# CienaLLM: Generative Climate-Impact Extraction from News Articles with Autoregressive LLMs

Javier Vela-Tambo[1,2] Jorge Gracia[3] Fernando Dominguez-Castro[1]

[1]Pyrenean Institute of Ecology (IPE-CSIC)

[2]Worcester Polytechnic Institute

[3]Aragon Institute of Engineering Research, Universidad de Zaragoza

September 25, 2025

# Contents

# Introduction

# Motivation: From Unstructured News to Insights

- **Extreme weather events** (e.g. floods, droughts, heatwaves) **are increasing** (WMO).
- Complete **extreme impact DBs** are not available.
- Impacts are scattered in **unstructured news articles**.
- **Traditional methods** (manual, supervised ML) **fail to scale** across hazards and languages.


(WMO), W. M. O. (2021). WMO Atlas of Mortality and Economic Losses from Weather, Climate and Water Extremes (1970–2019). Technical report, WMO, Geneva

# Motivation: From Unstructured News to Insights

- **Extreme weather events** (e.g. floods, droughts, heatwaves) **are increasing** (WMO).
- Complete **extreme impact DBs** are not available.
- Impacts are scattered in **unstructured news articles**.
- **Traditional methods** (manual, supervised ML) **fail to scale** across hazards and languages.


We need **scalable, adaptable methods** to automatically extract structured information.

(WMO), W. M. O. (2021). WMO Atlas of Mortality and Economic Losses from Weather, Climate and Water Extremes (1970–2019). Technical report, WMO, Geneva

# Our Solution: CienaLLM

- **Generative Information Extraction (GenIE)**: LLMs generate structured data.
- **Zero-Shot Learning**: Natural language prompt-only, no retraining.
- **Schema-Guided**: A predefined JSON schema ensures structured, consistent output.


## Key Idea

A **generalizable and adaptable** approach to extract climate impacts across hazards, datasets, and languages.

# Tasks and Datasets

- **Drought as a Case Study**: costliest hazard, highly relevant in Spain, and difficult to capture (diverse and dispersed impacts).
- **Tasks**: relevance detection (binary classification), impact extraction, and location extraction (multi-label classification).
- **Datasets**: adapted from Lopez Otal et al. (2025)'s **SeqIA**; reannotated + new dataset for location task.

| Task | Dataset | Source(s) | Size |
|---|---|---|---|
| Impact extraction | DID | *El País*, *Grupo Zeta* | 438 |
| Relevance | DRD | *El País* | 2,240 |
| Location | DILD | *El País* | 100 |

López-Otal, M., Domínguez-Castro, F., Latorre, B., Vela-Tambo, J., and Gracia, J. (2025). SeqIA: A Python framework for extracting drought impacts from news archives.
*Environmental Modelling & Software*, 187:106382 Lopez Otal, M., Domínguez-Castro, F., Latorre Garcés, B., Vela Tambo, J., and Gracia del Río, J. (2025). SeqIA - Annotated drought-related news articles

# Methodology and Framework

# Methodology

- **LLM Families**: Llama, Gemma, and Qwen.
- **Model Sizes**: S (<7B), M (7-25B), and L (>25B).
- **Precision Regimes**: Full-precision (`fp16`) vs. 4-bit quantization (`q4_K_M`).
- **Prompt Strategies**:
  - Summarization (`SUM`)
  - Chain of Thought (`CoT`)
  - Self-Criticism (`SC`)
  - Impact Descriptions (`DESC`)
- **Parsing Methods** (`RPARSE`): Single- vs. two-step.

