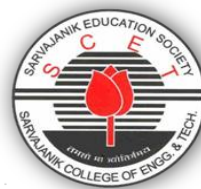


Sarvajanik College of Engineering and Technology
Department of Computer Engineering

Expert Talk organized by
Department of Computer Engineering
under
R & D Cell



Objectives of the Talk: To make students well-aware about the Code LLMs, their significance in addressing code quality challenges, popular use cases, community progress, recent advancements, and future prospects from a renowned industry expert.

Date & Day: 5th April, 2024 (Friday)

Time: 2:00 PM to 3:30 PM

Venue: Online Google Meet

Coordinator: Prof. (Dr.) Nirali Nanavati, Department of Computer Engineering

Speaker Profile:

Ms. Hima Patel is the Senior Technical Staff Member and Research Manager at **IBM Research**. She is with IBM Research since 7 years. Prior to IBM, Ms. Patel has worked as Analytics Manager, **Visa**. She was also the Data Scientist with **Shell Research** and Research Engineer and Image Analyst at **GE Global Research and GE Health Care** respectively. She is an alumni of **DAIICT**, Gandhinagar. Ms. Hima Patel is a seasoned AI practitioner with hands on experience in developing algorithms using machine learning, computer vision and NLP techniques for problems in healthcare, oil and gas and enterprise AI domains.

Summary of the Event:

The Department of Computer Engineering organized an expert talk online under R&D Cell, SCET in association with SSIP & IIC Cell of SCET on 5th April 2024. The talk was attended by approximately 40 students from 3rd and 4th year. The expert talk by Ms. Hima Patel commenced with an exploration of the challenges associated with code quality and the role of Code LLMs in mitigating these challenges. It provided attendees with a deep dive into the world of Code LLMs, covering various aspects such as development lifecycle, use cases, community progress, and recent advancements.

The session discussed the fundamental principles behind LLMs and their applications in software development. Various use cases of Code LLMs were explored, including code completion, bug detection, refactoring, documentation generation, and code summarization. Real-world examples were presented to illustrate the practical applications of LLM technology. The talk delved into the collaborative efforts driving the progress of Code LLMs within the research and developer community. It highlighted key milestones, breakthroughs, and initiatives contributing to the advancement of LLM technology.

Participants were provided with insights into recent advancements in Code LLM technology, including the release of the Claude model, which showcased superior performance compared to previous models. Discussions also revolved around the potential for further improvements and future prospects in LLM development. The talk concluded with an interactive Q&A session, allowing participants to engage with the speaker and delve deeper into topics of interest. Questions regarding code retrieval, context window size considerations, and the trajectory of LLM development were addressed during this session.

Further, the speaker Ms. Hima Patel informed the students about internship opportunities at IBM research and the details of the process and the work they would be undertaking.

Acknowledgement:

We extend our sincere gratitude to our honorable principal Dr. Hiren Patel for permitting us to organize this talk. His constant support and guidance inspire us to organize such enlightening events for our fellow students and colleagues. We are also thankful to Dr. Utpal Pandya, Dean, R&D, Dr. Dipali Kasat, HOD, Department of Computer Engineering, Dr. Mayuri Mehta, R&D Representative for encouragement and prompt support always. We would also like to thank Ms. Hima Patel, IBM Research for accepting our invitation and sparing her valuable time to be with us for the session. Sincere thanks to all the student volunteers for making necessary arrangements for this session. Last but not the least, we are very thankful to all the participants for their active presence during the session.

Photos :



The poster is for an "Expert Session: An Insight into Code LLMs". It features a light blue background with abstract geometric shapes in shades of blue and purple. At the top left are the logos for Sarvajani University and Sarva Engineering College. The title "Expert Session: An Insight into Code LLMs" is prominently displayed in a bold, dark blue font. Below the title, it says "by R & D Cell in collaboration with IIC Cell & SSIP, SCET". A central box contains a portrait of the speaker, Hima Patel, with her name and title "STSM & Research Manager, IBM Research" next to it. Below the speaker information, the date and time "5 April, 2024 2:15 PM Onwards" are listed, along with the online mode "Google Meet" and a URL. A QR code is provided for joining the session. At the bottom, contact information for Prof. (Dr.) Nirali Nanavati is given.

