

# **Data Challenge SoSe2024**

## **NLP**

Expanding the Ground Truth for Relation Extraction

by Anton Fink and Jasper Forth; presented: July 11th, 2024

# Intro

# **Intro**

## **Objective and Idea**

### **Objective**

Develop a pipeline  
to expand the Relation Extraction (RE) database

### **Idea**

Utilize next-gen NLP techniques (LLMs)  
for big data processing

# Intro

## Key Components

### NER:

F1-Score  $\approx 99.67\%$

Ground Truth

~22,333 labeled samples

### RE:

F1-Score  $\approx 83\%$

Ground Truth

~1,297 labeled samples

hand OBJECT  
Horse ANIMAL prancing right; below, monogram.  
Fecunditas, draped OBJECT, standing left, holding  
cornucopia OBJECT; at her foot OBJECT, child  
PERSON  
Genius PERSON, standing left, holding patera OBJECT in  
extended right hand OBJECT and two corn PLANT -  
ear PLANT downwards in left

design\_id, s, subject\_class, o, p, object\_class  
11, "Genius", "PERSON", "holding", "patera", "OBJECT"

Background

# **Background**

## **Motivation**

### **Ground truth for ML Model training**

RE data is used to train an ML model  
for searching the numismatic database.

### **Current Ground truth**

The ground truth is small and  
potentially imperfect due to labeling inconsistencies.

### **Using next gen tech**

Our project leverages recent advancements in NLP  
to enhance data processing capabilities.

# Background

## Data

design_en_orig	id	annotations_orig	design_en	annotations
Diademed head of deified Alexander the Great with horn of Ammon, right. Border of dots.	1	[(9, 13, 'OBJECT'), (25, 44, 'PERSON'), (50, 54, 'OBJECT')]	Diadem head of deified Alexander the Great with horn of Ammon, right. Border of dots.	[(0, 6, OBJECT), (7, 11, OBJECT), (23, 42, PERSON), (48, 52, OBJECT)]
Altar, lighted and garlanded.	6	[(0, 5, 'OBJECT')]	Altar, lighted and garland.	[(0, 5, OBJECT), (19, 26, OBJECT)]
Prize amphora on ornamental stand; within linear square and incuse square.	8	[(6, 13, 'OBJECT')]	Prize amphora on ornamental stand; within linear square and incuse square.	[(6, 13, OBJECT)]
Amphora with ribbed surface and crooked handles containing two ears of corn and poppy.	9	[(0, 7, 'OBJECT'), (40, 47, 'OBJECT'), (63, 67, 'PLANT'), (71, 75, 'PLANT'), (80, 85, 'PLANT')]	Amphora with ribbed surface and crooked handleholding two corn and poppy.	[(0, 7, OBJECT), (58, 62, PLANT), (67, 72, PLANT)]

Preprocessing

design_id	design_en	s	subject_class	p	o	object_class
9	Amphora with ribbed surface and crooked handles containing two ears of corn and poppy.	amphora	OBJECT	holding	poppy	PLANT

# Approach and Methodology



# Approach

## Overview

### **Self Attention Generative Pre-Trained Transformer (GPT):**

using the whole design description in parallel  
with the learned/generalized rules  
to predict the next token

