



DIPARTIMENTO  
DI INGEGNERIA  
DELL'INFORMAZIONE



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

PhD Application

# Document-level Event Extraction with Knowledge Injection

Presented by: M. Martinelli  
June 20<sup>th</sup>, 2024



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA



DIPARTIMENTO  
DI INGEGNERIA  
DELL'INFORMAZIONE

# Introduction



## Contemporary Data Landscape

Digital text sources are always-increasing in terms of volume and complexity.



## Information Overload

Information overload challenges our ability to process and utilize data effectively.



## Need for Effective IE

Information Extraction (IE) is essential for extracting information from unstructured text



## Contemporary Data Landscape

Digital text sources are always-increasing in terms of volume and complexity.



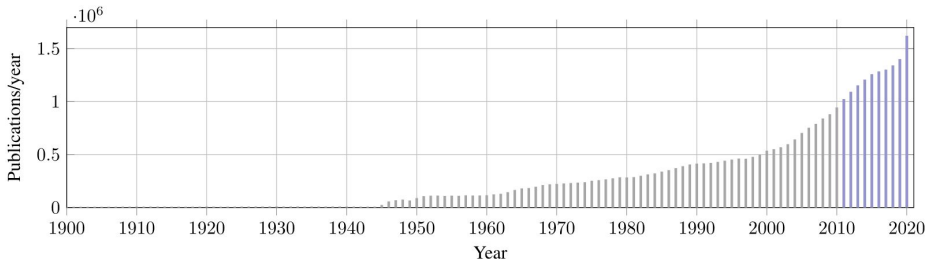
## Information Overload

Information overload challenges our ability to process and utilize data effectively.



## Need for Effective IE

Information Extraction (IE) is essential for extracting information from unstructured text



The annual rate of biomedical publications registered on PubMed





**Event Extraction (EE)** transforms unstructured text into a structured description of what happened, answering the who, what, when, where, why, and how questions about an occurrence in a structured manner.

## Possible domains of application:

Literature  
Processing



LLMs Training

Healthcare



News Reporting





**Event Extraction (EE)** transforms unstructured text into a structured description of what happened, answering the who, what, when, where, why, and how questions about an occurrence in a structured manner.

**In January 2023, protests in Atlanta, Georgia, escalated following the fatal shooting of an environmental activist, known as Tortuguita, by authorities during a raid, with protesters throwing rocks and Molotov cocktails at officers.**



**Event Extraction (EE)** transforms unstructured text into a structured description of what happened, answering the who, what, when, where, why, and how questions about an occurrence in a structured manner.

In **January 2023**, **protests** in **Atlanta, Georgia**, escalated following the **fatal shooting of an environmental activist, known as Tortuguita**, by **authorities** during a raid, with **protesters throwing rocks and Molotov cocktails at officers**.

Event Extraction ("Protest" event)	
Event Schema	Event Argument
<b>Date</b>	<b>January 2023</b>
<b>Location</b>	<b>Atlanta, Georgia</b>
<b>Protesters</b>	<b>Environmental activists</b>
<b>Cause</b>	<b>Fatal shooting by the authorities of an environmental activist, known as Tortuguita</b>
<b>Method</b>	<b>Throwing rocks and Molotov Cocktails at officers</b>



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA



DIPARTIMENTO  
DI INGEGNERIA  
DELL'INFORMAZIONE

# Approaches to EE

## “” Sentence-level EE (SEE)

In January 2023, protests in Atlanta, Georgia, escalated following the fatal shooting of an environmental activist, known as Tortuguita, by authorities during a raid, with protesters throwing rocks and Molotov cocktails at officers.

The protesters, opposing the construction of the 'Cop City' training complex in the forest area of Dekalb County, intensified their actions by setting fire to some police cruisers.

Governor Brian Kemp declared a state of emergency and authorized the deployment of National Guard troops in response to the growing unrest, leading to the arrest of a dozen protesters charged with domestic terrorism.



