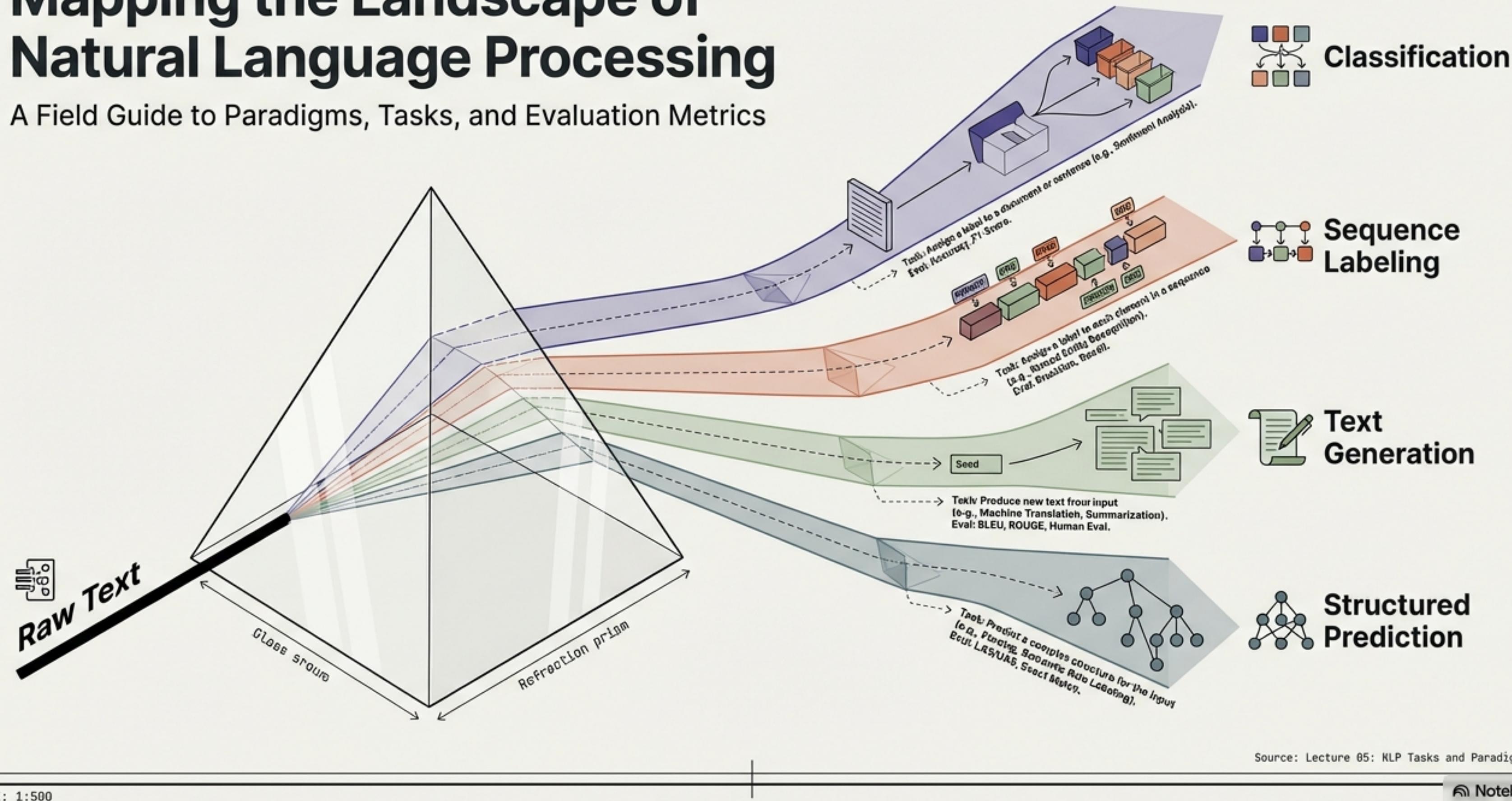


Mapping the Landscape of Natural Language Processing

A Field Guide to Paradigms, Tasks, and Evaluation Metrics



The Four Pillars of NLP

Most real-world problems map to one of these four machine learning paradigms.

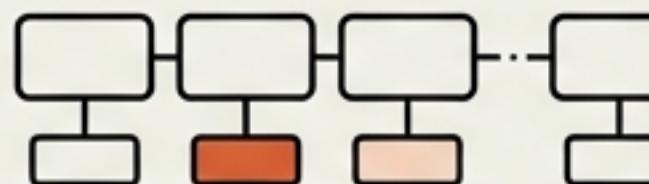
1. Text Classification



Assigning a label to a specific unit of text.

- Sentiment Analysis, News Grouping, Spam Detection.