### **Specific Few Shot Prompting**

Tailored prompts  
with semantic and syntactic relevant examples

### **Minimal Response**

Specific Data only  
to reduce API costs

# Methodology

## Overview

### Iterative Development Process

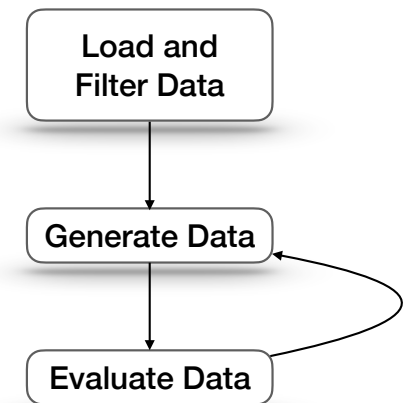
Continuous prompt refinement

### Evaluation and Feedback Loop

Adjustments based on validation results

### Goal of Pipeline

Handle variations, deliver results



# Workflow and Processing

# Workflow

## Overview

## Data Preparation

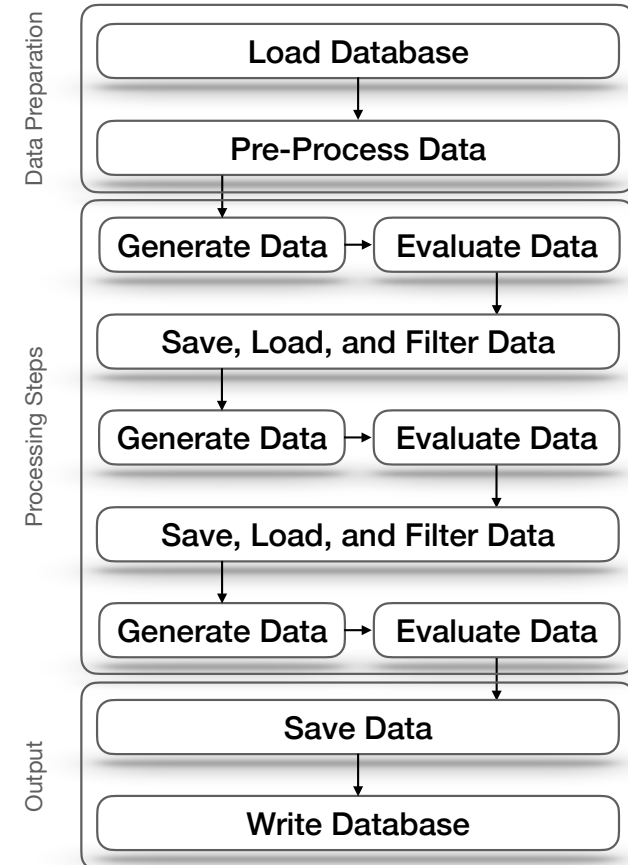
SQL database reading and preprocessing

## Processing Steps

Prompts for generating and evaluating data

## Output

JSON files with merged results



# Processing

## Steps

### Step 0: Refinement of NER Data

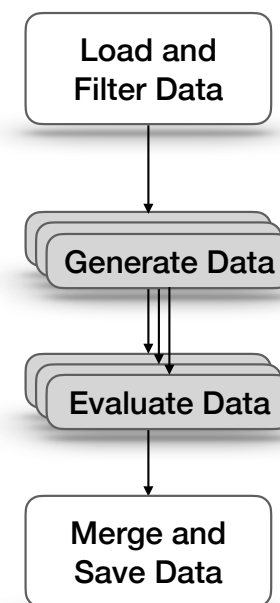
Refine entities identification for RE use

### Step 1: Identification of Entity Pairs

Identify possible subject-object pairs

### Step 2: Relation Extraction

Extract relations (predicates) between identified pairs



# Processing

## Details

### ∇ Processing Steps

**Prompts** with examples and input data

Responses, cleaned and merged with main data frame

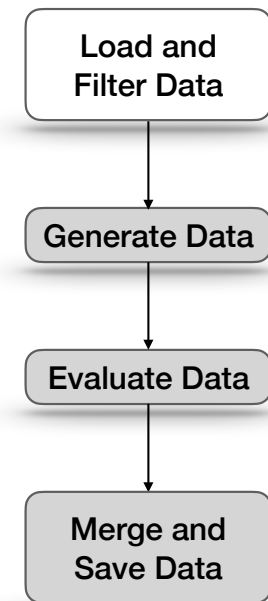
### Generation and Evaluation

**Generation:** Use specific prompts to generate desired data

**Evaluation:** Use prompts to evaluate results — rating and reasoning

### Output Handling

Clean and merge evaluation results with main data



# Prompting and Data Generation

# Prompts Example

Example:

```
{
  design_id: 101, // Unique identifier of the design
  Original Design: "Asclepius resting on left, on wing serpent to right.",
  Original List of Strings: [("Asclepius", "PERSON"), ("serpent", "ANIMAL")],
  Enhanced List of Strings: [("Asclepius", "PERSON"), ("serpent", "ANIMAL"), ("wing", "OBJECT")]
}
```

Now, enhance the following designs:

```
"""
for _, entry in batch.iterrows():
    prompt += f"""
    {{
      design_id: {entry['id']}, // Unique identifier of the design
      Original Design: "{entry['design_en']}",
      Original List of Strings: {entry['list_of_strings']}
    }}
    """
    prompt += """
```

