Bachelor Recherche – Saskia Jürgens

Paper	Datum	Hypothese	Scenario	Ergebnis
Usefulness of data flow diagrams and large language models for security threat validation: a registered report W. Mbaka, Katja Tuma	15. August 2024, arXiv.org	Der tatsächliche Nutzen unterscheidet sich zwischen Gruppen mit und ohne Zusatzmaterial. Die Bereitstellung zusätzlichen Materials (DFD oder LLM) soll die Leistung der Teilnehmenden verbessern, wobei die Kombination beider Materialien deren Fähigkeit zur Identifikation realistischer Bedrohungen erhöhen soll.	Teilnehmende mit Hintergrund in Softwareentwicklung und Cybersicherheitsrisikomanagement. Ein Think-Aloud-Protokoll. 4,5-stündiges Training zu Threat Modeling, Bedrohungen und STRIDE. 1. Zufällige Zuteilung zu einer von vier Gruppen (A, B, C, D). 2. Jeder Teilnehmende erhielt zwei zufällig zugewiesene Szenarien: GitHub-Repository-Update und Kubernetes-Pod-Deployment. 3. Bedrohungsbewertung: Einschätzung von 10 Bedrohungen (5 real, 5 falsch) und Markierung realistischer Fälle. 4. LLM-Interaktion: Gruppen A und C nutzten ChatGPT-3.5 Turbo zur Bedrohungsvalidierung. Interaktionen wurden protokolliert.	Gruppen mit LLM-Unterstützung zeigten höhere Trefferquoten bei der Identifikation realistischer Bedrohungen (A: 8,1; C: 9,4), aber auch mehr Falsch-Positive (A: 7,4; C: 4,7). Gruppe D (ohne Zusatzmaterial) meldete wenige Falsch-Positive, erkannte jedoch ebenfalls viele reale Bedrohungen korrekt. LLM-Nutzer tendierten dazu, erste Vorschläge des Modells zu übernehmen, was insbesondere bei weniger erfahrenen Teilnehmenden zu mehr Fehlbewertungen führen konnte.
A Knowledge Graph Approach to Cyber Threat Mitigation Derived from Data Flow Diagrams Andrei Chis, Oliviu Ionut Stoica, Ana-Maria Ghiran, R. Buchmann.	14. Juni 2024 IC-AQTR	Kein Zugriff auf vollständige Veröffentlichung.	Ein Design-Science-Ansatz nutzt Wissensgraphen zur frühzeitigen Erkennung von Risiken und zur Analyse bestehender Systeme. DFDs werden mithilfe von LLMs in semantisch auswertbare Graphen überführt, was fundierte Entscheidungen unterstützt.	DFDs erleichtern die Analyse und Gestaltung von Unternehmenssystemen – insbesondere bei Cloud-Migrationen – und helfen, Prozesse sowie sicherheitsrelevante Schnittstellen zu verstehen.
A modeling approach to cyber threat mitigation Andrei Chis, Oliviu Ionut Stoica, Ana-Maria Ghiran	2024, BIR Workshops	Die Anwendung von „Secure by Design“-Prinzipien in der Entwicklungsphase kann dazu beitragen, Bedrohungen frühzeitig zu erkennen und zu reduzieren Die Verwendung von Modellierungssystems, Datenflussdiagrammen (DFDs) und Bedrohungsmodellen, wie STRIDE, soll dabei helfen, sowohl technische als auch nicht-technische Sicherheitsbedrohungen zu identifizieren	Online-Shop-Szenario: Modellierung der Geschäftsprozesse (z. B. Produkteinführung, Anmeldung, Bestellung) in einem DFD. Jeder Prozess wird anhand der angewendeten Mitigationsmaßnahmen auf seine Sicherheit geprüft, z. B. beim DoLogin-Prozess gegen Spoofing.	DFDs und Bedrohungsmodelle unterstützen die Identifikation von Risiken und die Entwicklung von Sicherheitsmaßnahmen in der Entwurfsphase. Die Kombination erlaubt eine objektive Bewertung durch Sicherheits-Scores. Aber: Nicht alle Bedrohungen werden erkannt; hohe Komplexität bei mehreren Bedrohungsmodellen. Mit dem proposierten Modellierungsansatz können Unternehmen ihre Sicherheitsstrategie kontinuierlich verbessern.
