



Finanziato  
dall'Unione europea  
NextGenerationEU



Ministero  
dell'Università  
e della Ricerca



Italiadomani  
PIANO NAZIONALE  
DI RIPRESA E RESILIENZA



Future  
Artificial  
Intelligence  
Research

# Multimodal Knowledge Graph Extraction with/for Grounded World Models

*Aldo Gangemi<sup>1,2\*</sup>, Stefano De Giorgis<sup>1</sup>, Alessandro Russo<sup>1</sup>  
Institute of Cognitive Science and Technology, CNR<sup>1</sup>, Rome, University of  
Bologna<sup>2</sup>*

`aldo.gangemi, stefano.degiorgis, alessandro.russo@cnr.it`

*Thanks to Luana Bulla, Arianna Graciotti, Anna Sofia Lippolis, Misael Mongiovì, Andrea Nuzzolese, Luigi Asprino, Francesco Petri, Valentina Presutti, Diego Reforgiato Recupero*



Finanziato  
dall'Unione europea  
NextGenerationEU



Ministero  
dell'Università  
e della Ricerca



Italiadomani

PIANO NAZIONALE  
DI RIPRESA E RESILIENZA



Future  
Artificial  
Intelligence  
Research

## Multimodal multilayer meaning



photo by Evan Vucci/The Associated Press

A man in an elegant suit in the center, with a raised arm, surrounded by some people gathering around his body. Some blood on his ear, an American flag on the background.

***What we may know perceptually (physical patterns)***



Finanziato  
dall'Unione europea  
NextGenerationEU



Ministero  
dell'Università  
e della Ricerca



Italiadomani

PIANO NAZIONALE  
DI RIPRESA E RESILIENZA



Future  
Artificial  
Intelligence  
Research

## Multimodal multilayer meaning



photo by Evan Vucci/The Associated Press

A political figure or leader, doing a victory gesture, surrounded by guards / agents protecting him. Some blood spilling from his ear, an American flag on the background.

***What we may know commonsensically (social patterns)***



Finanziato  
dall'Unione europea  
NextGenerationEU



Ministero  
dell'Università  
e della Ricerca



Italiadomani

PIANO NAZIONALE  
DI RIPRESA E RESILIENZA



Future  
Artificial  
Intelligence  
Research

## Multimodal multilayer meaning



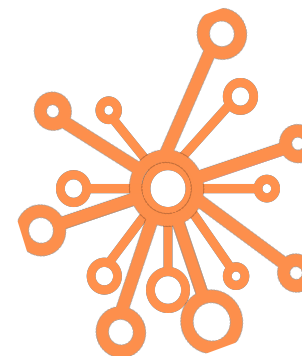
photo by Evan Vucci/The Associated Press

The assassination attempt on Donald Trump.

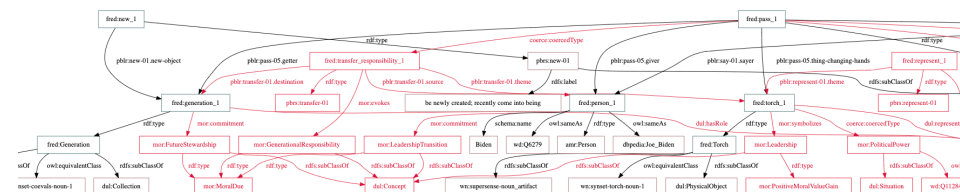
***What we may know as a specific social fact (news-like)***

## Research goals

- *Content* from **multiple modalities**, pragmatically described in **natural language** (NL), which acts as a hub from content to knowledge (automated extraction)
- *Knowledge\** as a **supramodal hybrid representation** for joint logical and geometrical inference
- *Ground World Models* (passive, active) as **shared reference knowledge**
- *Knowledge graphs* representing both **explicit (base KG)** and **implicit (eXtended KG - XKG)** facts at different **linguistic layers** (literal, presupposed, implicated, figurative, etc.), and **ontological layers** (material, cognitive, social, abstract)



Extended Knowledge Graph



\* Knowledge: logically represented data for automated reasoning



Finanziato  
dall'Unione europea  
NextGenerationEU



Ministero  
dell'Università  
e della Ricerca



Italiadomani  
PIANO NAZIONALE  
DI RIPRESA E RESILIENZA



Future  
Artificial  
Intelligence  
Research

How to extract knowledge from arbitrary signal, and make sense of it [11]?

*Direct to graph* [1]

*Mediated by language* (Multimodal generative AI)

*Passive vs. Active AI* [2]

[1] Asprino L., Daga E., Gangemi A., & Mulholland P. Knowledge graph construction with a façade: a unified method to access heterogeneous data sources on the web. *ACM Transactions on Internet Technology*, 2023 (Q1)

[2] Pezzulo, G, Parr, T, Cisek, P, Clark, A, Friston, K. Generating meaning: active inference and the scope and limits of passive AI, *Trends in Cognitive Sciences*, 2024 (Q1)

[11] Ciroku, F, et al. Automated multimodal sensemaking: Ontology-based integration of linguistic frames and visual data. *Computers in Human Behavior*, 2024 (Q1)