16

```
def enhance_objects_in_designs(data: pd.DataFrame, batch_size: int) -> list:
    prompts = []
    for i in range(0, len(data), batch_size):
        batch = data.iloc[i:i + batch_size]
        prompt = """
        You are an expert extraction algorithm for numismatic design descriptions.
        Your goal is to enhance the list of identified objects in the following designs.
        You will be provided with a design description and a list of objects, and you will output JSON objects containing the following information:

        {
          design_id: int, // Unique identifier of the design
          new_list_of_strings: [(string, string)] // Enhanced list of objects in the form of tuples: (entity, class)
        }

        Focus on identifying all semantically meaningful objects within each design,
        according to the categories "PERSON", "OBJECT", "ANIMAL", "PLANT".
        Do not include terms that describe the coin itself or are redundant.
        Only consider significant elements of the design.
        Consider each design as distinct.
        Remove any objects that are less significant or redundant in the context of the design description,
        but mention each entity at least once if it contributes to the overall meaning.

        For example, in a design of "Nude Aphrodite standing facing, head right, holding her breast with right hand and pudenda with left hand":
        - "hand" is less significant because it is a part of the actions (holding) rather than a standalone meaningful object in the design, but
        - "head" is also less significant as it is a common part of the figure and does not add unique semantic value in this context unless it is specified.

        Example:
        {
          design_id: 36, // Unique identifier of the design
          Original Design: "Nude Aphrodite standing facing, head right, holding her breast with right hand and pudenda with left hand; to left",
          Original List of Strings: [("Aphrodite", "PERSON"), ("head", "OBJECT"), ("breast", "OBJECT"), ("hand", "OBJECT"), ("hand", "OBJECT"), ("pudenda", "OBJECT")],
          Enhanced List of Strings: [("Aphrodite", "PERSON"), ("breast", "OBJECT"), ("pudenda", "OBJECT"), ("Eros", "PERSON"), ("dolphin", "ANIMAL")]
        }

        Example:
        {
          design_id: 8, // Unique identifier of the design
          Original Design: "Prize amphora on ornamental stand; within linear square and incuse square.",
          Original List of Strings: [("amphora", "OBJECT")],
          Enhanced List of Strings: [("amphora", "OBJECT"), ("stand", "OBJECT")]
        }

        Example:
        {
          design_id: 101, // Unique identifier of the design
          Original Design: "Asclepius resting on left, on wing serpent to right.",
          Original List of Strings: [("Asclepius", "PERSON"), ("serpent", "ANIMAL")],
          Enhanced List of Strings: [("Asclepius", "PERSON"), ("serpent", "ANIMAL"), ("wing", "OBJECT")]
        }

        For each design, provide the enhanced list of strings in the form of tuples: [(entity, class), ...].

        Respond only with the following fields for each design:
        {
          design_id: int, // Unique identifier of the design
          new_list_of_strings: [(string, string)] // Enhanced list of objects in the form of tuples: (entity, class)
        }

        Now, enhance the following designs:
        """
        for _, entry in batch.iterrows():
            prompt += f"""
            {{
              design_id: {entry['id']}, // Unique identifier of the design
              Original Design: "{entry['design_en']}",
              Original List of Strings: {entry['list_of_strings']}
            }}
            """
        prompt += """
        Notes:
        - Objects should be atomic and not compound terms. For example, horn of ammon should be represented as the key "horn" with the class "ANIMAL".
        - Persons should be named as they are, like Alexander the Great or Antoninus Pius, and not as "Alexander", "Great" or "Antoninus", "Pius".

        Respond with the design_id and the enhanced list of strings in valid JSON format for each design, like this:
        {
          "design_id": 36,
          "new_list_of_strings": [("Aphrodite", "PERSON"), ("breast", "OBJECT"), ("pudenda", "OBJECT"), ("Eros", "PERSON"), ("dolphin", "ANIMAL")]
        }
        """
        prompts.append(prompt)
```



# Data Generation

Step 0: Refinement of NER Data



## Objective

Clean and refine entity identification

design\_id, design\_en  
11, "Athena ... holding apple ..."



design\_id, entities  
[("Athena", "PERSON"), ("apple", "OBJECT"),...]

## Prompt

Enhance identified objects in design descriptions  
by focusing on significant elements and removing redundancies

# Data Generation

## Step 1: Identification of Entity Pairs

Generate Data

Evaluate Data

design\_id, design\_en, entities  
11, "Athena ... Apollo", [{"Athena", "PERSON"}, ...]

