



# Integrate a Large Language Model with the xeus-cpp Jupyter kernel

**Mentor/s : Anutosh Bhat, Johan Mabilie, Aaron Jomy, David Lange,  
Vassil Vassilev**

## About Me:

---

**Name** : Tharun Anandh  
**University** : National Institute of Technology, Tiruchirapalli  
**Github** : <https://github.com/tharun571>  
**LinkedIn** : <https://www.linkedin.com/in/tharun-anandh-8b624b199/>  
**Location** : Chennai, Tamil Nadu, India (Hometown)  
Hyderabad, Telangana, India (Current Residence)  
**Slack** : [@Tharun Anandh](#)  
**Email** : [atharun05@gmail.com](mailto:atharun05@gmail.com),  
[atharun571@gmail.com](mailto:atharun571@gmail.com)  
**Resume** : [TharunAnandh\\_Resume](#)  
**Phone no** : +91 80562 03425  
**Timezone** : IST (GMT +5:30)

## Background / Technical Skills:

---

I am Tharun Anandh, fresh out of college, currently working at Oracle as an Application Developer. I completed my Bachelor of Technology degree at NIT Trichy, which is one of the top 10 engineering colleges in the country. I am a software development enthusiast with the motto of learning everything about something and something about everything.

My journey into computer science began during my 11th and 12th grade when I delved into learning C++. I was captivated by its vastness and how it provided logical challenges. This initial exposure fueled my passion for the field. Upon entering college, I ventured into Android Native development, creating numerous applications. Additionally, I explored my interest in data visualization by merging my passion for football with coding. These experiences led me to join the premier coding club of my college, Delta Force, where the distinction between senior and junior members blurred, fostering an environment of continuous learning. Here, I honed my skills in high-scale frontend and backend frameworks, alongside rigorous training in Git.

One of the standout projects during my college tenure was Code Character([Frontend PR](#) | [Backend PR](#)), an AI programming challenge developed for the ISO certified annual tech fest, Pragyam. In this challenge, players utilized a provided high-level API to command troops and seize territories in a real-time strategy simulation game. Additionally, I took on a mentoring role for a project in Delta Winter of Code(a program similar to GSoC), overseeing the creation of an automated [Twitter Bot](#). This bot predicts the number of likes a parent tweet will receive when commented upon. Subsequently, I embarked on an internship at a startup, where I developed an application leveraging Flask, OpenCV, TensorFlow, Pandas, NumPy, Faster R-CNN, and VGG-NET to ascertain the bone age from hand X-rays.

In my final year, I delved into research and authored a paper on "Detection and Avoidance of IoT Routing Attacks using Machine Learning Techniques." This paper has been accepted for publication in the International Journal of

Information Technology. These diverse experiences culminated in securing a position at Oracle, a leading global MNC. Here, I serve as an application developer, contributing to the development of internal tools in Human Capital Management and resolving real-time bugs in Supply Chain management systems. Looking ahead, I am keen on pursuing a master's degree in computer science to further deepen my knowledge and fulfill my aspiration of learning everything about something.

## Why GSoC?:

---

The reason I aspire to participate in GSoC is because it presents an excellent opportunity to immerse myself in the realm of open source. Ever since my college days, I have consistently engaged in various activities, and this program appears to be a fantastic chance to continue expanding my knowledge and enhancing my software skills.

In my pursuit of learning something of everything, I've found compiler design intriguing, yet it's an area I didn't have a chance to delve deeply during my college years. Furthermore, considering the continual advancement of language models, this endeavor promises to not only enhance my understanding of compiler design but also broaden my expertise in the realm of LLMs. Now, amidst the routine of office work, I see an opportunity to embark on a fresh and stimulating journey that merges two of my interests into one project.