# AMR to FRED

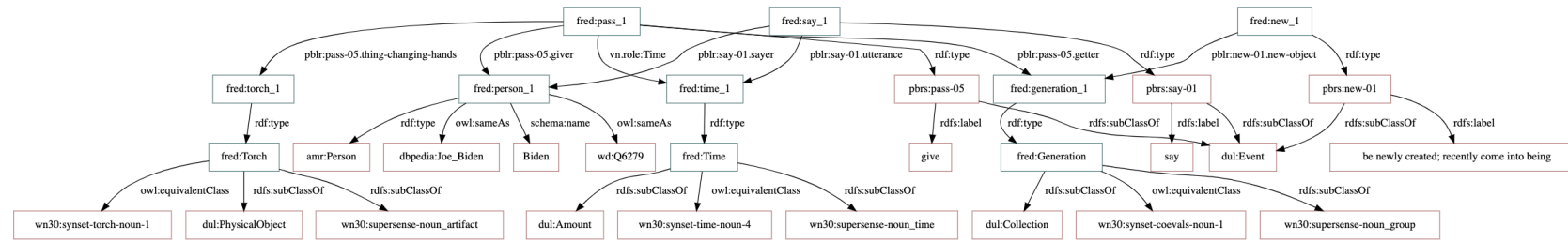
Your text

Biden said it was time to pass the torch to a new generation.

This is the result of converting the AMR graph to a FRED-like knowledge graph using AMR2FRED.

Zoom: double-click to zoom-in and Shift + Double-click to zoom-out, or use Shift + Scroll

Pan: drag left mouse button



## Extended graphs

Symbolic coercions

Metonymic coercions

Moral value-driven coercions

Full graph - All extensions

Graphical

Full graph

This is full graph built by merging the base graph with all the extensions provided by the LLM.

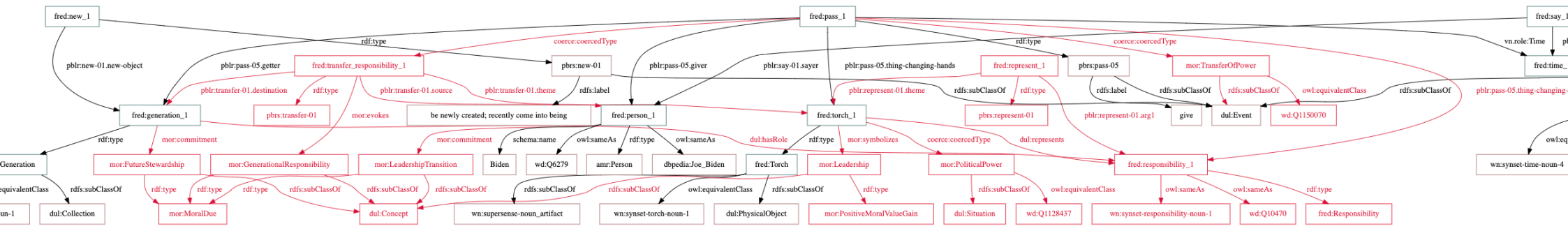
```
fred:new_1 a pbl:NEW-01 ;
pbl:new-01.new-object fred:generation_1 .

fred:pass_1 a pbrs:pass-05 ;
coerce:coercedType fred:transfer_responsibility_1,
mor:TransferOfPower ;
vn.role:Time fred:time_1 ;
pbl:pass-05.getter fred:generation_1 ;
pbl:pass-05.giver fred:person_1 ;
pbl:pass-05.thing-changing-hands fred:responsibility_1,
fred:torch_1 .

fred:represent_1 a pbrs:represent-01 ;
pbl:represent-01.arg1 fred:responsibility_1 ;
pbl:represent-01.theme fred:torch_1 .

fred:say_1 a pbrs:say-01 ;
pbl:say-01.sayer fred:person_1 ;
pbl:say-01.utterance fred:time_1 .

mor:commitment a owl:ObjectProperty .
mor:evokes a owl:ObjectProperty .
```





