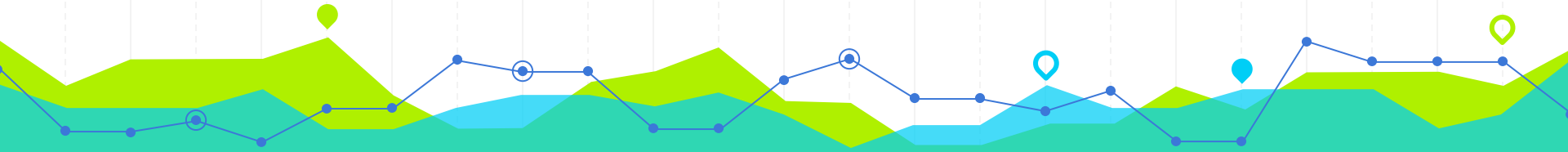


Multilingual Text Analytics for Healthcare by Terminators

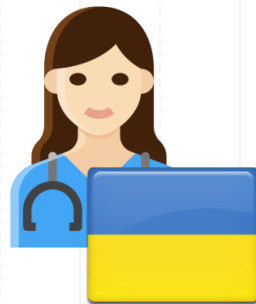


NoLIT

No(t as much) Loss in Translation

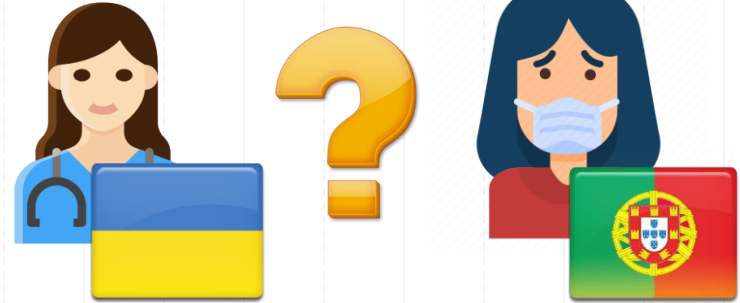
The Problem

- Doctor and patient might not share a common language



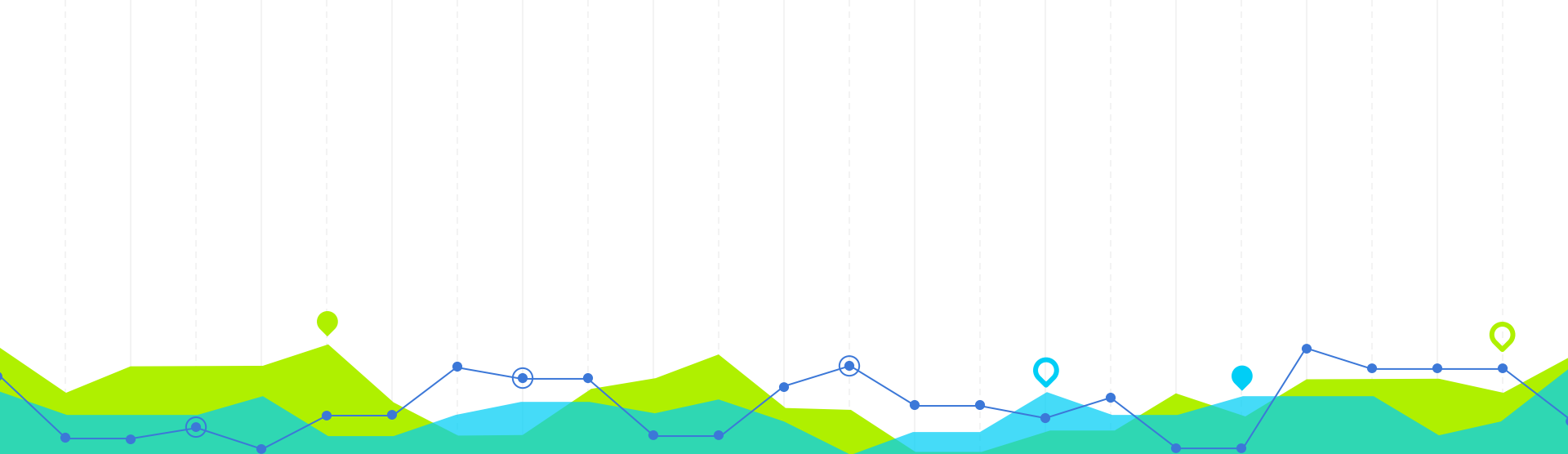
The Problem

- Doctor and patient might not share a common language



- Diagnosis might be overwhelmingly long and confusing





Baseline

Architecture as seen in the introduction

1

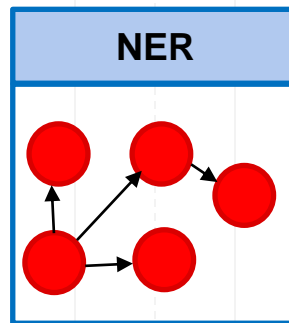
Baseline

Diagnosis
(Foreign Language)

Translate

Diagnosis
(English)