## Document-level EE (DEE)

## “” Sentence-level EE (SEE)



Document-level EE (DEE)

In **January 2023**, protests in **Atlanta, Georgia**, escalated following the **fatal shooting of an environmental activist, known as Tortuguita**, by authorities during a raid, with protesters **throwing rocks and Molotov cocktails at officers**.

The protesters, opposing the **construction of the 'Cop City' training complex** in the forest area of **DeKalb County**, intensified their actions by **setting fire to some police cruisers**.

Governor Brian Kemp declared a state of emergency and authorized the deployment of National Guard troops in response to the growing unrest, leading to the **arrest of a dozen protesters charged with domestic terrorism**.

Event Type: Protest	
Date	January 2023
Location	Atlanta, Georgia
Protesters	Environmental activists
Cause	Fatal Shooting of an environmental activist, known as Tortuguita
Method	Throwing rocks and Molotov Cocktails at officers

Event Type: Protest	
Location	DeKalb County
Cause	Construction of the 'Cop City' training complex
Damaged Property	Some police cruisers set to fire

Event Type: Protest	
Government Reaction	<ul style="list-style-type: none"><li>• State of emergency declared</li><li>• National Guard deployed</li></ul>
Arrested	A dozen protesters charged with domestic terrorism

“” Sentence-level EE (SEE)



## Document-level EE (DEE)





In **January 2023**, protests in **Atlanta, Georgia**, escalated following the **fatal shooting of an environmental activist, known as Tortuguita**, by authorities during a raid, with protesters **throwing rocks and Molotov cocktails at officers**.

The protesters, opposing the **construction of the 'Cop City' training complex** in the forest area of Dekalb County, intensified their actions by **setting fire to some police cruisers**.





Governor Brian Kemp **declared a state of emergency and authorized the deployment of National Guard troops** in response to the growing unrest, leading to the **arrest of a dozen protesters charged with domestic terrorism**.

Event Type: Protest	
Date	January 2023
Location	Atlanta, Georgia
Protesters	Environmental activists
Cause	Construction of the 'Cop City' training complex
Method	Throwing rocks and Molotov Cocktails at officers
Arrested	A dozen protesters charged with domestic terrorism
Government Reaction	<ul style="list-style-type: none"><li>• State of emergency declared</li><li>• National Guard deployed</li></ul>
Casualties and Losses	Fatal Shooting of an environmental activist, known as Tortuguita
Damaged Property	Some police cruisers set to fire

## 1 Maintain extraction effectiveness

-  Many lexical and syntactical possibilities
-  Coreference resolution
-  Event merging
-  Entity disambiguation

## 2 Scarce availability of training data

-  Annotated data are expensive to produce
-  Under-representation of certain event types
-  Under-representation of abbreviations and acronyms
-  Under-representation of domain-specific terminology

