



AAAI-25 / IAAI-25 / EAAI-25
FEBRUARY 25 – MARCH 4, 2025 | PHILADELPHIA, USA

Text2Zinc: A Cross-Domain Dataset for LLM Modeling Assistants

A. Singirikonda¹, S. Kadioğlu^{1,2}, K. Uppuluri¹



Serdar Kadioğlu

¹ Dept. of Computer Science, Brown University

² AI Center of Excellence, Fidelity Investments



skadio.github.io



BROWN



Decision Making in the Era of Large-Language Models

1

Decision Making & Optimization

- Constraint-solving techniques are powerful and have many applications.
- The cognitive barrier of translating problem descriptions into formal constraint models persists.

2

Large-Language Models

- LLMs have found success in many fields recently.
- However, they still face challenges in generating constraint models from free-form natural language text.

Our Contributions

Ner4Opt

A principled approach to extracting components of optimization models such as the objective, variables, and constraints from free-form natural language text.

[Hugging Face Ner4Opt Demo](#)

Text2Zinc

A unified dataset curated to work with LLMs and an associated leaderboard to evaluate strategies to generate MiniZinc models from natural language text.

[Hugging Face Text2Zinc Dataset](#)

Ner4Opt: A Principled Approach

1

Problem Formalization

Formalize this problem and show how it differs from the classical NER (CPAIOR'23)

2

Performance Study

Principled approach to study the performance of solutions

- Baseline methods from classical NLP
- Semantic methods from modern NLP
- Hybrid methods with data augmentation
- Fine-tuned domain-specific transformers

3

Automated In-context Learning

Holy Grail 2.0 (CP-PTHG'23) with Ner4Opt annotations (Constraints'24) achieves **50% boost** in execution accuracy of MiniZinc models.

