

# Enhancing LLM4S with Multimodal Capabilities

Anshuman Awasthi  
[GSoC 2025 Contributor]



# About Me

Namaste, This is Anshuman Awasthi

From: Indian Institute of  
Information Technology Lucknow, India



LinkedIn:



GitHub:



# Introduction

- LLM4S = large language Models for Scala
- Enables Scala developers to interact with AI models directly from their applications
- Open-source, maintained by the Scala Center & community
- Scala Center - Non-profit organization dedicated to the education and advocacy, of the Scala programming language

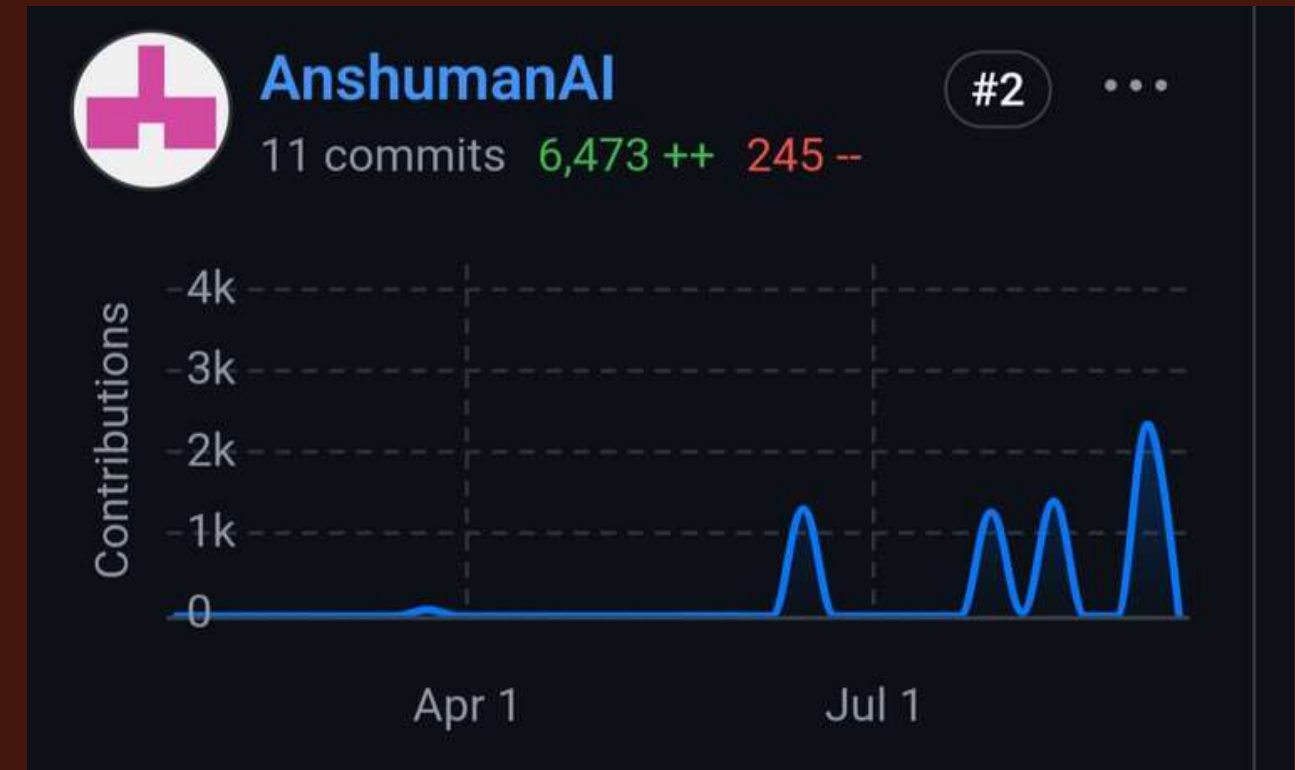
# Project Overview

- **Project Name:** Support image, voice and other  $\text{LLM}$  Modalities
- **Goal:** Extend  $\text{LLM4S}$  beyond text to support speech recognition, text-to-speech, and image generation
- **Purpose:** Enable Scala developers to build rich, interactive AI-powered applications

- **Key Deliverables:** Automatic Speech Recognition (ASR) integration ,  
Text-to-Speech (TTS) Module Integration,  
Image Generation Support, Image Processing Support
- **Impact:** Expands the scope of LLM4S, making it more versatile and developer-friendly
- **Tech Stack:** Scala, Huggingface/OpenAI APIs, Whisper, TTS Engines,  
Image generation models(Stable Diffusion)

# Contributions

- 13 PRs, 8 Effective Merges, LoC-
- Issues Solved (Non PR): 4
- Delivered Image Generation Capabilities
- Delivered Image Processing Capabilities
- Delivered Text to Speech and Speech to Text Capabilities



# Impact & Future Work

- Library can now be used to create Rich Apps, good example being Szork.
- Szrok is DND style Gen AI game .
- Future work : Work on Additional Issues and functionalities.

# Reflection and Acknowledgment

Thanks to the community, Mentors (Rory Graves and Kannupriya Kalra) and Scala Center.