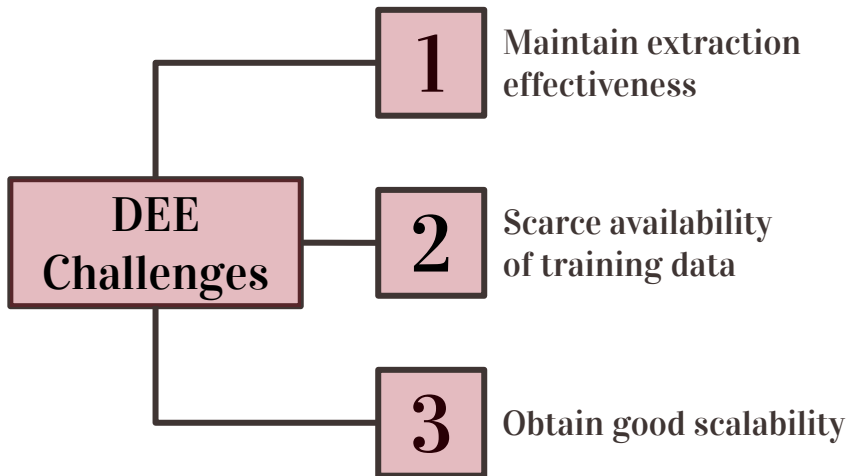




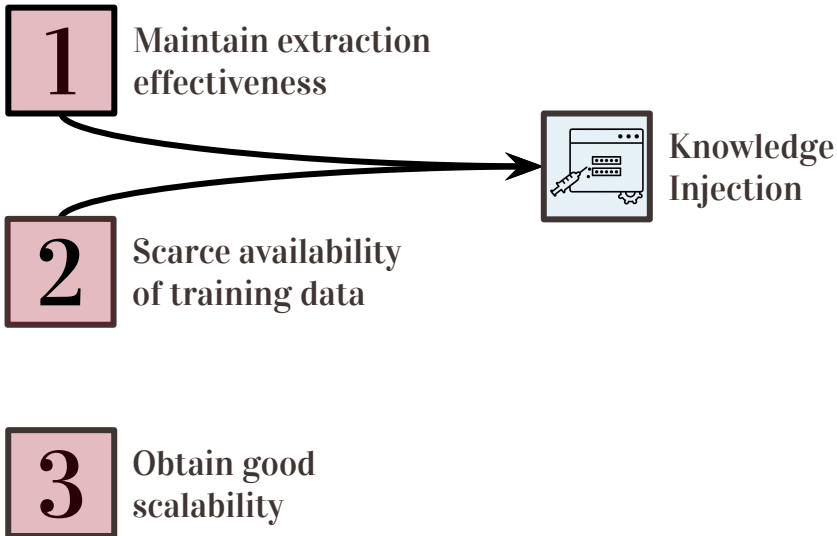
UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA



# Research Directions



# Research Directions





## Knowledge Injection (KI)

In January 2023, protests in Atlanta, Georgia, escalated following the fatal shooting of an environmental activist, known as **Tortuguita**, by authorities during a raid, with protesters throwing rocks and Molotov cocktails at officers.

The protesters, opposing the construction of the '**Cop City**' training complex in the forest area of Dekalb County, intensified their actions by setting fire to some police cruisers.

Governor Brian Kemp declared a state of emergency and authorized the deployment of National Guard troops in response to the growing unrest, leading to the arrest of a dozen protesters charged with domestic terrorism.

### Cop City:

- Description: Police training facility in **Atlanta, Georgia, USA**
- Also known as: **Atlanta Public Safety Training Center**

### Tortuguita:

- Description: Activist
- Also known as: **Manuel Teran**
- Member of: **Defend the Atlanta Forest**
- Date of death: **18 January 2023**
  - Age of subject at event: **26 years old**

### Defend the Atlanta Forest:

- Description: Social movement
- Slogan: **Defend the forest, Stop cop city**



## Knowledge Injection (KI)

### Cop City:

- Description: Police training facility in **Atlanta, Georgia, USA**
- Also known as: **Atlanta Public Safety Training Center**

### Tortuguita:

- Description: Activist
- Also known as: **Manuel Teran**
- Member of: **Defend the Atlanta Forest**
- Date of death: **18 January 2023**
  - Age of subject at event: **26 years old**

### Defend the Atlanta Forest:

- Description: Social movement
- Slogan: **Defend the forest, Stop cop city**

### Event Type: Protest

<b>Date</b>	<b>Subsequent to 18th of January 2023</b>
<b>Location</b>	<b>Atlanta, Georgia, USA</b>
<b>Protesters</b>	<b>"Defend the Atlanta Forest" environmental activists</b>
<b>Cause</b>	<b>Construction of the 'Cop City' training complex, officially known as 'Atlanta Public Safety Training Center'</b>
<b>Slogan</b>	<b>• Defend the forest • Stop cop city</b>
<b>Method</b>	<b>Throwing rocks and Molotov Cocktails at officers</b>
<b>Arrested</b>	<b>A dozen protesters charged with domestic terrorism</b>
<b>Government Reaction</b>	<b>• State of emergency declared • National Guard deployed</b>
<b>Casualties and Losses</b>	<b>Fatal Shooting of the 26 years old environmental activist Manuel Teran, known as Tortuguita</b>
<b>Damaged Property</b>	<b>Some police cruisers set to fire</b>