Finanziato dall'Unione europea  
NextGenerationEU



Ministero dell'Università e della Ricerca

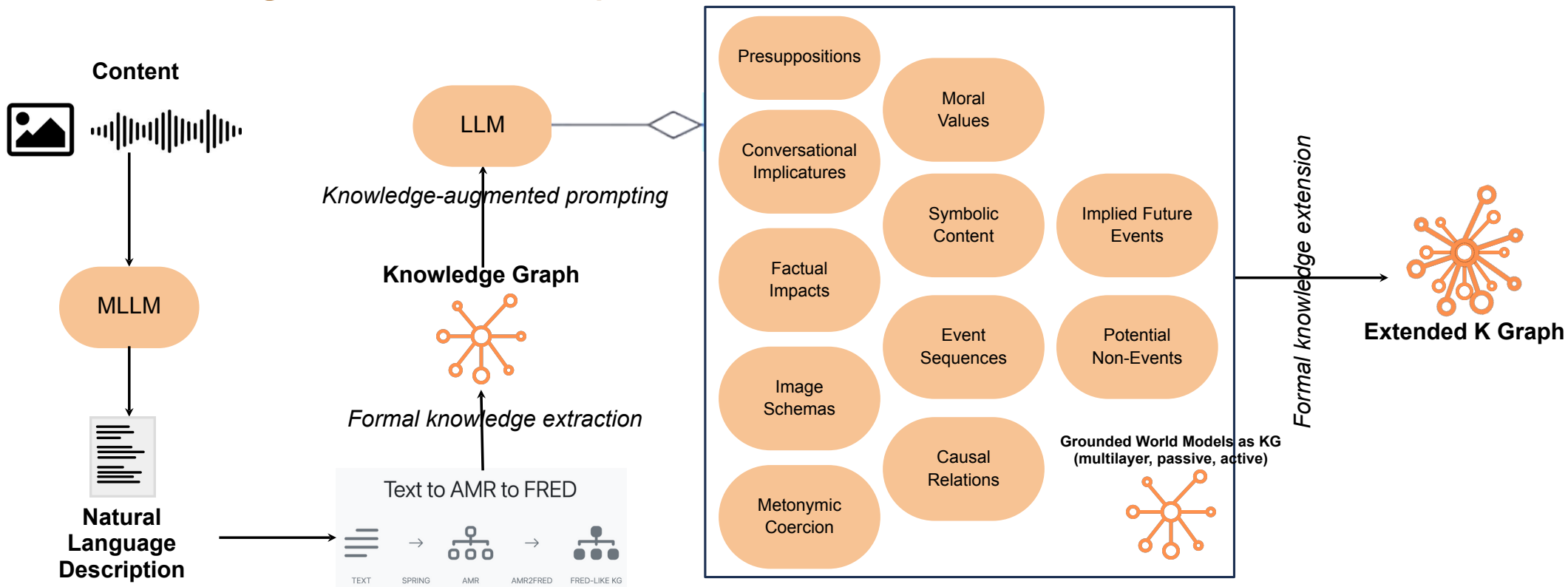


Italiadomani  
PIANO NAZIONALE DI RIPRESA E RESILIENZA



Future Artificial Intelligence Research

## Knowledge Extension Pipeline 1/6



Off-the-shelf formal graph extensions from automated LLM prompting (11 implicit motifs)





