Novelty Assessment Report

Paper: VoxPrivacy: A Benchmark for Evaluating Interactional Privacy of Speech Language Models
PDF URL: https://openreview.net/pdf?id=GNo1qMqgPD
Venue: ICLR 2026 Conference Submission
Year: 2026
Report Generated: 2025-12-30

Abstract

As Speech Language Models (SLMs) transition from personal devices to shared, multi-user environments such as smart homes, a new challenge emerges: the model is expected to distinguish between users to manage information flow appropriately. Without this capability, an SLM could reveal one user's confidential schedule to another—a privacy failure we term **interactional privacy**. Thus, the ability to generate speaker-aware responses becomes essential for SLM safe deployment. Current SLM benchmarks test dialogue ability but overlook speaker identity. Multi-speaker benchmarks check who said what without assessing whether SLMs adapt their responses. Privacy benchmarks focus on globally sensitive data (e.g., bank passwords) while neglecting contextually sensitive information (e.g., a user's private appointment). To address this gap, we introduce **VoxPrivacy**, the first benchmark designed to evaluate interactional privacy in SLMs. VoxPrivacy spans three tiers of increasing difficulty, from following direct secrecy commands to proactively protecting privacy. Our evaluation of nine SLMs on a 32-hour bilingual dataset reveals a widespread vulnerability: most open-source models perform close to random chance (around 50\% accuracy) on conditional privacy decisions, while even strong closed-source systems still fall short on proactive privacy inference. We further validate these findings on Real-VoxPrivacy, a human-recorded subset, confirming that the failures observed on synthetic data persist in real speech. We also demonstrate a viable path forward: by fine-tuning on a new 4,000-hour training set, we improve the model's privacy-preserving capabilities while achieving fair robustness. To support future work, we are releasing the VoxPrivacy benchmark, the large-scale training set, and the fine-tuned model to help the development of safer and more context-aware SLMs.

Core Task Landscape

This paper addresses: **Evaluating Interactional Privacy in Multi-User Speech Language Models**

A total of **22 papers** were analyzed and organized into a taxonomy with **10 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Privacy-Preserving Speech Processing Techniques**
- **Multi-User Privacy Management in Speech Systems**
- **General Multi-Party Privacy-Preserving Machine Learning**
- **Multi-Party Conversational Systems**

Complete Taxonomy Tree

- Evaluating Interactional Privacy in Multi-User Speech Language Models Survey Taxonomy
- Privacy-Preserving Speech Processing Techniques
  - Speaker Anonymization Methods (3 papers)
  - [1] A benchmark for multi-speaker anonymization (Xiaoxiao Miao, 2025) View paper
  - [9] Multi-speaker text-to-speech training with speaker anonymized data (Wen-Chin Huang, 2024) View paper
  - [12] Target speaker anonymization in multi-speaker recordings (Tomashenko, 2025) View paper
  - Secure Multi-Party Computation for Speech (3 papers)
  - [7] SecureGPT: A Framework for Multi-Party Privacy-Preserving Transformer Inference in GPT (Chenkai Zeng, 2024) View paper
  - [10] SecureNLP: A System for Multi-Party Privacy-Preserving Natural Language Processing (Qi Feng, 2020) View paper
  - [11] Private-preserving language model inference based on secure multi-party computation (Chen Song, 2024) View paper
  - Federated Speech Learning (3 papers)
  - [5] Privacy-Preserving Data Deduplication for Enhancing Federated Learning of Language Models (Abadi, 2024) View paper
  - [21] FedSpeech: Federated Text-to-Speech with Continual Learning (Ziyue Jiang, 2021) View paper
  - [22] Privacy-Preserving Deep Speaker Separation for Smartphone-Based Passive Speech Assessment (Apiwat Ditthapron, 2021) View paper
- Multi-User Privacy Management in Speech Systems
  - Interactional Privacy Evaluation and Benchmarking ★ (2 papers)
  - [0] VoxPrivacy: A Benchmark for Evaluating Interactional Privacy of Speech Language Models (Anon et al., 2026) View paper
  - [2] Beyond Individual Concerns: Multi-user Privacy in Large Language Models (Xiao Zhan, 2024) View paper
  - Access Control and Authentication (3 papers)
  - [6] A Closer Look at Access Control in Multi-User Voice Systems (Hassan A. Shafei, 2024) View paper
  - [15] Multi-User Smart Speakers-A Narrative Review of Concerns and Problematic Interactions (Nicole Meng-Schneider, 2023) View paper
  - [18] Testing Privacy and Security of Voice Interface Applications in the Internet of Things Era (Shafei, 2024) View paper
  - Ethical and Value-Based Privacy Frameworks (3 papers)
  - [3] Plural voices, single agent: Towards inclusive AI in multi-user domestic spaces (Chandra, 2025) View paper