### Objective

Identify subject-object pairs  
using refined NER data.

design\_id, so\_id, s, subject\_class, o, object\_class  
11, "a", "Athena", "PERSON", "apple", "OBJECT"  
11, "b", "Apollo", "PERSON", "dolphin", "ANIMAL"

### Prompt

Extract semantically meaningful subject, object pairs  
of entities from entities list based on design descriptions

# Data Generation

## Step 2: Relation Extraction

Generate Data

Evaluate Data

design\_id, so\_id, design\_en, s, subject\_class, o, object\_class  
11, "a", "Athena ...", [("Athena", "PERSON"), ("apple",...),...]

design\_id, so\_id, predicate  
11, "a", "Athena", "holding"

### Objective

Extract predicates — relationships  
between identified subject-object pairs.

### Prompt

Identify the most likely predicates for pairs  
explicitly mentioned in the design descriptions.

# Data Evaluation

# Data Evaluation

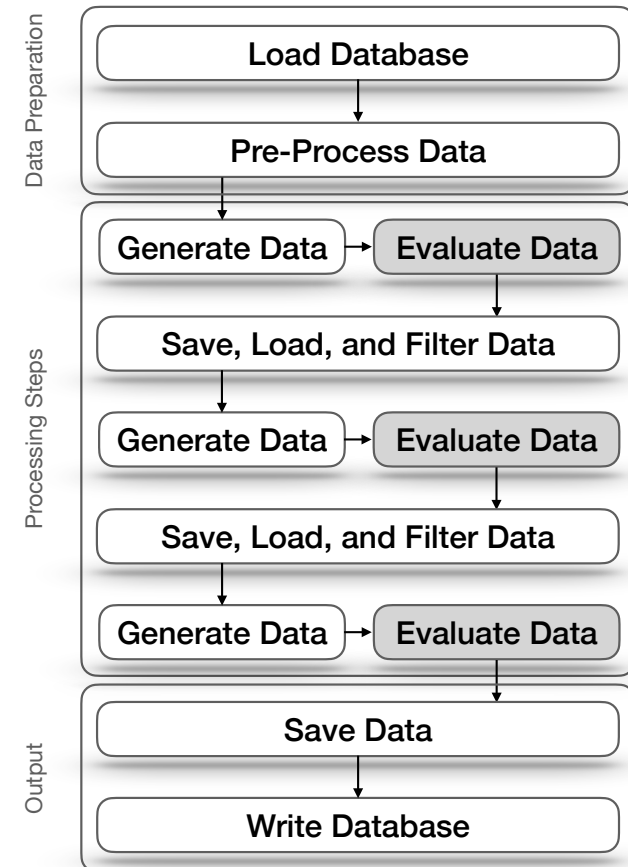
## Overview

### Purpose

Obtain a rating of the results  
by a different model / LLM instance  
Filter out bad and edge cases

### Application

For each of the three data generation steps  
with tailored prompting



# Data Evaluation

## Rating

### Relevance and Correctness

Prompts rate the enhanced lists of objects  
and subject-object pairs

### Predicate Validity

Validity of extracted subject-predicate-object triples  
and propose - implicit - predicate necessary

### Rating System

Use a scale of [-1, 0, 1]  
and provide reasoning comment for each rating

design\_id, design\_en, entities  
11, "Athena ...", [{"Athena", "PERSON"}, ...]

design\_id, so\_id, s, subject\_class, o, ...  
11, "a", "Athena", "PERSON", "apple", ...  
11, "c", "Athena", "PERSON", "horse", ...

design\_id, so\_id, predicate  
11, "a", "Athena", "holding"  
11, "c", "Athena", "wearing"

design\_id, so\_id, vote,  
comment, propose  
11, "a", 1, "based", "NULL"  
11, "c", -1, "wtf?", "seated\_on"

# Development Process

# Development Process

Data and API

## Filtering and preprocessing data

Adapting existing code

## OpenAI API

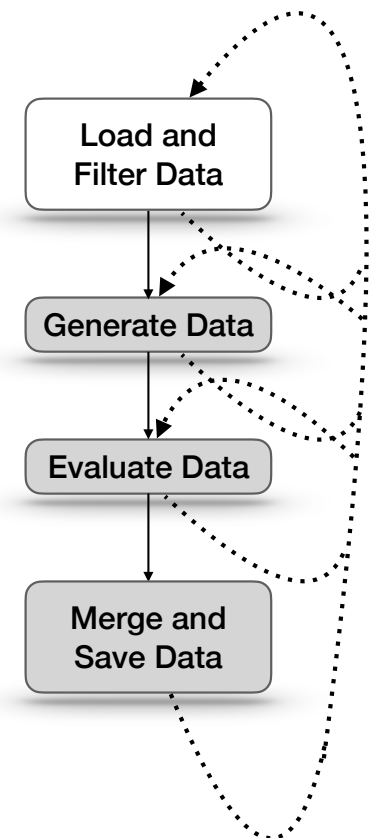
Calculation of tokens and cost management