Finanziato dall'Unione europea  
NextGenerationEU



Ministero dell'Università e della Ricerca

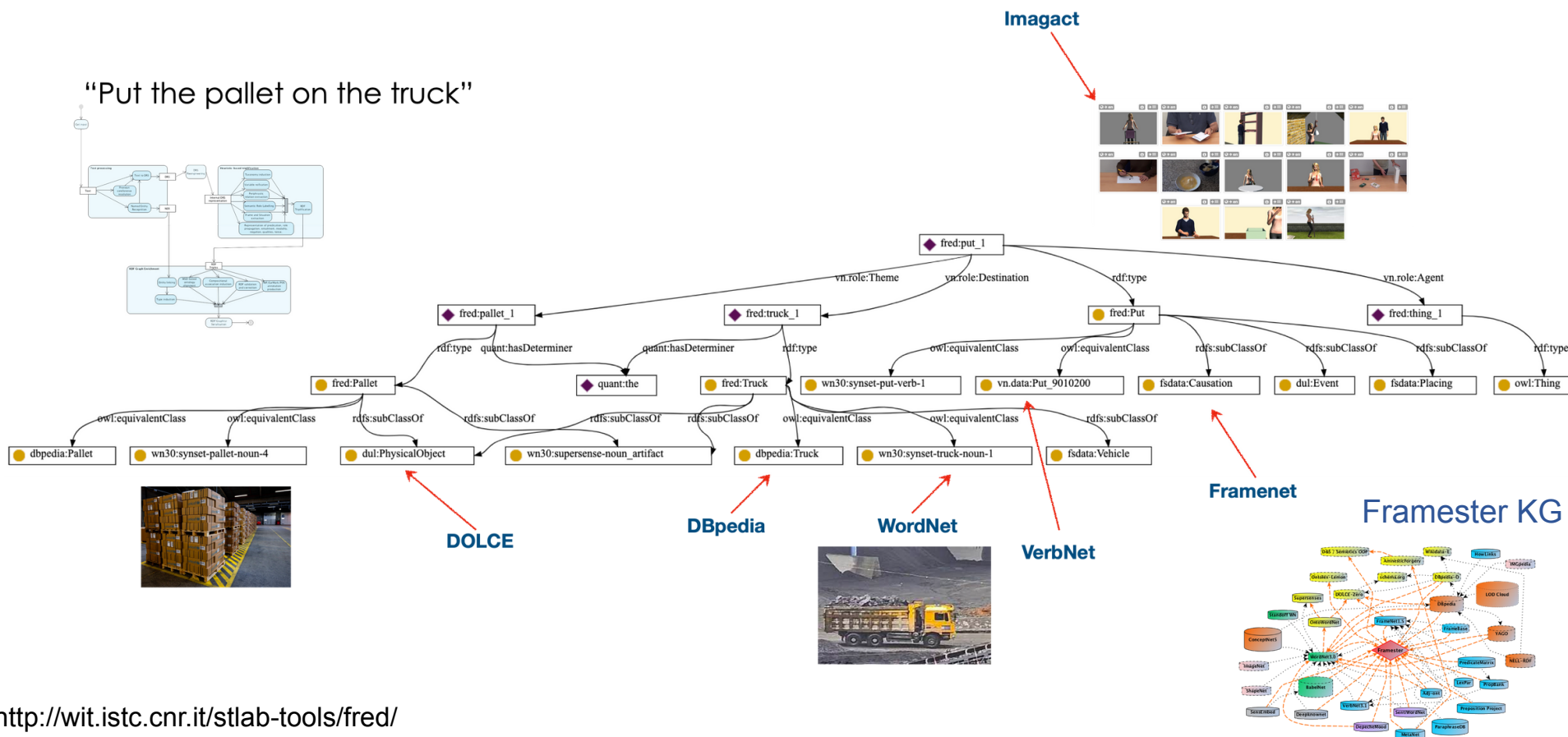


Italiadomani  
PIANO NAZIONALE DI RIPRESA E RESILIENZA

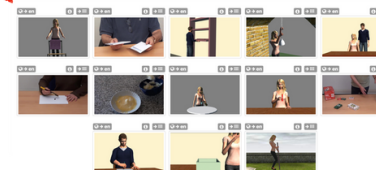


Future Artificial Intelligence Research

"Put the pallet on the truck"

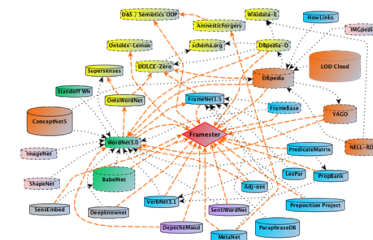


Imagact



Framester

Framester KG



KG Extraction and Alignment to GWM [3][4]

<http://wit.istc.cnr.it/stlab-tools/fred/>

[3] Gangemi et al. Semantic Web Machine Reading, *Semantic Web Journal*, 2017 (Q1)

[4] Gangemi et al. Identifying motifs for evaluating open knowledge extraction on the Web. *Knowledge-Based Systems*, 2016 (Q1)



Finanziato  
dall'Unione europea  
NextGenerationEU



Ministero  
dell'Università  
e della Ricerca



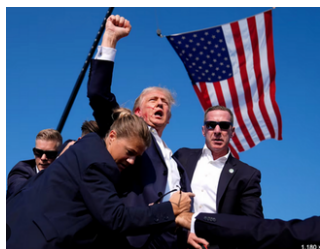
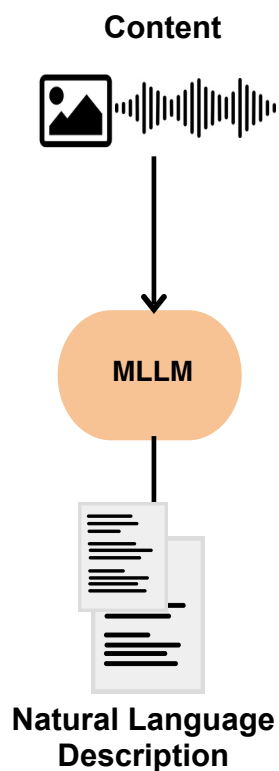
Italiadomani

PIANO NAZIONALE  
DI RIPRESA E RESILIENZA



Future  
Artificial  
Intelligence  
Research

## Knowledge Extension Pipeline – 2/6



A high-profile political figure raises his fist in defiance, surrounded by vigilant security personnel during a public rally, with a large American flag prominently displayed in the background, emphasizing themes of patriotism and protection.

Clever Prompts:

1. Describe the overt, implicit and symbolic content of the image. Do not use full stops, Do not use "this image depict"-like expressions, describe directly the content of the image
2. Describe this photo in detail, including gestures, participants, public roles, situations, bodily position, and general implied meaning.
3. ...



Finanziato  
dall'Unione europea  
NextGenerationEU



Ministero  
dell'Università  
e della Ricerca



Italiadomani  
PIANO NAZIONALE  
DI RIPRESA E RESILIENZA



Future  
Artificial  
Intelligence  
Research

Test it! [5]



<https://arco.istc.cnr.it/itaf/>

The screenshot shows the web application interface for 'Text to AMR to FRED'. At the top, there are two buttons: 'FROM TEXT' (selected) and 'FROM IMAGE'. Below this is the title 'Text to AMR to FRED'. A process flow diagram shows the following steps: TEXT (represented by a hamburger menu icon) → SPRING → AMR (represented by a tree graph icon) → AMR2FRED → FRED-LIKE KG (represented by a tree graph icon) → LLM → EXTENDED KG (represented by a tree graph icon). Below the diagram is a text input field containing the sentence: 'A high-profile political figure raises his fist in defiance, surrounded by vigilant security personnel during a public rally, with a large American flag prominently displayed in the background, emphasizing themes of patriotism and protection.' Below the input field is a 'Settings' button and a large blue 'Convert text' button.

[5] Gangemi et al., Text2AMR2FRED, a Tool for Transforming Text into RDF/OWL Knowledge Graphs via Abstract Meaning Representation, International Semantic Web Conference, 2023 (A)



