrated several benefits in analyzing neural data:

1. **Improved decoding**: Using LFADS-inferred latent factors improved neural decoding accuracy by 40% compared to raw neural data.
2. **Single-trial analysis**: By inferring the underlying dynamics from noisy spike trains, LFADS enables meaningful analysis of individual trials rather than requiring trial averaging.
3. **Identification of dynamics**: LFADS successfully recovered the underlying dynamical structure in both simulated and real neural populations.

20.5.4 Neuroscience Connection

The LFADS model connects to neuroscience theories in several ways:

- **Low-dimensional dynamics**: LFADS is built on the neuroscience insight that high-dimensional neural activity often reflects low-dimensional latent dynamics.
- **Temporal constraints**: The recurrent generator mirrors the continuous-time dynamics of neural circuits.
- **Initial condition encoding**: The model's focus on initial state mirrors theories about how neural trajectories are initialized based on sensory inputs.

20.5.5 Limitations and Future Directions

Key challenges and future directions include:

1. **Model complexity**: The full LFADS model is computationally intensive to train.
2. **Interpretability**: The biological meaning of extracted latent factors requires careful interpretation.
3. **Future work**: Ongoing research is exploring extensions to multi-area recordings and incorporating more detailed biophysical constraints.

20.6 Lessons from Successful NeuroAI Integration

Across these case studies, several patterns emerge that highlight successful strategies for integrating neuroscience and AI:

20.6.1 Common Patterns of Success

1. **Focus on computational principles**: Successful NeuroAI implementations focus on computational principles rather than precise biological details.
2. **Iterative refinement**: The most successful projects involved multiple iterations between neuroscience insights and AI implementations.
3. **Cross-disciplinary teams**: Projects typically involved researchers with expertise in both neuroscience and AI working closely together.
4. **Translation flexibility**: Successful implementations allowed for flexible translation of neuroscience principles to match the constraints of deep learning architectures.

20.6.2 Practical Implementation Strategies

Based on these case studies, several practical strategies emerge:

```python
def neuroai_implementation_framework(neuroscience_principle, existing_ai_system):
    """
    A framework for implementing neuroscience principles in AI systems

    Parameters:
    -----------
    neuroscience_principle : dict
        Description of the neuroscience principle to implement
    existing_ai_system : object
        The AI system to enhance

    Returns:
    --------
    enhanced_system : object
        The enhanced AI system
    """
    # Step 1: Extract the computational essence of the neuroscience principle
    computational_essence = extract_computational_essence(neuroscience_principle)

    # Step 2: Analyze compatibility with existing AI system
    compatibility_analysis = analyze_compatibility(computational_essence, existin

    # Step 3: Implement a minimal version to test the principle
    prototype = implement_minimal_version(computational_essence, existing_ai_syst

    # Step 4: Evaluate and iterate
    evaluation_results = evaluate_prototype(prototype)
    enhanced_system = iterative_refinement(prototype, evaluation_results)

    # Step 5: Scale up implementation
    enhanced_system = scale_implementation(enhanced_system)

    return enhanced_system

def extract_computational_essence(neuroscience_principle):
    """
    Extract the core computational principle from neuroscience findings
    """
    # Focus on functional aspects, not biological implementation
    # Identify the information processing role
    # Abstract away biological details
    # Identify the computational advantage
    pass

def analyze_compatibility(computational_essence, existing_ai_system):
    """
    Analyze how compatible the principle is with existing AI
    """
    # Identify integration points
    # Assess computational overhead
    # Determine architectural modifications needed
    # Evaluate training implications
    pass
```

```python
def implement_minimal_version(computational_essence, existing_ai_system):
    """
    Implement a minimal version to test the principle
    """
    # Focus on core functionality
    # Implement the simplest version that could work
    # Ensure measurable outcomes
    # Document assumptions and simplifications
    pass

def evaluate_prototype(prototype):
    """
    Evaluate the prototype against baselines
    """
    # Compare to baseline
    # Test on simplified tasks
    # Analyze failure modes
    # Identity promising directions
    pass

def iterative_refinement(prototype, evaluation_results):
    """
    Refine implementation based on evaluation
    """
    # Address failure modes
    # Optimize computational efficiency
    # Reduce complexity where possible
    # Enhance successful components
    pass

def scale_implementation(enhanced_system):
    """
    Scale up implementation for real-world use
    """
    # Optimize for computational efficiency
    # Address edge cases
    # Add necessary complexity for general use
    # Document implementation details
    pass
```

