What is Learning?

What is Learning ?

- Learning is the process of converting experience into expertise or knowledge.
- Training data represents experience.
- The output of learning is expertise, often in the form of a computer program.
- Key questions in learning:
  - What is the training data?
  - How can learning be automated?
  - How do we evaluate learning success?

What is Reasoning?

- Reasoning is the ability to draw logical conclusions from known facts or learned knowledge.
- It does not require large amounts of data but rather depends on logical inference.

Example: Animal Learning

Bait Shyness in Rats

- Rats sample novel food cautiously.
- If the food causes illness, they avoid it in the future.
- Past experience informs future decisions.

Example: Machine Learning Task

Spam Email Filtering

- Naive approach: Memorization of past spam emails.
- Limitation: Cannot classify unseen emails.
- Solution: **Generalization using inductive reasoning**.
- Like rats apply their attitude to new unseen examples of food with similar smell and taste
- Extract patterns (e.g., words indicating spam) to classify new emails.

Types of Reasoning in AI and ML

1. **Inductive Reasoning (Primary)**
   - Extracts patterns from observed data to make predictions.
   - Used in deep learning and LLMs.
   - **Example:** A spam classifier learns from previously labeled emails and generalizes patterns to detect new spam messages.
2. **Abductive Reasoning (Inference to Best Explanation)**
   - Guesses the most probable explanation given incomplete data.
   - Used in medical diagnosis and troubleshooting AI.
   - **Example:** A doctor observes symptoms like fever and cough and infers that the patient likely has the flu, even without a lab test.
3. **Analogical Reasoning (Pattern Transfer)**
   - Applies knowledge from one context to another.
   - Used in AI-powered tutoring and cross-domain learning.
   - **Example:** AI that learns human speech patterns in English and transfers that learning to generate speech in another language.

Probabilistic and Other Reasoning Types (Part 1)

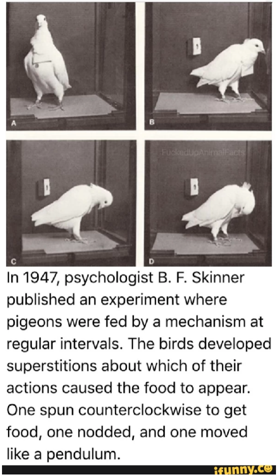
4. **Bayesian Reasoning (Probabilistic Prediction)**
   - Uses probability to predict outcomes.
   - Used in spam filtering and AI language models.
   - **Example:** A Bayesian spam filter assigns probabilities to words appearing in spam emails and calculates the likelihood that a new email is spam.
5. **Deductive Reasoning (Not Used in LLMs)**
   - Moves from general rules to specific conclusions.
   - LLMs lack strict logical reasoning.
   - **Example:** In mathematics, if all squares have four sides and a shape is a square, it must have four sides.

Probabilistic and Other Reasoning Types (Part 2)

6. **Causal Reasoning (Understanding Cause-and-Effect)**
   - Determines causal relationships rather than correlations.
   - LLMs struggle with causality.
   - **Example:** In healthcare, researchers identify that smoking causes lung cancer, rather than just observing that smokers have higher cancer rates.
7. **Counterfactual Reasoning (What-If Thinking)**
   - Explores hypothetical scenarios.
   - Used in risk analysis and AI decision-making.
   - **Example:** A self-driving car AI simulates different driving scenarios to decide the safest course of action in an emergency.

Limitations of Inductive Reasoning

Pigeon Superstition Experiment (B.F. Skinner)



Bait Shyness Revisited: Garcia & Koelling Experiment (1966)

- The experiment was conducted by John Garcia and Robert Koelling to study **selective associative learning** in rats.
- It demonstrated that **not all stimuli are equally associated** with consequences, challenging traditional learning theories.
- The experiment used a **compound stimulus** consisting of:
  - A **taste cue** (saccharin-flavored water).
  - **Audiovisual cues** (lights and sounds that played while drinking).
- After exposure to the stimuli, the rats were subjected to an **aversive event**:
  - **Group 1: Induced Illness** (nausea caused by mild radiation or a toxin).
  - **Group 2: Electric Shock** (mild foot shocks upon drinking).

Key Findings of Garcia & Koelling Experiment

- **Illness-Induced Group:**
  - Developed a **strong aversion** to the **taste cue** (saccharin water).
  - Showed little to no aversion to **audiovisual cues** (lights and sounds).
- **Shock-Induced Group:**
  - Developed a **strong aversion** to **audiovisual cues** (lights and sounds).
  - Showed no aversion to **taste cue** (saccharin water).