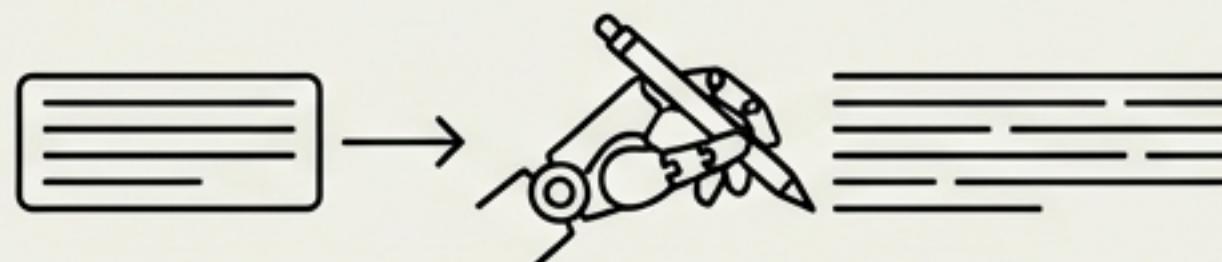
2. Sequence Labeling



Assigning a specific label to each individual word/token in a sequence.

- Named Entity Recognition (NER), POS Tagging.

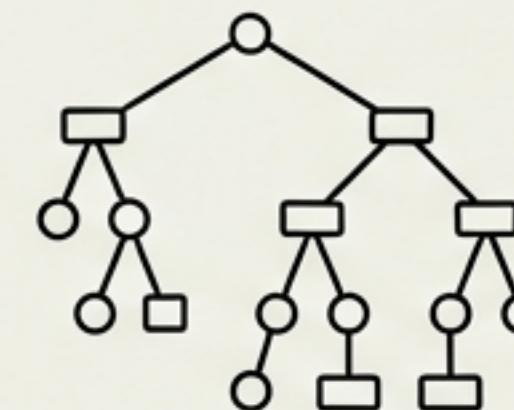
3. Text Generation



Creating new text sequences based on history and context.

- Machine Translation, Summarization, Chatbots.

4. Structured Prediction



Generating complex structures like graphs or trees.

- Dependency Parsing, Semantic Parsing.

Text Classification

The task of categorizing a specific unit of text into a predefined set of classes.



Document,
Sentence, or
Snippet

Classifier
Function

Output (C)

One class from
set {C₁...C_j}

Real-World Applications



Movie/Restaurant
Reviews
(Positive/Negative)



Product Launches
(iPhone Sentiment)



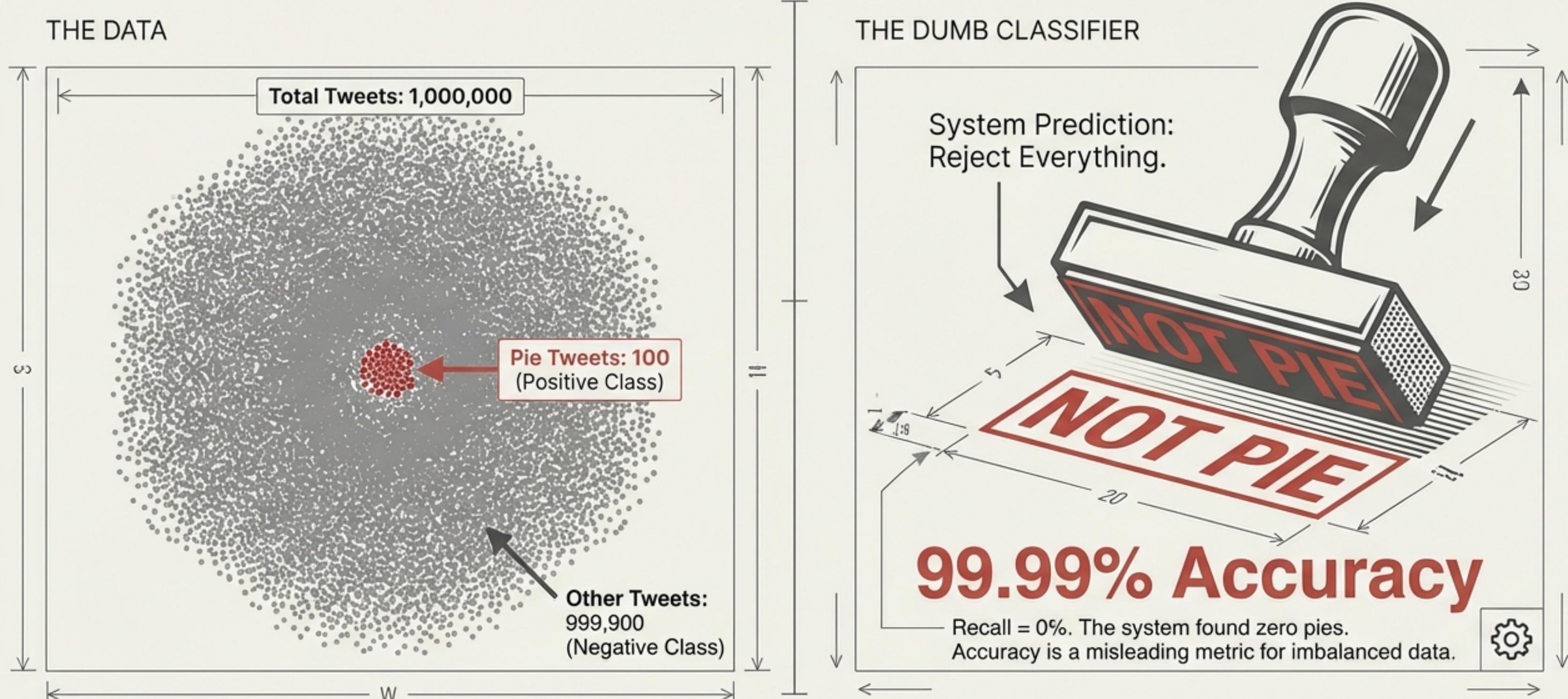
Political Monitoring
(Election Trends)



