

Rechercheprotokoll

Autor: Noah Zepner • Datum: 30.09.2025

Ziel und Fragestellung

Ziel: LLM-basierte Informationsextraktion aus polizeilichen Freitexten (DE/EN), Fokus 2021–2025, mit Schwerpunkt auf prototypischer Systemintegration (Workflow/Pipeline) in @rtus.

Leitfragen:

- Welche LLM-Strategien zur strukturierten Ausgabe (JSON, schema-guided) sind belegt?
- Welche domänennahen Arbeiten existieren (DE/EN)?
- Welche methodisch nahen Arbeiten gibt es?
- Welche Integrationsmuster sind praktikabel?

Rahmenbedingungen

- **Datenbanken/Quellen:** Google Scholar, Semantic Scholar
- **Zeitraum/Filter:** 2021–2025 (LLM); ältere Arbeiten für klassische Polizeibericht-IE als Kontext/Baseline zulassen.
- **Sprachen:** DE, EN.
- **Dokumenttypen:** Peer-reviewte Konferenz-/Journal-Beiträge; Preprints separat gekennzeichnet.
- **Fokus/Scope:** Freitext aus Polizeikontext; Informationsextraktion (IE/NER/NLP); Integration in System.
- **Abgrenzung:** Nur textbasierte unstrukturierte Daten (keine Bilder/Audio/Video, kein PDF-OCR). Kein breit angelegter Modell-Leistungsvergleich. Datenschutz wird als vorgeprüft angenommen; keine umfassende Compliance-Analyse.

Suchstrings

Domäne:

- “police report” AND “information extraction” AND LLM
- “incident report” AND (NLP OR “information extraction”) AND LLM
- “law enforcement” AND (NLP OR “information extraction”) AND LLM

Deutsch (Domäne)

- “Polizeibericht” AND (Informationsextraktion OR NLP)
- “Einsatzbericht” AND (NLP OR Extraktion)
- “Vorgangsbeschreibung” AND (Informationsextraktion OR NLP)
- “Strafanzeige” AND (NLP OR extraktion) AND LLM

Methodik LLM/Extraktion

- “LLM-based information extraction” OR (“large language model” AND “information extraction”)
- “structured output” AND (LLM OR transformer) AND (JSON OR XML)
- “schema-guided extraction” OR (“JSON schema” AND (LLM OR transformer))
- “few-shot” AND “information extraction” AND LLM
- “retrieval-augmented” AND (extraction OR “structured output”) AND LLM

Integration/Architektur

- (“workflow” OR “pipeline”) AND (LLM OR “language model”) AND integration
- “enterprise integration” AND LLM

Suchläufe

Datum	Suchstring	Filter	Treffer	Notizen
27.09.2025	“police report” AND “information extraction” AND LLM	“2021–2025; EN/DE”	4	
27.09.2025	“incident report” AND (NLP OR “information extraction”) AND LLM	“2021–2025; EN/DE”	0	
30.09.2025	“law enforcement” AND (NLP OR “information extraction”) AND LLM	“2021–2025; EN/DE”	1	
30.09.2025	“Polizeibericht” AND (Informationsextraktion OR NLP)	“2021–2025; EN/DE”	0	
30.09.2025	“Einsatzbericht” AND (NLP OR Extraktion)	“2021–2025; EN/DE”	0	
30.09.2025	“Strafanzeige” AND (NLP OR extraktion) AND LLM	“2021–2025; EN/DE”	0	
30.09.2025	“LLM-based information extraction” OR (“large language model” AND “information extraction”)	“2021–2025; EN/DE”	2	
30.09.2025	“structured output” AND (LLM OR transformer) AND (JSON OR XML)	“2021–2025; EN/DE”	3	
30.09.2025	“few-shot” AND “information extraction” AND LLM	“2021–2025; EN/DE”	1	
30.09.2025	“retrieval-augmented” AND (extraction OR “structured output”) AND LLM	“2021–2025; EN/DE”	2	
30.09.2025	(“workflow” OR “pipeline”) AND (LLM OR “language model”) AND integration	“2021–2025; EN/DE”	2	
30.09.2025	“enterprise integration” AND LLM	“2021–2025; EN/DE”	2	

Inklusions- und Exklusionskriterien

Inklusion:

- IE/Strukturierung aus unstrukturiertem Text (NER/NLP).
- LLM/Transformer-basiert (ab 2021) ODER klassische IE zu Polizeiberichten als Kontext/Baseline.
- Empirische Methode/Evaluation oder nachvollziehbare Pipeline/Implementierung.
- Bezug zu Integration/Workflow erwünscht.

Exklusion:

- Reine Klassifikation ohne Extraktion.
- Nicht-textuelle Modalitäten (Bilder, Audio, Video) und PDF-OCR-Spezialfälle.
- Meinungs-/Positionspapiere ohne Methode/Evaluation.
- Kein Volltext oder nicht zugänglich.
- Breite Modell-Leistungsvergleiche ohne Integrationsbezug.

Screening-Prozess

1. Titel/Abstract-Screening
2. Volltextprüfung der geeigneten Beiträge.
3. Snowballing: Vorwärts/Rückwärts über Zitationen/Referenzen.

Limitierungen

- Keine Modalitäten jenseits von Text; keine OCR-Spezialfälle.
- Kein umfassender Modell-Leistungsvergleich (Fokus Integration/Workflow).
- Zeitlich begrenzt; Zugriff auf lizenzierte Datenbanken ggf. eingeschränkt.