20.6.3 Interdisciplinary Collaboration Best Practices

The case studies highlight the importance of effective collaboration between neuroscientists and AI researchers:

1. **Establish shared vocabulary**: Develop a common language that bridges neuroscience and AI concepts.

2. **Focus on translatable insights**: Prioritize neuroscience findings with clear computational implications.
3. **Prototype and iterate**: Build small-scale prototypes to test neuroscience concepts before large-scale implementation.
4. **Mutual education**: Invest time in cross-disciplinary education to ensure deep understanding of both fields.

20.7 Practical Exercise: Implementing a Neuroscience-Inspired AI Component

This exercise guides you through implementing a simplified hippocampal-inspired memory system for reinforcement learning:

```python
import numpy as np
from collections import deque
import random

class EpisodicMemoryBuffer:
    """
    A simple episodic memory buffer inspired by hippocampal function
    """

    def __init__(self, capacity=1000, similarity_threshold=0.8):
        """
        Initialize the episodic memory buffer

        Parameters:
        -----------
        capacity : int
            Maximum number of episodes to store
        similarity_threshold : float
            Threshold for determining similar experiences
        """
        self.buffer = deque(maxlen=capacity)
        self.similarity_threshold = similarity_threshold

    def add_experience(self, state, action, reward, next_state, done):
        """
        Add an experience to the buffer

        Parameters:
        -----------
        state : np.ndarray
            Current state
        action : int
            Action taken
        reward : float
            Reward received
        next_state : np.ndarray
            Next state
        done : bool
            Whether the episode is done
        """
        experience = (state, action, reward, next_state, done)
        self.buffer.append(experience)

    def find_similar_experiences(self, query_state, k=5):
        """
        Find experiences with similar states

        Parameters:
        -----------
        query_state : np.ndarray
            State to compare against
        k : int
            Number of similar experiences to retrieve
```

```
        Returns:
        --------
        similar_experiences : list
            List of similar experiences
        """
        similarities = []

        for experience in self.buffer:
            state = experience[0]
            # Compute cosine similarity
            similarity = np.dot(query_state, state) / (np.linalg.norm(query_state
            similarities.append((similarity, experience))

        # Sort by similarity
        similarities.sort(reverse=True, key=lambda x: x[0])

        # Filter by threshold and get top k
        similar_experiences = [exp for sim, exp in similarities if sim >= self.si

        return similar_experiences

    def sample_batch(self, batch_size=32, include_similar=True, query_state=None)
        """
        Sample a batch of experiences

        Parameters:
        -----------
        batch_size : int
            Size of the batch to sample
        include_similar : bool
            Whether to include similar experiences
        query_state : np.ndarray or None
            State to find similar experiences for

        Returns:
        --------
        batch : list
            Sampled batch of experiences
        """
        # Regular random sampling
        if len(self.buffer) <= batch_size:
            return list(self.buffer)

        # Regular random batch
        random_batch = random.sample(self.buffer, batch_size - 5 if include_simil

        if include_similar and query_state is not None:
            # Find similar experiences
            similar_experiences = self.find_similar_experiences(query_state, k=5)

            # Combine random and similar experiences
            combined_batch = random_batch + similar_experiences
```

```python
            return combined_batch

        return random_batch

class EpisodicReinforcementLearningAgent:
    """
    A reinforcement learning agent with episodic memory
    """

    def __init__(self, state_dim, action_dim, learning_rate=0.001, gamma=0.99):
        """
        Initialize the agent

        Parameters:
        -----------
        state_dim : int
            Dimension of the state space
        action_dim : int
            Dimension of the action space
        learning_rate : float
            Learning rate for the model
        gamma : float
            Discount factor
        """