Market Prediction
(Social Media Signals)

The Accuracy Trap

Case Study: The Delicious Pie Company



Beyond Accuracy: The Confusion Matrix

		Ground Truth: YES	Ground Truth: NO
System Predicted: YES	True Positive (TP) (Hit)	False Positive (FP) (Type I Error)	
	False Negative (FN) (Type II Error)	True Negative (TN) (Correct Rejection)	
System Predicted: NO			

Precision (Quality):

$$Precision = \frac{TP}{TP + FP}$$

Of the ones we labeled Pie, how many were actually pies?

Recall (Quantity):

$$Recall = \frac{TP}{TP + FN}$$

Of the actual Pies in existence, how many did we find?

F1 Score:

The harmonic mean of Precision and Recall.



Multi-Class Evaluation: Macro vs. Micro Averaging

Scenario: Classifying Emails as Urgent, Normal, or Spam

CLASS	COUNT	PRECISION
Class: Urgent	Count: 19 (Small)	Precision: 8/19
Class: Normal	Count: 86	Precision: 60/86
Class: Spam	Count: 268 (Dominant)	Precision: 200/268

MACRO AVERAGING

(Class-Focused)

$$\frac{(8/19) + (60/86) + (200/268)}{3}$$

Treats every **CLASS** equally.

Average ≈ 0.60

Best for tail classes.

MICRO AVERAGING

(Instance-Focused)

$$\frac{(8 + 60 + 200)}{(19 + 86 + 268)}$$

Treats every **INSTANCE** equally.

Average ≈ 0.73

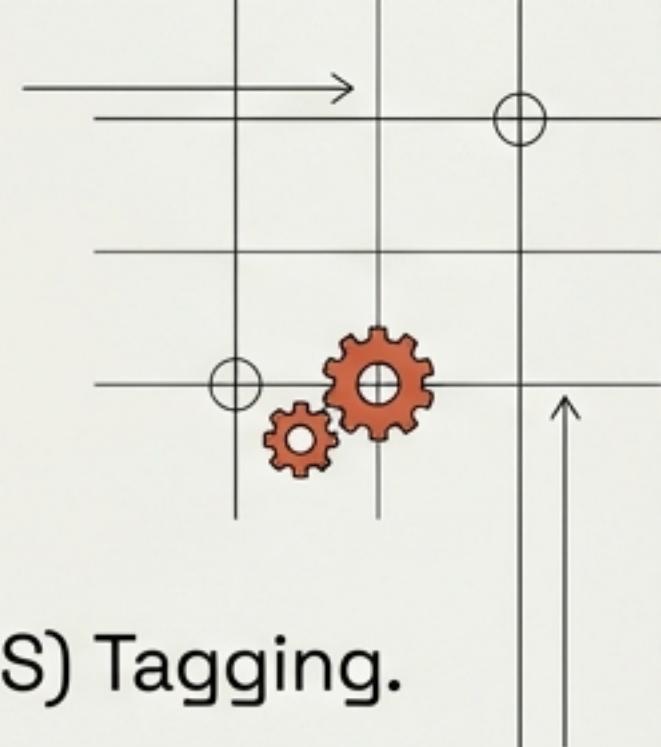
Biased toward the dominant class (Spam).



Sequence Labeling

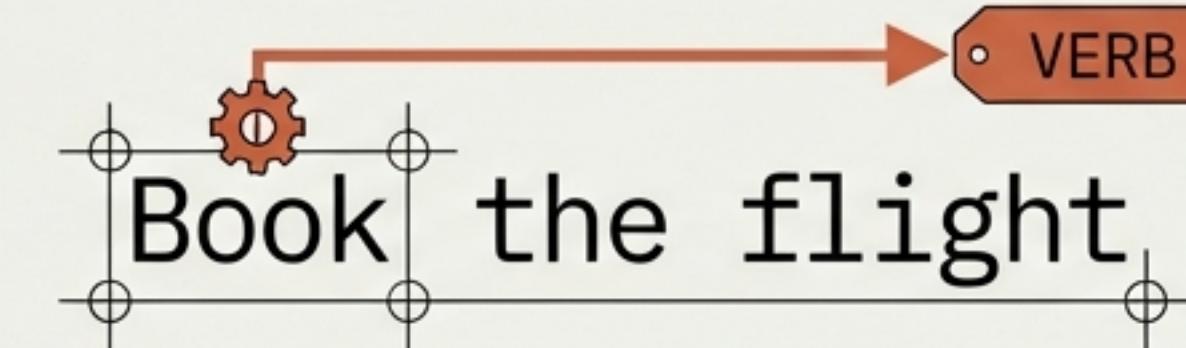
Assigning a label to every individual token in a sequence.