Human Aspect of Threat Analysis: A Replication	2. August 2022, arXiv.org	Menschliche Faktoren wie Geschlecht, Herkunft und Nationalität können Einfluss auf die	Kontrollierte, explorative Experimente im akademischen Umfeld mit Replikation bestehender STRIDE-Studien zur	Die Studie wurde ausführlich geplant, jedoch nicht durchgeführt – es liegen daher keine Ergebnisse, sondern nur Annahmen vor.

Katja Tuma, W. Mbaka		Ergebnisse von Bedrohungsanalysen haben..	Erfassung menschlicher Einflussfaktoren.	(Relatet Work! Worth mentioning)
Towards Automating a Risk-First Threat Analysis Technique Karanveer Singh, Margit Saal, Andrius Sakalas	18. November 2019	Die Automatisierung der eSTRIDE-Methodik eSTRIDE uses extended Data Flow Diagrams (eDFDs)) durch ein Prototyp-Tool kann den manuellen Aufwand verringern, die Produktivität steigern sowie Präzision und Recall der Bedrohungsanalyse verbessern.	Prototyp-Tool nach Design-Science-Ansatz entwickelt. Das Tool sollte den Nutzern helfen, erweiterte Datenflussdiagramme (eDFDs) zu erstellen oder zu ändern und die eSTRIDE-Bedrohungsanalyse durchzuführen Evaluation in zwei Workshops mit je fünf Studierenden ohne Vorerfahrung. Die Teilnehmenden führten die Analyse einmal manuell und einmal mit Tool durch.	Tool-Nutzer erreichten im Durchschnitt 74,8 % Präzision (vs. 60 % bei manuellen Methoden) und 58,5 % Recall (vs. 62 % bei STRIDE). Produktivität: 1,4 Bedrohungen/Min. mit Tool vs. 1,2/Std. manuel.
Security and Privacy Threat Analysis for Solid Omid Mirzamohammadi, Kristof Jannes, Laurens Sion, Dimitri Van Landuyt, Aysajan Abidin, Dave Singelée	8. November 2023 IEEE Cybersecurity Development	Kein Zugriff auf vollständige Veröffentlichung.	Analyse eines realistischen Finanzanwendungsfalls mit Tools wie SPARTA, STRIDE und LINDDUN zur Bedrohungsmodellierung.	Mehrere kritische, besonders datenschutzbezogene Schwachstellen wurden erkannt. Ziel ist eine bessere Priorisierung zukünftiger Maßnahmen.
Two Architectural Threat Analysis Techniques Compared Katja Tuma, R. Scandariato	24 September 2018	Vergleich der Effektivität, Kosten und des Zeitaufwands zweier STRIDE-Varianten (isoliert vs. interaktiv).	Zwei STRIDE-Ansätze: Variante 1: Einzelne Systemkomponenten. Variante 2: Interaktionen zwischen Komponenten.	Die interaktive Variante war deutlich zeitintensiver. (Paper nicht vollständig verfügbar)
Automating the early detection of security design flaws Katja Tuma, Laurens Sion, R. Scandariato, Koen Yskout	16. Oktober 2020	Kann Teilautomatisierung Sicherheitsanalysen im Designstadium effizienter und zuverlässiger machen.	Einsatz von Modellabfragen zur Identifikation sicherheitskritischer Schwachstellen, z. B. unsichere Datenfreigabe in großen Softwaremodellen.	Vollständige Automatisierung ist nicht möglich, aber gezielte Unterstützung durch automatisierte Hinweise auf Schwachstellen ist realistisch.
A descriptive study of Microsoft's threat modeling technique Riccardo Scandariato, **Kim Wuyts, Wouter Joosen**	1 Juni 2015	Evaluation von Effektivität, Präzision, Recall und Produktivität bei Verwendung von STRIDE. [1]	Eine deskriptives (beschreibendes) Erststudium zur Effektivität von STRIDE die über drei Jahre hinweg läuft. 57 Informatik-Masterstudierende führten STRIDE-basierte Bedrohungsanalysen eines verteilten Systems durch. Dokumentation von Annahmen, Aufwand und Schwierigkeit. Teilnehmer mussten Bedrohungen anhand von DFD-Elementen und STRIDE-Kategorien identifizieren (gemäß einer Checkliste mit Bedrohungsbäumen) und diese als Misuse Cases dokumentieren. Die Teams arbeiteten eigenständig, dokumentierten ihren Aufwand (in Stunden), und bewerteten die Schwierigkeit	STRIDE ist erlernbar und prinzipiell anwendbar. Aber: Zeitintensiv, Recall unter 80 %, Präzision ca. 81 %. 1,2–1,8 korrekte Bedrohungen/Stunde. Viele Teilnehmende unterschätzten übersehene Bedrohungen (FN). T-I-D-Kategorien (Tampering, Information Disclosure, Denial of Service) häufiger erkannt als S-R-E (Spoofing, Repudiation, Elevation of Privilege),. Mehr investierte Zeit führte nicht zwangsläufig zu besseren Ergebnissen. Teilnehmer hatten oft kein realistisches Bild davon, wie viele Bedrohungen sie übersehen hatten (FN), aber es gab mehr korrekte (True Positives, TP) als FP.