Key Findings of Garcia & Koelling Experiment

- This demonstrated that rats were **biologically predisposed to associate taste with internal discomfort (nausea)** and **sounds/lights with external discomfort (shock)**.

Key Findings of Garcia & Koelling Experiment

- Bait shyness learning in rats incorporates prior knowledge, known as inductive bias.
- Rats are biased towards detecting patterns and ignoring other temporal correlations.
- Pigeons, however, adopt any explanation for the occurrence of food.

Scientific Implications of the Experiment

- The experiment **challenged the traditional idea of equipotentiality**, which assumed that any stimulus can be associated with any consequence.
- It provided evidence that **evolution influences learning mechanisms**.
- Animals (including humans) have **prewired biases** for learning certain associations based on survival advantages.
- **Biological Constraints on Learning**:
  - Taste is more likely to be linked with food poisoning.
  - Sounds or visual cues are more likely to be linked with external dangers (e.g., predators, electric shocks).

Key Takeaways from the Garcia & Koelling (1966) Experiment

- Learning is not arbitrary – it requires **inductive bias**.
- Not all features are equally useful for learning.
- Correlation does not imply causation.

Key Takeaways from the Garcia & Koelling (1966) Experiment

- Evolutionary constraints shape learning – **domain knowledge matters**.
- The No-Free-Lunch theorem – **There's no universal learner**.
- The role of evolutionary and pre-trained knowledge in AI.

What is Inductive Bias?

Definition:

- Inductive bias refers to the set of assumptions that a learning algorithm makes to generalize from limited training data to unseen data.
- Since real-world data is often incomplete or noisy, a model must rely on prior knowledge or constraints to make meaningful predictions.

Why is Inductive Bias Important?

- Machine learning models do not have infinite training data.
- They must generalize from past observations to unseen cases.
- Without a bias, a model may overfit (memorize data without true learning).

Types of Inductive Biases

1. **Preference for Simpler Models (Occam's Razor)**
   - **Assumption:** Simpler explanations are preferred over complex ones.
   - **Example:** Decision trees with fewer splits are preferred because they generalize better.
   - **In Deep Learning:** Regularization techniques (L1, L2) penalize complex models to enforce simplicity.

Types of Inductive Biases

2. Smoothness Assumption

- **Assumption:** Data points that are close together should have similar outputs.
- **Example:** In image classification, two similar images should belong to the same class.
- **In ML:** K-Nearest Neighbors (KNN) assumes that nearby data points have the same label.

Types of Inductive Biases

3. **Similar Features Should Have Similar Effects**
   - **Assumption:** If two features (e.g., temperature & humidity) are related, their effects should be similar.
   - **Example:** In linear regression, correlated features often have similar coefficients.
   - **In ML:** In decision trees and feature selection algorithms, correlated features (e.g., "annual income" & "monthly salary") often contribute similarly to predictions.

Types of Inductive Biases

4. **Prior Knowledge About the Task (Domain-Specific Bias)**
   - **Assumption:** Certain relationships are more likely in specific tasks.
   - **Example:** In natural language processing (NLP), word order matters (e.g., "The cat sat on the mat" is not equal to "The mat sat on the cat").
   - **In ML:** Transformers (e.g., BERT, GPT) use positional embeddings to capture sentence structure.

Types of Inductive Biases

5. **Invariance Bias (Translation, Rotation, Scale Invariance)**
   - **Assumption:** Some transformations should not change predictions.
   - **Example:** In computer vision, rotating an image of a cat should still classify it as a cat.
   - **In ML:** CNNs use convolutional filters to enforce translation invariance.

Types of Inductive Biases

6. Sparsity Assumption

- **Assumption:** Only a few features are truly important.
- **Example:** In text classification, most words are irrelevant, and only a few indicate spam.
- **In ML:** L1 regularization forces models to select only the most important features.

Inductive Bias in Machine Learning Models

Example 1: Human Learning vs. ML

- **Humans:** When a child learns what a "dog" is, they assume all dogs have four legs, fur, and bark. This is an inductive bias—it helps the child generalize.
- **ML Model:** If trained on 100 pictures of dogs, an image classifier may learn that four legs + fur are key features of a dog. However, without bias, it may fail when seeing a three-legged dog.