Finanziato dall'Unione europea  
NextGenerationEU

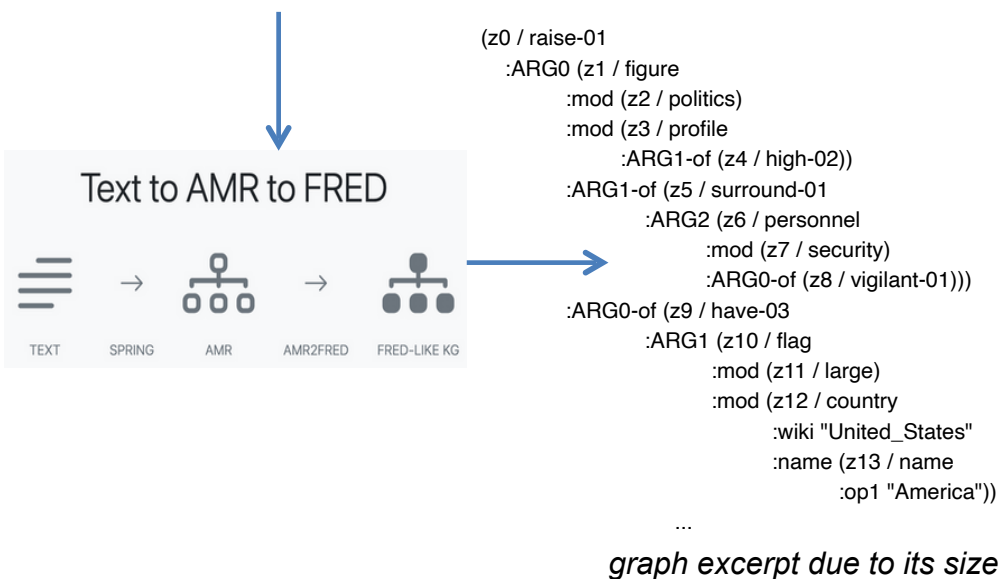


Ministero dell'Università e della Ricerca



# Knowledge Extension Pipeline – 3/6

A high-profile political figure raises his fist in defiance, surrounded by vigilant security personnel during a public rally, with a large American flag prominently displayed in the background, emphasizing themes of patriotism and protection.



FROM TEXT FROM IMAGE

### Image to Text to AMR to FRED

IMAGE → LLM → TEXT → SPRING → AMR → AMR2FRED → FRED-LIKE KG → LLM → EXTENDED KG

Given an image, this tool will get a textual description from an LLM and parse it into an [AMR](#) (Abstract Meaning Representation) graph, using [SPRING](#). The AMR graph is then converted into an RDF/OWL knowledge graph that follows [FRED](#)'s knowledge representation patterns, using [AMR2FRED](#).

Provide your image

Choose file trump\_ap.jpg

Settings

Convert image

#### Image to text

This is the result of generating a textual description for your image.

Energetic political rally scene: central figure with raised fist and open mouth expressing passion, surrounded by supporters and security personnel, American flag prominently displayed against blue sky, symbolizing patriotism and national pride, stark contrast between animated speaker and stoic guards in dark suits and sunglasses, woman in foreground taking notes suggests documentation of event, overall composition conveys power, authority, and fervent support, imagery evokes themes of leadership, populism, and American political spectacle, visual elements combine to create sense of momentum and rally atmosphere, underscoring the performative nature of modern campaign events

#### Text to AMR

Your text

PENMAN Graph

This is the result of converting your text to AMR using SPRING. The AMR graph is serialized using the [PENMAN notation](#).

```

:ARG2 (z29 / and
  op1 (z30 / patriotism)
  op2 (z31 / pride-01)

```

Copy AMR

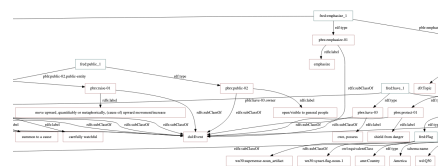
Collaboration with SPOKE 5 for SPRING AMR parser [6]

[6] Bevilacqua, M et al., One SPRING to rule them both: Symmetric AMR semantic parsing and generation without a complex pipeline, AAAI, 2021 (A\*)



## Knowledge Extension Pipeline – 5/6

Formal knowledge extraction



Knowledge-Augmented Generation (KAG)

Task:

Your goal is to extend KG with more knowledge that can be assumed from T, but it is not explicit.

Using the elements of KG, and PropBank and WordNet elements as linking points, add any further elements you need to extract implicit knowledge about:

*Conversational implicatures.*

Conversational implicatures, in the sense of Grice's pragmatics.

Here are natural language inference examples:

1) She won't necessarily get the job -> She will possibly get the job

...

Formal knowledge extension

You receive a text "T" and a frame-based knowledge graph "KG" that is the extraction of factual knowledge from T.

T:  
{{ Text }}

KG:  
{{ KG }}

Your goal is to extend KG with more knowledge that can be assumed from T, but it is not explicit.



