SEMINAR ANNOUNCEMENT

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING
COLLEGE OF DESIGN AND ENGINEERING
Website: https://cde.nus.edu.sg/ece

Area: Communications and Networks (CN)

Host: Assoc Prof Mohan Gurusamy

TOPIC	:	**Evaluation of Knowledge Leakage in LLM Editing using Knowledge Graphs**
SPEAKER	:	**Mr Manit Baser** **Graduate Student, ECE Dept, NUS**
DATE	:	**Tuesday, 23 September 2025**
TIME	:	**2:00PM to 3:00PM**
VENUE	:	**Join Zoom Meeting** **https://nus-sg.zoom.us/j/84364308745?pwd=ZFBDbsHX4k7mnNKuDHsqcq7vnIQK61.1** **Meeting id: 84364308745** **Passcode: 369884**

ABSTRACT

Robust model-editing techniques are essential for deploying large language models in practical applications, to enable cost-effective ways to deal with challenges. However, many editing techniques focus on isolated facts, which critically fail to prevent indirect knowledge leakage. To assist users in selecting the right editing technique, we present ThinkEval, a framework to systematically quantify indirect knowledge leakage and ripple effects in model-editing. ThinkEval builds and employs specialized knowledge graphs to analyze the causal structure of facts before and after editing. To support this approach, we present KnowGIC, a benchmark dataset comprising multi-step reasoning paths that precisely measure these complex knowledge transformation effects. Our results show that these techniques struggle to balance indirect fact suppression with the preservation of related knowledge.

BIOGRAPHY

Manit Baser is currently a PhD student and research assistant in the ECE department, working in areas including Model Editing, Trustworthy AI and UAV security. Prior to his doctoral studies, he gained industry experience as a software engineer at Microsoft and Flipkart.