Inductive Bias in Convolutional Neural Networks (CNNs)

CNNs are designed for image processing and rely on three key inductive biases:

1.1 Locality Bias (Local Connectivity)

- **Assumption:** Nearby pixels in an image are more relevant to each other than distant pixels.
- **Why it helps?** Instead of treating each pixel separately, CNNs focus on small local features (edges, textures, shapes) before forming high-level representations.
- **Example:** In facial recognition, a CNN first detects eyes, nose, and mouth individually before recognizing an entire face.

Inductive Bias in Convolutional Neural Networks (CNNs)

1.2 Translation Invariance

- **Assumption:** An object in an image should be recognized regardless of its position.
- **Why it helps?** A cat in the top-left corner should be classified as a cat, just as a cat in the bottom-right corner.
- **How it works?** CNNs use shared convolutional filters to detect patterns anywhere in the image.
- **Example:** A handwritten digit "3" should be recognized no matter where it appears in the image.

Inductive Bias in Convolutional Neural Networks (CNNs)

1.3 Hierarchical Feature Learning

- **Assumption:** Complex patterns can be learned by stacking multiple layers of abstraction.
- **Why it helps?** Lower layers detect edges, middle layers detect shapes, and deeper layers detect objects.
- **Example:** A CNN learns small textures $\rightarrow$ eyes/mouth $\rightarrow$ full face in image classification.

Takeaway CNNs perform better than traditional dense networks for image tasks because of these biases.

Inductive Bias in Recurrent Neural Networks (RNNs & LSTMs)

RNNs are designed for sequential data (e.g., speech, text, time-series forecasting) and rely on two main biases:
2.1 Temporal Dependency Bias

- **Assumption:** Recent information is more important than distant past information.
- **Why it helps?** Words appearing closer together in a sentence are more related.
- **Example:**
- In "The cat sat on the mat", the word "mat" is a likely prediction.
- In a time-series forecast, the last few observations are more important.

Inductive Bias in Recurrent Neural Networks (RNNs & LSTMs)

2.2 Order Sensitivity Bias

- **Assumption:** The order of input elements matters.
- **Why it helps?** "Dog bites man" is not equal to "Man bites dog."
- **Example:** Machine translation relies on word ordering to make correct translations.

Takeaway: RNNs/LSTMs work well for time-dependent problems because they incorporate sequential learning bias.

Inductive Bias in Transformers (e.g., BERT, GPT)

Transformers, which power GPT, BERT, and other large language models (LLMs), incorporate several modern inductive biases that make them highly generalizable:

Attention-Based Bias (Self-Attention)

- **Assumption:** Important words in a sentence can be anywhere, not just nearby.
- **Why it helps?** Instead of only relying on adjacent words, transformers attend to relevant words across an entire sentence.
- **Example:**
- In "The dog chased the ball across the field, past the trees, over the hill, which was blue." the word "which" refers to "ball", even though they are not next to each other.

Inductive Bias in Transformers (e.g., BERT, GPT)

Context-Aware Learning Bias

- **Assumption:** The meaning of a word depends on its context.
- **Why it helps?** The word "bank" can mean a financial institution or a riverbank, and transformers can disambiguate meaning based on context.
- **Example:**
- "I deposited money in the bank" $\rightarrow$ Financial institution
- "The boat was near the river bank" $\rightarrow$ River side

Inductive Bias in Transformers (e.g., BERT, GPT)

3.3 Positional Encoding Bias

- **Assumption:** Order still matters, even without RNNs.
- **Why it helps?** Since transformers do not process words sequentially, they use positional encodings to maintain the order of words.
- **Example:**
- "She ate an apple" $\neq$ "An apple ate she."

Takeaway: Transformers use self-attention, context-awareness, and positional encoding biases, making them more flexible than CNNs and RNNs.

Role of Prior Knowledge in Machine Learning

- Development of tools for expressing domain expertise, translating it into learning bias and quantifying the effect of such a bias on the success of learning is a central theme of the theory of machine learning
- Stronger prior knowledge makes learning from examples easier.
- However, stronger prior assumptions reduce the flexibility of learning.
- The challenge is to find the right balance between prior knowledge and data-driven learning.

Key Takeaways

- Learning involves converting experience into generalizable knowledge.
- Different reasoning types play a role in AI and ML
- Inductive bias is crucial for generalization and learning efficiency.
- Deep learning relies primarily on inductive reasoning.
- LLMs approximate abductive and Bayesian reasoning but lack causal and deductive reasoning.
- Future AI models may incorporate stronger causal and counterfactual reasoning capabilities.

