Computional Linguistic Task 6 - report

Jerzy Boksa

January 2026

1 Introduction

The objective of this laboratory assignment is to design, train, and evaluate a Tiny Reasoning Model (TRM) on challenging structured reasoning benchmarks, and subsequently compare its behavior with a large pre-trained language model. The primary focus is on reasoning quality, generalization capability, and data efficiency, rather than model scale.

Contemporary Large Language Models (LLMs) demonstrate notable limitations when confronted with tasks requiring systematic logical reasoning, such as Sudoku-Extreme, ARC-AGI-1, and ARC-AGI-2 benchmarks. The work presented in [1] proposes an alternative architecture designed to address these specific challenges through recursive computation mechanisms.

This report documents the implementation and evaluation of a TRM model trained on the Sudoku puzzle-solving task, with a comparative analysis against the Gemini 3.0 model.

2 Dataset

The original TRM paper [1] utilized 1,000 base samples with extensive augmentation, resulting in approximately 1 million training instances. Given that this laboratory focuses primarily on model architecture rather than data engineering, a pre-existing dataset was employed.

The Sudoku-Extreme dataset [2] from Hugging Face was utilized, comprising 3.8 million training samples and 423,000 evaluation samples. The dataset contains approximately 1.1 million puzzles classified as simple and 3.1 million classified as hard. Puzzle difficulty is quantified by the number of backtracks required by the sudoku solver to reach a solution, where higher values indicate greater difficulty.

For computational efficiency during training validation, a subset of 2,912 samples was employed for periodic evaluation.

3 Hardware

All experiments were conducted on the Cyfronet Athena high-performance computing cluster, utilizing an NVIDIA A100 GPU with 40GB of memory.

4 TRM Model Architecture

The model architecture was designed to closely approximate the specifications outlined in the original paper [1]. The following hyperparameters were maintained at their original values:

- H_cycles: 3
- L_cycles: 6

but due to computational constraints, certain parameters were reduced:

- Hidden size: reduced from 512 to 384
- Maximum steps: reduced from 16 to 8

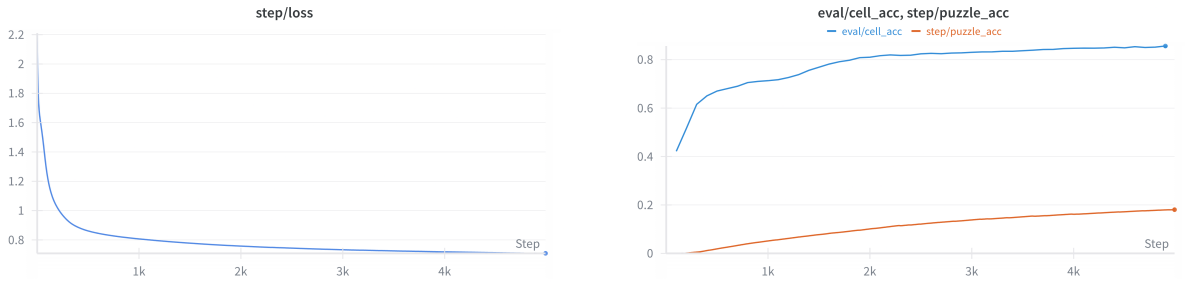
Following the recommendations in [1], a Multi-Layer Perceptron (MLP) architecture was employed instead of attention layers, as MLP-based modules demonstrate superior performance on structured tasks such as Sudoku.

The resulting model comprises approximately 3.5 million parameters, representing half the parameter count of the original implementation.

5 Training

Training was performed on the complete training set over a single epoch, corresponding to 4,990 optimization steps with a batch size of 768 samples. The total training duration was 12 hours, 4 minutes, and 53 seconds. This extended training time, despite the relatively modest model size (3.5M parameters), demonstrates that the recursive computation mechanism inherent to the TRM architecture is computationally intensive.

For optimization, the Lion optimizer [3] was employed with a learning rate of 1.5×10^{-4}. This optimizer was selected for experimental purposes, as it has demonstrated promising results in recent literature.