Refactoring workflow for (new) batch API

for lower cost and controlled processing of large data

## Cleaning and formatting responses for JSON output

Testing and re-adjusting for edge cases





# Development Process

Prompt Creation and Adjustment

## Continuous refinement of prompts

### Balancing restrictions

Overly restrictive prompts degrade results

### Refinement of prompts

### Cover variations in designs

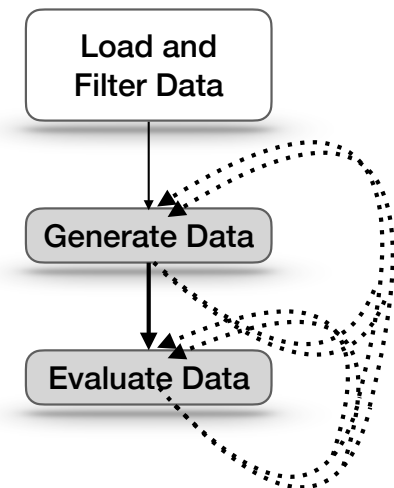
Again, refine prompts

### Evaluation of low ratings and?

Yes, refine the prompts...

## Continuous refinement of prompts

### Balancing restrictions



## Results and Findings

# Results

## Overview

### Processed 3200 Designs

Pipeline successfully tested for processing many designs

All 3200 designs were processed in one run → 6941 triples  
using the OpenAI *batch* API with GPT-4o