Expert Session:
An Insight into Code LLMs

by R & D Cell in collaboration with IIC Cell & SSIP, SCET

Speaker

 **Hima Patel**
STSM & Research Manager,
IBM Research
[in linkedin/in/patelhima](https://www.linkedin.com/in/patelhima)

5 April, 2024
2:15 PM Onwards

Online Mode - Google Meet
<https://meet.google.com/frn-bpgn-agu>

For more information contact:

Prof. (Dr.) Nirali Nanavati
Associate Professor, Computer Engg. Dept.
nirali.nanavati@scet.ac.in

Scan To Join:



Bringing 75 Years of Leading Innovation

3,000 Researchers

- 2018 Summit and Sierra: World's Fastest Supercomputers
- 2017 Commercial Quantum Computing
- 2016 World's first quantum computer on the cloud
- 2015 Watson Genomic Analytics for Personalized Cancer Treatment
- 2014 Synapse: Biologically Inspired Neural Architecture
- 2013 Antimicrobial Polymers
- 2012 Atomic Imaging (Charge Distribution, Bond Order)
- 2011 Watson Wins Jeopardy!
- 2009 Nanoscale Magnetic Resonance Imaging (MRI)
- 2008 World's First Petalip Supercomputer
- 2007 Web-scale Mining
- 2005 Cell Broadband Engine
- 2004 Blue Gene/L
- 2003 5 Stage Carbon Nanotube Ring Oscillator
- 2000 Java Performance
- 1998 Silicon on Insulator (SOI)
- 1997 Copper Interconnect Wiring
- 1997 Deep Blue
- 1994 Silicon Germanium (SiGe)
- 1990 Chemically Amplified Photoresists
- 1987 High-Temperature Superconductivity (Nobel Prize)
- 1986 Scanning Tunneling Microscope (Nobel Prize)
- 1980 Reduced Instruction Set Computing (RISC)
- 1979 Thin Film Recording Heads
- 1978 Winchester Disk Drive
- 1971 Speech Recognition
- 1970 Relational Database
- 1967 Fractals
- 1966 One-Device Memory Cell
- 1957 FORTRAN
- 1956 Random Access Memory Accounting Machine (RAMAC)

- 6 Nobel Laureates
- 10 Medals of Technology
- 5 National Medals of Science
- 6 Turing Awards

[illegible]

Hima Patel (Presenting)

Let's zoom into code data preparation pipeline

```
graph LR; A[Raw Code Data] --> B[Parquet Creation/Schema Conversion]; B --> C[Exact Dedup]; C --> D[Partition the data to 150-200MB]; D --> E[Prog Lang Detection and Filter]; E --> F[PII Detection]; E --> G[License]; E --> H[HAP]; E --> I[Code Quality Filters]; E --> J[Fuzzy Dedup]; E --> K[Malware Scanning]; F --> L[PII Redaction]; G --> L; H --> L; I --> L; J --> L; K --> L; L --> M[Merge/Sift output of different filters]; M --> N[Tokenisation];
```

Open-source version of this pipeline – coming soon!

10

Code LLM Development Lifecycle

```
graph TD; A[Application Development (Fine Tuning, RAG patterns..)] --- B[Evaluation (Accuracy, Security, Speed..)]; B --- C[Model Building (Pretraining, Model Alignment)]; C --- D[Data Processing and Engineering (Code repositories to code tokens)]; A --- S[Scalable Processing]; B --- S; C --- S; D --- S;
```

Scalable Processing

Compiled by: Dr. Nirali Nanavati