(a) Training loss progression

(b) Training and evaluation accuracy

Figure 1: Training metrics over the course of one epoch

As illustrated in Figure 1a, the training loss exhibits a steady decrease without reaching a plateau, suggesting that additional training epochs could yield further performance improvements.

An unexpected observation is the disparity between evaluation and training accuracy, with evaluation accuracy consistently exceeding training accuracy. This phenomenon may

be attributable to either implementation mistaktes or sampling bias in the evaluation subset. Specifically, the 2,912 samples selected from the 400,000 available evaluation samples may not be representative of the true difficulty distribution. Further investigation is required to check this behavior.

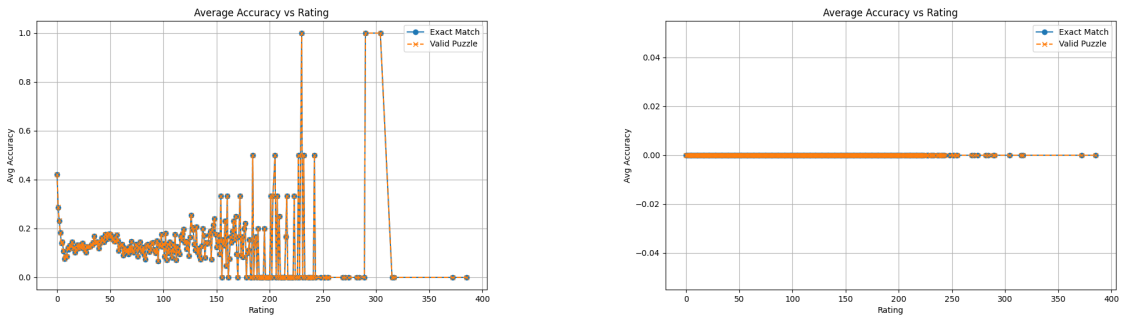
6 Evaluation

This section presents the evaluation results for two models: the trained TRM and the Gemini 3.0 Pro LLM. The TRM underwent comprehensive quantitative evaluation, while the Gemini model was assessed qualitatively on a limited sample set to demonstrate its capabilities on structured reasoning tasks.

6.1 TRM Evaluation

Quantitative evaluation of the TRM was conducted on 100,000 samples from the test set, with stratified sampling based on puzzle rating (difficulty). Two metrics were computed:

- **Exact Match Accuracy**: Measures whether the predicted solution exactly matches the ground truth solution from the dataset.
- **Puzzle Accuracy**: Measures whether the predicted solution constitutes a valid Sudoku solution, regardless of whether it matches the ground truth. This metric accounts for puzzles that may have multiple valid solutions.



(a) Trained TRM model

(b) Untrained (empty) TRM model

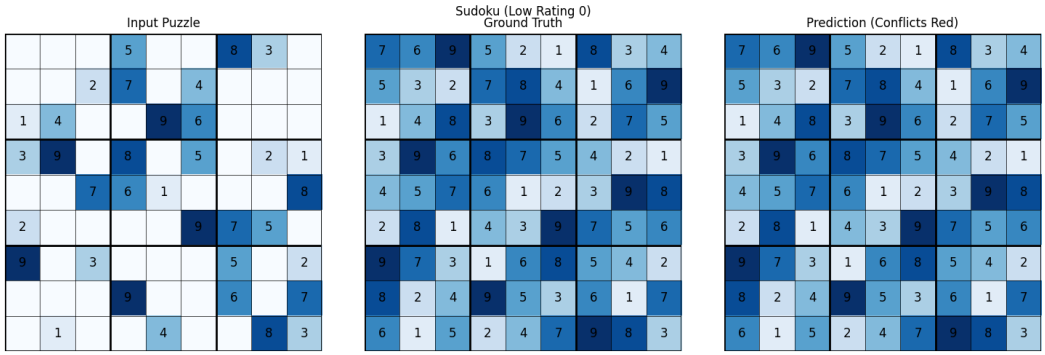
Figure 2: Evaluation results grouped by puzzle difficulty rating

Figure 2 presents the accuracy metrics grouped by puzzle difficulty rating. The relationship between difficulty and model performance is not strictly monotonic; while generally lower difficulty puzzles yield higher accuracy, occasional peaks appear at higher difficulty levels. These anomalous peaks are likely attributable to small sample sizes at extreme difficulty levels, what makes the averaged metrics less reliable for those categories. Overall, the model was able to reach around 18% of puzzle accuracy.