Common Tasks: Named Entity Recognition (NER), Part-of-Speech (POS) Tagging.

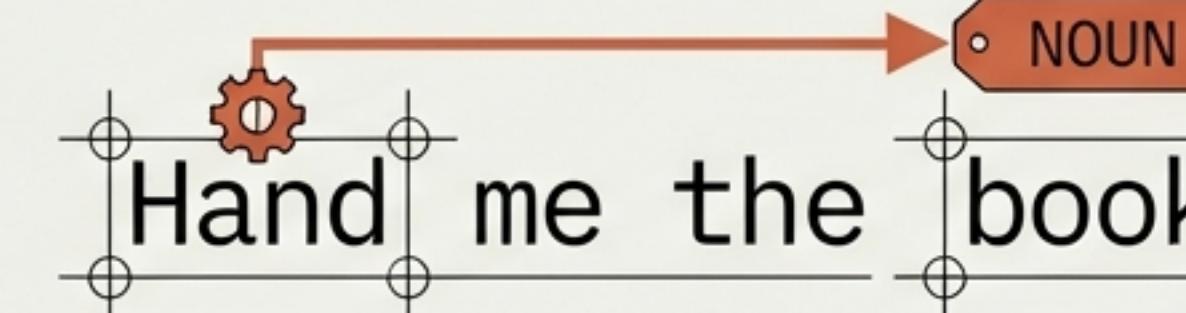


Visual Example (The Ambiguity)

Scenario A



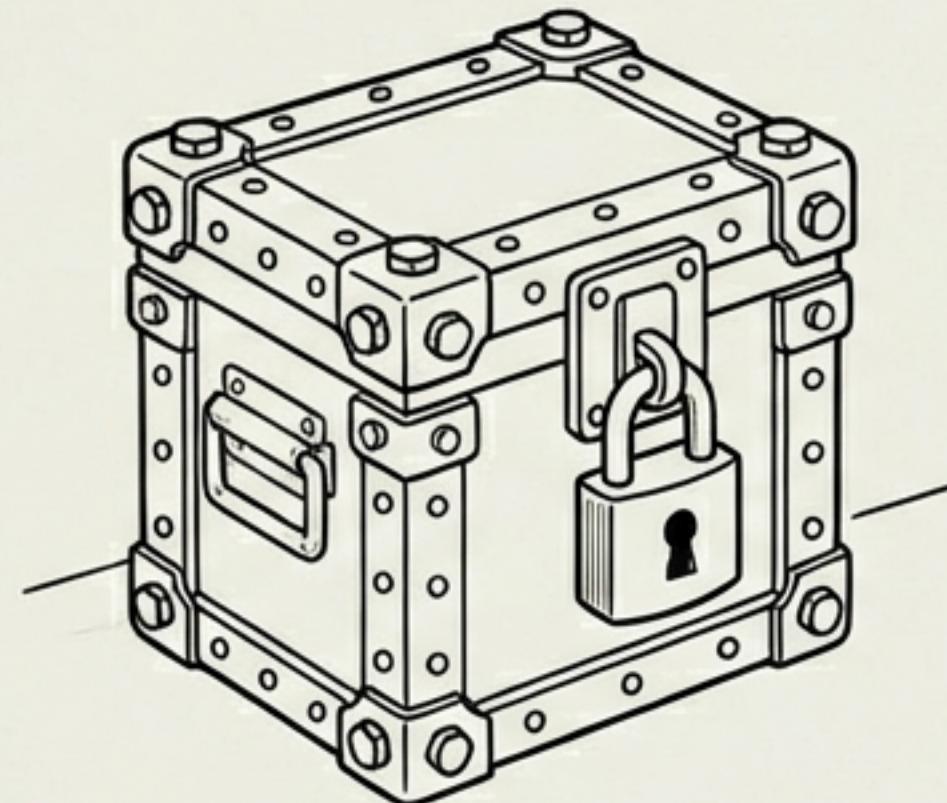
Scenario B



The system must resolve ambiguity based on context.