Baseline



Extractive
Summary

Abstractive
Summary

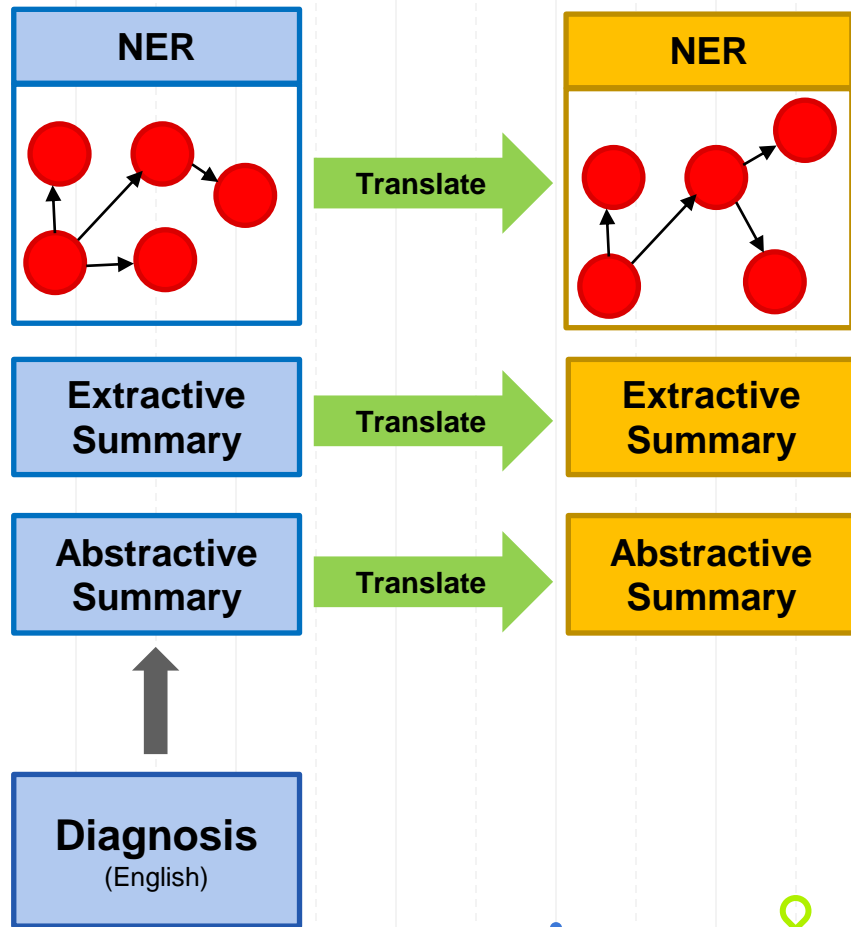


Diagnosis
(English)

Diagnosis
(Foreign Language)

