



VIT

Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

School of Computer Science and Engineering

Two-Day Workshop in Collaboration with NVIDIA: Developing and Deploying TensorRT-Optimized Gen AI Models - From the Ground Up

📍 Silver Jubilee Tower 507 , VIT Vellore

📅 October 3rd & 4th 2024 | 10 AM to 5 PM

Registration Fee

Internal participants: Rs. 500

External participants: Rs. 750

Link : <https://events.vit.ac.in/>

(Participants are requested to arrange their own accommodation)

DAY 1: Intersection of Generative and Responsible AI – Models and Methods

DAY 2: Model Optimization and Deployment with NVIDIA TensorRT and Triton Server

Target Audience

Students, Research
Scholars, Faculty Members



Elite
Partner



**Om Ashish
Mishra**

Product Data Scientist
Deloitte



Vishal K

Data Scientist
DigitalTrack Solutions
Pvt Ltd
Nvidia Partner Network

About VIT

VIT was established with the aim of providing quality higher education on par with international standards. It persistently seeks and adopts innovative methods to improve the quality of higher education on a consistent basis. The campus has a cosmopolitan atmosphere with students from all corners of the globe. Experienced and learned teachers are strongly encouraged to nurture the students. The global standards set at VIT in the field of teaching and research spur us on in our relentless pursuit of excellence.

About School of Computer Science and Engineering (SCOPE)

The School of Computer Science and Engineering was conceived in the year 2009 with the mission to strive persistently for achieving excellence in computing disciplines and an aim to carve a niche for itself in Computer Science and allied domains. The school strives to become a pioneering world-class centre of excellence in education and research through collaborative, consultative and participatory approaches. It nurtures effective capabilities for the development of high- quality technical and scientific manpower to meet the challenges of the knowledge era, and provide cost-effective Information and Communication Technology (ICT) - based solutions and value-added services to a variety of organisations and to meet the expectations of stakeholders.

About the Workshop

Generative AI and Large Language Models have gained significant attention in recent years. Gen AI technology is capable of creating new, diverse outputs ranging from images to music and text with Large Language Models (LLMs). This workshop is designed to walk the participants through the essential steps to unlock the full potential of Gen AI technology with the top-notch tools such as NVIDIA TensorRT and NVIDIA Triton. While TensorRT library helps to optimize the models for faster inference speeds. NVIDIA Triton Inference Server, an open-source software helps to deploy, run, and scale AI models on any GPU- or CPU-based infrastructure, cloud, data center or edge.



Elite
Partner

Convenors

Dr. Ramesh Babu . K
Dean, SCOPE

Dr. Saleem Durai M A
Associate Dean, SCOPE

Co-Convenors

Dr. Swathi J N
Head, Dept. of
Computational Intelligence

Dr. Rajkumar S
Head, Dept. of Analytics

Coordinators

Dr. Geraldine Bessie Amali D
Associate Professor
📞 9489134212
✉ geraldine.amali@vit.ac.in

Dr. Varalakshmi M
Associate Professor
📞 8870529139
✉ mvaralakshmi@vit.ac.in



If The Data Matters The Partner Matters



If The Data Matters The Partner Matters