### Triples (6941) Rating

5845 (0.84) correct → Rating 1

706 (0.10) not sure → Rating 0

390 (0.06) wrong → Rating -1

# **Findings**

## **Challenges**

### **Prototype Performance**

Worked well with the RE database

masking and inferring predicates based on subject-object pairs

### **Challenges with Preprocessed NER Data**

Differences encountered with real, unfiltered preprocessed data

“One-step” predicate inference did not deliver

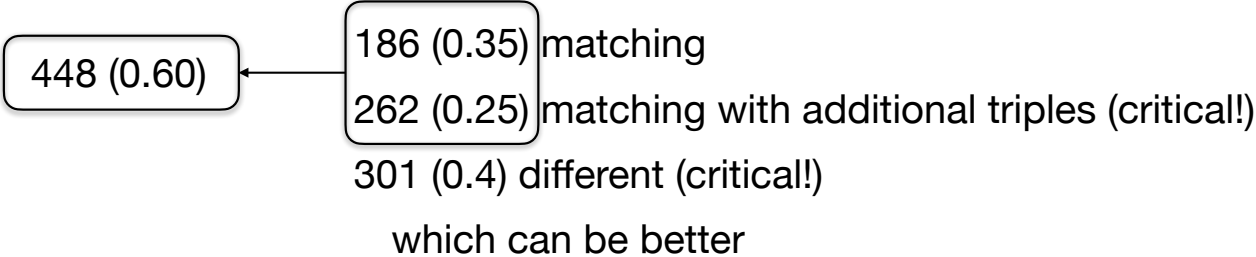
“Multi-step” pipeline

Two “pre-steps” to create subject-subject-object pairs

# Results and Findings

Results Compared to Ground Truth

## Compared 749 Designs



or worse

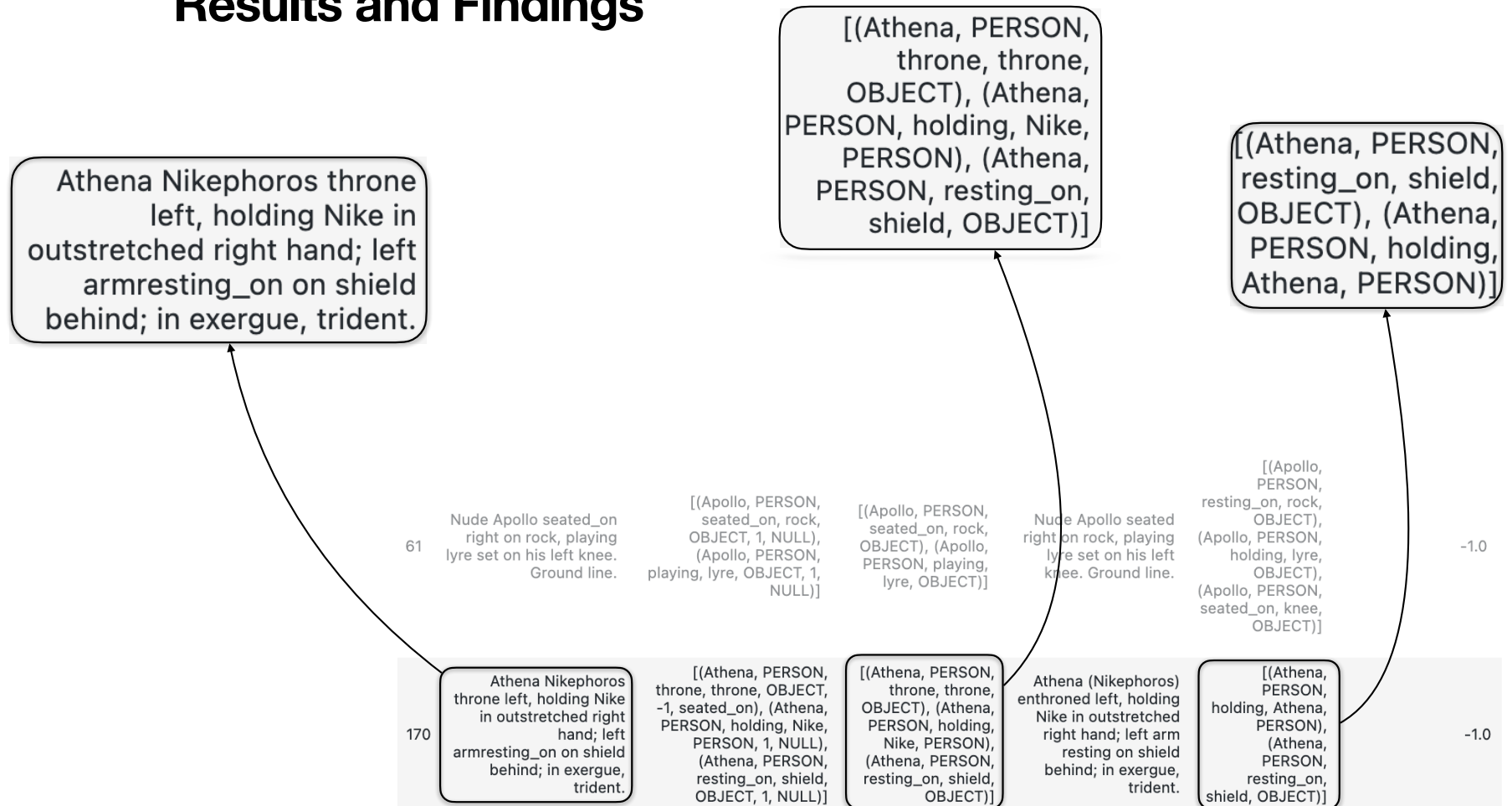
61	<div>Nude Apollo seated_on right on rock, playing lyre set on his left knee. Ground line.</div>	[(Apollo, PERSON, seated_on, rock, OBJECT, 1, NULL), (Apollo, PERSON, playing, lyre, OBJECT, 1, NULL)]	<div>[(Apollo, PERSON, seated_on, rock, OBJECT), (Apollo, PERSON, playing, lyre, OBJECT)]</div>	Nude Apollo seated right on rock, playing lyre set on his left knee. Ground line.	<div>[(Apollo, PERSON, resting_on, rock, OBJECT), (Apollo, PERSON, holding, lyre, OBJECT), (Apollo, PERSON, seated_on, knee, OBJECT)]</div>	-1.0
170	Athena Nikephoros throne left, holding Nike in outstretched right hand; left armresting_on on shield behind; in exergue, trident.	[(Athena, PERSON, throne, throne, OBJECT, -1, seated_on), (Athena, PERSON, holding, Nike, PERSON, 1, NULL), (Athena, PERSON, resting_on, shield, OBJECT, 1, NULL)]	[(Athena, PERSON, throne, throne, OBJECT), (Athena, PERSON, holding, Nike, PERSON), (Athena, PERSON, resting_on, shield, OBJECT)]	Athena (Nikephoros) enthroned left, holding Nike in outstretched right hand; left arm resting on shield behind; in exergue, trident.	[(Athena, PERSON, holding, Athena, PERSON), (Athena, PERSON, resting_on, shield, OBJECT)]	-1.0

# Results and Findings

Same same but different



## Results and Findings



# Results and Findings

Ground *un*-Truth

Demeter standing left,  
holding corn and poppy in  
right hand and torch in left  
arm.