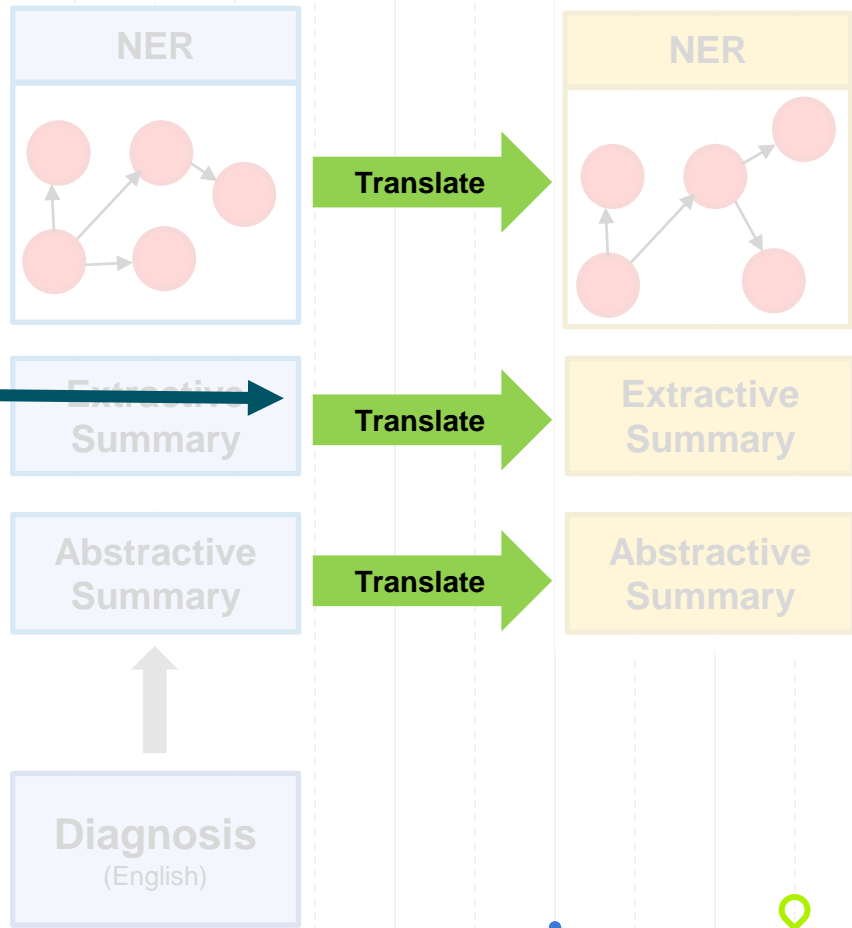


Baseline



Baseline

Compounding Errors



Our Goals

Reduce Error Compounding

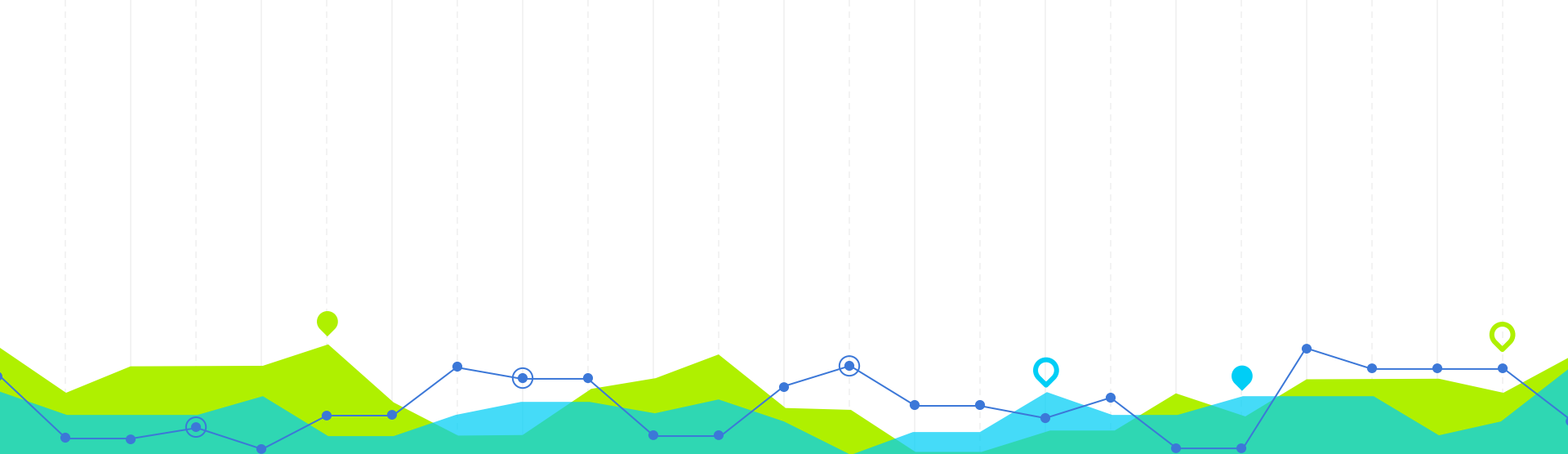


Confidence Feedback



Improve Translation



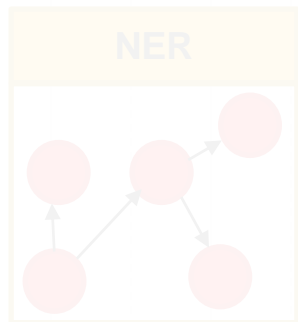


NoLIT

Not as much **Loss** In **Translation**!

2

Observation

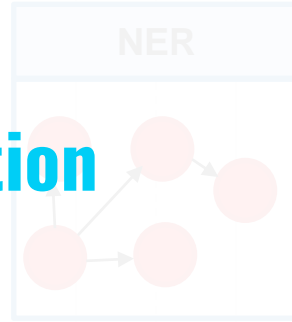


**Extractive
Summary**

**Abstractive
Summary**

Diagnosis
(Foreign Language)

Foreign language
**Small / Non-existent
Dataset**



**Extractive
Summary**

**Abstractive
Summary**

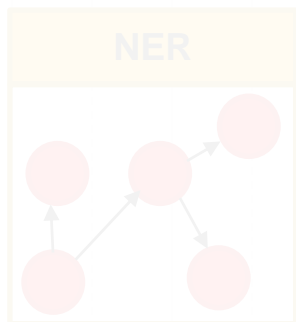
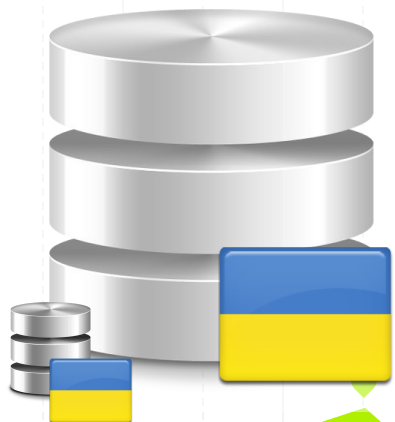
Diagnosis
(English)

English
Large Dataset



Approach

Foreign language
Augmented Dataset

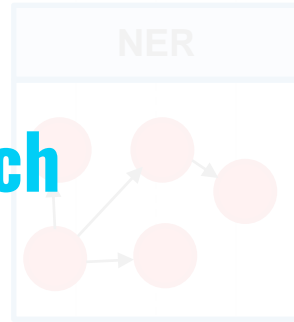


Extractive Summary

Abstractive Summary

Diagnosis
(Foreign Language)

Translate



Extractive Summary

Abstractive Summary

Diagnosis
(English)

Translate

English
Large Dataset

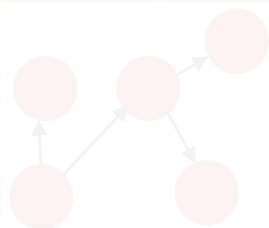


Approach

Foreign language
Augmented
Dataset

Independent
Translations

NER



Extractive
Summary

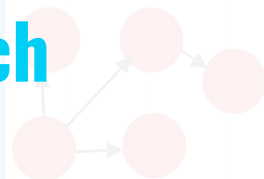
Abstractive
Summary

Diagnosis
(Foreign Language)

Translate

Translate

NER



Extractive
Summary

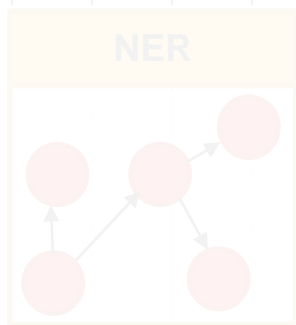
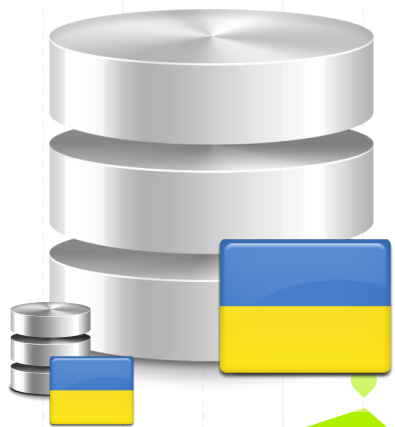
Abstractive
Summary

Diagnosis
(English)

English
Large English



Foreign language
**Augmented
Dataset**

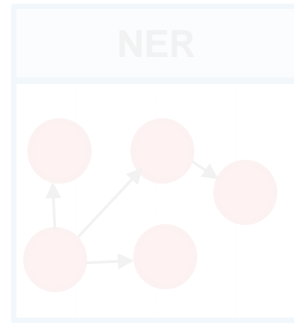


**Extractive
Summary**

**Abstractive
Summary**

Diagnosis
(Foreign Language)

