Final Project: Semiconductor Simulation Code Generation

CSC 375/575 - Generative AI | Fall 2025

Prof. Rongyu Lin, Quinnipiac University

Project Overview

Goal: Build a language model system (≤1B parameters) to generate semiconductor device simulation code from natural language circuit design specifications.

About the Simulation Platform: This project uses Silvaco TCAD (Technology Computer-Aided Design), an industry-standard semiconductor simulation platform for device modeling and circuit analysis. You will train models to generate SPICE-compatible simulation code that describes semiconductor device structures and electrical characteristics.

Possible Approaches: You can explore various techniques such as fine-tuning (LoRA, QLoRA), prompt engineering, retrieval-augmented generation (RAG), chain-of-thought prompting, or any combination that works best for your solution. The choice of methodology is completely open.

Format: Individual or team (2-3 students)

Final Presentation: December 3 (10 minutes per team)

Final Submission: December 12, 11:59 PM

Dataset

Download:

Download Dataset (24 MB)

Contents:

- `silvaco_dataset_train.json` - 713 instruction-code pairs for training
- `Silvaco_Examples_Student.zip` - 726 reference .in files + 76 .lib files
- `README.md` - Complete dataset documentation

> **Important:** An additional 97 samples are reserved as a hidden test set for evaluation. Focus on generalization, not memorization.

> ### Dataset Usage Restrictions:
>
> - This dataset is for **CSC 375/575 course use only**
> - **Prohibited:** Sharing, distributing, or publishing this dataset outside of this course
> - **Prohibited:** Using this dataset for other projects, publications, or commercial purposes
> - Violation of these restrictions may result in academic penalties

Model Constraints

> **CRITICAL:** You must use models with **≤1B parameters**.

Allowed models:

- GPT-2 (117M-762M): gpt2, gpt2-medium, gpt2-large
- Llama 3.2 (1B): Llama-3.2-1B
- TinyLlama (1.1B): TinyLlama-1.1B

- DistilGPT-2 (82M): distilgpt2
- Qwen3 (0.6B): Qwen/Qwen3-0.6B-Base (Recommended - Latest 2025 model with strong multilingual support)
- T5 (60M-770M): t5-small, t5-base, t5-large
- BERT (110M-340M): bert-base, bert-large
- Any other pre-trained model ≤1B parameters

Grading Rubric (100 Points Total)

Component	Points	Description
Model Performance	30	Automated evaluation on hidden test set
Implementation & Methodology	30	Training approach and technical implementation
Presentation	20	Live demonstration and explanation
Documentation & Code Quality	20	Technical report and code organization
Total	**100**	

Graduate students (CSC 575): Higher expectations for methodology sophistication, literature review, and analysis depth.

Deliverables

1. **Trained Model:** Model weights and tokenizer (Hugging Face format preferred)
2. **Code:** Training scripts, data preprocessing, evaluation code
3. **Technical Report (maximum 4 pages):**
   - Model selection and justification

    - Training methodology and hyperparameters
    - Results and analysis
    - Failure case analysis

4. **Presentation (10 minutes):** Live demo, methodology, results, Q&A
5. **README:** Setup instructions and usage guide

Submission

Submit via course website:

- All code and model files in a single archive
- Technical report (PDF)
- README with clear setup and usage instructions

Deadline: December 12, 11:59 PM

Academic Integrity

Allowed:

- Pre-trained models ≤1B from Hugging Face
- Standard libraries (transformers, PyTorch, TensorFlow)
- Data augmentation and preprocessing
- ChatGPT/Claude for debugging

Not Allowed:

- Models >1B parameters
- Copying code between teams
- Using external code generation APIs or services
- **Sharing or distributing the course dataset** outside of CSC 375/575

Back to Course Page