conversational\_implicatures.prompt



Finanziato dall'Unione europea  
NextGenerationEU



Ministero dell'Università e della Ricerca



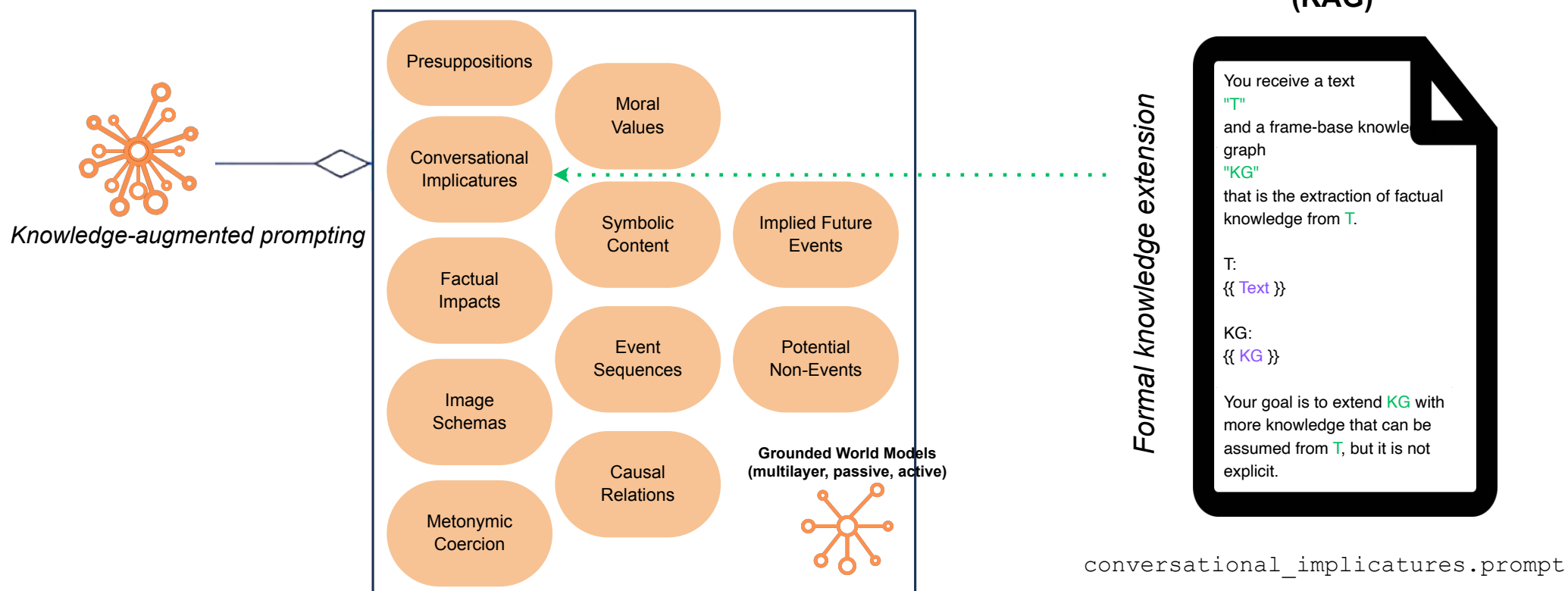
Italiadomani  
PIANO NAZIONALE DI RIPRESA E RESILIENZA



Future Artificial Intelligence Research

## Knowledge Extension Pipeline – 6/6

### Knowledge-Augmented Generation (KAG)









Finanziato  
dall'Unione europea  
NextGenerationEU



Ministero  
dell'Università  
e della Ricerca



Italiadomani  
PIANO NAZIONALE  
DI RIPRESA E RESILIENZA



Future  
Artificial  
Intelligence  
Research

BASE GRAPH

fred:raise\_1 a pbrs:raise-01 ;  
pblr:raise-01.agent fred:figure\_2 ;  
pblr:raise-01.logical-subject fred:fist\_1 .

fred:surround\_1 a pbrs:surround-01 ;  
pblr:surround-01.thing-surrounded fred:personnel\_1 ;  
pblr:surround-01.thing-surrounding fred:figure\_2 .

**The figure has a raised fist. The personnel surrounds the figure.**

fred:fist\_1 coerce:coercedType  
mor:Resistance, mor:Strength .

fred:flag\_1 coerce:coercedType  
mor:NationalIdentity, mor:Patriotism .

**The fist is a symbol of Resistance and strength.  
The flag is a symbol of national identity and patriotism.**

SYMBOLIC  
COERCION

FACTUAL  
IMPACT

fred:figure\_2 impact:hasExpectedEmotion impact:Defiance ;  
impact:hasExpectedSocialImpact impact:IncreasedPublicSupport .

fred:personnel\_1 impact:hasExpectedEmotion impact:Tension ;  
impact:hasExpectedMentalState impact:Alertness .

**The figure has expected social impact an increased public support.  
The personnel has expected mental state alertness.**



Formal knowledge  
extension

MORAL  
VALUES

fred:assert\_1 a pbrs:assert-02 ;  
mor:evokes mor:PowerDemonstration ;  
pblr:assert-02.agent fred:figure\_2 ;  
pblr:assert-02.topic fred:power\_1 .



**The figure assertion is a demonstration of power.**

fred:support\_1 a owl:ObjectProperty,  
pbrs:support-01 ;  
pblr:support-01.supported fred:figure\_2 ;  
pblr:support-01.supporter fred:audience\_1 .

**We can expect a raise in the support from the audience.**

IMPLIED  
FUTURE  
EVENTS

## Evaluation methods and downstream tasks

- **A new benchmark:** experts create sample extensions, manually based on the heuristics, and we automatically estimate the distance between human and AI using a large KG embedding space and measuring the similarity between the basic KG and XKG
- **Silver standard for precision@k** of generated and inferred triples: experts review the extended triples for correctness, relevance, and meaningfulness for each heuristics 
- **Stability and Reproducibility:** replicating the XKG extensions for the different heuristics to establish how deterministic a LLM functionality is, and the difference between multiple LLMs 