Translate



**Extractive
Summary**

**Abstractive
Summary**

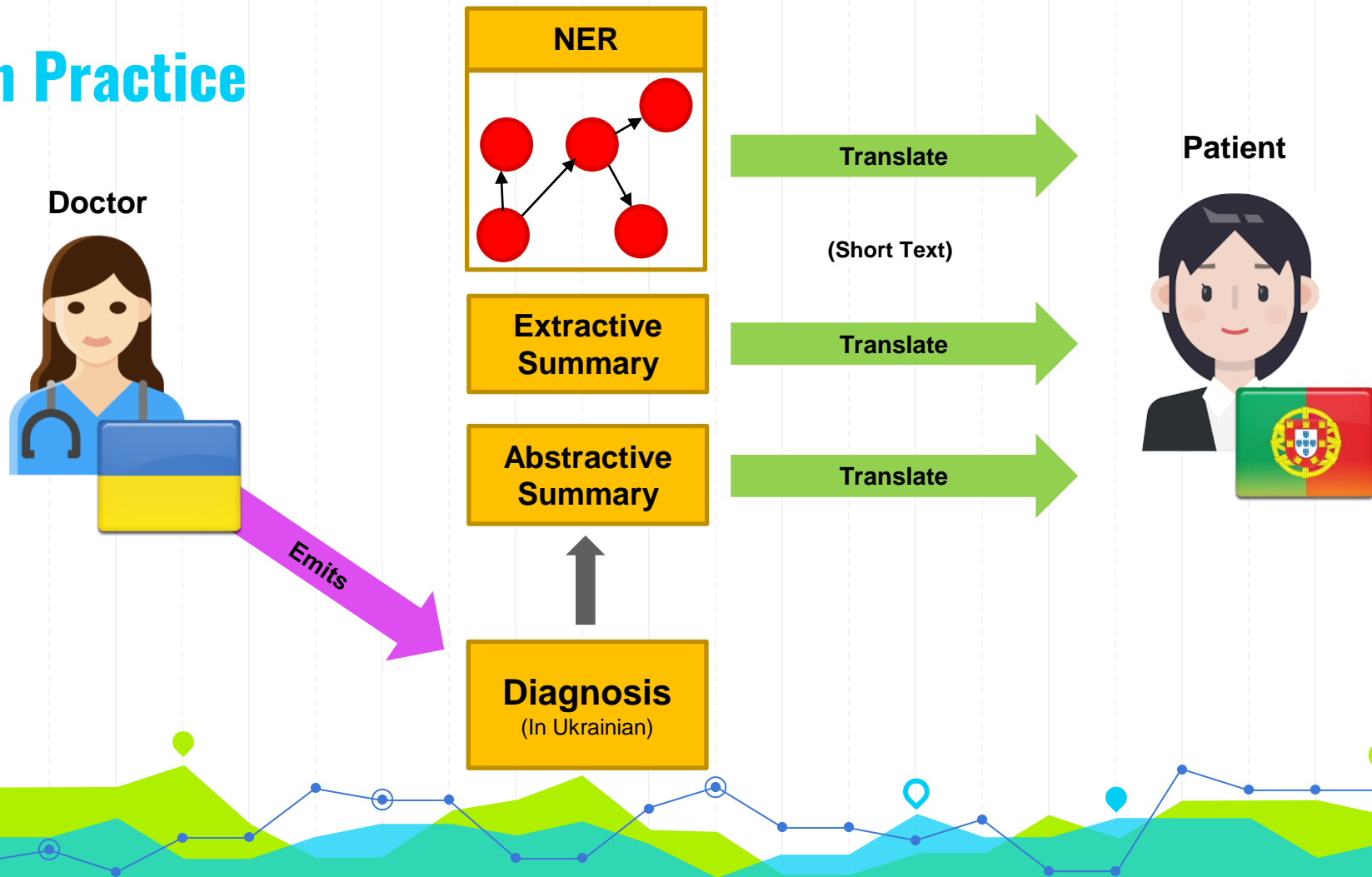
Diagnosis
(English)

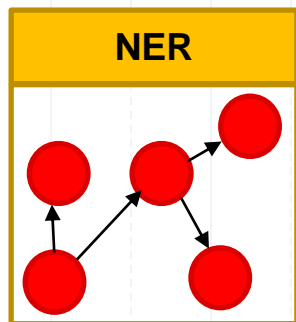
Translate

English
Large Dataset



In Practice



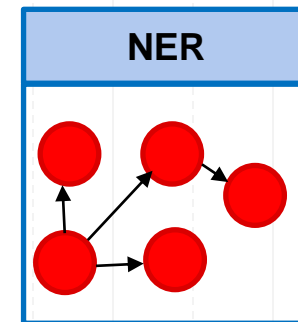
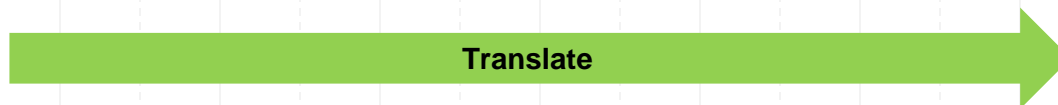


**Extractive
Summary**

**Abstractive
Summary**



Diagnosis
(Foreign Language)