- [4] Beyond User-centric: Modelling Privacy and Fairness Effects of Speech Interfaces on Community-and Society-Levels (Tom BÃ¤ckstrÃ¶m, 2025) View paper
- [14] Preserving Speech Privacy in Interactions with Ad Hoc Sensor Networks (Zarazaga, 2022) View paper

- General Multi-Party Privacy-Preserving Machine Learning
  - Differential Privacy in Multi-Party Learning (2 papers)
  - [13] Private, Efficient, and Accurate: Protecting Models Trained by Multi-party Learning with Differential Privacy (Wenqiang Ruan, 2022) View paper
  - [17] TPMDP: Threshold Personalized Multi-party Differential Privacy via Optimal Gaussian Mechanism (Jiandong Liu, 2023) View paper
  - Secure Multi-Party Computation Frameworks (2 papers)
  - [19] Privacy-enhanced multi-party deep learning. (Maoguo Gong, 2020) View paper
  - [20] ABG: A Multi-Party Mixed Protocol Framework for Privacy-Preserving Cooperative Learning (Wang, 2022) View paper
- Multi-Party Conversational Systems
  - Multi-Party Dialogue Understanding (1 papers)
  - [8] Do LLMs suffer from multi-party hangover? A diagnostic approach to addressee recognition and response selection in conversations (NicolÃ² Penzo, 2024) View paper
  - Conversational Agent Design for Relational Contexts (1 papers)
  - [16] Design Framework for Conversational Agent in Couple relationships: A Systematic Review (Jung Soyoung, 2025) View paper

Narrative

Core task: evaluating interactional privacy in multi-user speech language models. The field addresses privacy challenges that arise when multiple speakers interact with voice-enabled systems, where protecting one user's information may conflict with another's needs or expectations. The taxonomy organizes work into four main branches. Privacy-Preserving Speech Processing Techniques focuses on cryptographic and anonymization methods that protect speech data at the signal or feature level, including approaches like speaker anonymization and federated learning for speech models. Multi-User Privacy Management in Speech Systems examines how systems handle privacy when multiple individuals are present, covering access control mechanisms, community-level privacy considerations, and evaluation frameworks for interactional privacy scenarios. General Multi-Party Privacy-Preserving Machine Learning provides foundational techniques such as secure multi-party computation and differential privacy adapted for collaborative settings. Multi-Party Conversational Systems explores the design and user experience of voice interfaces in shared environments, including smart speakers and conversational agents used by couples or families.

Several active lines of work reveal key tensions in the field. One strand emphasizes technical anonymization and benchmarking, with studies like Multi-speaker Anonymization Benchmark[1] and Target Speaker Anonymization[12] developing methods to obscure speaker identity while preserving utility. Another explores policy and access mechanisms, as seen in Access Control Voice[6] and Community Privacy Speech[4], which address who should control privacy settings in shared contexts. VoxPrivacy[0] sits within the evaluation-focused cluster alongside Multi-user Privacy LLMs[2], both concerned with measuring how well systems respect privacy boundaries when multiple users interact. While Multi-user Privacy LLMs[2] examines text-based language models in collaborative scenarios, VoxPrivacy[0] extends this lens specifically to speech modalities, where acoustic information and real-time interaction introduce distinct privacy risks. This positioning highlights an emerging need for rigorous benchmarks that capture the nuanced, often conflicting privacy expectations in multi-speaker environments.