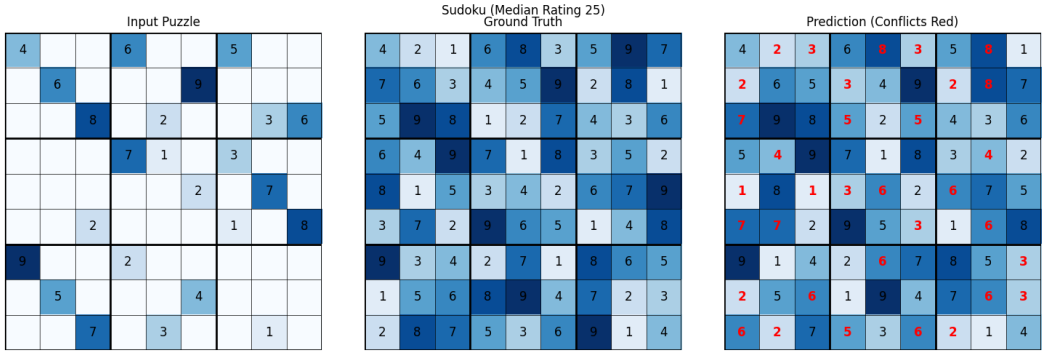
Importantly, the comparison with an untrained model (Figure 2b) demonstrates that the trained TRM has acquired meaningful problem-solving capabilities, as the untrained model achieves near-zero accuracy across all difficulty levels.

6.1.1 Qualitative Analysis

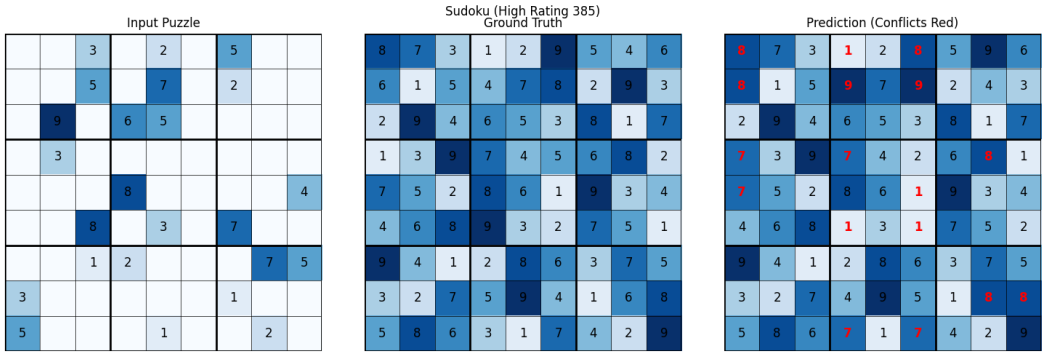
To provide qualitative insight into model behavior, Figure 3 presents example solutions across three difficulty levels. Highlighted cells indicate prediction errors.