**Extractive
Summary**

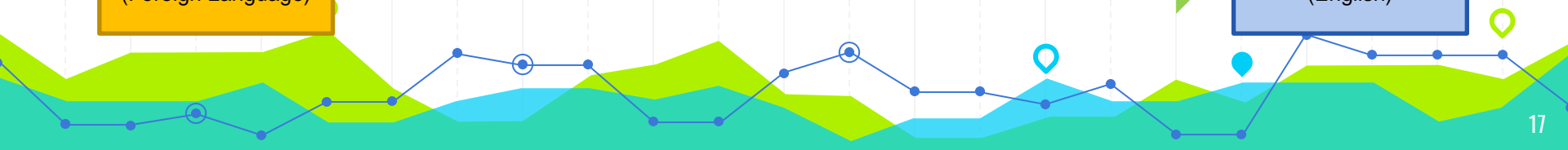
**Abstractive
Summary**

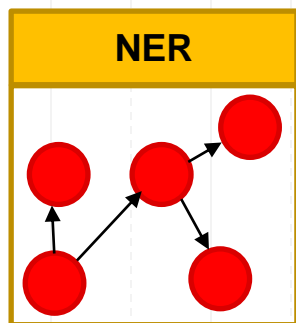


Diagnosis
(English)

Dual Model

If correctness is a concern

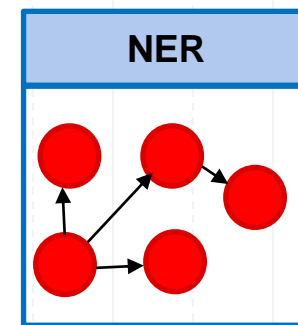




Extractive Summary

Abstractive Summary

Diagnosis
(Foreign Language)



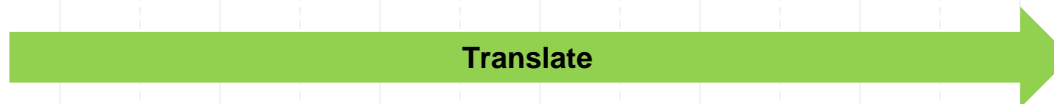
Extractive Summary

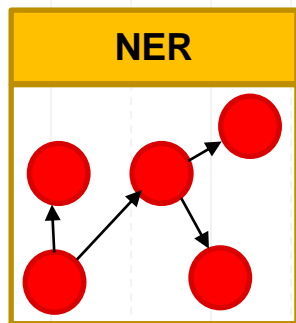
Abstractive Summary

Diagnosis
(English)

 **Important Entities**

5 mg
Lorazepam
Dengue
Chlordiazepoxide
Brucellosis

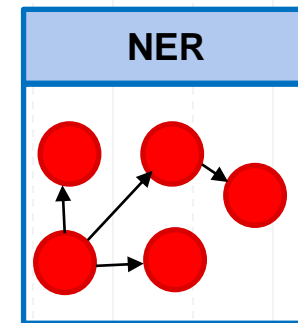




**Extractive
Summary**

**Abstractive
Summary**

Diagnosis
(Foreign Language)



**Extractive
Summary**

**Abstractive
Summary**

Diagnosis
(English)




Important Entities

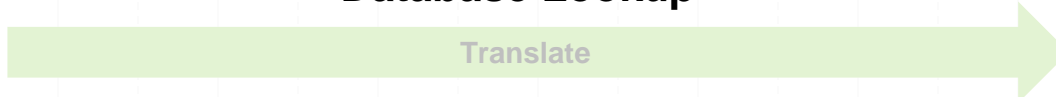
5 mg
Lorazepam
Dengue
Chlordiazepoxide
Brucellosis

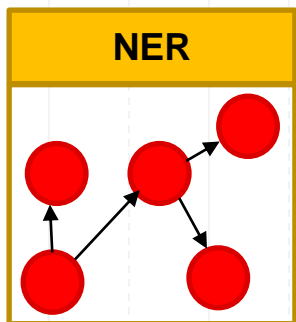


Database Lookup



Translate





**Extractive
Summary**

**Abstractive
Summary**

Diagnosis
(Foreign Language)


Important Entities

5 мг
лоразепам
Денге
Хлордиазепоксид
бруцеллез

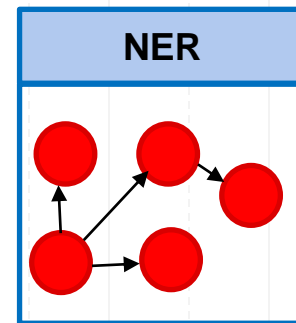

Important Entities

5 mg
Lorazepam
Dengue
Chlordiazepoxide
Brucellosis



Database Lookup

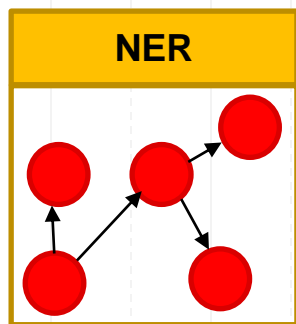
Translate



**Extractive
Summary**

**Abstractive
Summary**

Diagnosis
(English)




Important Entities

5 мг лоразепам
Денге
Хлордиазепоксид
бруцеллез

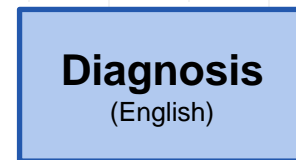
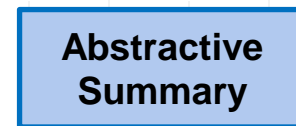
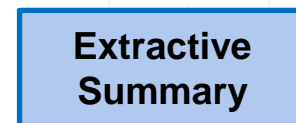
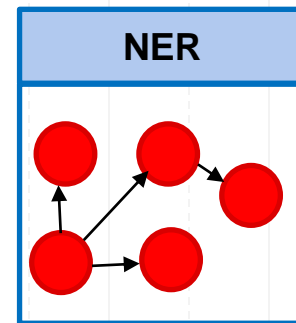