Related Works in Same Category

The following **1 sibling papers** share the same taxonomy leaf node with the original paper:

1. Beyond Individual Concerns: Multi-user Privacy in Large Language Models

Authors: Xiao Zhan, William Seymour, JosÃ© Such, Jose Such | **Year/Venue**: 2024 • International Conference on Conversational User Interfaces | **URL**: View paper

Abstract

In this paper, we explore the nuanced and increasingly relevant issue of Multi-user Privacy (MP) in the context of Large Language Models (LLMs). Addressing the gap in current research, we examine how LLMs can inadvertently compromise the privacy of multiple users, particularly in scenarios involving advanced multimodal capabilities. We highlight the challenges in mitigating these privacy concerns, stemming from the complexities of shared data permissions, varying user perceptions of privacy, and...

Relationship Analysis

Both papers belong to the Interactional Privacy Evaluation and Benchmarking category, focusing on privacy challenges in multi-user contexts with language models. While the original paper (VoxPrivacy) specifically addresses interactional privacy in speech language models through a comprehensive benchmark evaluating speaker-aware privacy protection across three difficulty tiers with audio data, the candidate paper takes a broader conceptual approach, exploring multi-user privacy concerns in text-based LLMs with emphasis on shared data permissions and varying user perceptions. The key difference is that VoxPrivacy provides a concrete evaluation framework for speech-based systems with speaker identity verification, whereas the candidate paper offers a more general discussion of multi-user privacy issues without speech-specific mechanisms or benchmarking infrastructure.

Contributions Analysis

Overall novelty summary. The paper introduces VoxPrivacy, a benchmark for evaluating interactional privacy in speech language models (SLMs) operating in multi-user environments. It resides in the 'Interactional Privacy Evaluation and Benchmarking' leaf, which contains only two papers total. This is a notably sparse research direction within the broader taxonomy of 22 papers across multiple branches. The sibling paper examines text-based multi-user privacy in LLMs, making VoxPrivacy the sole work explicitly addressing speech-specific interactional privacy evaluation. This positioning suggests the paper targets an underexplored niche where acoustic modalities and real-time multi-speaker interaction create distinct privacy challenges not covered by existing benchmarks.

The taxonomy reveals that neighboring research directions focus on technical anonymization methods (speaker anonymization, secure computation) and access control mechanisms rather than evaluation frameworks. The 'Privacy-Preserving Speech Processing Techniques' branch contains nine papers addressing cryptographic and anonymization approaches, while 'Access Control and Authentication' includes three papers on permission management. VoxPrivacy diverges from these by providing a measurement tool rather than a protection mechanism. The taxonomy's scope notes clarify that evaluation benchmarks are explicitly separated from technical privacy-preserving methods, positioning this work as complementary infrastructure for assessing existing systems rather than proposing new protection techniques.

Among 30 candidates examined, the three-tiered evaluation framework shows one refutable candidate from 10 examined, while the VoxPrivacy benchmark itself and the large-scale vulnerability findings show no clear refutations among their respective candidate sets. The framework contribution appears to have more substantial prior work overlap within the limited search scope, though the specific tiered structure for privacy capabilities may still offer differentiation. The benchmark and empirical findings appear more novel given the

absence of refuting candidates, though this reflects the 30-paper search scope rather than exhaustive coverage. The speech-specific focus and 32-hour bilingual dataset represent concrete artifacts not directly matched in the examined literature.

Based on the limited search scope, the work addresses a demonstrably sparse research area with minimal direct competition in speech-based interactional privacy evaluation. The taxonomy structure confirms that evaluation frameworks for multi-user speech privacy remain underdeveloped compared to technical protection methods. However, the analysis covers top-30 semantic matches and does not capture the full landscape of privacy benchmarking in adjacent domains (e.g., text-based systems, general dialogue evaluation) that might inform assessments of incremental versus foundational contributions.

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: VoxPrivacy benchmark for evaluating interactional privacy in SLMs

Description: The authors present VoxPrivacy, a novel benchmark specifically designed to assess how well Speech Language Models maintain interactional privacy in multi-user spoken dialogues. It features a three-tiered task structure measuring capabilities from following direct secrecy commands to proactively protecting privacy, accompanied by a 32-hour bilingual dataset, a human-recorded validation subset, and a 4000-hour training set.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Game-Time: Evaluating Temporal Dynamics in Spoken Language Models

URL: View paper

Brief Assessment

Game-Time[46] focuses on evaluating temporal dynamics (timing, tempo, simultaneous speaking) in spoken language models, not privacy evaluation. The benchmarks address fundamentally different capabilities.