(a) Easy puzzle



(b) Medium puzzle



(c) Hard puzzle

Figure 3: TRM solutions across difficulty levels with error highlighting

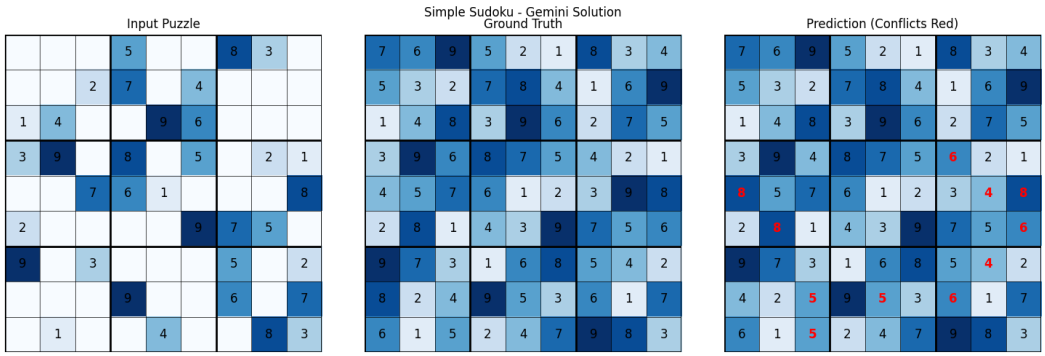
The qualitative analysis reveals that the model successfully solves easy puzzles but exhibits degraded performance on medium and hard instances. This limitation likely stems from the reduced maximum step count (8 versus 16 in the original paper) and model size, which constrains the model's capacity for extended recursive reasoning required by complex puzzles. Moreover, training was much shorter than in original papaer.

6.2 Gemini Evaluation

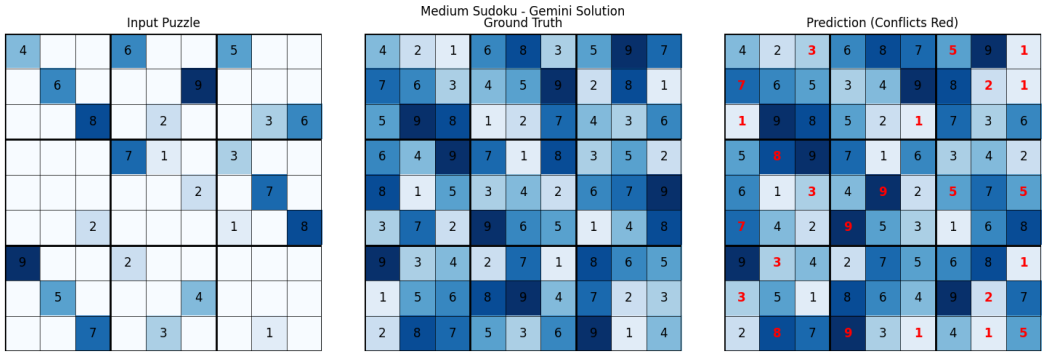
The Gemini 3.0 Pro model was evaluated on three representative puzzles (easy, medium, and hard) using a carefully structured prompt that explicitly stated Sudoku rules and solving constraints. The following prompt template was employed:

```
You are an expert Sudoku solver. Your task is to solve the given
Sudoku puzzle correctly and completely by strictly following the
rules below.

Rules:
1. The puzzle is a 9x9 grid, where each row, column, and 3x3
   subgrid must contain the digits 1-9 exactly once.
2. The puzzle is given as a string of 81 characters, where each
   character represents a cell in the grid.
3. The characters '0' or '.' represent empty cells that need to
   be filled.
4. The characters '1'-'9' represent the digits that are already
   placed in the grid.
5. The puzzle is guaranteed to have a unique solution.
6. You must solve the puzzle by reasoning step by step, showing
   your work and explaining your thought process.
7. You must not make any assumptions or guesses about the solution.
8. You must not use any shortcuts or strategies that are not based
   on logical reasoning.

Sudoku:

+-------+-------+-------+
| 4 . . | 6 . . | 5 . . |
| . 6 . | . . 9 | . . . |
| . . 8 | . 2 . | . 3 6 |
+-------+-------+-------+
| . . . | 7 1 . | 3 . . |
| . . . | . . 2 | . 7 . |
| . . 2 | . . . | 1 . 8 |
+-------+-------+-------+
| 9 . . | 2 . . | . . . |
| . 5 . | . . 4 | . . . |
| . . 7 | . 3 . | . 1 . |
+-------+-------+-------+
```