## Document-level EE (DEE)

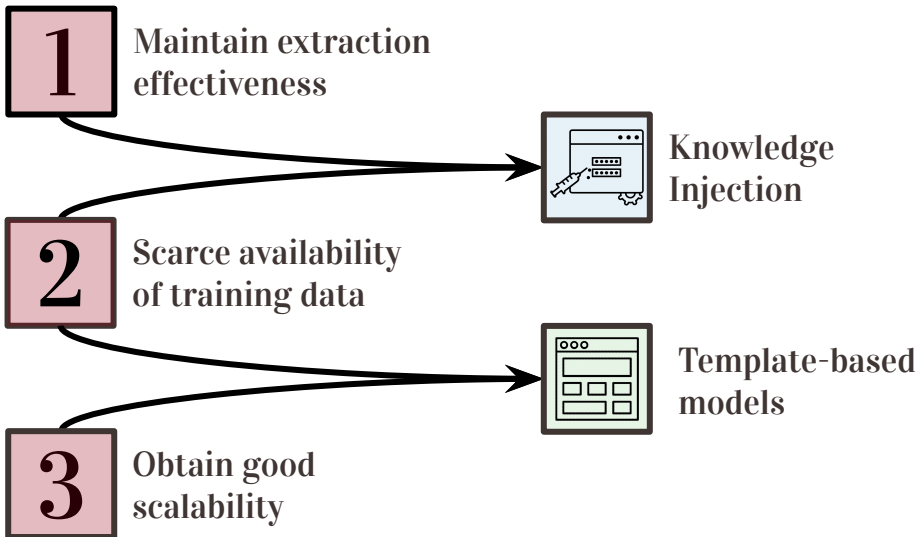


## Knowledge Injection

Event Type: Protest	
Date	January 2023
Location	Atlanta, Georgia
Protesters	Environmental activists
Cause	Construction of the 'Cop City' training complex
Method	Throwing rocks and Molotov Cocktails at officers
Arrested	A dozen protesters charged with domestic terrorism
Government Reaction	<ul style="list-style-type: none"><li>State of emergency declared</li><li>National Guard deployed</li></ul>
Casualties and Losses	Fatal Shooting of an environmental activist, known as Tortuguita
Damaged Property	Some police cruisers set to fire

Event Type: Protest	
Date	Subsequent to 18th of January 2023
Location	Atlanta, Georgia, USA
Protesters	"Defend the Atlanta Forest" environmental activists
Cause	Construction of the 'Cop City' training complex, officially known as 'Atlanta Public Safety Training Center'
Slogan	<ul style="list-style-type: none"><li>Defend the forest</li><li>Stop cop city</li></ul>
Method	Throwing rocks and Molotov Cocktails at officers
Arrested	A dozen protesters charged with domestic terrorism
Government Reaction	<ul style="list-style-type: none"><li>State of emergency declared</li><li>National Guard deployed</li></ul>
Casualties and Losses	Fatal Shooting of the 26 years old environmental activist Manuel Teran, known as Tortuguita
Damaged Property	Some police cruisers set to fire

# Research Directions





## Template-based (TB) models

### 1. Template Creation



The protest took place on **[Date]** in **[Location]**, organized by **[Protesters]**.

The primary cause of this demonstration was **[Cause]**, with participants unified under the slogan **[Slogan]**.

Protest methods employed included **[Method]**, and **[Damaged Property]** was reported as damaged property.

In response, the government **[Government Reaction]**.

The event led to **[Arrested]** being arrested and to **[Casualties and Losses]**.





