# TUTSTI @ ISWC2024

The tutorial TUTSTI introduces the topic of Semantic Table Interpretation (STI), covering theoretical and practical considerations. In particular, the tutorial will provide a comprehensive analysis of how the approaches to STI have evolved from heuristic-based to ML-based, to the most recent LLM-based approaches. The analysis will consider the specific characteristics of these different classes, providing insights into their respective advantages and limitations to identify the contexts of use. The final part will describe a case study to demonstrate the application of two state-of-the-art approaches. A comprehensive survey of all STI approaches developed up to October 2024 is available here.

## Objectives

### L1

Present key aspects of STI: the role of semantic annotation of tabular data; why STI remains relevant in the generative AI era; phases of STI design and deployment; and a review of existing solutions, highlighting their strengths and limitations

### L2

Present a link and extend paradigm for sti as a unifying abstraction to develop semantic data enrichment solutions

## Semantic Table Interpretation: from Heuristic to LLM-based approaches

The tutorial will be a half-day tutorial and will be split into two slots:

**Slot 1:** where we discuss the main concepts and review SOTA (L1, L2, L3 and L5);

**Slot 2:** where we present two SOTA approach and use it in a hands-on session, by walking the audience through a use case (L4)

50 MINUTES
### Semantic Table Interpretation
**Topics:** Definitions; Tasks; Objectives, SemTab Challenge; (slides)

50 MINUTES
### State-of-the-art
**Topics:** Key Challenges, SOTA; (slides)

30 MINUTES
### Break

60 MINUTES
### Impact of (L)LMs on STI
**Topics:** From heuristic approaches to generalistic table interpretation and manipulation approaches; (slides)

40 MINUTES
### Hands-on session
Fine-Tuning a LLM on CEA task (material & slides).

## Intended audience

The tutorial type falls into the category of an introductory tutorial within a specific domain, catering to an intermediate level of proficiency. Intended attendees are:

**i) researchers** with expertise in semantic technologies, and, in particular in their application to data integration problems (e.g., ontology matching, semantic reconciliation, table annotation), who will discover new industry-driven, real-world application scenarios

**ii) young researchers** that recently joined the semantic web community (e.g., PhD students and postdocs), who will be exposed to challenging problems and possible solutions; iii) data scientists and data engineers, who will learn how semantic technologies can be exploited to support data-driven innovation

**iii) data scientists and data engineers**, who will learn how semantic technologies can be exploited to support data-driven innovation

## Presenters

### MATTEO PALMONARI
Professor at the University of Milan-Bicocca



He has been a coordinator and partner in projects about data enrichment, and he is particularly interested in combining machine learning and human-in-the-loop mechanisms to support knowledge-based applications.

### FABIO D'ADDA
Research Assistant at the University of Milan-Bicocca

He specialises in the application of ML techniques in the Semantic Web. He is chair of SemTab 2024, and organiser of the "STI vs LLMs Track"

### MARCO CREMASCHI
Assistant Professor at the University of Milan-Bicocca

He specialises in the application of ML techniques in the Semantic Web. He is chair of SemTab 2024, and organiser of the "STI vs LLMs Track"

### ERNESTO JIMENEZ-RUIZ
Lecturer in Artificial Intelligence and Senior Tutor for Research at City, University of London affiliated to the Research Center for Adaptive Computer Systems and Machine Learning

His current research interests focus on applying Semantic Technology to Data Science workflows and combining Knowledge Representation and Machine Learning techniques. He is one of the founders of the SemTab challenge