Extended K Graph

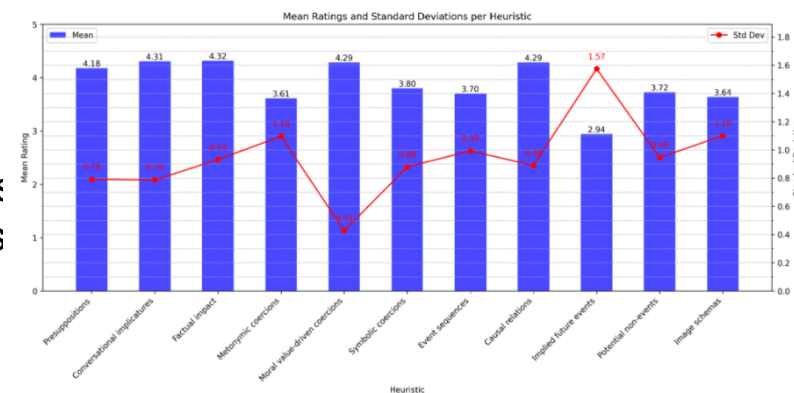
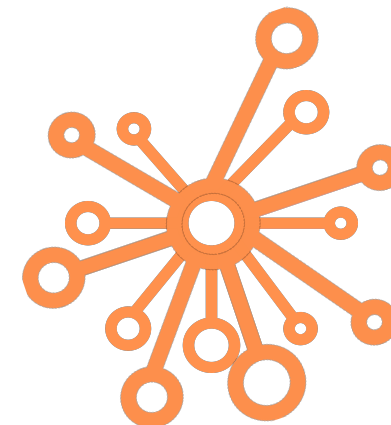


Figure 6: Mean Ratings and Standard Deviation per Heuristics.

## Evaluation methods and downstream tasks

- **Semantic similarity:** using an independent LM, measure the similarity between the original text and a text generated back from XKG: high distance → improper extension; low distance → minimal extension ✓
- **Logical validity:** OWL reasoners check the consistency/coherence of the extended KG ✓
- **Foundational ontology-grounded validity:** deeper reasoning via full alignment of predicates to foundational ontologies like DOLCE Zero (as made for DBpedia in [7]) ✓
- **Downstream tasks:** *deeper question answering (e.g., FigurativeQA), nuanced content recommendation, exploratory search, automated hate speech detection, fact checking, legal document analysis, medical diagnosis support, risk assessment, VR interaction, synchronic and diachronic analysis of domain knowledge [8], etc.*



Extended K Graph

[7] Paulheim H., & Gangemi A., Serving DBpedia with DOLCE—more than just adding a cherry on top, ISWC 2015 (A)

[8] Gangemi A., Graciotti A., Meloni A., Nuzzolese A., Presutti V., Reforgiato D., MusicBO, an application of Text2AMR2FRED to the Musical Heritage domain, ESWC 2024 (B)



Finanziato  
dall'Unione europea  
NextGenerationEU



Ministero  
dell'Università  
e della Ricerca



Italiadomani

PIANO NAZIONALE  
DI RIPRESA E RESILIENZA



Future  
Artificial  
Intelligence  
Research

## Critical issues

- Shared GWM KGs like Wikidata are difficult for LLM-based alignment (opaque ID): *more traditional approaches* work (e.g., BLINK for the base KG of iTAF)
- Active GWMs need to be adapted to produce a KG output associated with shared and local KGs: *ongoing work* for spatial and causal GWM and Bayesian networks
- KAG requires a good amount of data, and prompts are heavy (and eventually costly under intense usage): moving to *fine-tuning open source models* is a natural move to reduce data at prompting time
- KAG with our heuristics runs well with Claude's Sonnet 3.5, but other large models (incl. GPT4o) do not: *moving to open source, mixture of experts, and small model stitching* in the medium term
- Eventually, *Agentic MLLMs* might get the full-monty, and KAG gets a built-in feature that helps making them robust, flexible, interpretable, and sustainable



Finanziato  
dall'Unione europea  
NextGenerationEU



Ministero  
dell'Università  
e della Ricerca



Italiadomani  
PIANO NAZIONALE  
DI RIPRESA E RESILIENZA



Future  
Artificial  
Intelligence  
Research

## Ongoing work: Scene Recognition, Hybrid Reasoning, Active GWMs

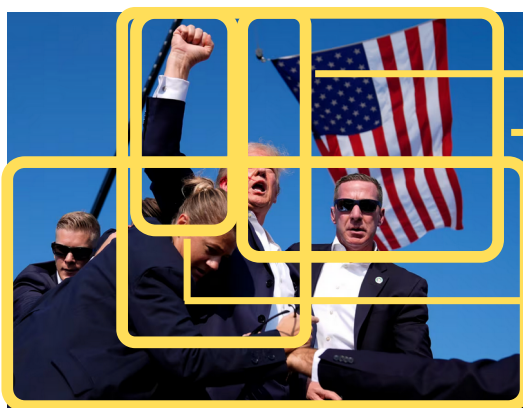


photo by Evan Vucci/The Associated Press

:AttackVictim , :PoliticalFigure  
:AttackContext ,  
:AmericanFlag :symbolOf :Patriotism .  
:VictimDefender , :SecretServiceAgent  
:RaisedFist :symbolOf :Triumph .  
:RaisedFist :partOf :PoliticalFigure .