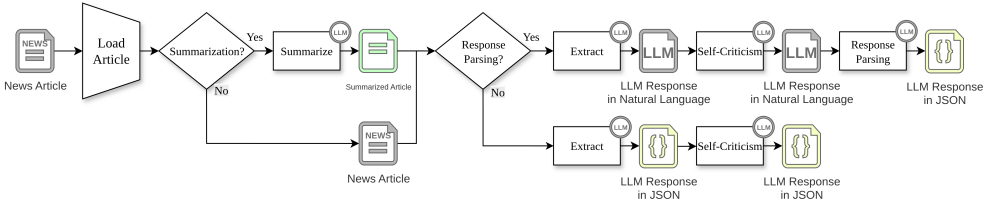
# Methodology

- **LLM Families**: Llama, Gemma, and Qwen.
- **Model Sizes**: S (<7B), M (7-25B), and L (>25B).
- **Precision Regimes**: Full-precision (`fp16`) vs. 4-bit quantization (`q4_K_M`).
- **Prompt Strategies**:
  - Summarization (`SUM`)
  - Chain of Thought (`CoT`)
  - Self-Criticism (`SC`)
  - Impact Descriptions (`DESC`)
- **Parsing Methods** (`RPARSE`): Single- vs. two-step.

**Systematic Evaluation**

Total of 384 configurations tested.

# The CienaLLM Framework

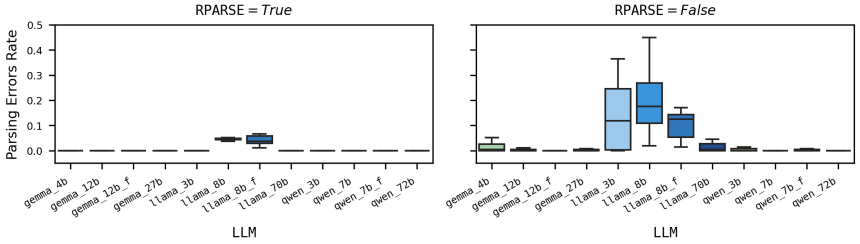



- **Input**:
  - **News Article**
  - **Base Prompt**: `You are an expert in environmental analysis...`
  - **Configuration**: `qwen_7b + SUM + CoT + RPARSE`
- **LLM Inference**: LLM zero-shot extraction using a custom prompt.
- **Structured Parsing**: The response is formatted into JSON.

# Results and Discussion

# Key Finding 1: The Importance of Response Parsing

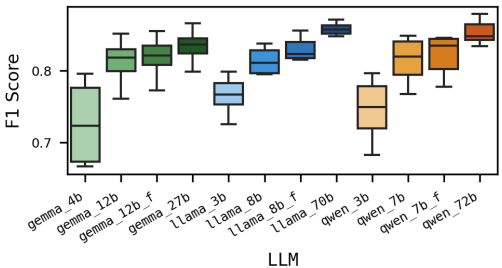
- LLMs can produce **malformed JSON outputs**.
- If parsing fails, the **article's information is lost**.
- A two-step response parsing process **nearly eliminates formatting errors**.
- This adds a **small increase in latency** but ensures reliability.




## When to use **Response Parsing**?

Use if necessary to reduce parsing errors. Skip for newer models that produce less errors.

# Key Finding 2: Scale and Precision Drive Performance




- **Model Size**: Larger models achieve higher accuracy and greater stability.
- **Quantization**: Offers a substantial reduction in latency (up to 40%) with a modest drop in accuracy.

**How big LLM** should I choose?

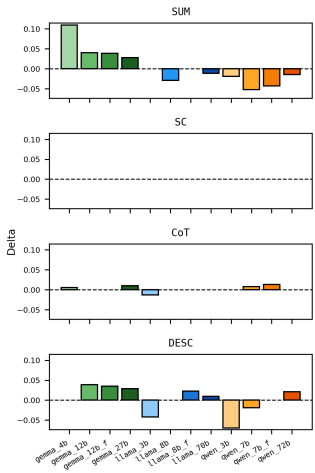
Adapt to hardware and try to use smaller, quantized models if possible.

# Key Finding 3: Prompt Strategies Are Not One-Size-Fits-All

- **No universal recipe**: No single prompt strategy works best for all models.
- **Model-specific effects**:
  - `SUM` **helped Gemma** models but hurt Llama and Qwen.
  - `DESC` worked best for **larger models**.
  - `CoT` **and** `SC` **had minimal impact** for many models.