			einzelner Schritte per Fragebogen. Die Ergebnisse wurden von zwei Sicherheitsexperten bewertet.	[Betonen wurden methodische Herausforderungen in der Sicherheitsbewertung und regen weitere Forschung an, um Ursachen von Fehlern zu verstehen und das Threat Modeling zu verbessern.]
PILLAR: An AI-POWERED PRIVACY THREAT MODELING TOOL Majid Mollaeefar, Andrea Bissoli a, and Silvio Ranise	11. Oktober 2024 Departmen t of Mathemati cs, University of Trento, Italy	Evaluation, ob das LLM-basierte Tool PILLAR Datenschutzbedrohung en automatisiert erkennen und priorisieren kann – im Vergleich zu manuellen Methoden.	Automatische **DFD-Analyse** (Automatische Erkennung von Datenflüssen, Akteuren, Prozessen und Sicherheitsmaßnahmen), Bedrohungserkennung via STRIDE, **Risikobewertung** (Bewertung jeder Bedrohung mit Risikoklassifizierung), Integration von Standards (z. B. Unterstützung von GDPR, NIST, ISO 27001), **maschinelles Lernen (**Kontinuierliches Lernen aus neuen Sicherheitsvorfällen) für kontinuierliche Verbesserung:	Falsch-Positive/-Negative (Relevante Bedrohungen werden übersehen oder harmloses Verhalten als Gefahr eingestuft), begrenztes Kontextverständnis, fehlende Nachvollziehbarkeit, Datenschutzrisiken und teils unpassende Sicherheitsvorschläge.
THREATMODE LING-LLM: Automating Threat Modeling using Large Language Models for Banking System Shuiqiao Yang, Tingmin Wu, Shigang Liu, David Nguyen, Seung Jang, A. Abuadbba	26. November 2024, arXiv.org	Wie effektiv sind LLMs zur automatisierten Erstellung, Analyse und Pflege von Bedrohungsmodellen – mit Fokus auf Effizienzsteigerung und Aufwandreduktion.	**Automatisierte Erstellung von DFDs:** LLMs analysieren Systemdokumentationen und Quellcode, um automatisch Datenflussdiagramme zu generieren. **Analyse von DFDs:** LLMs identifiziert potenzieller Angriffsvektoren. Klassifikation von Risiken (z. B. Hoch-, Mittel-, Niedrigrisiko). **Bedrohungsszenarien:** LLMs generieren automatisch Bedrohungsszenarien auf Basis der DFDs zur Unterstützung menschlicher Sicherheitsanalysten.	Automatisierung spart Zeit und Ressourcen, LLMs erkennen viele gängige Bedrohungen zuverlässig. Aber: Fehlerhafte DFDs, schlechte Erkennung domänenspezifischer Bedrohungen, Overfitting und keine vollständige Ersetzbarkeit menschlicher Experten.
A Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications Pranab Sahoo, Ayush Kumar Singh, Sriparna Saha, Vinija Jain, S. Mondal, Aman Chadha	5. Februar 2024, arXiv.org	Eine sorgfältige Gestaltung von Eingabeaufforderungen (Prompts) kann den Modellen helfen, Aufgaben erfolgreich zu bewältigen, ohne umfangreiche Nachschulungen oder Anpassungen der Modellparameter vorzunehmen.	Es wird untersucht, wie verschiedene Techniken des Prompt Engineerings – von Zero-Shot und Few-Shot Prompts bis hin zu fortgeschrittenen Methoden wie Chain-of-Thought (CoT) und LogiCoT – die Leistung von LLMs und VLMs in unterschiedlichen Anwendungsbereichen verbessern können. Die Studie betrachtet speziell, welche Ansätze sich für komplexe Aufgaben wie mathematische Problemlösungen, logisches Denken und Fragebeantwortung am besten eignen. [2]	CoT und Auto-CoT erzielten signifikante Verbesserungen in der Genauigkeit, insbesondere bei komplexen Aufgaben wie Mathematik und Logik. Self-Refine zeigte große Fortschritte bei der Verbesserung der Modellantworten hinsichtlich Kontextualisierung und Genauigkeit. Die Graph-of-Thought (GoT)-Methode steigerte die Genauigkeit bei Aufgaben, die komplexe und nicht-lineare Denkvorgänge erforderten. Einige Techniken, wie Chain-of-Symbol und System2Attention, hatten jedoch Probleme mit der Skalierbarkeit und der Generalisierbarkeit auf breitere Anwendungsbereiche. Die Few-Shot Prompting-Technik war aufgrund des Token-Bedarfs bei langen Eingabedaten kostspielig und ineffizient. Herausforderungen wie Bias, Präzision und Interpretierbarkeit bleiben bestehen.

				Ethik ist ein wesentlicher Aspekt in der Entwicklung von LLMs.
LLM-Adapters: An Adapter Family for Parameter-Efficient Fine-Tuning of Large Language Models Zhiqiang Hu, Yihuai Lan, Lei Wang, Wanyu Xu, Ee-Peng Lim, R. Lee, Lidong Bing, Soujanya Poria	2023	Adapterbasierte, parameter-effiziente Fine-Tuning-Methoden (PEFT) ermöglichen es, kleinere Open-Source-Sprachmodelle (wie LLaMA, BLOOM, GPT-J) so leistungsstark oder sogar leistungsfähiger zu machen als große Modelle (z. B. GPT-3.5 oder GPT-4) bei deutlich geringerem Rechenaufwand und weniger zu trainierenden Parametern.	Es wird untersucht, welche Adapter-Typen, -Platzierungen und -Konfigurationen die beste Leistung erzielen. PEFT-Methoden wie der Series Adapter, Parallel Adapter, LoRA (Reparametrisierung) und Prompt-basierte Methoden werden verglichen. 14 Datensätze aus den Bereichen „Mathematisches Denken“ (Arithmetic Reasoning) und „Alltagslogik“ (Commonsense Reasoning) wurden verwendet, um die Adapter in verschiedenen Testszenarien zu evaluieren.	LLaMA-13B mit LoRA übertraf GPT-3.5 (175B) bei mehreren mathematischen Aufgaben (z. B. MultiArith, AddSub, SingleEq). LLaMA-13B + Adapter schlug ChatGPT bei Aufgaben des Alltagsverstands, wenn mit passenden In-Distribution-Daten feinjustiert wurde. Adapter stellen eine leistungsstarke und kostengünstige Alternative zu großen Modellen dar. Es gibt jedoch keine „einheitlich beste Lösung“ für alle Aufgaben. Leistung bei Out-of-Distribution-Daten war schwächer.
The Accuracy and Appropriateness of ChatGPT Responses on Nonmelanoma Skin Cancer Information Using Zero-Shot Chain of Thought Prompting Ross O'Hagan, D. Poplausky, Jade N Young, N. Gulati, Melissa Levoska, Benjamin Ungar, J. Ungar	2023, JMIR Dermatology	Es wird untersucht, wie genau ChatGPT medizinische Informationen zu nicht-melanotischem Hautkrebs (NMSC) liefert und wie sich die Genauigkeit von Standard-Prompting im Vergleich zu Zero-Shot Chain of Thought (ZS-COT) Prompting verhält.	Es werden 25 Fragen zu NMSC in vier Kategorien (Allgemeines, Diagnose, Behandlung, Risikofaktoren) gestellt und sowohl mit Standard-Prompting als auch ZS-COT-Prompting beantwortet. Drei Dermatologen bewerten die Genauigkeit und Eignung der Antworten für Websites und EHR-Nachrichten.	Die Genauigkeit beider Methoden war hoch (Durchschnitt 4,89 von 5). Es gab keinen signifikanten Unterschied zwischen den beiden Methoden. Beide Methoden waren für Websites geeignet, jedoch war ZS-COT nicht signifikant besser als Standard-Prompting.