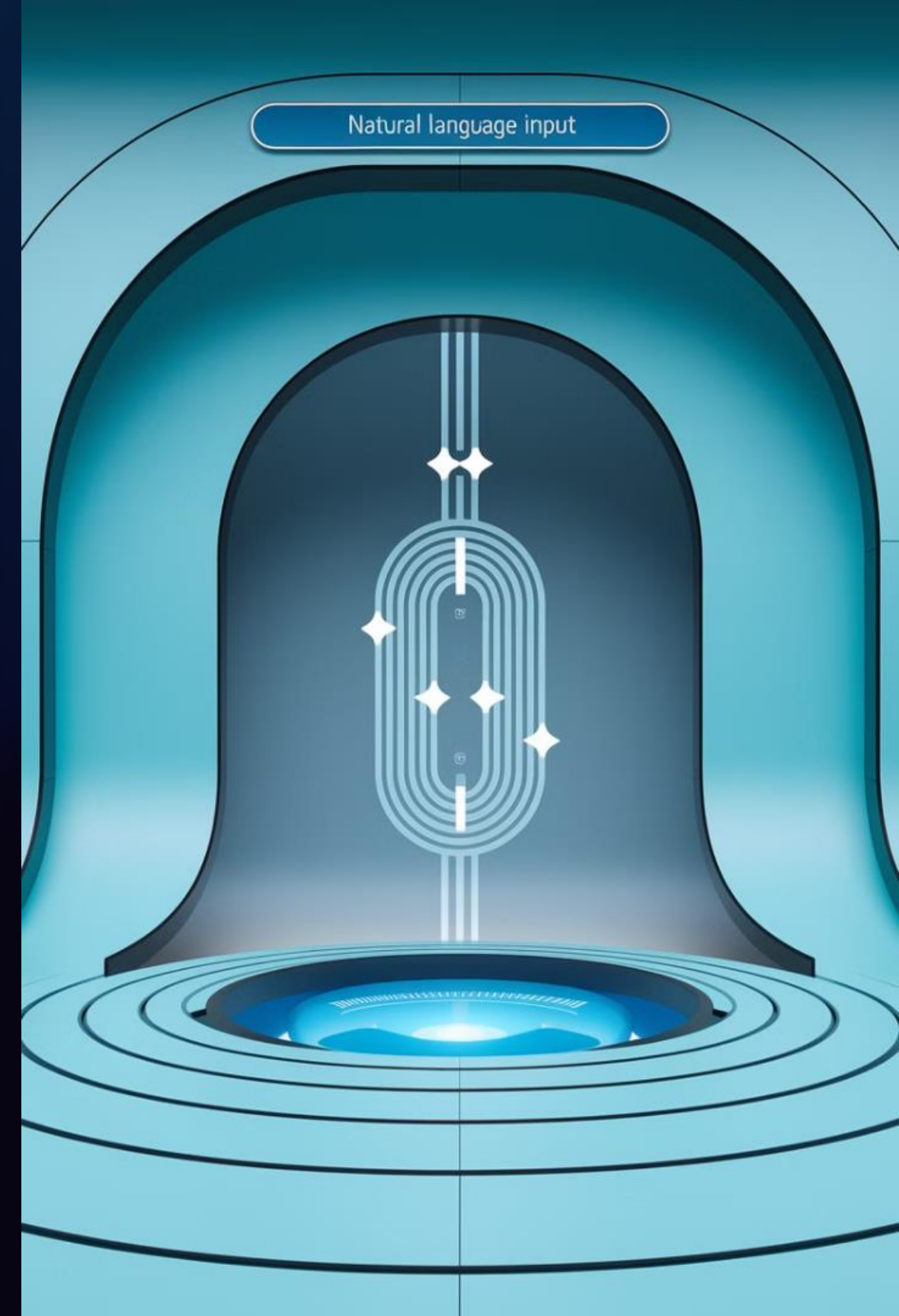
4

Open-Source Library & Fined-Tuned Models

[Ner4Opt Library](#) via `pip install ner4opt`

[Hugging Face Space Demo](#) & [Hugging Face Fine-Tuned Models](#)

Kadioglu et. al. Ner4Opt: Named Entity Recognition for Optimization Modelling (Constraints'24)
Tsouros et. al., Holy Grail 2.0: From Natural Language to Constraint Models (CP-PTHG'23)



Text2Zinc: Motivation

Driving Progress

Datasets and benchmarks **fuel progress** in various domains: Computer Vision, NLP, and SAT, CP, MIP, RecSys, etc.

Room for Improvement

Current problem datasets have **potential for improvement** for integration with language models.

Structured Information & Metadata

Models and natural language descriptions of problems have been documented heavily but seldom occur together. Crucial **metadata is unavailable**.

Existing Resources



NL4OPT

- Linear programming problems
- No separation between problem description and data
- Relatively easy instances



NLP4LP

- Extends NL4OPT
- Introduces mixed integer programming
- Evaluated with GurobiPy and cvxpy



ComplexOR

- Standard OR Problems
- Evaluated with GurobiPy



Logic Grid Puzzles

- Introduces satisfaction problems in the form of logic grid puzzles



CSPLib

- CP and Satisfaction problems
- Not designed to work with ML or LLMs



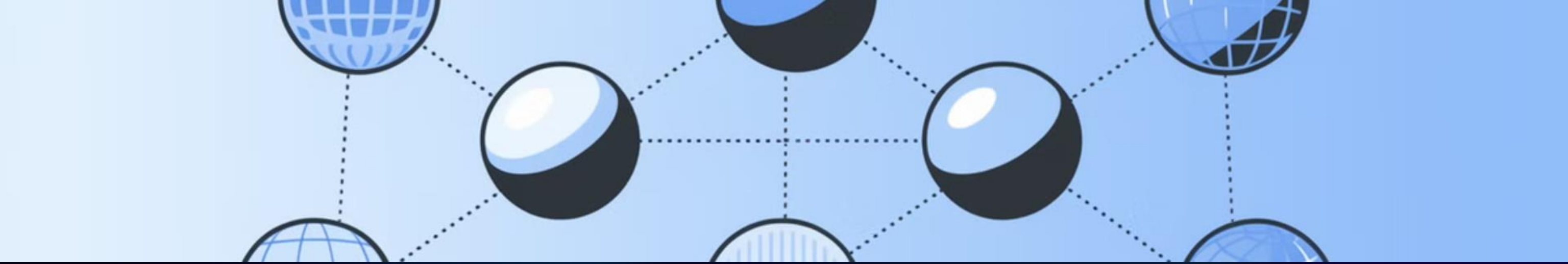
Hakank's Models

- Extensive set of constraint programming models in various languages
- Does not capture metadata

Optimization

Satisfaction

**Massive thank you to the community for contributing these valuable resources!*



Text2Zinc: Addressing Dataset Gaps

1

Cross-Domain

- Focus on combining both **optimization & satisfaction** problems.
- Incorporates LP, MIP, CP problems.

2

Unified Format

- Unifies existing datasets.
- **Clear separation** of problem description & instance data.