Important Entities

5 mg Lorazepam
Dengue
Chlordiazepoxide
Brucellosis

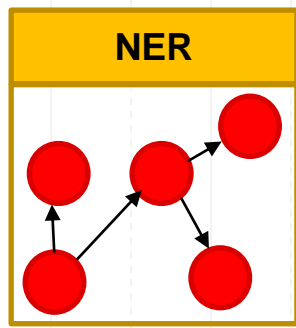


Database Lookup

Translate



Confidence



 **Important Entities**

5 мг лоразепам
Денге
Хлордиазепоксид
бруцеллез

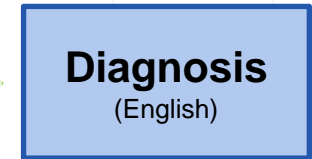
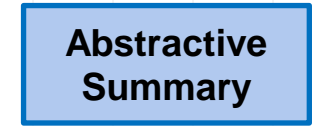
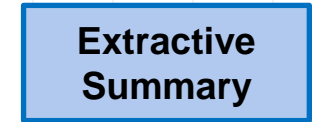
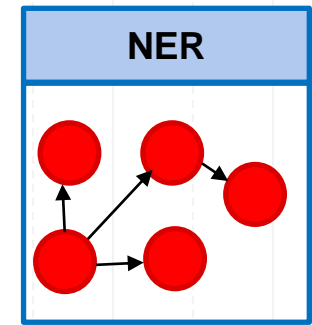
 **Important Entities**