## Template-based (TB) models

### 1. Template Creation



The protest took place on [Date] in [Location], organized by [Protesters].  
The primary cause of this demonstration was [Cause], with participants unified under the slogan [Slogan].  
Protest methods employed included [Method], and [Damaged Property] was reported as damaged property.  
In response, the government [Government Reaction].  
The event led to [Arrested] being arrested and to [Casualties and Losses].

### 2. Prompt to LLM



Perform the Event Extraction task by completing the template below by extracting relevant information about a "Protest" event from the provided news article.  
Focus on accurately filling in the specific details within square brackets to summarize the key points of the event.  
If you cannot find the match for a specific field, fill with the symbol 'X'

Template to complete:

...

News Article:

...



## Template-based (TB) models

### 1. Template Creation



The protest took place on [Date] in [Location], organized by [Protesters].  
The primary cause of this demonstration was [Cause], with participants unified under the slogan [Slogan].  
Protest methods employed included [Method], and [Damaged Property] was reported as damaged property.  
In response, the government [Government Reaction].  
The event led to [Arrested] being arrested and to [Casualties and Losses].

### 2. Prompt to LLM



Perform the Event Extraction task by completing the template below by extracting relevant information about a "Protest" event from the provided news article.  
Focus on accurately filling in the specific details within square brackets to summarize the key points of the event.  
If you cannot find the match for a specific field, fill with the symbol 'X'

Template to complete:

...  
News Article:  
...

### 3. LLM Output



The protest took place on [January 2023] in [Atlanta, Georgia], organized by [environmental activists].  
The primary cause of this demonstration was [opposition to the construction of the 'Cop City' training complex in Dekalb County forest area], with participants unified under the slogan [X].  
Protest methods employed included [throwing rocks and Molotov cocktails at officers, setting fire to police cruisers], and [police cruisers] was reported as damaged property..  
In response, the government [declared a state of emergency and authorized the deployment of National Guard troops].  
The event led to [a dozen protesters] being arrested and to [the fatal shooting of the environmental activist Tortuguita].



## Knowledge Injection



Controlling the introduction of noise



Maintaining the accuracy of injected knowledge



Avoiding the sub-optimal injection of knowledge



## Template-based models



Adapting to variations in events descriptions



Learning from data and improving over time



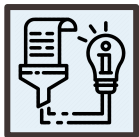
Managing syntactic and semantic variations