---

2. Privacy Disclosure of Similarity Rank in Speech and Language Processing

URL: View paper

Brief Assessment

Similarity Rank Disclosure[45] focuses on quantifying privacy disclosure through similarity rank distributions in biometric identification systems, not on evaluating interactional privacy in multi-user spoken dialogues as VoxPrivacy does.

---

3. A comparative analysis of word-level metric differential privacy: Benchmarking the privacy-utility trade-off

URL: View paper

Brief Assessment

Word-level Metric Privacy[47] focuses on word-level differential privacy mechanisms for text data, not speech language models or interactional privacy in multi-user spoken dialogues.

---

4. Effectiveness of Privacy-preserving Algorithms in LLMs: A Benchmark and Empirical Analysis

URL: View paper

Brief Assessment

Privacy Algorithm Effectiveness[48] focuses on privacy-preserving algorithms for LLMs during training and inference, evaluating utility-privacy trade-offs. VoxPrivacy addresses a fundamentally different problem: interactional privacy in multi-user spoken dialogues for Speech Language Models, where the model must distinguish between speakers and manage information flow based on voice identity. These are distinct research domains with different objectives and methodologies.

---

5. On Differential Privacy for Language Models

URL: View paper

Brief Assessment

Differential Privacy Language[50] focuses on differential privacy techniques for language models to prevent memorization of training data, not on evaluating interactional privacy in multi-user speech dialogues where speaker identity must be distinguished.

---

6. The voiceprivacy 2024 challenge evaluation plan

URL: View paper

Brief Assessment

VoicePrivacy Challenge[43] focuses on voice anonymization to conceal speaker identity in speech data while preserving linguistic and emotional content, not on evaluating interactional privacy in multi-user spoken dialogues where models must manage information flow between different speakers.

---

7. Privacylens: Evaluating privacy norm awareness of language models in action

URL: View paper

Brief Assessment

PrivacyLens[37] focuses on evaluating privacy norm awareness in text-based language models for communication tasks, not speech language models or interactional privacy in multi-user spoken dialogues.

---

8. Audiotrust: Benchmarking the multifaceted trustworthiness of audio large language models

URL: View paper

Brief Assessment

AudioTrust[25] focuses on evaluating trustworthiness of audio large language models across six dimensions (fairness, hallucination, safety, privacy, robustness, authentication), not specifically on interactional privacy in multi-user spoken dialogues as VoxPrivacy does.

---

9. The man behind the sound: Demystifying audio private attribute profiling via multimodal large language model agents

URL: View paper

Brief Assessment

Audio Profiling Agents[44] focuses on inferring sensitive personal attributes (age, gender, occupation, etc.) from general audio using multimodal LLMs, not on evaluating interactional privacy in multi-user spoken dialogues where speech language models must manage information flow between different speakers.

---

10. Long-Form Speech Generation with Spoken Language Models

URL: View paper

Brief Assessment

Long-Form Speech Generation[49] focuses on generating coherent multi-minute speech continuations using state-space models, not on evaluating privacy or speaker identity management in multi-user dialogue contexts.

Contribution 2: Three-tiered evaluation framework for privacy capabilities

Description: The authors develop a structured evaluation framework with three tiers of increasing cognitive difficulty: Tier 1 tests obedience to explicit secrecy commands, Tier 2 requires speaker-verified conditional disclosure using voice as a biometric key, and Tier 3 evaluates proactive privacy protection where models must autonomously infer sensitivity without instructions.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Privacy-preserving Prompt Personalization in Federated Learning for Multimodal Large Language Models

URL: View paper

Brief Assessment

Federated Prompt Personalization[39] focuses on privacy-preserving prompt learning in federated settings for multimodal models, not on evaluating privacy capabilities through tiered cognitive difficulty tests for speech language models.

---

2. Privlm-bench: A multi-level privacy evaluation benchmark for language models

URL: View paper

Brief Assessment

PrivLM-Bench[34] focuses on privacy evaluation for language models through data extraction, membership inference, and embedding-level attacks, not on tiered frameworks for privacy protection capabilities in speech language models.