356

Demeter standing left, holding corn and  
poppy in right hand and torch in left  
arm.

[(Demeter, PERSON, holding, corn,  
PLANT, 1, NULL), (Demeter, PERSON,  
holding, poppy, PLANT, 1, NULL),  
(Demeter, PERSON, holding, torch,  
OBJECT, 1, NULL)]

[(Demeter, PERSON,  
holding, corn, PLANT),  
(Demeter, PERSON,  
holding, poppy, PLANT),  
(Demeter, PERSON,  
holding, torch,  
OBJECT)]

[(Demeter, PERSON, holding,  
corn, PLANT), (Demeter,  
PERSON, holding, poppy,  
PLANT), (Demeter, PERSON,  
holding, torch, OBJECT)]

[(Demeter, PERSON,  
holding, double  
spear, OBJECT),  
(Demeter, PERSON,  
holding, poppy,  
PLANT), (Demeter,  
PERSON, holding,  
torch, OBJECT),  
(Demeter, PERSON,  
holding, Salonina,  
PERSON)]

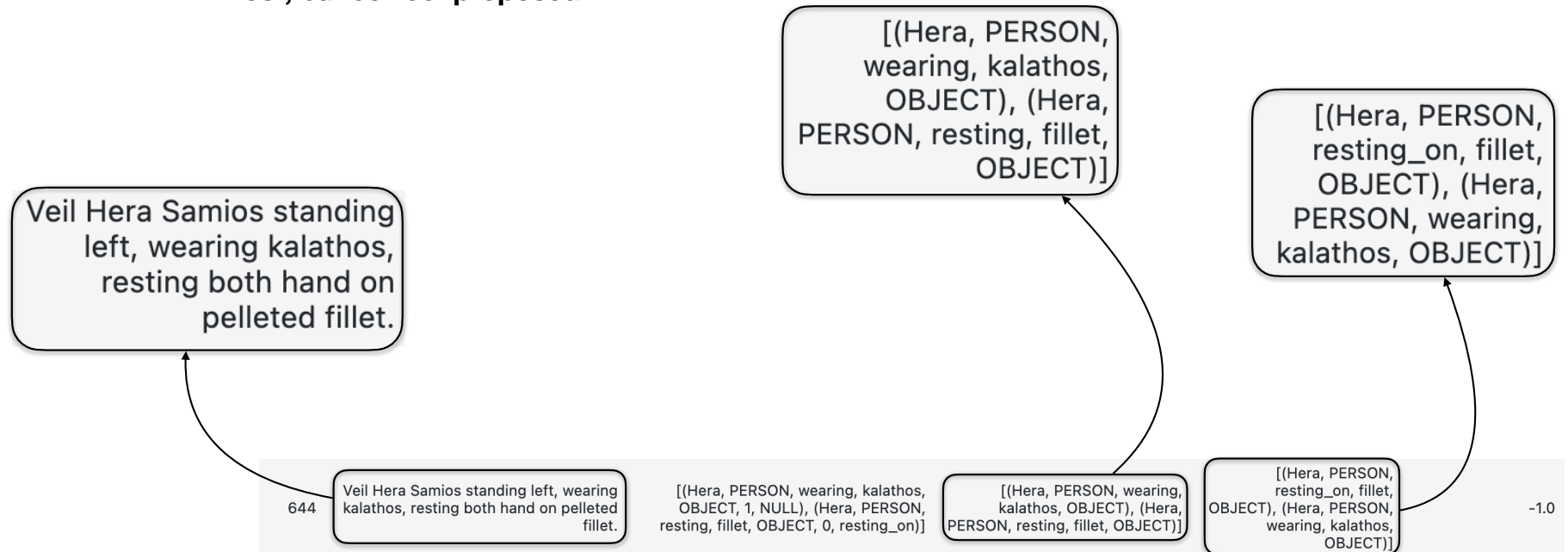
[(Demeter, PERSON,  
holding, double spear,  
OBJECT), (Demeter,  
PERSON, holding, poppy,  
PLANT), (Demeter,  
PERSON, holding, torch,  
OBJECT), (Demeter,  
PERSON, holding,  
Salonina, PERSON)]

-1.0



# Results and Findings

Almost; but correct proposed



# Results and Findings

More is more

Emperor Caracalla in military attire, standing facing, head left, wearing cuirass, flattering paludamentum and boot, holding Nike on globe with laurel wreath and branch in right hand, resting raised left on scepter; to left, captive, seated\_on left, wearing Phrygian cap.