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA



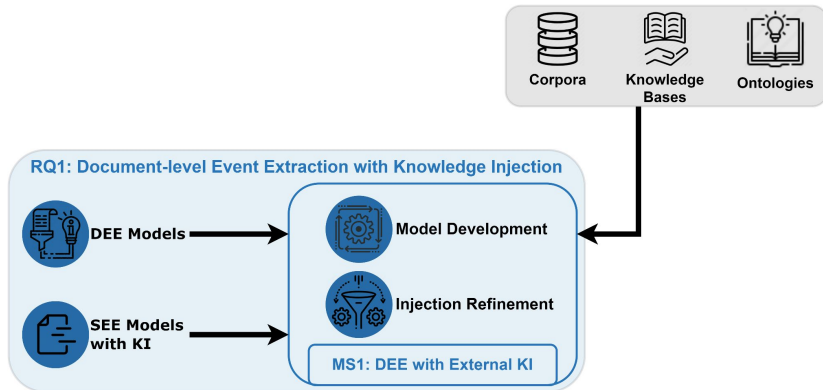
# Research Questions



RQ1

## Knowledge Injection in DEE Models

To improve the extraction of entities and relations





## Knowledge Injection in DEE Models

To improve the extraction of entities and relations

RQ1



Fine-tuning a LM on externally annotated corpus



Augmenting word embeddings with additional features



Enriching input data before model processing

DBpedia spotlight



PubTator

OpenTapioca



cTakes



RQ1

## Knowledge Injection in DEE Models

To improve the extraction of entities and relations



RQ2

## Knowledge Injection in TB Models

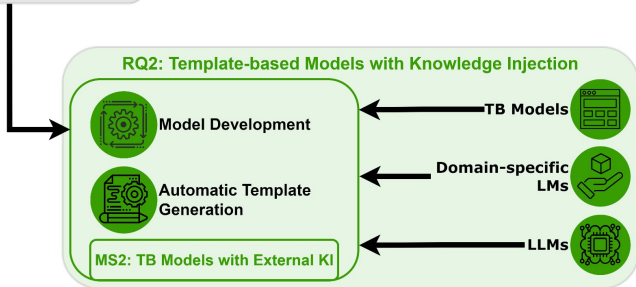
To enhance the robustness and flexibility



RQ2

## Knowledge Injection in TB Models

To enhance the robustness and flexibility







RQ2

## Knowledge Injection in TB Models

To enhance the robustness and flexibility



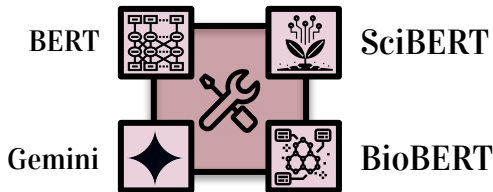
Augmenting event templates with external knowledge



Enhancing the accuracy of augmented templates using techniques from automatic KBs enrichment



Leveraging LMs to automate template generation





RQ1

## Knowledge Injection in DEE Models

To improve the extraction of entities and relations



RQ2

## Knowledge Injection in TB Models

To enhance the robustness and flexibility



RQ3

## Performance Evaluation

To evaluate the impact of the injection approaches on effectiveness and efficiency



RQ3

## Performance Evaluation

To evaluate the impact of the injection approaches on effectiveness and efficiency

### RQ3: Models Evaluation



Evaluation Metric Analysis



Ablation Study



Computational Trade-off  
Evaluation

MS3: DEE Evaluation Framework



## RQ3

### Performance Evaluation

To evaluate the impact of the injection approaches on effectiveness and efficiency



Evaluating effectiveness using precision, recall, and F1-score as metrics



Conducting ablation studies to understand the impact of the various injection approaches



Evaluating computational load to determine the most scalable approaches

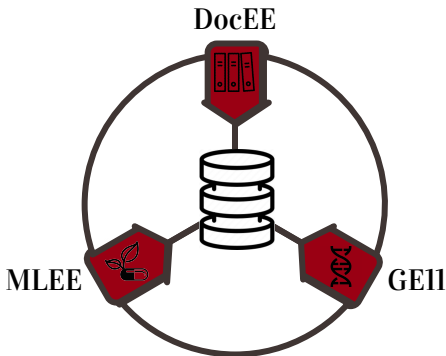


RQ3

## Performance Evaluation

To evaluate the impact of the injection approaches on effectiveness and efficiency

## Test Collections



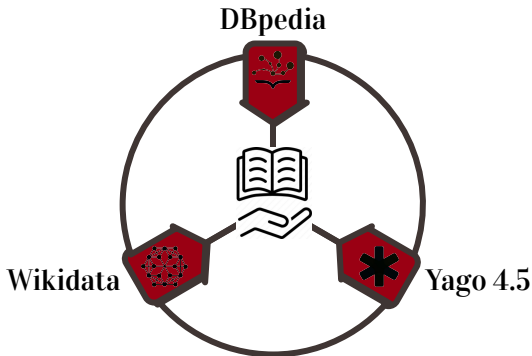


RQ3

## Performance Evaluation

To evaluate the impact of the injection approaches on effectiveness and efficiency

