PriMod4AI: Lifecycle-Aware Privacy Threat Modeling for AI Systems Using LLMs

Gautam Savaliya, Robert Aufschläger, Prof. Dr. Michael Heigl
(Deggendorf Institute of Technology)

JUST BETTER DATA

Motivation & Problem

AI systems process sensitive data across multiple lifecycle stages, creating significant privacy risks and challenges for GDPR compliance [1,2,3]. Traditional frameworks like LINDDUN [4] focus on data-centric risks but fail to capture modern model-driven threats such as membership inference, model inversion, and data leakage. As a result, existing approaches cannot jointly analyze data-level and model-level privacy risks.

Research Question

Can privacy threats in autonomous system architectures be automatically identified across both classical LINDDUN-based and AI-specific risks, while ensuring that retrieval-augmented LLM reasoning provides reliable and explainable threat identification?

Primod4AI Framework

We propose PriMod4AI, a lifecycle-aware privacy threat modeling framework for AI systems, as shown Figure 1:

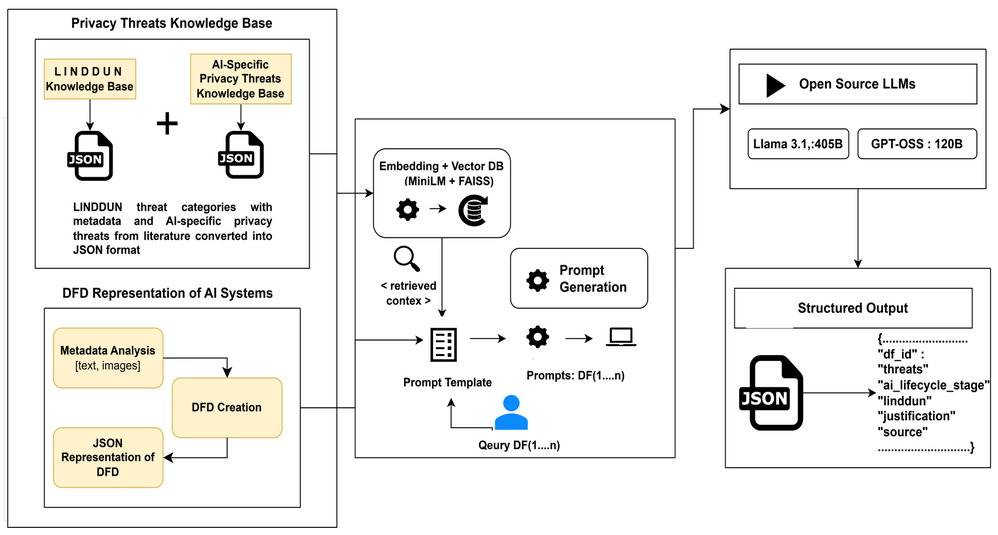

Figure 1: Architecture of the proposed PriMod4AI framework for automated privacy threat modeling in AI systems. The framework integrates a LINDDUN + AI-specific privacy threat knowledge base, DFD representation of AI systems, and open-source LLMs for prompt-based threat identification, producing structured JSON outputs. (© Deggendorf Institute of Technology)

Knowledge & System Modeling: LINDDUN taxonomy and AI privacy threats are converted into structured JSON knowledge bases, while the target AI system is modeled using Data Flow Diagrams (DFDs) to represent data flows, system components, and sensitive information across different stages of the given Autonomous System in Figure 2.

Retrieval & Prompt Generation: Semantic retrieval (MiniLM + FAISS) extracts the most relevant threat knowledge, which is combined with DFD metadata to generate data-flow-specific prompts for analysis.

LLM-based Threat Reasoning: Open-source LLMs (GPT-OSS, LLaMA) perform schema-constrained inference to identify privacy threats across lifecycle stages with provided Prompt and context.

Structured Output Generation: The framework produces explainable and structured JSON outputs including threat ID, threat name, justification, LINDDUN category, lifecycle stage, and source reference, ensuring clear traceability along with improved interpretability.

