

AI Exploration Challenge – Discover, Analyze & Present the Power of Language Models!

Are you curious about how artificial intelligence is transforming technology, businesses, and everyday life? Here's your chance to explore cutting-edge Large Language Models (LLMs) and Small Language Models (SLMs), learn how they work, understand their strengths and limitations, and identify where they can make a real impact in technology, education and student life!

We invite you to be part of this fast-paced, hands-on challenge (stage 1) that combines learning, experimentation, and problem-solving — without the need to build complex systems or write extensive code. Top selected teams in Stage 1 will be invited for Stage 2 challenge.

In a follow-up challenge (stage 2), we will invite you to build a more involved application to solve a problem of your choice, after the End Semester exam!

Who Can Participate?

- Undergraduate and Postgraduate students from COEP
- Work individually or in teams (up to 3 members)
- Students eager to learn how AI models can solve real-world problems

What Will You Do?

1. *Select a language model* (Large or Small) from available research papers, benchmarks, or tools.
2. *Study its design, claimed capabilities, and limitations* using official documentation and published results.
3. *Analyze benchmark findings* — accuracy, speed, cost, bias, and any other metric that you think is relevant. You are expected to run inference on a small, representative subset of benchmarks suitable for your chosen model and available infrastructure. Full benchmark reproduction is not required.
4. *Identify practical use cases* where the model could be applied very impactfully in software, business, education, student life, or social good. Any hands-on for the use-case(es) is not mandatory, but will carry more weightage.
5. *Prepare two key deliverables:*
 - A slide deck summarizing your study, to be presented in no more than 15 minutes before a jury panel.
 - A poster highlighting your findings, which will be displayed on notice boards across the CSE department — an opportunity for broader recognition!

Exciting Example Use Cases You Can Explore

For Software & Industry

- ✓ Automating code completion, debugging hints, and documentation
- ✓ Summarizing technical reports, logs, or emails
- ✓ Customer support through intelligent chatbots
- ✓ Sentiment analysis and user feedback interpretation
- ✓ Social media moderation and content creation
- ✓ Translation services for global communication

For Students & Learning

- ✓ Generating personalized study plans and practice problems
- ✓ Summarizing lecture notes and textbooks
- ✓ Drafting essays, lab reports, and research summaries
- ✓ Preparing for exams with mock questions and quizzes
- ✓ Explaining difficult concepts with analogies and examples

For Teaching, Faculty & College Administration

- ✓ Automating question paper and assignment generation
- ✓ Attendance tracking and student feedback analysis
- ✓ FAQs and helpdesk chatbots for administrative queries
- ✓ Placement preparation tools and resume suggestions
- ✓ Analyzing student participation and recommending interventions
- ✓ Creating newsletters, event content, and announcements

For Social Good & Innovation

- ✓ Enabling accessibility for students with disabilities
- ✓ Translating regional languages of India to improve inclusivity
- ✓ Assisting mental health support services with empathetic conversations
- ✓ Detecting misinformation and abusive content online

What You Don't Need to Do

- ✗ No need to build complete applications or AI pipelines
- ✗ No requirement for advanced programming skills
- ✗ No need for deep machine learning knowledge

All you need is curiosity, analytical thinking, and a passion for exploring new technologies and new applications!

Awards & Recognition

-  Top 3 Teams (at the end of Stage 2) – ₹3,000 each
-  All Participants – “Dare-to-Try” certificates recognizing your initiative and effort
-  Poster Display – Selected posters will be showcased across the CSE department to inspire fellow students

 Presentation Opportunity – Present your findings to a panel of judges and receive feedback from experts

Key Dates - Stage 1

-  Registration Opens: 24 Nov 2025
-  Study & Analysis Period: 7 weeks from registration
-  Submission Deadline: 16 Jan 2026
-  Presentations (15 mins per team) & Poster Display: 19/20 Jan 2026 (tentative)
-  Result Announcement & Certificates: 23 Jan 2026 (tentative)

Evaluation Criteria - Stage 1

- ✓ Depth of understanding of the chosen model's architecture and capabilities
- ✓ Insightfulness in interpreting and analyzing benchmark and replicated results
- ✓ Creativity, relevance in identifying use cases, Hands-on experiments
- ✓ Clarity, structure, and impact of the slides and articulation
- ✓ Appeal and informativeness of the poster

Support Provided

- ✓ Mentorship sessions to guide your study and refine your analysis (First by TCS Research on 28 Nov 2025 - slides shared with you, Second in early Jan 2026)
- ✓ Pointers to research papers, benchmark datasets, and platforms
- ✓ Template and tips for preparing presentations and posters
- ✓ Opportunity to interact with peers and share ideas

How to Register

While you have already registered, if any of your friends still want to register, ask them to get in touch with Prof. A.M.Abhijit. Teams of up to 3 members can collaborate and submit one slide deck and one poster.

For queries, contact: A.M.Abhijit, R.D.Naik

Models you may consider

Amongst the SLMs that you can potentially run on your laptops or freely infra, you can consider the following, though you are free to consider any other models from the leaderboards:

Model	Params	Remarks
Phi-2	2.7B	Excellent reasoning, well-benchmarked
Gemma-2B	2B	Clean architecture, strong NLP
Qwen-1.8B	1.8B	Multilingual, efficient
TinyLlama	1.1B	Very fast, ideal for experimentation
DistilGPT-2	82M	Great for quick tests & baselines
Mistral-7B (4-bit)	7B	Requires 8–12 GB VRAM
LLaMA-7B (GGUF)	7B	Slow on CPU, ok on good GPU
Qwen-7B (quantized)	7B	Use very small benchmark subsets

LLMs that you can consider are:

Model	Size	Platform	Remarks
LLaMA-3-8B	8B	Groq / Colab	Strong reasoning, modern
Mistral-7B-Instruct	7B	Groq / Colab	Excellent instruction following
Gemma-7B-IT	7B	Colab	Clean Google model
Qwen-7B-Chat	7B	Colab	Multilingual & structured output
LLaMA-3-70B	70B	Groq	Use very small subsets
Mixtral-8x7B	7B	Groq	Focus on qualitative analysis

Benchmarks

Multiple indicative benchmarks (subset) you can consider:

1. Reasoning: GSM8K

2. Knowledge and Comprehension: MMLU
3. Common Sense: HellaSwag
4. Practical benchmarks: News articles for Summarization, Sentiment or spam dataset for Classification, Short code snippets for Code Understanding, Admin Rules / Syllabus text / Job Description to Resume match for College Use-cases

References

1. SLM-Bench: A Comprehensive Benchmark of Small Language Models (2025),
<https://arxiv.org/pdf/2508.15478>
2. A Comprehensive Survey of Small Language Models in the Era of Large Language Models (2025), <https://dl.acm.org/doi/10.1145/3768165>
3. Small Language Models: Architectures, Techniques, and Practical Inference (2025),
<https://arxiv.org/html/2505.19529v1>