4161

Emperor Caracalla in military attire, standing facing, head left, wearing cuirass, flattering paludamentum and boot, holding Nike on globe with laurel wreath and branch in right hand, resting raised left on scepter; to left, captive, seated\_on left, wearing Phrygian cap.

[(Emperor Caracalla, PERSON, in, military attire, OBJECT, 1, NULL), (Emperor Caracalla, PERSON, wearing, cuirass, OBJECT, 1, NULL), (Emperor Caracalla, PERSON, wearing, paludamentum, OBJECT, 1, NULL), (Emperor Caracalla, PERSON, wearing, boot, OBJECT, 1, NULL), (Emperor Caracalla, PERSON, holding, Nike, PERSON, 1, NULL), (Emperor Caracalla, PERSON, holding, globe, OBJECT, 1, NULL), (Emperor Caracalla, PERSON, holding, laurel wreath, OBJECT, 1, NULL), (Emperor Caracalla, PERSON, resting, scepter, OBJECT, 1, NULL), (captive, PERSON, wearing, Phrygian cap, OBJECT, 1, NULL)]

[(Emperor Caracalla, PERSON, in, military attire, OBJECT), (Emperor Caracalla, PERSON, wearing, cuirass, OBJECT), (Emperor Caracalla, PERSON, wearing, paludamentum, OBJECT), (Emperor Caracalla, PERSON, wearing, boot, OBJECT), (Emperor Caracalla, PERSON, holding, Nike, PERSON), (Emperor Caracalla, PERSON, holding, globe, OBJECT), (Emperor Caracalla, PERSON, holding, laurel wreath, OBJECT), (Emperor Caracalla, PERSON, resting, scepter, OBJECT), (captive, PERSON, wearing, Phrygian cap, OBJECT)]

[(Emperor Caracalla, PERSON, in, military attire, OBJECT), (Emperor Caracalla, PERSON, wearing, cuirass, OBJECT), (Emperor Caracalla, PERSON, wearing, paludamentum, OBJECT), (Emperor Caracalla, PERSON, wearing, boot, OBJECT), (Emperor Caracalla, PERSON, holding, Nike, PERSON), (Emperor Caracalla, PERSON, holding, globe, OBJECT), (Emperor Caracalla, PERSON, holding, laurel wreath, OBJECT), (Emperor Caracalla, PERSON, resting, scepter, OBJECT), (captive, PERSON, wearing, Phrygian cap, OBJECT)]

[(Caracalla, PERSON, wearing, cuirass, OBJECT), (Caracalla, PERSON, holding, Athena, PERSON), (Caracalla, PERSON, wearing, boot, OBJECT), (Caracalla, PERSON, wearing, paludamentum, OBJECT)]

[(Caracalla, PERSON, wearing, cuirass, OBJECT), (Caracalla, PERSON, holding, Athena, PERSON), (Caracalla, PERSON, wearing, boot, OBJECT), (Caracalla, PERSON, wearing, paludamentum, OBJECT)]

-1.0

# **Conclusion**

## **Summary**

### **Background and Approach**

### **Workflow, Processing and Prompting**

### **Data Generation and Evaluation**

### **Development and Findings**

# Conclusion

## Results

### Summary Result

Used *state of the art* NLP techniques  
Processed *huge* amount of data  
Extracted *meaningful* relations  
Created larger ground-*truth*



84% rated good  
10% rated indifferent  
6% rated wrong

### Summary Weakness

Proper measuring or metric, as  
ground truth comparison seems confusing  
Domain knowledge

Different → 301 (0.4)

Matching → 448 (0.6)

# **Impact**

## **Contribution and Future Work**

### **Contributions**

- Larger, more accurate RE database
- Improved ground truth for model training
- Potential applications in other datasets

### **Future Enhancements**

- Refining the pipeline
- Domain-specific improvements
  - Specialized prompts based on the comparison
  - Fine-tuned model approach
  - Quantifiable metrics

(Not? The End!

Thank you!

Questions?

Walkthrough?

Demo?