Use Case

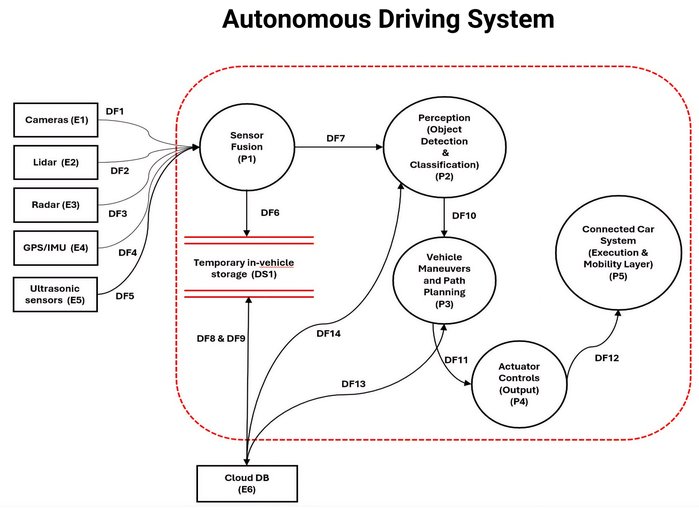

Figure 2: DFD of the autonomous driving system, representing a simplified edge-platform from the jbDATA project, showing key processes, data stores, and data flows (DF1–DF14). (© Deggendorf Institute of Technology)

14 data flows across lifecycle stages

Evaluation Results

Evaluation Aspects	Metric	Value	Interpretation
LINDDUN Comparison: Against PILLAR [5] for classical threats	PILLAR - Recall	85.2%	**High alignment with reference**
	Category Coverage	81.1%	**Broad threat coverage**
	Jaccard Similarity	0.726	**Strong overlap with ground truth**
Cross-Model Agreement: LLaMA vs GPT-OSS consistency	Cohen's K [6]	0.69	**Moderate agreement**
	Observed Agreement	71.5%	**Consistent predictions**
Model-Centric Threats: AI-specific risks beyond LINDDUN	No. of AI- Specific Threats Category Covered	11	**Captures model-based risks and Beyond classical LINDDUN**

Table 1: Evaluation results of PriMod4AI across LINDDUN alignment, cross-model agreement, and AI-specific threat coverage.

Result Interpretation & Key Insights

- High recall (~85%) and coverage (~81%) → Most relevant LINDDUN threats are correctly identified with broad category coverage across system flows.
- Strong similarity (Jaccard ~0.72) and agreement (κ ~0.69, ~70%) → High overlap with reference results and consistent predictions across LLMs.
- Stable robustness (R ~0.61) with 11 AI threat categories → Captures model-centric privacy risks beyond classical LINDDUN.

Conclusions: PriMod4AI demonstrates strong effectiveness in privacy threat modeling by achieving high alignment with LINDDUN, broad category coverage, and consistent results across different LLMs. The framework reliably identifies both classical and AI-specific privacy threats, including model-centric risks, while maintaining stable performance across complex system architectures. These results highlight its capability to provide accurate, explainable, and lifecycle-aware privacy threat analysis for modern AI systems.

References

[1] Liu et al., *Generative AI Model Privacy Survey*, AI Review, 2025.
[2] Carlini et al., *Extracting Training Data from LLMs*, USENIX Security, 2021.
[3] Shahriar et al., *Privacy Risks in AI Lifecycle*, IEEE Access, 2023.
[4] Wuyts et al., *LINDDUN GO*, IEEE Euro S&P Workshop, 2020.
[5] Mollaeefar et al. PILLAR LLM-based LINDDUN, 2025.
[6] Cohen, *A Coefficient of Agreement for Nominal Scales*, Educ. Psychol. Meas., 1960

For more information regarding the poster, contact:

gautam.savaliya@th-deg.de

linkedin.com/company/ki-familie
www.justbetterdata.de

just better DATA is a project of the KI Familie. It was initiated and developed by the VDA Leitinitiative autonomous and connected driving and is funded by the Federal Ministry for Economic Affairs and Energy.

Supported by: Federal Ministry for Economic Affairs and Climate Action
Funded by the European Union NextGenerationEU
on the basis of a decision by the German Bundestag