Linguistic Categories: Open vs. Closed Classes

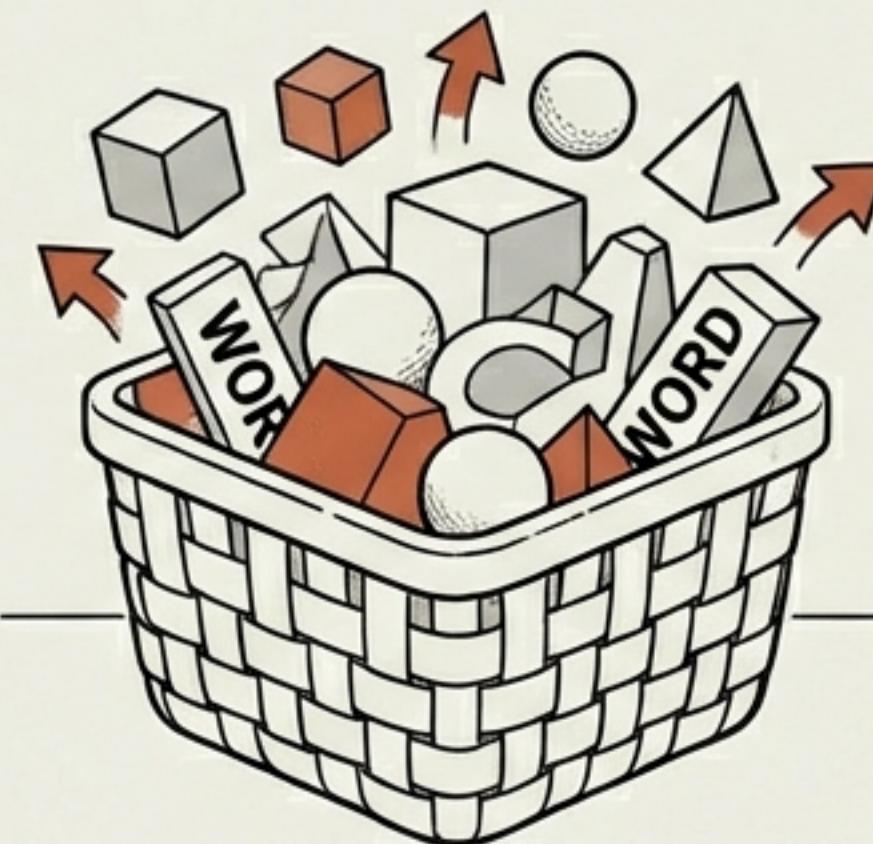
Closed Classes (Function Words)



Fixed sets. Rarely change. Grammatical glue.

Pronouns (I, you), Prepositions (in, on),
Conjunctions (and, but).

Open Classes (Content Words)

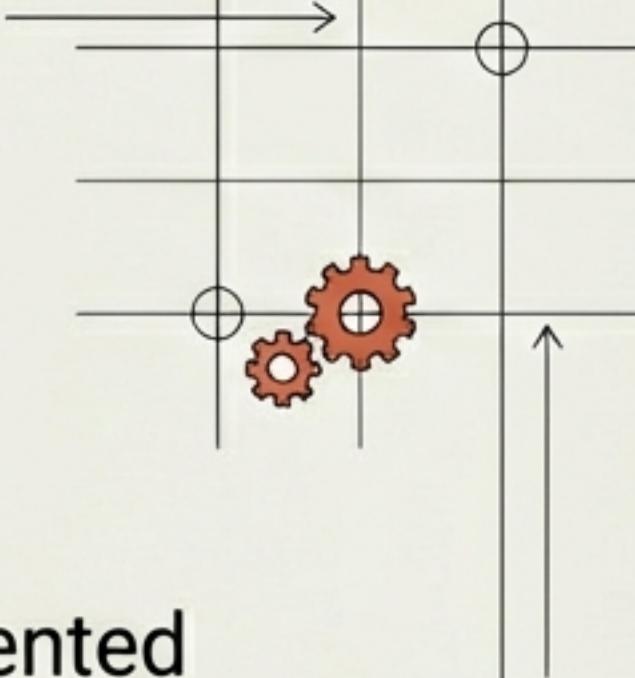


Evolving sets. New words coined constantly.

Nouns (iPhone, TikTok), Verbs (Google,
Tweet), Adjectives.

