

# TO COMMENT OR NOT TO COMMENT?

Author  
Vidas Bacevičius

Supervisors

Assistant Professor: **Annibale Panichella**  
Supervisor: **Leonhard Applis**

Analysing the Impact of Inline Comments for the Task of Code Captioning

## 1. Introduction

- **AI-assisted development tools** use Machine Learning models to generate code captions
- These models are usually trained with code **without inline comments**
- We want to see what happens if we **include code comments** in training data

## 2. Background

- **Code2seq**<sup>[1]</sup> is a model that performs natural language generation tasks
- **Code Captioning** generates natural language descriptions for code snippets

## 3. Hypothesis

Including code comments in training data will **improve** code2seq's performance for the task of Code Captioning.

## 4. Methodology

Code2seq works by transforming input code into **Abstract Syntax Trees (ASTs)**, and then walking through the AST predicting the most fitting word sequence. We found a way to meaningfully include code comments into the processed AST path:

```
void methodName() {  
    // This is a comment!  
    int a = 1;  
}
```

Figure 1: A code snippet

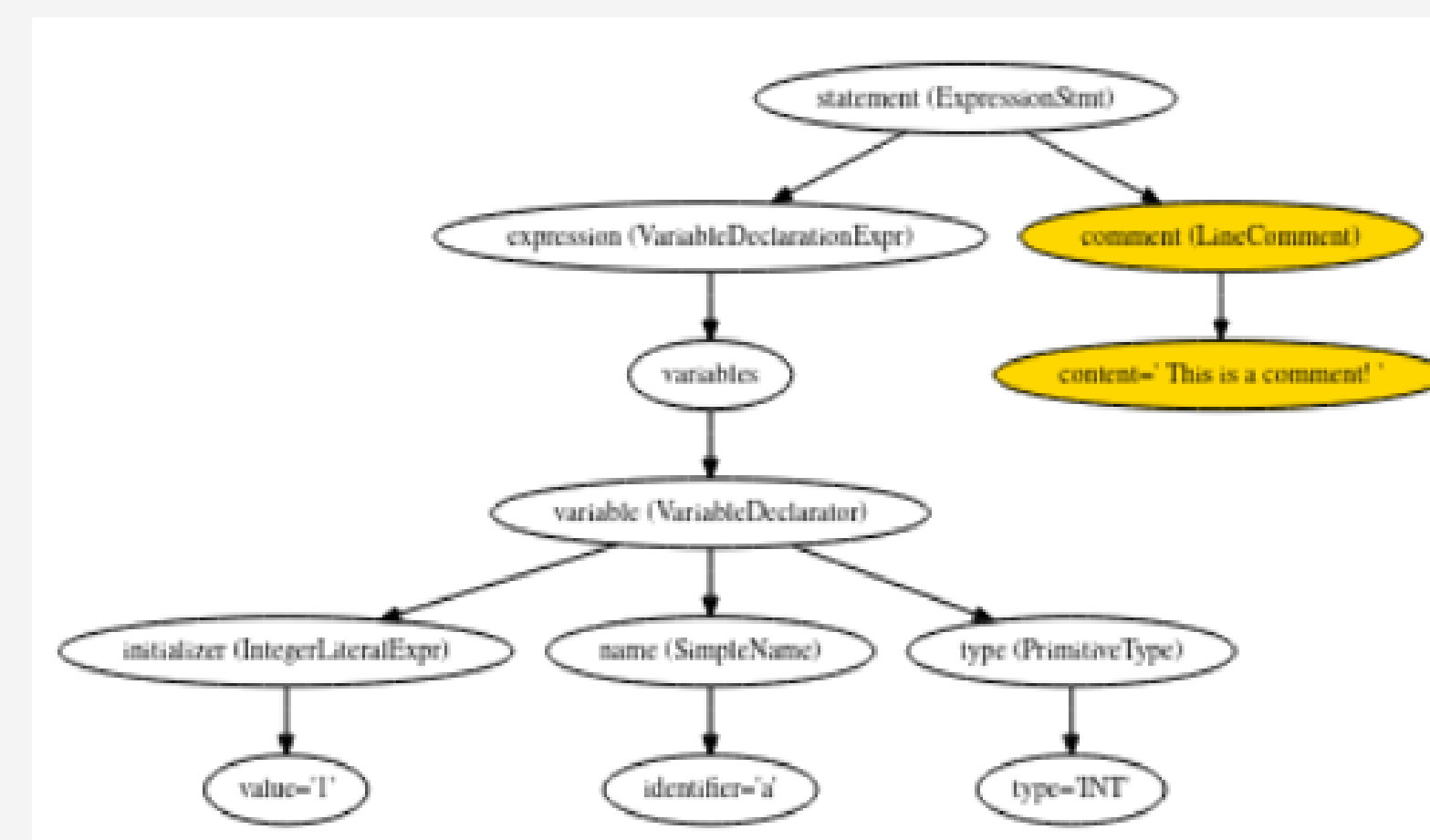


Figure 2: Generated AST with comments

## 5. Experiment

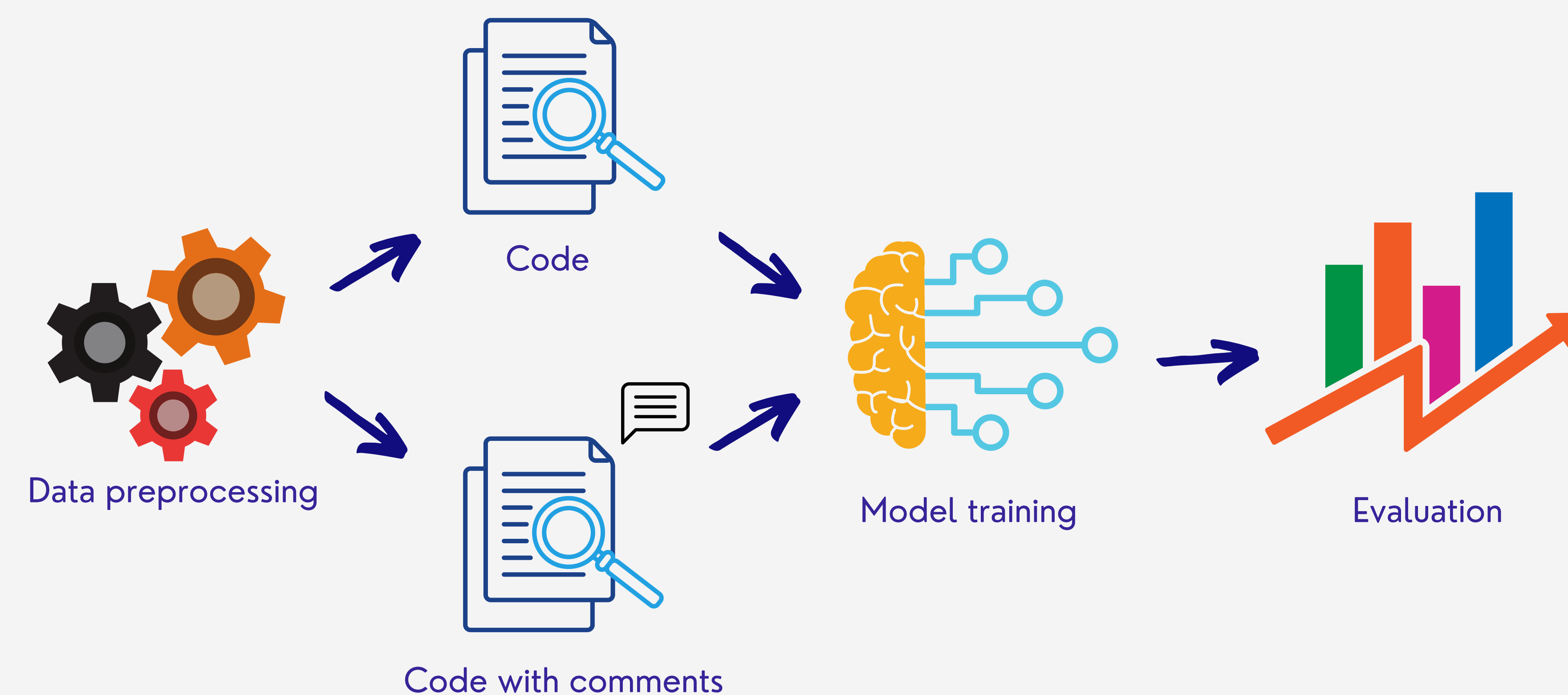


Figure 3: Experimental setup

## 6. Results

	BLEU	F1
No Comments	15.35	0.442
Comments	14.98	0.461

Figure 4: BLEU and F1 scores: code2seq with inline comments and without

	Training Time	Epochs
No Comments	72h	40
Comments	24h	12

Figure 5: Training time and epochs: code2seq with inline comments and without

## 7. Conclusion

- Comments model:
  - Reaches a **higher F1 score**
  - Reaches good scores **3x faster**
  - Produces **more verbose** captions
- No Comments model reaches a slightly higher **BLEU score**

## References

[1] U. Alon, S. Brody, O. Levy, and E. Yahav, "Code2seq: Generating sequences from structured representations of code," in International Conference on Learning Representations, 2019. [Online]. Available: <https://openreview.net/forum?id=H1gKY09tX>