## Should I use **Prompt Engineering**?

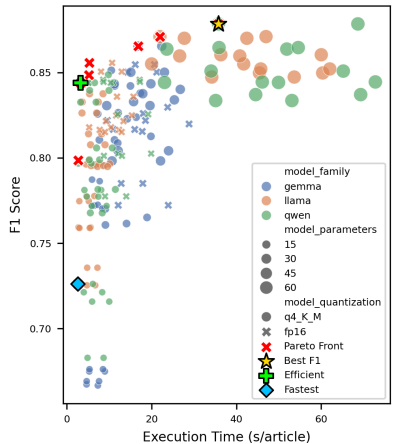
Adapt prompt length to model size, and provide available details to larger models.



# Performance vs. Efficiency

- **Pareto-front analysis** resulting in 3 selected configurations:

| Configuration | F1 Val (Test) | Exec. Time |
|---|---|---|
| *Best-F1* `qwen_72b + CoT + DESC` | **0.878** (**0.873**) | 35.8 |
| *Fastest* `qwen_3b + DESC` | 0.726 (0.646) | **2.6** |
| *Efficient* `qwen_7b` | 0.844 (0.808) | 3.2 |



# Comparison with SeqIA (Supervised Baseline)

- **SeqIA** López-Otal et al. (2025): supervised classifiers trained on annotated data.

**Drought Impact Extraction (E2E)**

| | F1 | Time (s) |
|---|---|---|
| *Best-F1* | **0.782** | 30.07 |
| *Fastest* | 0.452 | 1.81 |
| *Efficient* | 0.645 | 2.54 |
| SeqIA | 0.769 | **1.54** |

**Drought Relevance Classification (DRD)**

| | F1 | Time (s) |
|---|---|---|
| *Best-F1* | **0.968** | 11.28 |
| *Efficient* | 0.929 | 1.74 |
| *Fastest* | 0.779 | 1.36 |
| SeqIA | **0.961** | **0.18** |

## Key Trade-off

SeqIA is much faster, but CienaLLM achieves similar or better accuracy and adapts to new hazards, schemas, and languages *without retraining*.

López-Otal, M., Domínguez-Castro, F., Latorre, B., Vela-Tambo, J., and Gracia, J. (2025). SeqIA: A Python framework for extracting drought impacts from news archives.
*Environmental Modelling & Software*, 187:106382

# Conclusion

# Conclusions

- **GenIE is practical and effective**: enables structured information extraction from unstructured news.
- **CienaLLM is adaptable**: zero-shot and schema-driven, generalizing to new hazards, languages, and schemas without retraining.
- **CienaLLM improves with better LLMs**: its performance increases as newer, stronger models become available.
- **Performance drivers**: larger models improve accuracy and stability; quantization reduces cost with modest trade-offs.
- **Prompt design is model-specific**: `SUM` helps some models; `DESC` benefits larger ones; `SC` and `CoT` often have little effect.

# Future Work

- **Integrate**: combine CienaLLM with fast classifiers or keyword filters for hybrid pipelines.
- **Scale up**: extract actionable insights from large corpora.
- **Extend**: apply the framework to other hazards (floods, hail, heatwaves).
- **Experiment**: explore new paradigms (few-shot prompting, retrieval-augmented, LLM ensembles).

UIMP
Universidad Internacional Menéndez Pelayo

AEPIA
ASOCIACIÓN ESPAÑOLA PARA LA INTELIGENCIA ARTIFICIAL

Escuela de Ingeniería y Arquitectura
Universidad Zaragoza

Instituto Pirenaico de Ecología
CSIC

# Thank you for your attention

**Javier Vela-Tambo**

jvela@wpi.edu

javiervelatambo.com

# References

López-Otal, M., Domínguez-Castro, F., Latorre, B., Vela-Tambo, J., and Gracia, J. (2025). SeqIA: A Python framework for extracting drought impacts from news archives. *Environmental Modelling & Software*, 187:106382.

Lopez Otal, M., Domínguez-Castro, F., Latorre Garcés, B., Vela Tambo, J., and Gracia del Río, J. (2025). SeqIA - Annotated drought-related news articles.

(WMO), W. M. O. (2021). WMO Atlas of Mortality and Economic Losses from Weather, Climate and Water Extremes (1970–2019). Technical report, WMO, Geneva.