## Knowledge Sources – General Purpose





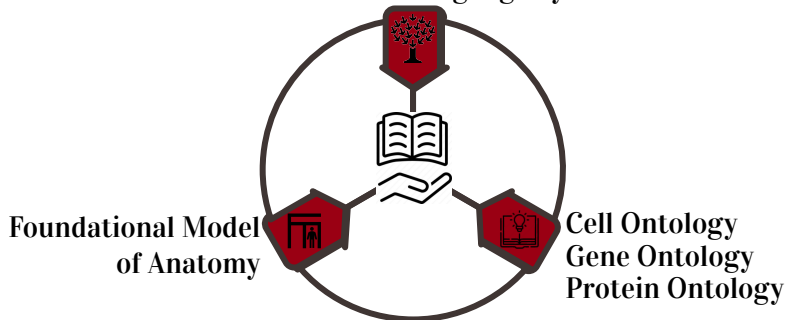
RQ3

## Performance Evaluation

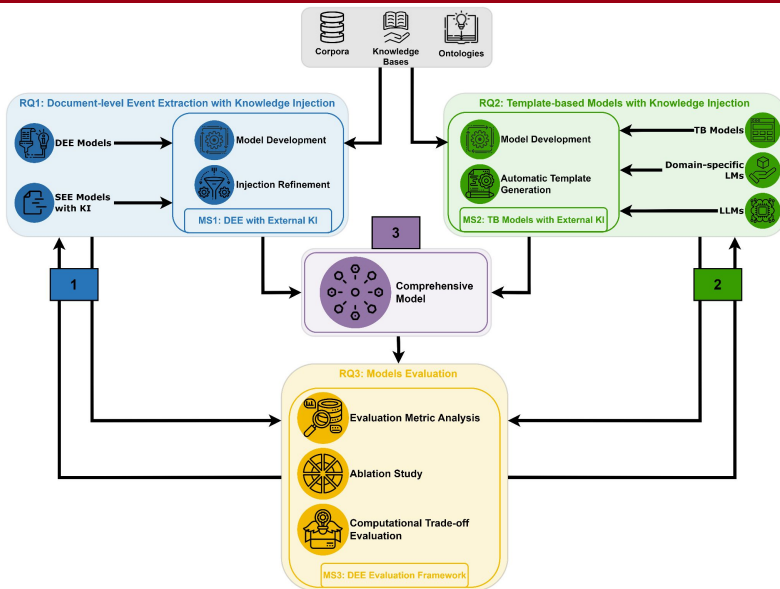
To evaluate the impact of the injection approaches on effectiveness and efficiency

## Knowledge Sources – Biomedical Domain

### Unified Medical Language System



# Research Framework







UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA



# Research Timeline

## WP3 – MS2

Develop an efficient  
TB model with KI for  
DEE

## WP5

Communication  
and  
Dissemination

## MS1

Literature  
Review

## WP2 – MS1

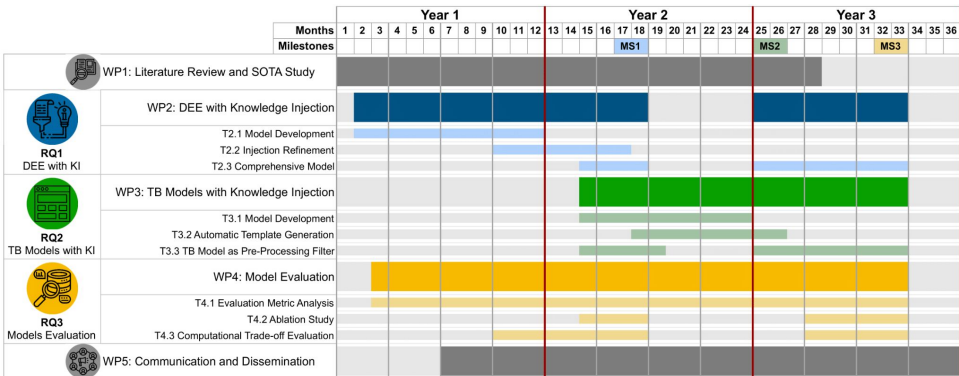
Develop an effective  
DEE model with KI

## WP4 – MS3

Develop an  
evaluation  
framework for  
DEE models



# Gantt Chart





# Thank you for the attention!

# Questions?