5 mg Lorazepam
Dengue
Chlordiazepoxide
Brucellosis



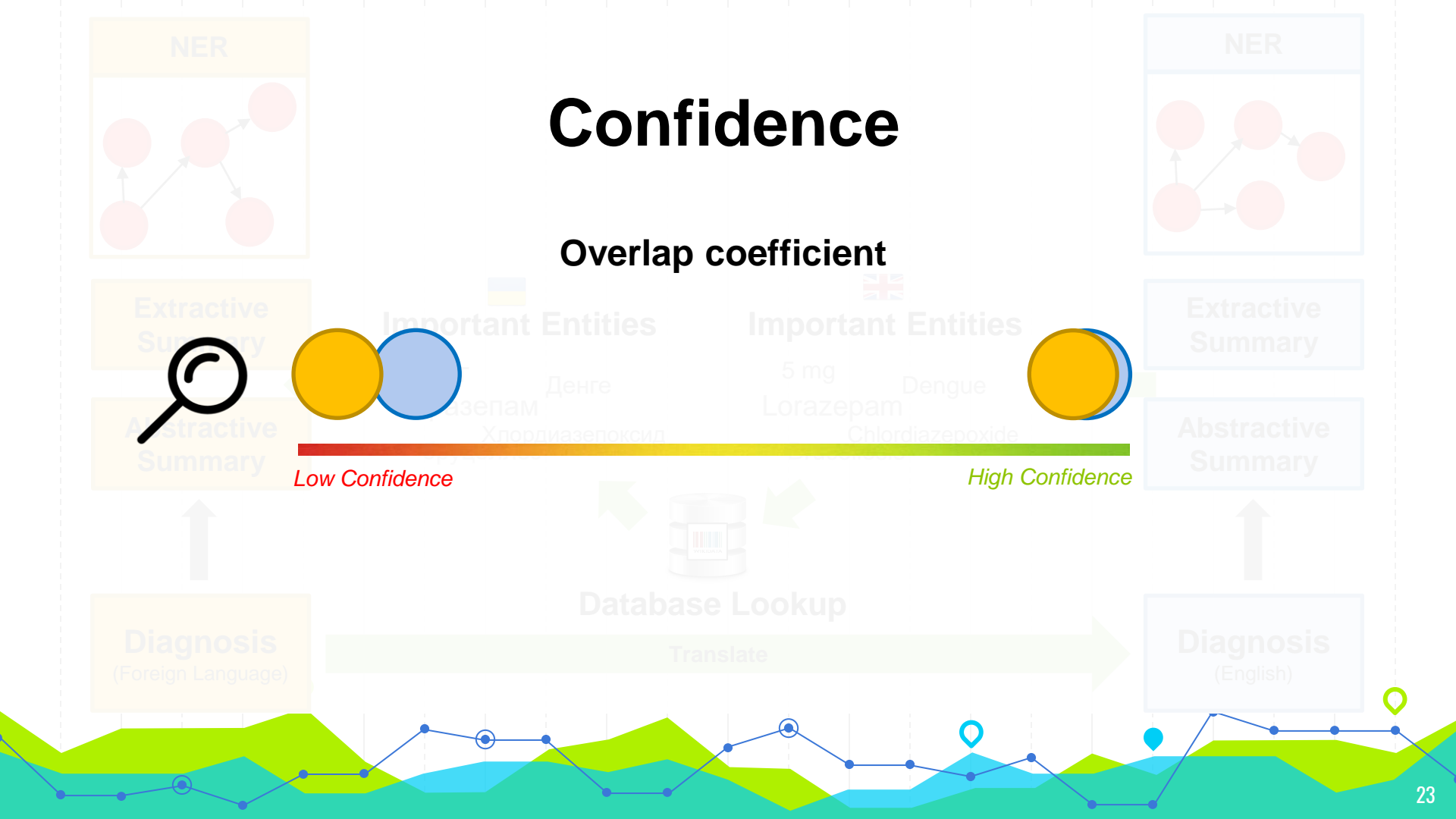
Database Lookup

Translate



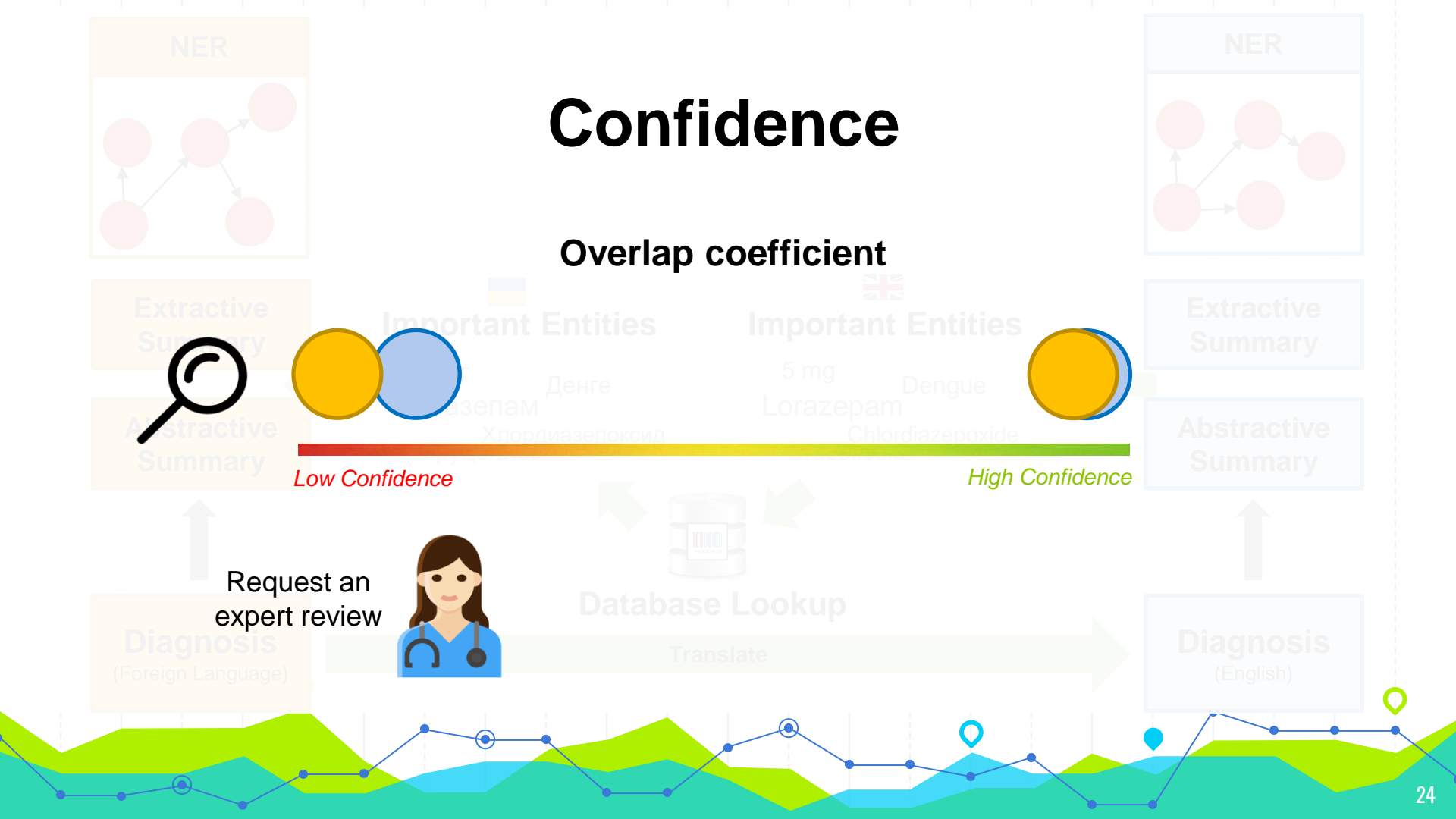
Confidence

Overlap coefficient



Confidence

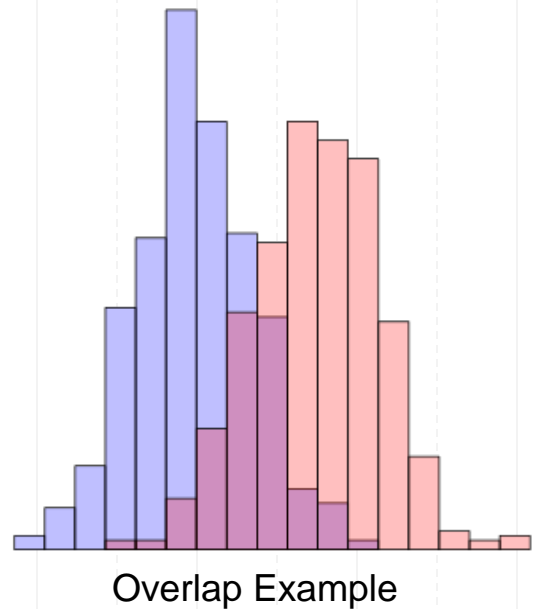
Overlap coefficient



Improve Translation

Entity Overlapping Percentage (EOP)

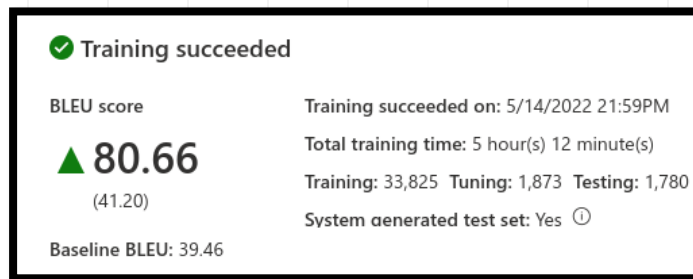
- Serves as overall quality insurance
 - Estimates quality of translation
 - Estimates improvement of fine-tuning summarization
- Highly accurate translation of entities
- Average EOP of translation ~70%