        self.state_dim = state_dim
        self.action_dim = action_dim
        self.learning_rate = learning_rate
        self.gamma = gamma

        # Create episodic memory
        self.episodic_memory = EpisodicMemoryBuffer()

        # Simple Q-table for this example
        self.q_table = np.zeros((state_dim, action_dim))

    def select_action(self, state, epsilon=0.1):
        """
        Select an action using epsilon-greedy policy with episodic memory

        Parameters:
        -----------
        state : np.ndarray
            Current state
        epsilon : float
            Exploration rate

        Returns:
        --------
        action : int
            Selected action
        """
        if random.random() < epsilon:
            # Random exploration
            return random.randint(0, self.action_dim - 1)
```

```python
        else:
            # Check episodic memory for similar states
            similar_experiences = self.episodic_memory.find_similar_experiences(s

            if similar_experiences and random.random() < 0.3:  # 30% chance to us
                # Use action from a similar experience with high reward
                similar_experiences.sort(key=lambda x: x[2], reverse=True)  # Sor
                return similar_experiences[0][1]  # Return action from highest-re
            else:
                # Use Q-table
                return np.argmax(self.q_table[self.discretize_state(state)])

    def discretize_state(self, state):
        """
        Discretize continuous state (simplification for this example)
        """
        # This is a placeholder; in a real implementation,
        # you would properly discretize the state space
        return int(sum(state) * 10) % self.state_dim

    def store_experience(self, state, action, reward, next_state, done):
        """
        Store experience in episodic memory
        """
        self.episodic_memory.add_experience(state, action, reward, next_state, do

    def learn(self, state, action, reward, next_state, done):
        """
        Update Q-table based on experience
        """
        # Discretize states for Q-table
        state_idx = self.discretize_state(state)
        next_state_idx = self.discretize_state(next_state)

        # Q-learning update
        best_next_action = np.argmax(self.q_table[next_state_idx])
        td_target = reward + (1 - done) * self.gamma * self.q_table[next_state_id
        td_error = td_target - self.q_table[state_idx, action]
        self.q_table[state_idx, action] += self.learning_rate * td_error

        # Store experience in episodic memory
        self.store_experience(state, action, reward, next_state, done)

    def train_from_episodic_memory(self, batch_size=32):
        """
        Train using experiences from episodic memory
        """
        # Sample batch from episodic memory
        batch = self.episodic_memory.sample_batch(batch_size)

        # Learn from each experience
        for state, action, reward, next_state, done in batch:
            self.learn(state, action, reward, next_state, done)
```

```python
# Example usage
def run_episodic_memory_example():
    """
    Run a simple example of episodic memory in reinforcement learning
    """
    # Create environment (simplified for this example)
    state_dim = 100
    action_dim = 4

    # Create agent
    agent = EpisodicReinforcementLearningAgent(state_dim, action_dim)

    # Run episodes
    num_episodes = 100
    max_steps = 200

    for episode in range(num_episodes):
        # Reset environment
        state = np.random.rand(10)  # 10-dimensional state
        total_reward = 0

        for step in range(max_steps):
            # Select action
            action = agent.select_action(state)

            # Take action (simplified environment dynamics)
            next_state = state + 0.1 * np.random.randn(10)
            next_state = np.clip(next_state, 0, 1)

            # Get reward (simplified)
            reward = 1.0 if np.sum(next_state) > np.sum(state) else -0.1
            done = step == max_steps - 1 or np.sum(next_state) >= 9.0

            # Learn from experience
            agent.learn(state, action, reward, next_state, done)

            # Update state and total reward
            state = next_state
            total_reward += reward

            if done:
                break

        # Train from episodic memory
        agent.train_from_episodic_memory()

        # Print progress
        if episode % 10 == 0:
            print(f"Episode {episode}, Total Reward: {total_reward:.2f}")

if __name__ == "__main__":
    run_episodic_memory_example()
```

