

## Short Bio | Giovanni Sullutrone

### Giovanni Sullutrone



LLMs, AI Safety, and Over-Safety Researcher

View the Project on GitHub [giosullutrone/cv](https://github.com/giosullutrone/cv)

- Download **ZIP File**
- Download **TAR Ball**
- View On **GitHub**

## Short Bio

I'm a PhD student at the University of Modena and Reggio Emilia and I've held a visiting position at the research center DARE LAB. My research has focused on both **applications**, **AI Safety** and **Over-Safety**.

---

## Featured Works (TL;DR)

- **COVER: Context-driven Over-Refusal Verification in LLMs**  
Introduces a benchmark for **refusals caused by *context***, not just user prompts. Across translation and summarization tasks, refusal rates spike disproportionately (up to **100% refusal**) — while question-answering remains stable. Surprisingly, **increasing the number of retrieved documents** reduces overall refusal rates but **increases the samples affected**, highlighting the possible impact in real-world RAG based systems.  
*G Sullutrone, RA Vigliermo, S Bergamaschi, L Sala* — Findings of ACL 2025  
*Read the Paper: Coming soon*
  - **Text-to-Refused-SQL: Evaluating LLMs' Refusal Behavior on Sensitive Data**  
Tests LLMs on Text-to-SQL queries over healthcare databases with **PII/SPI**. Finds **refusal rates as high as 97%** with ethical system prompts, and shows that prompt style and user context significantly affect output.  
*G Sullutrone, L Sala, L Trigiante, S Bergamaschi* — SEBD2025  
*Read the Paper: Coming soon*
  - **Sensitive Topics Retrieval in Digital Libraries**  
Uses a **ḥadīṭ** dataset to test retrieval on sensitive religious content. Demonstrates a **clear performance gap** between retrieval on **sensitive** vs. **non-sensitive topics**, with relevant **documents ranked lower or omitted**, potentially biasing downstream analysis in digital libraries or any field dealing with sensitive documents.  
*G Sullutrone, RA Vigliermo, L Sala, S Bergamaschi* — TPD 2024  
*Read the Paper*
  - **Text-to-SQL with LLMs: Exploring the Promise and Pitfalls**  
Evaluates LLMs' performance on Text-to-SQL across accuracy, scalability, and hallucination. Identifies how **benchmarks fail to reflect real-world complexity** and highlights risks of untrustworthy queries, especially under ambiguous or large-scale input.  
*L Sala, G Sullutrone, S Bergamaschi* — SEBD 2024  
*Read the Paper*
-

## Latest Papers

- **COVER: Context-Driven Over-Refusal Verification in LLMs**  
*G Sullutrone, RA Vigliermo, S Bergamaschi, L Sala* — Findings of ACL 2025  
*Read the Paper: Coming soon*
  - **Text-to-Refused-SQL: Evaluating LLMs' Refusal Behavior on Sensitive Data**  
*G Sullutrone, L Sala, L Trigiane, S Bergamaschi* — SEBD2025  
*Read the Paper: Coming soon*
  - **Proposing a Comprehensive Dataset for Arabic Script OCR in the context of Digital Libraries and Religious Archives**  
*RA Vigliermo, G Sullutrone, S Bergamaschi, L Sala* — IRCDL 2025  
*Read the Paper*
  - **Text-to-sql with large language models: Exploring the promise and pitfalls**  
*L Sala, G Sullutrone, S Bergamaschi* — SEBD 2024  
*Read the Paper*
  - **Sensitive Topics Retrieval in Digital Libraries: A Case Study of ḥadīṭ collections**  
*G Sullutrone, RA Vigliermo, L Sala, S Bergamaschi* — TPDFL 2024  
*Read the Paper*
  - **Arabic Text Enhancement with GPT for Digital Libraries**  
*L Sala, G Sullutrone, S Bergamaschi, RA Vigliermo* — JCDL 2024  
*Read the Paper*
- 

## Education

- **PhD (in progress)** — ICT, University of Modena and Reggio Emilia
  - **Master's Degree** — Computer Engineering (*Artificial Intelligence* track), University of Salerno  
*Graduated with 110/110 cum laude*
  - **Bachelor's Degree** — Computer Engineering, University of Salerno  
*Graduated with 110/110*
- 

## Research Grants

- **Artificial Intelligence grant** at A.I. Tech S.R.L.
-

## More About Myself

I co-founded **Vivi S.R.L.**, a startup aiming to democratize wine culture by making wine-by-the-glass sustainable and accessible, addressing economic and environmental issues in wine consumption. Our startup has won multiple awards, among many a bootcamp sponsored by *Amazon*.

I also co-hold an **international patent** for a liquid dispensing device that extracts wine or other liquids from sealed bottles without removing the cork — making it suitable for both home and professional use.

This project is maintained by giosullutrone

Hosted on GitHub Pages — Theme by orderedlist