## Overview:

---

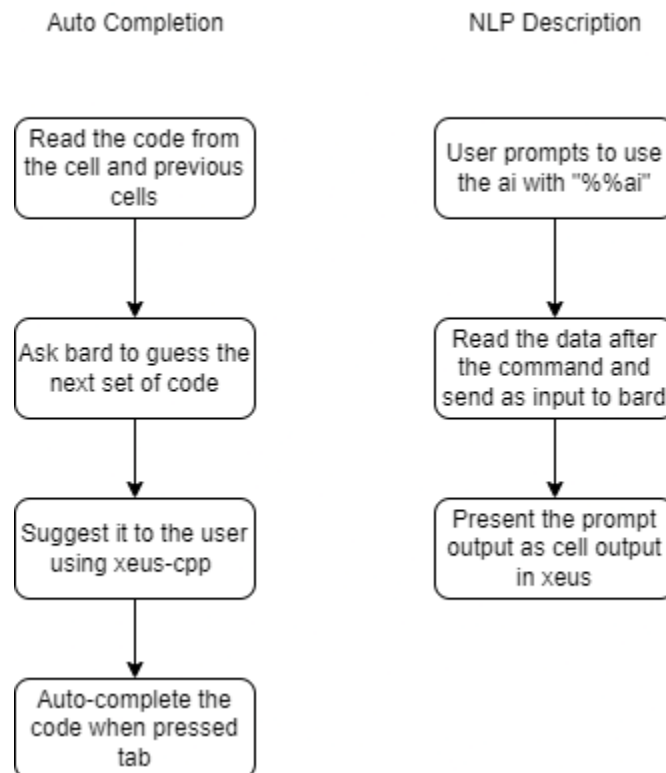
This project aims to integrate a large language model with the `xeus-cpp` Jupyter kernel. This integration will enable users to interactively generate and execute code in C++ leveraging the assistance of the language model.

## Goals:

---

Here are the basic objectives:

- Design and implement mechanisms to interface the large language model with the xeus-cpp kernel.
- Develop functionalities within the kernel to utilize the language model for code generation based on natural language descriptions and suggestions for auto completion.
- Comprehensive documentation and thorough testing/CI additions to ensure reliability.
- [Stretch Goal] After achieving the previous milestones, the student can work on specializing the model for enhanced syntax and semantic understanding capabilities by using xeus notebooks as datasets.



## Implementation:

---

- **Pre-Selection (April 3 - April 30)**
  - Learn more about the project's workflow and the main xeus-cpp repository.
- **Community Bonding (May 1 - May 26)**
  - Interact and Engage with the community.
  - Get to know more about the Mentors.
  - Get to know about the other selected projects and interact with fellow contributors.
  - Make possible changes and updates to the implementation if needed.
- **Milestone 1 (May 27 - June 9)**
  - Learn and write code that captures user input and transfers it from the kernel.
  - Develop the API endpoint to BARD and get the necessary output.
  - Present the completed milestone and get feedback.
- **Milestone 2 (June 10 - June 23)**
  - Present the BARD output to the xeus-cpp in an aesthetic manner.
  - Present the completed milestone and get feedback.
- **Milestone 2 (June 24 - July 7)**
  - Implement and adjust the Autocomplete and NLP Description feature.
  - Organize and categorize different results.
- **Mid-Term Evaluations (July 8 - July 12)**
  - Submit documented code for milestones 1, 2 to the git repository.
  - Discuss the required changes for the completed milestones with the mentors.
  - Write a blog post about the implemented features.
- **Milestone 3 (July 13 - July 21)**
  - Use the first evaluation results to optimize or correct previous tasks.

- Write extensive documentation and rigorous testing, including Continuous Integration (CI) enhancements to make sure the tool adjusts to every possible scenario.
- **Milestone 4 (July 22 - Aug 18)**
  - Optimize the model to accustom more of xeus-cpp code and possible training it with multiple notebooks and public github repos to make our own version of github copilot.
- **Final Evaluations (August 19 - August 26)**
  - Generate the pull requests for the completed milestones.
  - Final presentation of the overall work to the mentor.
  - Hand over all the observations, report, and code to the mentor for the final evaluation.

## Commitments:

---

I am committed to dedicating **30 - 35 hours per week** to my project, with the flexibility to increase my input as needed. Should personal or miscellaneous commitments arise, I will notify my mentor in advance, providing a timeframe for any additional hours required to meet deadlines.

### Working hours:

- 07:00 - 09:00 IST
- 19:30 - 23:30 IST

## Contribution:

---

Closed: <https://github.com/compiler-research/xeus-cpp/pull/37>

Merged: <https://github.com/compiler-research/xeus-cpp/pull/47>