20.8 Chapter Take-aways

- Successful NeuroAI implementations focus on computational principles rather than precise biological details
- The most effective implementations involve iterative refinement between neuroscience insights and AI implementations
- Key areas where neuroscience has informed AI include attention mechanisms, memory systems, predictive processing, and neural data analysis
- Effective cross-disciplinary collaboration requires establishing shared vocabulary and mutual education
- Implementing neuroscience principles in AI often requires creative adaptations to match the constraints of current deep learning frameworks
- The most successful projects demonstrate measurable improvements in performance, generalization, or sample efficiency

20.9 Interactive Materials and Exercises

To deepen your understanding of the case studies presented in this chapter, we've created several interactive examples and exercises. These materials allow you to explore key concepts through hands-on experimentation.

> **Tip**
>
> Access the interactive notebook to experiment with:
>
> 1. **PredNet Visualization**: Adjust parameters to see how predictive coding works in practice
> 2. **Prioritized Experience Replay**: Compare standard and prioritized replay in reinforcement learning
> 3. **Vision Transformer Attention**: Visualize attention mechanisms on different image patches
> 4. **Interactive Glossary**: Explore definitions with popup explanations of neural-AI connections

The interactive examples include sliders to adjust parameters, visualizations that update in real-time, and explanatory annotations to help you connect theoretical concepts with their practical implementations.

AI-Assisted Learning

We've also integrated Jupyter AI to enhance your learning experience. With Jupyter AI, you can:

> **Tip**
>
> Explore the AI-Assisted Learning notebook to:
>
> 1. **Generate Code**: Get implementation help for neuroscience-inspired AI models
> 2. **Receive Explanations**: Ask for clarification on complex concepts
> 3. **Debug Implementations**: Fix and improve your code
> 4. **Create Visualizations**: Generate custom visualizations for neural data

This integration of AI assistance allows for a more dynamic, personalized learning experience that adapts to your specific interests and questions about the case studies.

Presentation Materials

For educators and presenters, we've created a guide to developing slide presentations from the handbook content:

> **Tip**
>
> Check out our RISE presentation guide to learn how to:
>
> 1. **Create Interactive Slides**: Transform notebook content into polished presentations
> 2. **Execute Live Code**: Run code demonstrations during presentations
> 3. **Add Interactive Elements**: Include widgets and visualizations in slides
> 4. **Customize Styling**: Adjust themes and transitions for your audience

RISE (Reveal.js - Jupyter/IPython Slideshow Extension) allows you to create engaging presentations directly from Jupyter notebooks, perfect for teaching the concepts covered in this chapter.

20.10 Further Reading

- Hassabis, D., Kumaran, D., Summerfield, C., & Botvinick, M. (2017). Neuroscience-inspired artificial intelligence. *Neuron, 95*(2), 245-258.
- Kriegeskorte, N., & Douglas, P. K. (2018). Cognitive computational neuroscience. *Nature Neuroscience, 21*(9), 1148-1160.
- Zador, A. M. (2019). A critique of pure learning and what artificial neural networks can learn from animal brains. *Nature Communications, 10*(1), 1-7.
- Richards, B. A., et al. (2019). A deep learning framework for neuroscience. *Nature Neuroscience, 22*(11), 1761-1770.
- Marblestone, A. H., Wayne, G., & Kording, K. P. (2016). Toward an integration of deep learning and neuroscience. *Frontiers in Computational Neuroscience, 10*, 94.
- Botvinick, M., et al. (2020). Deep reinforcement learning and its neuroscientific implications. *Neuron, 107*(4), 603-616.