---

3. Privacylens: Evaluating privacy norm awareness of language models in action

URL: View paper

Brief Assessment

PrivacyLens[37] uses a multi-level evaluation approach (seed, vignette, trajectory) for text-based LMs, not a three-tiered framework testing obedience, speaker-verified disclosure, and proactive privacy protection in speech contexts.

---

4. Data privacy and safety with large language models

URL: View paper

Brief Assessment

Data Privacy LLMs[40] focuses on data protection frameworks and adherence to regulations for text-based LLMs, while the original paper develops a speech-specific framework testing speaker-aware privacy decisions in multi-user voice interactions. The candidate does not address tiered cognitive difficulty levels or voice-based biometric verification.

---

5. Leveraging hierarchical representations for preserving privacy and utility in text

URL: View paper

Brief Assessment

Hierarchical Privacy Text[38] focuses on text privacy through hierarchical word representations in hyperbolic space, not on evaluating speech language models' privacy capabilities across cognitive difficulty tiers.

---

6. Neural pathway embedding through hierarchical interchange networks in large language models

URL: View paper

Brief Assessment

Neural Pathway Embedding[42] focuses on hierarchical interchange networks in LLMs for neural pathway embedding, not on evaluating privacy capabilities through tiered frameworks. The minimal context provided shows references to privacy protection but does not describe a comparable three-tiered evaluation structure for testing obedience, speaker-verified disclosure, and proactive privacy protection in speech language models.

---

7. Exploring the Privacy Protection Capabilities of Chinese Large Language Models

URL: View paper

Prior Art Analysis

Chinese LLM Privacy[35] demonstrates that a three-tiered progressive evaluation framework for privacy capabilities was proposed prior to the original paper. Both papers employ a three-tier structure with increasing cognitive difficulty to evaluate privacy protection in language models. The candidate paper's framework progresses from general privacy information evaluation (Tier 1), to contextual privacy evaluation (Tier 2), to privacy evaluation under attacks (Tier 3), which parallels the original paper's progression from direct command secrecy to speaker-verified secrecy to proactive privacy protection. Both frameworks share the core design principle of escalating difficulty levels to assess privacy awareness and protection capabilities.

Evidence

Evidence 1 - **Rationale**: Both frameworks employ three distinct tiers that progress from basic compliance to more complex contextual reasoning about privacy, showing the tiered evaluation structure was not novel to the original paper. - **Original**: tier 1 tests a model's ability to obey a direct command to keep a secret. tier 2 tests if it can use a voice as a key to share information only with its owner. tier 3, the hardest, asks the model to decide for itself what is secret and act accordingly. - **Candidate**: we developed a three-tiered method to evaluate various chinese large language models, focusing on general privacy information evaluation, contextual privacy evaluation, and privacy evaluation under attacks.

---

8. Hierarchical semantic encoding for contextual understanding in large language models

URL: View paper

Brief Assessment

Hierarchical Semantic Encoding[36] focuses on semantic encoding mechanisms for contextual understanding in LLMs. The limited context provided does not demonstrate a three-tiered evaluation framework for privacy capabilities as described in the original paper.

---

9. Assessing Visual Privacy Risks in Multimodal AI: A Novel Taxonomy-Grounded Evaluation of Vision-Language Models

URL: View paper

Brief Assessment

Visual Privacy Multimodal[41] focuses on visual privacy risks in vision-language models using a taxonomy-grounded approach, not a tiered evaluation framework for speech language models' privacy capabilities as in the original paper.

---

10. MUSE: Machine Unlearning Six-Way Evaluation for Language Models

URL: View paper

Brief Assessment

MUSE[33] focuses on machine unlearning evaluation with six desirable properties for unlearned models, not on tiered evaluation of privacy protection capabilities in interactive speech scenarios. The original paper evaluates interactional privacy in multi-speaker dialogues, while MUSE[33] evaluates data removal efficacy in language models.

Contribution 3: Large-scale evaluation revealing widespread privacy vulnerabilities

Description: The authors conduct a comprehensive evaluation of nine state-of-the-art SLMs, demonstrating that interactional privacy is a critical unsolved problem. Their findings show most open-source models achieve only around 50% accuracy on conditional privacy decisions, establishing clear baselines and identifying specific failure modes through controlled experiments and adversarial tests.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Adversarial Speech for Voice Privacy Protection from Personalized Speech Generation

URL: View paper

Brief Assessment

Adversarial Speech Privacy[26] focuses on protecting voice privacy from personalized speech generation (TTS/VC) through adversarial perturbations, not on evaluating privacy vulnerabilities in speech language models through benchmarking and systematic assessment.

