

# myT-LLM - Design Document

Mahanth Yalla  
M.Tech AI,  
Vision & AI Lab (VAL),  
Indian Institute of Science, Bangalore.  
[mahanthya@iisc.ernet.in](mailto:mahanthya@iisc.ernet.in)

## Contents

<b>1</b>	<b>Transformer Architecture</b>	<b>2</b>
1.1	Multi Head Attention . . . . .	2
1.2	Activation Funcs . . . . .	2
<b>2</b>	<b>Pre-Training</b>	<b>2</b>
<b>3</b>	<b>Post-Training</b>	<b>2</b>

# 1 Transformer Architecture

## 1.1 Multi Head Attention

- GQA: Training Generalized Multi-Query Transformer Models from Multi-Head Checkpoints [1]

## 1.2 Activation Funcs

- SwiGLU [2]

# 2 Pre-Training

# 3 Post-Training

## References

- [1] Joshua Ainslie, James Lee-Thorp, Michiel de Jong, Yury Zemlyanskiy, Federico Lebrón, and Sumit Sanghai. Gqa: Training generalized multi-query transformer models from multi-head checkpoints. ArXiv : <https://arxiv.org/abs/2305.13245>, 2023.
- [2] Noam Shazeer. Glu variants improve transformer. ArXiv : <https://arxiv.org/abs/2002.05202>, 2020.