

Tolstoy

Machine learning to
read your text.

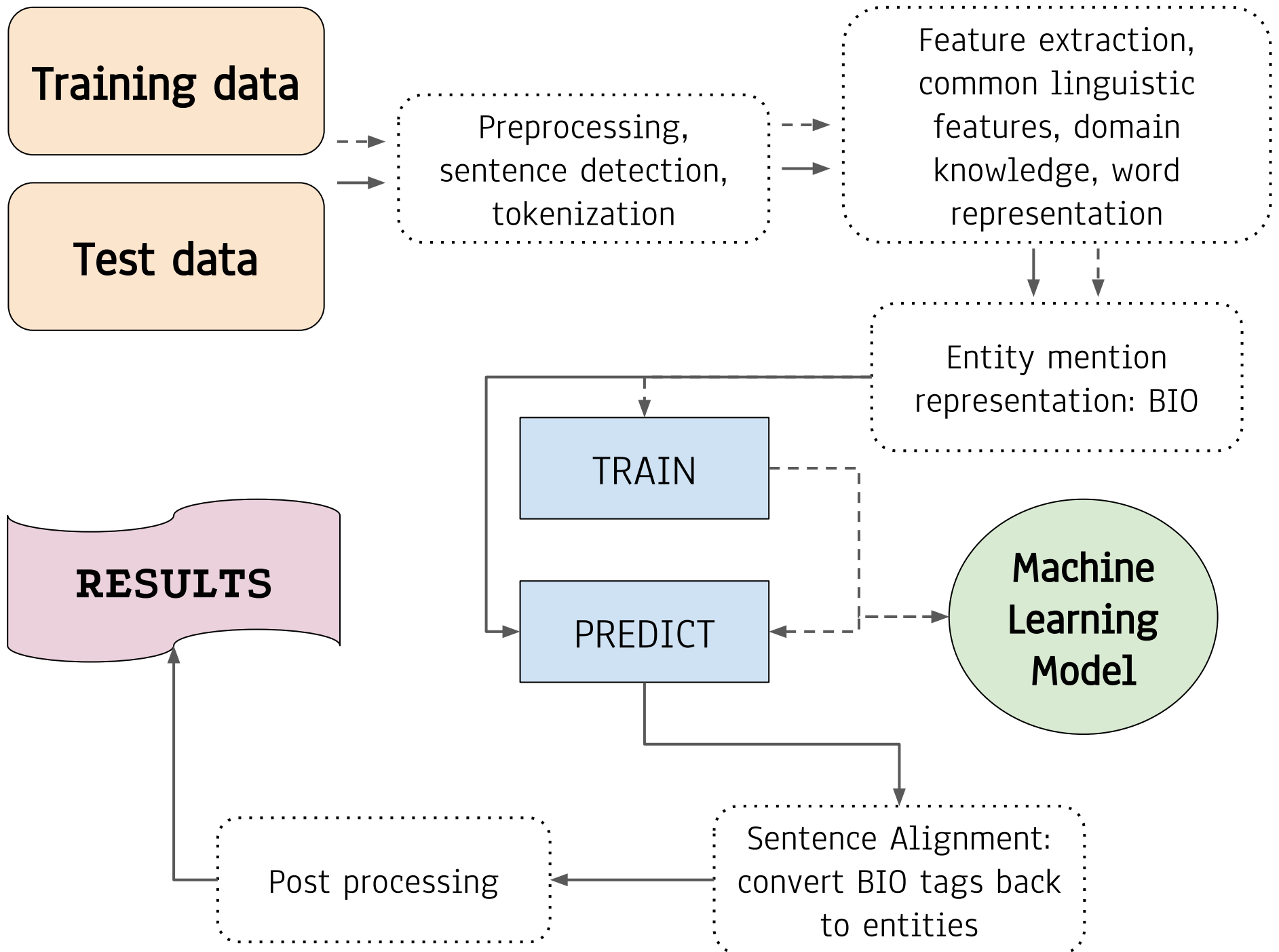
SUMMARY

- Tolstoy can read and index field notes.
- We use machine learning and natural language processing to do this.
- We're much faster than humans:
4 months of reading → 1 minute.

Named Entity Recognition on Transcribed Field Notes

- **AIM** Identify proper names of people, organizations, locations or other entities from natural language documents.
- **RESULTS** Reference resolution, other types of disambiguation, meaning representation (in maps, tables, etc.), cross- referencing, richer analytical frameworks in NLP applications.
- **NEXT STEPS** Question and answer systems, textual entailment, and document searching.

TECHNOLOGY



- **Named entity recognition** means finding spans of text that constitute proper names and then classifying the type of the entity.
- **This flowchart shows the basic NER pipeline**, which mainly includes extracting content (some low level formatting), preprocessing text, and ultimately recognizing entities.

Person

Apple CEO Tim Cook introduces 2 new,
larger iPhones, Smart Watch at
Cupertino Flint Center event.

Location

Organization

TYPES OF NER TOOLS

- **OLD WAY Rule-based/Regex methods** - Use manually constructed finite state patterns. *Time consuming, misses stuff.*
 - **NEW WAY Machine learning-based methods** - Use the classification approach. *More nuanced, much faster.*
 - **Hybrid methods** - Mixture of both. *Most powerful.*
- ⋮
▼
- **REALLY OLD WAY** People manually read and tag entities. *Too slow!*

NATURAL LANGUAGE PROCESSING

- Natural Language Processing (NLP) is the **automatic manipulation of natural language**, like speech and text, by software.
- It sits at the intersection of computer science, artificial intelligence, and computational linguistics
- NLP is primarily hard because **language is** highly ambiguous, **ever changing and evolving**.

2 p.m., at the Natural History Museum/Locations, in Balboa Park/Locations. Here in the interests of launching an MVZ "expedition" into Lower California/Locations.

Left Berkeley/Locations on "Owl" Friday night (Feb. 27/Date). Spent Saturday at the Los Angeles Museum/Locations, Exposition Park/Locations, and Saturday night with Lee Chambers/People.

Came to San Diego/Locations, Sunday forenoon; and since then have been meeting various people; and especially, waiting to see Jose Maria Gallegos/People, the Mexican official who is supposed to be mainly influential in the granting of permits for an entry into Lower California/Locations.

ASR

NLP

- named entity recognition (NER)

- part-of-speech tagging
- text categorization
- syntactic parsing
- coreference resolution
- machine translation

automatic
speech
recognition (ASR)

NLU

(nat. language understanding)

- relation extraction
- semantic parsing
- question answering
- paraphrase & natural language inference
- sentiment analysis
- dialogue agents
- summarization

text-to-speech
(TTS)

OTHER USE CASES

- World Bank - project data categorization
- EPA - public comments
- Give us your ideas!

CONTACT US



Rosa Lin
Founder & CEO
rosa@tolstoy.ai



Prerna Kashyap
ML Engineer
prerna@tolstoy.ai



Arjun Kumar
Cofounder, COO
arjun@tolstoy.ai

Check us out at <http://tolstoy.ai/> !