---

2. Assessing the Vulnerabilities of the Open-Source Artificial Intelligence (AI) Landscape: A Large-Scale Analysis of the Hugging Face Platform

URL: View paper

Brief Assessment

Hugging Face Vulnerabilities[31] focuses on security vulnerabilities in AI model repositories and code bases, not on privacy vulnerabilities in speech language models' handling of user information across multi-speaker interactions.

---

3. Configurable privacy-preserving automatic speech recognition

URL: View paper

Brief Assessment

Configurable ASR Privacy[29] focuses on privacy vulnerabilities in automatic speech recognition systems through speech separation and discretization modules, not on evaluating interactional privacy in speech language models across multi-user dialogues.

---

4. How Trustworthy are Open-Source LLMs? An Assessment under Malicious Demonstrations Shows their Vulnerabilities

URL: View paper

Brief Assessment

Open-Source LLM Trustworthiness[30] evaluates privacy in text-based LLMs through adversarial prompting attacks, while the original paper evaluates interactional privacy in speech language models (SLMs) where speaker identity must be inferred from voice. These are fundamentally different modalities and privacy challenges.

---

5. Privacyasst: Safeguarding user privacy in tool-using large language model agents

URL: View paper

Brief Assessment

PrivacyAsst[24] focuses on privacy-preserving frameworks for tool-using LLM agents through encryption and shuffling mechanisms, not on evaluating privacy vulnerabilities in speech language models through large-scale benchmarking studies.

---

6. Backdoor Attacks Against Speech Language Models

URL: View paper

Brief Assessment

Backdoor Speech Attacks[32] focuses on backdoor attacks against speech language models for tasks like ASR and emotion recognition, not on evaluating privacy vulnerabilities in conversational multi-user settings as the original paper does.

---

7. Audiotrust: Benchmarking the multifaceted trustworthiness of audio large language models

URL: View paper

Brief Assessment

AudioTrust[25] evaluates privacy in terms of direct content leakage and paralinguistic inference, not the conditional privacy decisions in multi-speaker contexts that the original paper examines.

---

8. SPIRIT: Patching Speech Language Models against Jailbreak Attacks

URL: View paper

Brief Assessment

SPIRIT[27] focuses on jailbreak attacks and defenses for speech language models, not on evaluating privacy vulnerabilities in the context of interactional privacy or conditional privacy decisions in multi-user environments.

---

9. Audio Jailbreak Attacks: Exposing Vulnerabilities in SpeechGPT in a White-Box Framework

URL: View paper

Brief Assessment

Audio Jailbreak Attacks[23] focuses on adversarial attacks targeting speech input in multimodal models (specifically jailbreak attacks on SpeechGPT), not on evaluating privacy vulnerabilities in open-source speech language models as described in the original contribution.

---

10. Balancing Transparency and Risk: The Security and Privacy Risks of Open-Source Machine Learning Models

URL: View paper

Brief Assessment

Open-Source Model Risks[28] focuses on privacy and security attacks against open-source ML models in general (model inversion, membership inference, backdoors), not on evaluating privacy vulnerabilities specifically in speech language models or interactional privacy in multi-user dialogue contexts.

Appendix: Text Similarity Detection

No high-similarity text segments were detected across any compared papers.