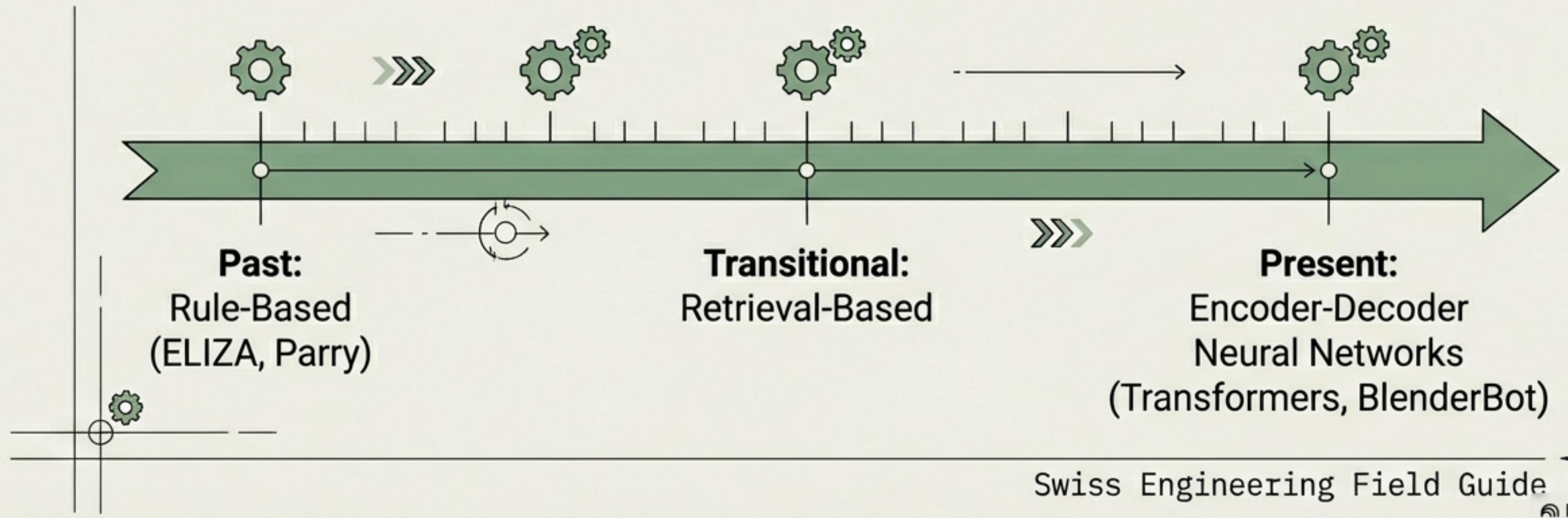
Text Generation

Generating coherent response sequences based on context.



Chatbots: Social, open-ended
(Therapy, Fun).

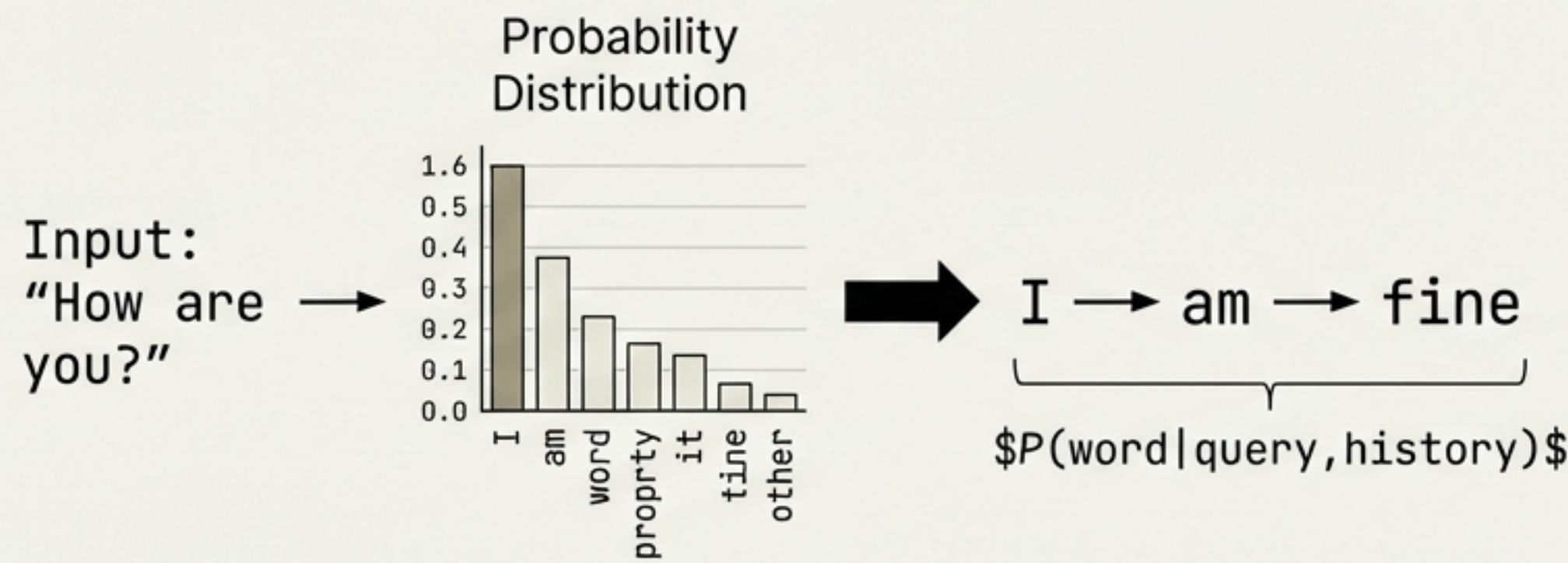
Dialogue Agents: Task-oriented
(Flight booking, Customer support).



