

Antonio Moreno Sandoval, Blanca Carbajo Coronado, Jordi Porta Zamorano, Yanco Torterolo Orta, Doaa Samy

## What are the FinCausal series?

### FinCausal series: evolution of the Causal relation task

Started in **2020**, within COLING (Barcelona) as two tasks:

- a binary classification task
- a relation extraction task

In **2021** and **2022** (LREC), the focus shifted to a **relation detection task**. The aim is to identify, in a causal sentence or text block, the causal elements and the consequential ones.

In **2023** (IEEE Big Data), the task was introduced in **Spanish** for the first time but remained an extraction task.

In **2025** (COLING), we changed the task to a **generative AI task**, where questions about causes or effects are posed, and the system responses are evaluated using exact matching (EM) and semantic answer similarity (SAS) metrics.

### FinCausal 2025 as a Generative QA

Dataset Structure

- Context: Original paragraph from annual reports.
- Question: Always abstractive, reflecting the content of the cause or effect.
- Answer: Extractive, extracted verbatim from the provided context.

Context	Question	Answer
In October 2016, we announced an implementation agreement to sell ACR to two Shenzhen government sponsored investment companies. <b>This approval process remains ongoing</b> and, as a result, we did not value ACR on an imminent sales basis as at 31 March 2017.	Why was ACR not valued on an imminent sales basis as of March 31, 2017?	This approval process remains ongoing

## Results

### English

Ranking	Team	SAS	Exact Match
1	Team nirvanatear (Jonathan Zhou, China)	0.9779 (1)	0.8798 (1)
2	TU Graz Data Team (Graz University of Technology, Austria)	0.9732 (2)	0.8637 (2)
3	Sarang (National Institute of Technology ,Trichy, India)	0.9674 (3)	0.7014 (7)
4	CLRG (n/a)	0.9604 (4)	0.7214 (6)
5	Semantists (Institute for Infocomm Research, Singapore)	0.9598 (5)	0.7435 (5)
5	LaithTeam (Copenhagen University, Denmark)	0.9598 (5)	0.7615 (4)
7	CUFE (Cairo University, Egypt)	0.9595 (7)	0.8277 (3)
8	OraGenAIOrganisation (Oracle, India)	0.9244 (8)	0.3527 (9)
9	RG IPT (India)	0.9086 (9)	0.5110 (8)
10	PresiUniv (Dpt. CSE, Presidency Univ, Bangalore, India)	0.8241 (10)	0.2244 (11)
11	Baseline (LLI-UAM, Spain)	0.7373 (11)	0.3287 (10)

Table 4: English results

### Spanish

Ranking	Team	SAS	Exact Match
1	TU Graz Data Team (Graz University of Technology, Austria)	0.9841 (1)	0.8703 (2)
2	Team nirvanatear (Jonathan Zhou, China)	0.9801 (2)	0.8782 (1)
3	LenguajeNatural.AI (Spain)	0.9787 (3)	0.8164 (4)
4	LaithTeam (Copenhagen University, Denmark)	0.9756 (4)	0.8084 (5)
5	CUFE (Cairo University, Egypt)	0.9755 (5)	0.8224 (3)
6	CLRG (n/a)	0.9607 (6)	0.7166 (7)
7	Semantists (Institute for Infocomm Research, Singapore)	0.9555 (7)	0.7525 (6)
8	OraGenAIOrganisation (Oracle, India)	0.9219 (8)	0.0898 (9)
9	RG IPT (India)	0.8987 (9)	0.0619 (10)
10	PresiUniv (Dpt. CSE, Presidency Univ, Bangalore, India)	0.7520 (10)	0.0140 (11)
11	Baseline (LLI-UAM, Spain)	0.7244 (11)	0.2515 (8)

Table 6: Spanish results

## Systems tipology

Team	Discriminative	Generative	Fine-tuning	Prompting	Quantization
Team Nirvanatear	✗	✓	✓	Simple	✗
OraGenAIOrganisation	✗	✓	✗	CoT	✗
AI Laith	✓	✗	✓	✗	✗
Sarang	✗	✓	✓	Simple	✓
RG IPT	✗	✓	✗	CoT+FS/FS	✗
TU Graz	✗	✓	✓	✗	✓
PresiUniv	✓	✗	✗	✗	✗
LenguajeNatural.AI	✗	✓	✓	Simple	✓
CLRG	✓	✗	✓	✗	✗

Table 5: Systems comparison. In Prompting, Simple means a simple prompt or instruction, CoT stands for Chain of Thoughts and FS stands for Few Shot.

## Error analysis

Types of Errors

- **Confusion with Purpose:** Cause-effect relationships are often mistakenly identified as purpose-based relationships.
- **Irrelevant Information:** Answers may include unnecessary contextual information, distracting from the core causal relationship. Low EM score.
- **Minimal Overlap:** Low overlap between the generated response and the expected answer. Low SAS score. Indicates potential issues in:
  - Identifying key causal elements.
  - Understanding the question's context.

Context	Question	Answer	Result	SAS	Exact match
In accordance with the Company's stated dividend policy, the Board recommends a further quarterly dividend of 3.57p per Ordinary Share, payable on 30 April 2018 to shareholders on the register on 6 April 2018. Total dividends paid for the year therefore amount to 14.04p per Ordinary Share equivalent to a dividend yield of 4.1 per cent at the year-end.	Why does the total dividends paid for the year amount to 14.04p per Ordinary Share, equivalent to a dividend yield of 4.1 per cent at the year-end?	the Board recommends a further quarterly dividend of 3.57p per Ordinary Share, payable on 30 April 2018 to shareholders on the register on 6 April 2018	In accordance with the Company's stated dividend policy, the Board recommends a further quarterly dividend of 3.57p per Ordinary Share, payable on 30 April 2018 to shareholders on the register on 6 April 2018	0.980	0

## Conclusions

### Key findings

- Fine-tuned LLMs and generative approaches demonstrate significant value for extractive question answering in financial contexts.
- Performance on Spanish financial texts highlights the importance of language-specific nuances and the need for tailored datasets and approaches in multilingual settings.

## Future directions

- **Hybrid Methodologies:** Explore combining the precision of fine-tuned models with the adaptability of generative models.
- **Complex Causality:** Address nested causal structures and handling of multiple potential causes. Develop clear annotation guidelines for complex scenarios.
- **Dataset Refinement:** Standardize formatting (spacing, hyphenation, punctuation). Remove unnecessary text elements.
- **Multilingual Expansion:** Extend FinCausal to include diverse languages, such as Chinese, Arabic, Hindi, Japanese, German, and French.

## Acknowledgements

This publication is part of the project GRESEL (PID2023-1512800B-C21) funded by the Spanish Ministry of Science and Innovation and Universities. We also gratefully acknowledge the financial support received by the second author through a FPU grant (FPU20/04007) awarded by the Spanish Ministry of Science, Innovation and Universities. We thank our dedicated annotator, Paula Gozalo, who contributed to creating the datasets.