TAMASOMA JYOTIRGAMAYA
Estd. 1995

VALLURUPALLI NAGESWARA RAO VIGNANA JYOTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

AnAutonomous,ISO21001:2018&QSI-GaugeDiamondRatedInstitute,AccreditedbyNAACwith‘A++’Grade
NBAAccreditationforB.Tech.CE,EEE,ME,ECE,CSE,EIE,IT,AME,M.Tech.STRE,PE,AMS,SWEProgrammes
Approved by AICTE, New Delhi, Affiliated to JNTUH, NIRF (2024) Rank band:151-200 in Engineering
Category CollegewithPotentialforExcellencebyUGC,JNTUH-RecognizedResearchCentres:CE,EEE,ME,ECE,CSE
VignanaJyothiNagar,PragathiNagar,Nizampet(S.O.),Hyderabad–500090,TS,India.
TelephoneNo:040-23042758/59/60,Fax:040-23042761
E-mail:postbox@vnrvjiet.ac.in,Website:www.vnrvjiet.ac.in

Department of CSE-(CyS,DS) and AI&DS

Team No: 09 **Date: 12-02-2026**

Sentio-RAG: Mitigating Information Leakage in RAG Systems via Span-Aware Differential Privacy

Abstract:

The integration of Large Language Models (LLMs) into Retrieval-Augmented Generation (RAG) systems over private corpora raises significant concerns regarding sensitive information leakage. Existing Differentially Private RAG (DP-RAG) approaches typically apply uniform perturbation at the document or embedding level, implicitly assuming equal sensitivity across all retrieved text. In practice, sensitive information is localized to specific spans, while most surrounding context is benign, leading coarse-grained privacy mechanisms to unnecessarily disrupt semantic structure and degrade downstream reasoning performance. We present **Sentio-RAG**, a span-aware privacy framework that selectively protects privacy-critical regions while preserving non-sensitive context. Sentio-RAG identifies sensitive spans using a sensitivity scorer, allocates privacy budgets adaptively via Rényi Differential Privacy composition, and perturbs only high-risk spans using the Exponential Mechanism through semantically consistent substitutions. This design provides formal document-level differential privacy guarantees while maintaining grammatical and logical coherence of retrieved content. We evaluate Sentio-RAG on Enron Email, WikiBio, and Privacy-Augmented Squad, benchmarking against document-level DP-RAG baselines. Experimental results demonstrate improved question-answering accuracy and significantly reduced Membership Inference Attack success under equivalent privacy budgets, indicating a superior privacy–utility tradeoff. These findings suggest that span-level privacy localization offers a practical and effective approach for deploying privacy-preserving RAG systems in real-world settings.

Keywords:

Retrieval-Augmented Generation (RAG), Differential Privacy, Span-Level Perturbation, Privacy-Utility Tradeoff, Large Language Models, Sensitivity Scoring

Signature of Guide
(with date)

Student Name & Rollno: Sign
1. J.Mani – 23071A6724
2. T.Venkat Vishnu – 23071A6761
3. B.Vivek – 23071A6764
4. K.Vyshnavi– 24075A6703