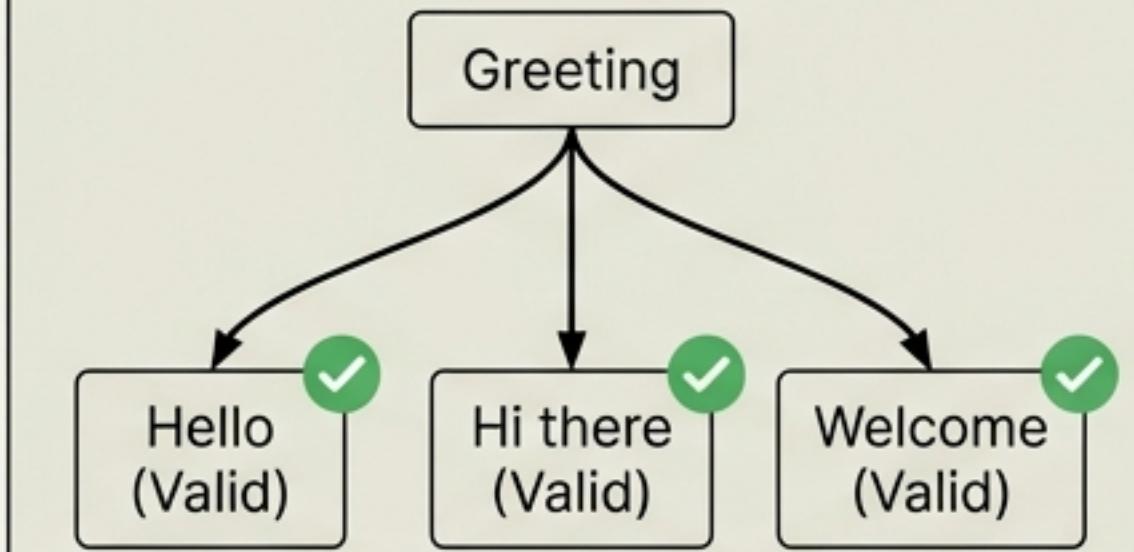
Generation Mechanics & Evaluation



Conditional Language Modeling



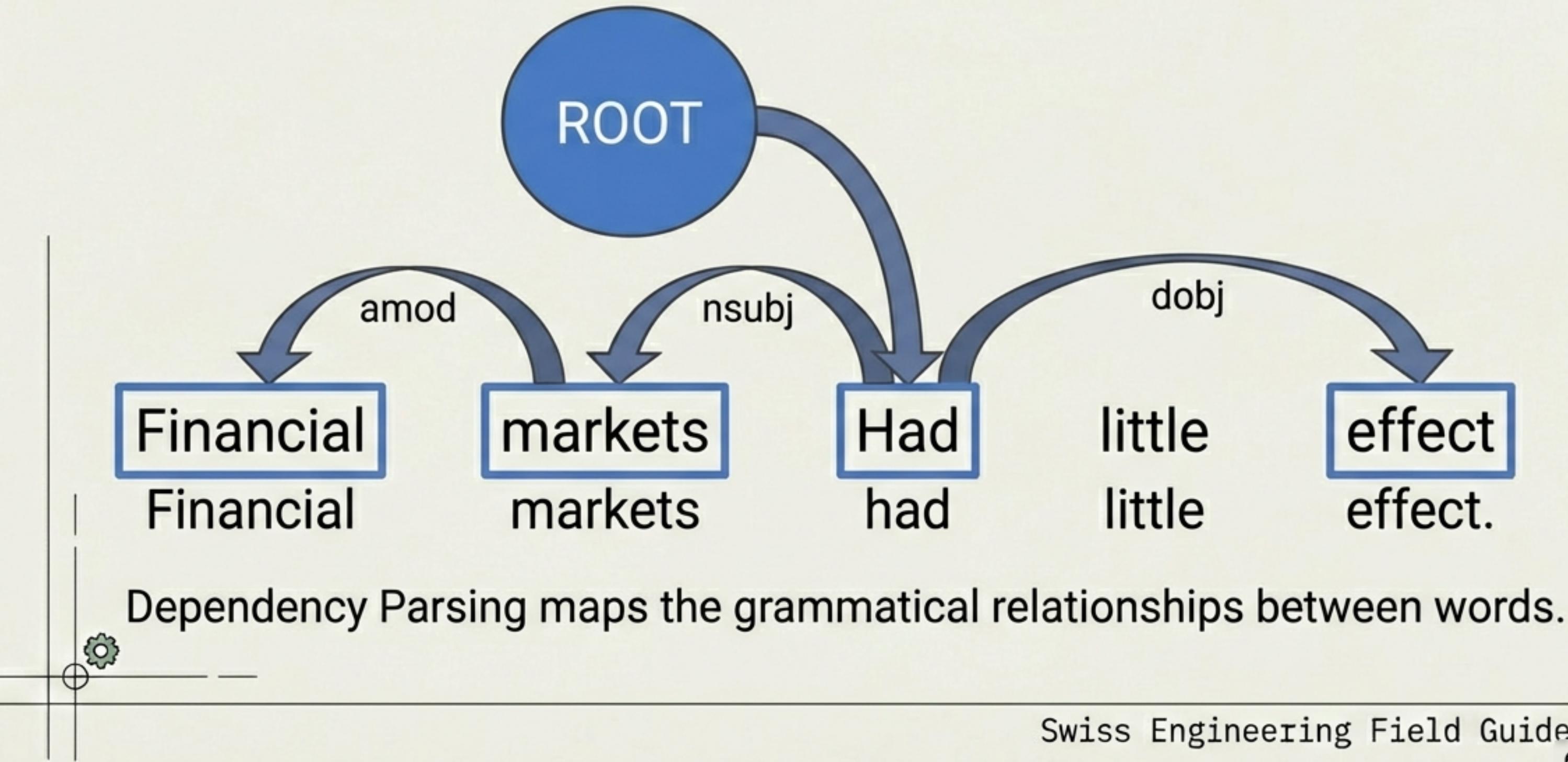
The One-to-Many Problem



Unlike classification, there is no single correct answer. Human evaluation is critical.

Structured Prediction

Predicting complex structures (graphs/trees) rather than flat labels.

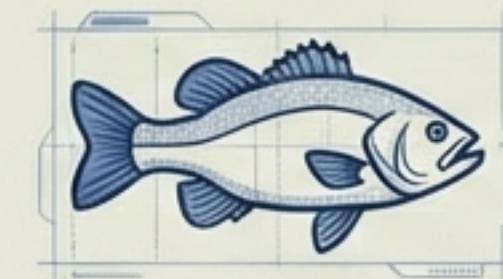


NLP Tasks by Granularity: Word Level



Word Sense Disambiguation (WSD)

Determining which meaning is invoked.



Context: "Lake"

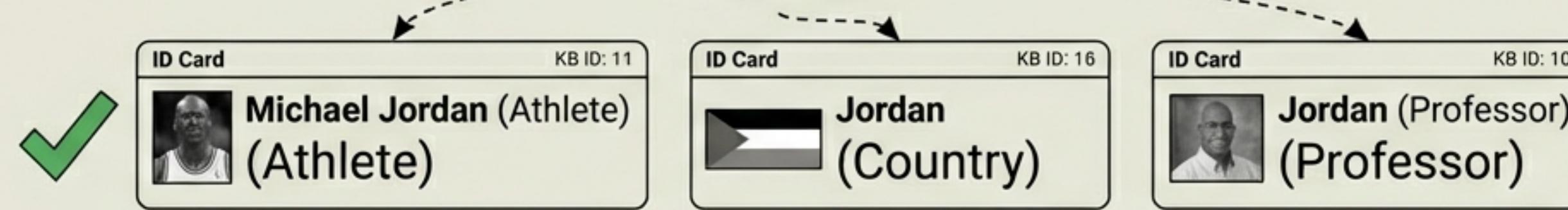


Context: "Music"

Entity Linking

Connecting a mention to a specific Knowledge Base ID.

Jordan



NLP Tasks by Granularity: Sentence Level

Sentence Similarity & Paraphrase Detection

Determining if two sentences have the same meaning or are duplicates.

Example 1 (Duplicate)

Box A

How to start a bakery



Box B

How can one start a bakery business?

Example 2 (Not Duplicate)

Box A

What are natural numbers?



Box B