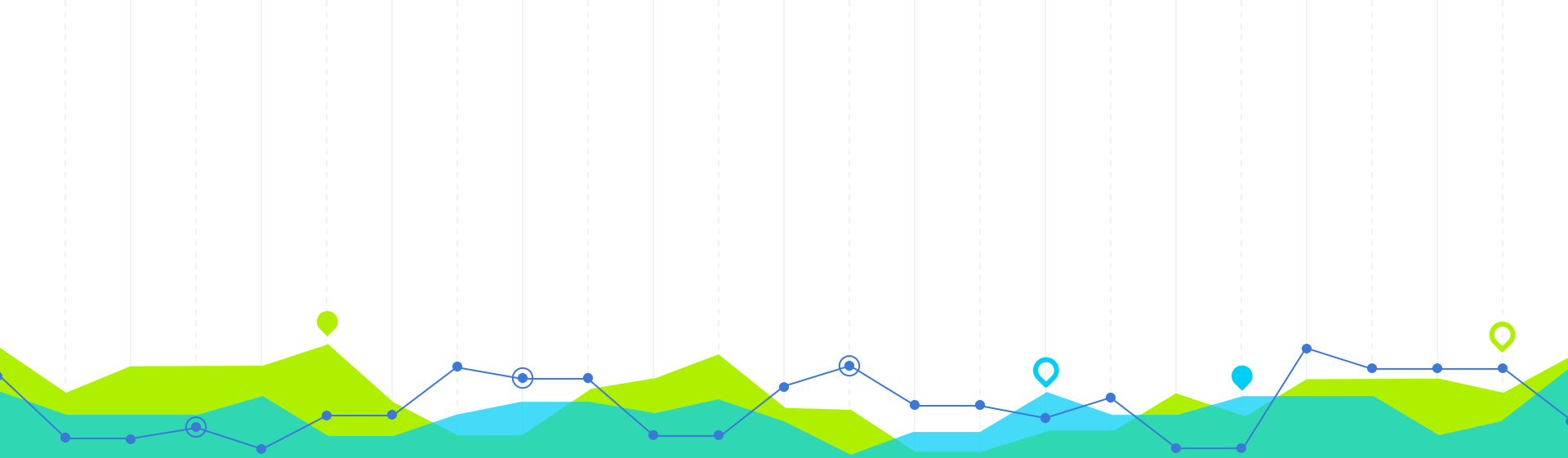


Improve Translation

Fine-Tuning on Medical Domain

- **Tuning Set**
Medical sentence pairs (~40.000)
- **Database Lookup**
1 to 1 conversion of medical terms
- **BLEU score from 39 to 81**
(super-human performance)





Benefits of our Approach

Taking a look at the closed feedback loop!

3

Biggest Benefits

● Shorter Pipeline

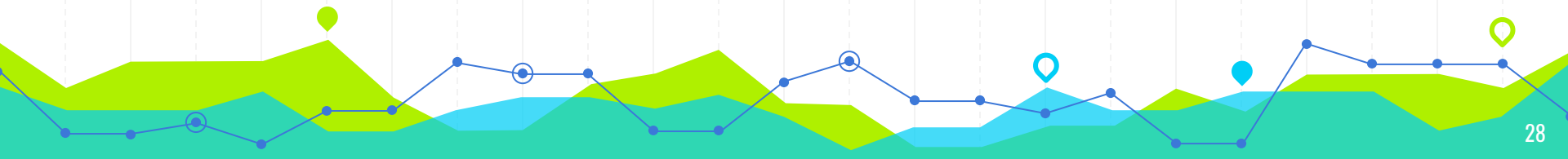
- Less loss in translation
- Lower inference costs (no back & forth translation)

● Confidence Feedback

- Discover Weaknesses
- Know when to seek expertise

● Augmented Translation

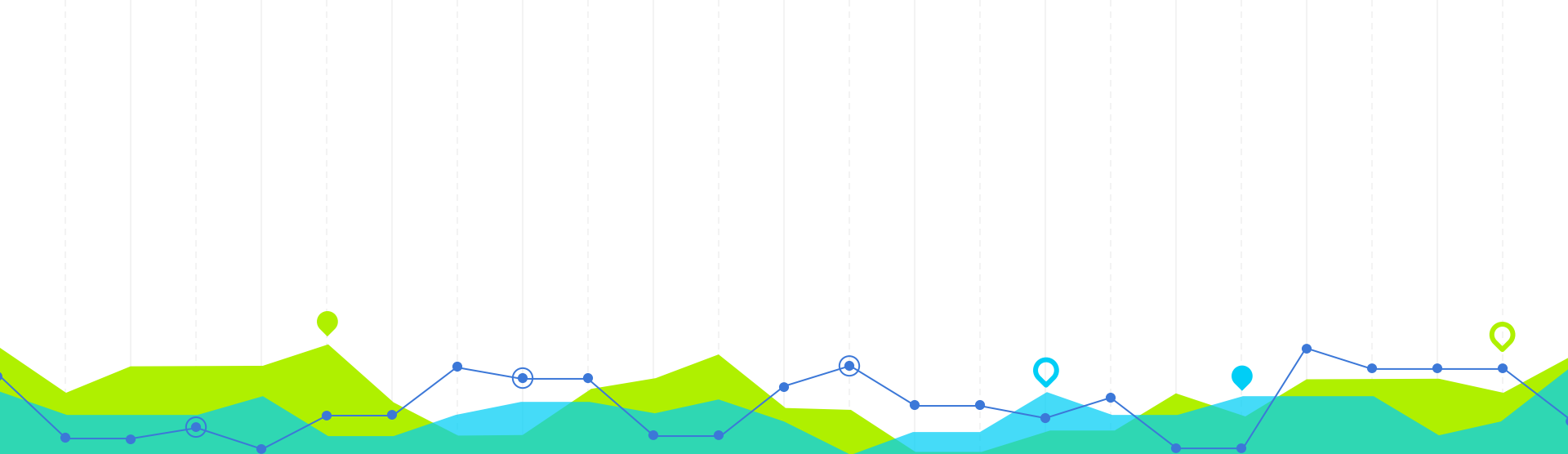
- Enables models for low resource languages



10.000s of sentence pairs & 1 dictionary for entity mapping

... is all you need!





Sebastien Foucher
Siddhant Ray
Rafael Sterzinger
Fatjon Zogaj

EXTRA RESOURCES

<https://cutt.ly/PHlKnok>, <https://users.ugent.be>