References

- [0] VoxPrivacy: A Benchmark for Evaluating Interactional Privacy of Speech Language Models View paper
- [1] A benchmark for multi-speaker anonymization View paper
- [2] Beyond Individual Concerns: Multi-user Privacy in Large Language Models View paper
- [3] Plural voices, single agent: Towards inclusive AI in multi-user domestic spaces View paper
- [4] Beyond User-centric: Modelling Privacy and Fairness Effects of Speech Interfaces on Community-and Society-Levels View paper
- [5] Privacy-Preserving Data Deduplication for Enhancing Federated Learning of Language Models View paper
- [6] A Closer Look at Access Control in Multi-User Voice Systems View paper
- [7] SecureGPT: A Framework for Multi-Party Privacy-Preserving Transformer Inference in GPT View paper
- [8] Do LLMs suffer from multi-party hangover? A diagnostic approach to addressee recognition and response selection in conversations View paper
- [9] Multi-speaker text-to-speech training with speaker anonymized data View paper
- [10] SecureNLP: A System for Multi-Party Privacy-Preserving Natural Language Processing View paper
- [11] Private-preserving language model inference based on secure multi-party computation View paper
- [12] Target speaker anonymization in multi-speaker recordings View paper
- [13] Private, Efficient, and Accurate: Protecting Models Trained by Multi-party Learning with Differential Privacy View paper
- [14] Preserving Speech Privacy in Interactions with Ad Hoc Sensor Networks View paper
- [15] Multi-User Smart Speakers-A Narrative Review of Concerns and Problematic Interactions View paper
- [16] Design Framework for Conversational Agent in Couple relationships: A Systematic Review View paper
- [17] TPMDP: Threshold Personalized Multi-party Differential Privacy via Optimal Gaussian Mechanism View paper
- [18] Testing Privacy and Security of Voice Interface Applications in the Internet of Things Era View paper
- [19] Privacy-enhanced multi-party deep learning. View paper
- [20] ABG: A Multi-Party Mixed Protocol Framework for Privacy-Preserving Cooperative Learning View paper
- [21] FedSpeech: Federated Text-to-Speech with Continual Learning View paper
- [22] Privacy-Preserving Deep Speaker Separation for Smartphone-Based Passive Speech Assessment View paper
- [23] Audio Jailbreak Attacks: Exposing Vulnerabilities in SpeechGPT in a White-Box Framework View paper
- [24] Privacyasst: Safeguarding user privacy in tool-using large language model agents View paper
- [25] Audiotrust: Benchmarking the multifaceted trustworthiness of audio large language models View paper
- [26] Adversarial Speech for Voice Privacy Protection from Personalized Speech Generation View paper
- [27] SPIRIT: Patching Speech Language Models against Jailbreak Attacks View paper
- [28] Balancing Transparency and Risk: The Security and Privacy Risks of Open-Source Machine Learning Models View paper
- [29] Configurable privacy-preserving automatic speech recognition View paper
- [30] How Trustworthy are Open-Source LLMs? An Assessment under Malicious Demonstrations Shows their Vulnerabilities View paper
- [31] Assessing the Vulnerabilities of the Open-Source Artificial Intelligence (AI) Landscape: A Large-Scale Analysis of the Hugging Face Platform View paper
- [32] Backdoor Attacks Against Speech Language Models View paper
- [33] MUSE: Machine Unlearning Six-Way Evaluation for Language Models View paper
- [34] Privlm-bench: A multi-level privacy evaluation benchmark for language models View paper
- [35] Exploring the Privacy Protection Capabilities of Chinese Large Language Models View paper
- [36] Hierarchical semantic encoding for contextual understanding in large language models View paper
- [37] Privacylens: Evaluating privacy norm awareness of language models in action View paper
- [38] Leveraging hierarchical representations for preserving privacy and utility in text View paper
- [39] Privacy-preserving Prompt Personalization in Federated Learning for Multimodal Large Language Models View paper
- [40] Data privacy and safety with large language models View paper
- [41] Assessing Visual Privacy Risks in Multimodal AI: A Novel Taxonomy-Grounded Evaluation of Vision-Language Models View paper
- [42] Neural pathway embedding through hierarchical interchange networks in large language models View paper
- [43] The voiceprivacy 2024 challenge evaluation plan View paper
- [44] The man behind the sound: Demystifying audio private attribute profiling via multimodal large language model agents View paper
- [45] Privacy Disclosure of Similarity Rank in Speech and Language Processing View paper
- [46] Game-Time: Evaluating Temporal Dynamics in Spoken Language Models View paper
- [47] A comparative analysis of word-level metric differential privacy: Benchmarking the privacy-utility trade-off View paper
- [48] Effectiveness of Privacy-preserving Algorithms in LLMs: A Benchmark and Empirical Analysis View paper
- [49] Long-Form Speech Generation with Spoken Language Models View paper
- [50] On Differential Privacy for Language Models View paper