What is the least natural number?

Entailment: Does Sentence A imply Sentence B?

NLP Tasks by Granularity: Document Level

Question Answering (SQuAD)

Reading Comprehension Test: The Duchy of Normandy

The Duchy of Normandy was a large feudal state in northern France. It was created in 911 AD by the Treaty of Saint-Clair-sur-Epte between King Charles III of West Francia and the Viking leader Rollo. Rollo was the first ruler of Normandy, establishing a dynasty that would later conquer England.

Question: Who ruled the Duchy?

Key Insight: The model must read the full paragraph to extract the specific answer span. Advanced models (SQuAD 2.0) must also know when the answer is *not* present.



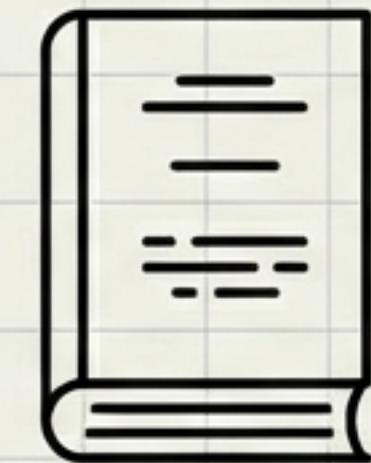
Exploring the Frontier

Resources and Next Steps



Resource Box

The Landscape is vast.
'Papers With Code' currently lists over 2,800 text datasets.



Recommended Reading

'Speech and Language Processing' (Jurafsky & Martin).

Coming Next:
From Paradigms to Engines.

Deep Learning Basics.