3

Solver Agnostic

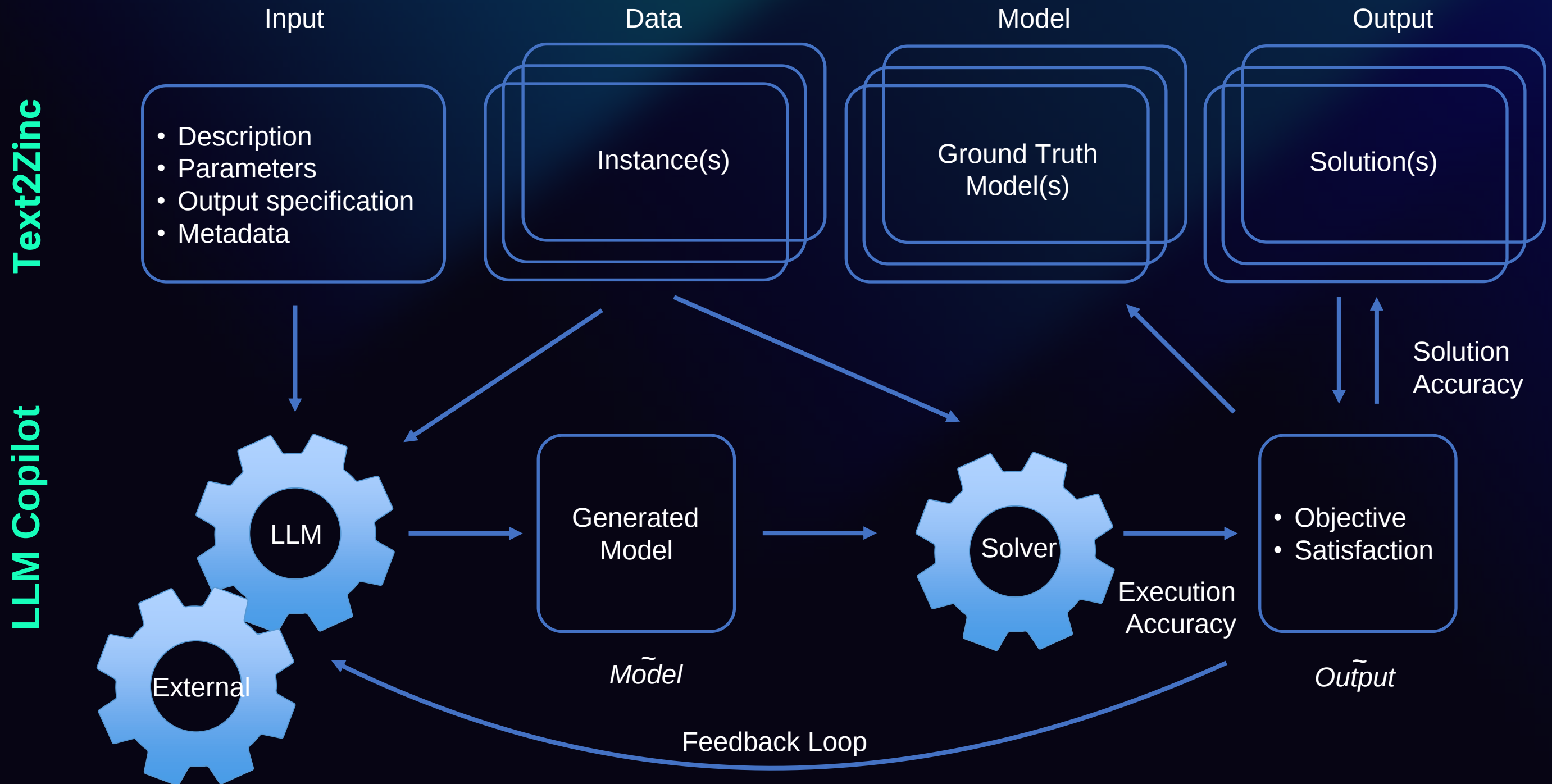
- Enables **solver agnostic** approaches.
- MIP, CP, SAT, LCG through MiniZinc.

4

Data Augmentation

- Clear and concise descriptions.
- Input and output specification.
- **Metadata** generation.
- Manual verification.

Text2Zinc: A Unified Approach



Text2Zinc Statistics

64

Linear Programming

Continuous variables
with linear constraints

31

Mixed Integer Programming

Continuous
and discrete variables

15

Constraint Programming

Global
constraints

Our dataset includes a mix of **LP, MIPs, and CP problems** across various domains
Providing a comprehensive benchmark for natural language to constraint model translation.

Text2Zinc Initial Approaches

Out-of-the-box LLM

Vanilla prompting, zero-shot,
few-shot performance
Single vs. Multi-Call

Evaluation

Assessment of execution and
solution accuracy



Chain-of-Thought

Improved reasoning through step-
by-step problem-solving

Knowledge Graph

Leveraging structured knowledge
as intermediary representation

Text2Zinc Initial Approaches

Out-of-the-box LLM

Vanilla prompting, zero-shot,
few-shot performance
Single vs. Multi-Call

Evaluation

Assessment of execution and
solution accuracy



Chain-of-Thought

Improved reasoning through step-
by-step problem-solving

Knowledge Graph

Leveraging structured knowledge
as intermediary representation

Text2Zinc Initial Results

Solution Approach	Execution Accuracy	Solution Accuracy
Out-of-the-box Prompting	0.1904	0.0634
+ Data & examples	0.3650	0.1904
+ Shape information	0.1904	0.1269
+ Knowledge Graph	0.3492	0.1111
Chain-of-Thought (COT)	0.4285	0.1746
+ In-context examples	0.5873	0.2539
+ Shape information	0.5555	0.2063
Multi-Call + Composition	0.6031	0.2222

[Hugging Face Text2Zinc Leaderboard](#)

What's Next?

1

Text2Zinc

- The Text2Zinc dataset is now available on Hugging Face
- Providing a valuable resource for researchers and practitioners in the field

2

Performance

- Explore other approaches
- Contribute to the leaderboard to establish comprehensive benchmarks

3

Contributions

We encourage contributions to the dataset through new problems

[Hugging Face Text2Zinc Dataset](#)