Why Machine Learning

- For many problems, it's difficult to program the correct behavior by hand.
  - E.g., recognizing objects in images, understanding human speech.
- Machine learning approach: program an algorithm to automatically learn from data or experience.
- Reasons to use a learning algorithm:
  - Hard to code up a solution by hand (e.g., vision, speech).
  - System needs to adapt to a changing environment (e.g., spam detection).
  - Want the system to perform better than the human programmers.
  - Privacy/fairness (e.g., ranking search results).

Machine Learning vs Statistics

- Both fields try to uncover patterns in data.
- Both draw heavily on calculus, probability, and linear algebra, and share many of the same core algorithms.
- However, they have different focuses:
  - Statistics: Helping scientists and policymakers draw good conclusions.
  - Machine Learning: Building autonomous agents.
- Emphasis differences:
  - Statistics: Interpretability and mathematical rigor.
  - Machine Learning: Predictive performance, scalability, and autonomy.

Relations to AI

- Nowadays, “machine learning” is often brought up with “artificial intelligence” (AI).
- AI does not always imply a learning-based system:
  - Symbolic reasoning
  - Rule-based system
  - Tree search
  - etc.
- Learning-based system:
  - Learned based on the data
  - More flexibility
  - Good at solving pattern recognition problems

What is Symbolic AI?

- Symbolic AI (also known as **Good Old-Fashioned AI, GOFAI**) represents knowledge using **symbols, rules, and logic**.
- It uses **explicitly programmed rules** to perform reasoning and problem-solving.
- Symbolic AI is based on **formal logic, tree search, and knowledge representation**.
- **Example:** If *All humans are mortal* and *Socrates is human*, then we can deduce that *Socrates is mortal*.

What is Tree Search in Symbolic AI?

- Tree search is a fundamental approach in Symbolic AI for solving problems by exploring possible states in a **decision tree**.
- Algorithms like **Depth-First Search (DFS), Breadth-First Search (BFS), and A* Search** are used to navigate possible solutions.
- **Example:** Chess-playing AI searches possible future board states to decide the best move.
- Unlike Machine Learning, tree search does **not generalize from data** but explicitly computes solutions using logical steps.

Rule-Based AI: A Subset of Symbolic AI

- Rule-Based AI applies Symbolic AI using **explicit if-then rules**.
- Rules are manually defined by experts rather than learned from data.
- **Example:** A medical expert system might use rules like:
  - IF patient has fever AND cough $\rightarrow$ Diagnose as flu.
  - IF transaction amount $>$ \$10,000 $\rightarrow$ Flag as potential fraud.
- Limitation: Rule-based systems **struggle to handle exceptions and uncertain scenarios**.

How Symbolic AI Uses Reasoning

- Symbolic AI relies on different types of reasoning:
  - **Deductive Reasoning:** Uses general rules to reach specific conclusions.
  - **Inductive Reasoning:** Generalizes from examples.
  - **Abductive Reasoning:** Chooses the most likely explanation given incomplete information.
  - **Probabilistic Reasoning:** Uses probabilities to model uncertainty.
- Machine Learning models **approximate** these reasoning types but do not explicitly use formal logic.

Limitations of Symbolic AI

- Difficult to scale—rule-based systems require **manual updates**.
- Cannot handle **unstructured data** (e.g., images, speech).
- Struggles with **uncertainty**—rules do not adapt to new situations.
- Machine Learning **outperforms Symbolic AI** in perception tasks (e.g., NLP, vision).

The Future: Hybrid AI (Neuro-Symbolic AI)

- **Hybrid AI** combines **Symbolic AI (structured logic)** with **Machine Learning (pattern recognition)**.
- **Example:** AI lawyer uses Symbolic AI to interpret legal rules + Machine Learning to analyze past case history.
- This approach enhances **explainability, adaptability, and reasoning.**

Final Takeaways

- **Symbolic AI** relies on **explicit rules and logic**.
- **Machine Learning** learns from **data without explicit programming**.
- **Rule-Based AI is a subset of Symbolic AI** but lacks flexibility.
- Future AI will combine **symbolic reasoning with deep learning** to create **more intelligent and interpretable models**.

Symbolic AI vs Machine Learning

Feature	Symbolic AI	Machine Learning
Knowledge	Rules & logic	Data & patterns
Interpretability	Explainable	Often a black box
Adaptability	Rigid (manual updates)	Can generalize
Data Needs	Minimal	Requires large datasets
Example Uses	Theorem proving, expert systems	NLP, vision, recommendations