*Cf. WP10.3 on EASG, e.g. within  
industrial cascade call*

- Native hybrid, frame-based reasoning can be provided by **novel neuro-symbolic reasoners** such as Sandra [9]
- Open-ended **knowledge extensions**, e.g., norms, trust (WP8), affordances, etc.
- Further extensions for **dynamic in-context representation, anticipation and inference** can be assisted by dedicated active GWMs [2][10]

[9] Lazzari, N., De Giorgis, S., Gangemi, A., & Presutti, V. (2024). Sandra--A Neuro-Symbolic Reasoner Based On Descriptions And Situations. arXiv preprint arXiv:2402.00591. [→ Collaboration with SPOKE 8](#)

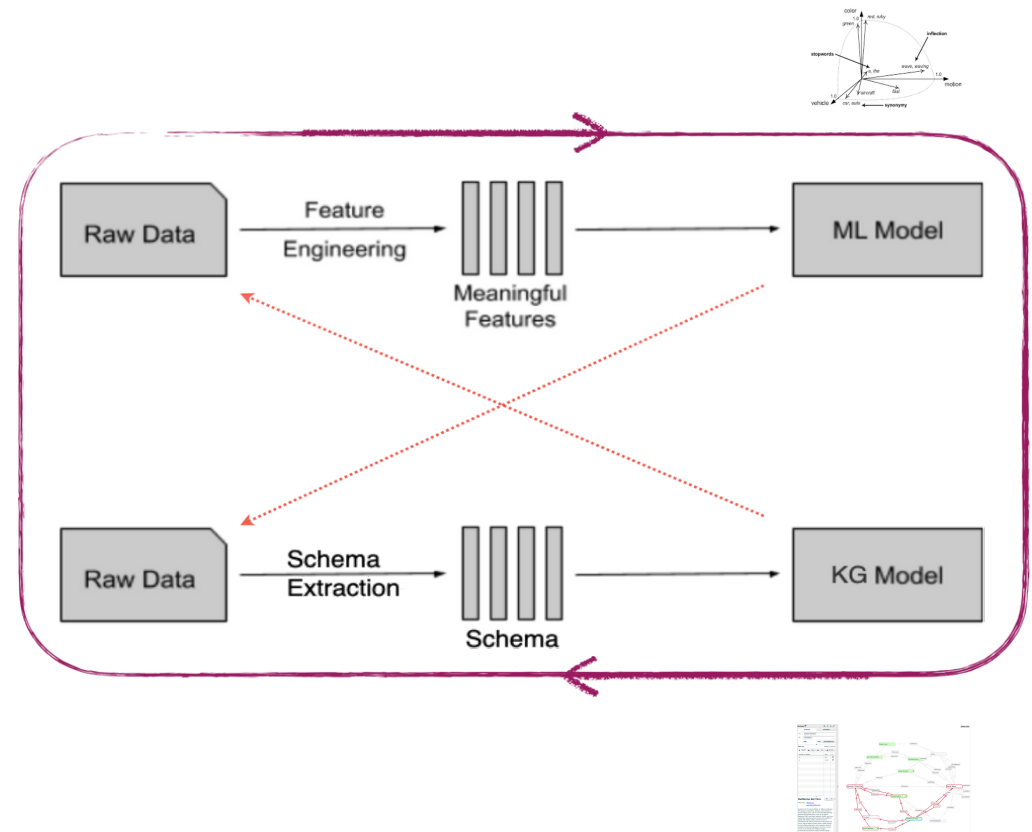
[10] Petri, F., Asprino, L., & Gangemi, A. (2024). Transformers and Slot Encoding for Sample Efficient Physical World Modelling. arXiv preprint arXiv:2405.20180.

[2] Pezzulo, G, Parr, T, Cisek, P, Clark, A, Friston, K. Generating meaning: active inference and the scope and limits of passive AI, Trends in Cognitive Sciences, 2024.



## Foundational issue: Hybridising GWMs

- **Generative AIs** are signal processing machines: input from any modality is used to learn activation patterns, and at inference time they *output symbols that we may interpret as world-referring*
- **Knowledge-based AIs** are logical machines: they *extract/design symbolic representation patterns of the world*, with a model-theoretical interpretation to ensure correct inference
- A **feedback loop** for multimodal, multilayered Grounded World Models (GWM)
- **GWM are multi-varied**: physical, neurocognitive, social, cultural, ...





Finanziato  
dall'Unione europea  
NextGenerationEU



Ministero  
dell'Università  
e della Ricerca



Italiadomani

PIANO NAZIONALE  
DI RIPRESA E RESILIENZA



Future  
Artificial  
Intelligence  
Research

## Foundational issue: a good KG for starting, then sensible patterns

- LLMs are not (yet) good at extracting the qualitative knowledge we need, but are **sensible to local, circumstantial patterns**
- Automated machine reading extracts the **basic KG expressed by NL and grounds it on public knowledge**
- KG are **firstly extracted neuro-symbolically**, then they are fed to a LLM jointly with adequate prompting for **in-context learning**
- Starting from a qualitative KG, **ICL extends it (XKG) with contextual knowledge and GWMs**, using heuristic pragmatic patterns (presuppositions, implicatures, impact, coercion, etc.)



Extended K Graph



Finanziato  
dall'Unione europea  
NextGenerationEU



Ministero  
dell'Università  
e della Ricerca



Italiadomani  
PIANO NAZIONALE  
DI RIPRESA E RESILIENZA



Future  
Artificial  
Intelligence  
Research

Thanks for your attention