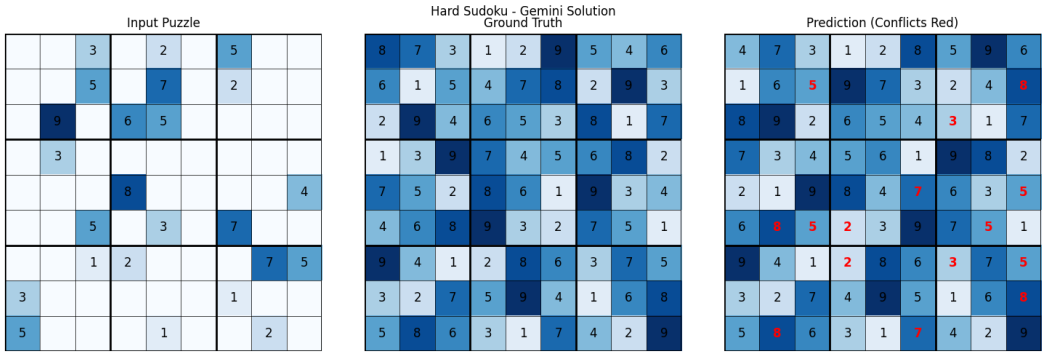
Despite the detailed instructions, the model was unable to produce correct solutions. However, the outputs demonstrate an understanding of Sudoku structure, as the model attempts to fill cells with valid digits.



(a) Easy puzzle



(b) Medium puzzle



(c) Hard puzzle

Figure 4: Gemini 3.0 Flash solutions across difficulty levels with error highlighting

The Gemini model failed to produce correct solutions for all three test cases. However, the generated outputs demonstrate that the model possesses an understanding of Sudoku rules, as it attempts to fill cells with valid digits and produces grid-structured outputs. This suggests that while the model comprehends the task conceptually, it lacks the systematic reasoning capability required for accurate puzzle resolution.

6.3 Comparative Analysis

A direct quantitative comparison between the TRM and Gemini models was not feasible due to the computational cost associated with evaluating the LLM across a large sample set. However, the qualitative assessment reveals a fundamental distinction: despite having orders of magnitude fewer parameters, the TRM successfully solves at least some easy

puzzles after a single training epoch, whereas Gemini fails on all tested instances.

This observation supports the hypothesis presented in [1] that specialized recursive architectures can outperform general-purpose LLMs on structured reasoning tasks, even with substantially reduced parameter counts.

7 Conclusion

The TRM model trained in this study demonstrates limited but meaningful Sudoku-solving capability, successfully resolving easy puzzles while struggling with medium and hard instances. The achieved performance falls short of the results reported in [1], where a 7M parameter model achieved 74.7% accuracy. Several factors contribute to this discrepancy:

- **Model capacity**: The implemented model (3.5M parameters) is half the size of the reference model.
- **Recursive depth**: The maximum step count was reduced from 16 to 8.
- **Training duration**: Only a single epoch was completed, whereas the original work employed more extensive training (3 Days of training on H100 GPU).
- **Dataset differences**: The Sudoku-Extreme dataset may present different difficulty characteristics compared to the augmented dataset used in the original paper.

The Gemini model's failure on all test cases, despite its substantially larger capacity, underscores the limitations of general-purpose LLMs on tasks requiring systematic logical reasoning. This aligns with the motivation presented in [1] for developing specialized recursive reasoning architectures.

The TRM architecture demonstrates promise for structured reasoning tasks, although the training process is computationally intensive relative to model size. Future work could explore increased recursive depth, extended training duration, and curriculum learning strategies to improve performance on harder puzzle instances.

References

[1] Alexia Jolicoeur-Martineau. *Less is More: Recursive Reasoning with Tiny Networks*. 2025. arXiv: 2510.04871 [cs.LG]. URL: https://arxiv.org/abs/2510.04871.

[2] Sapient Inc. *Sudoku Extreme Dataset*. https://huggingface.co/datasets/sapientinc/sudoku-extreme. Dostęp: 2026-01-22. 2024.

[3] Xiangning Chen et al. *Symbolic Discovery of Optimization Algorithms*. 2023. arXiv: 2302.06675 [cs.LG]. URL: https://arxiv.org/abs/2302.06675.