OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG
FAKULTÄT FÜR INFORMATIK

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

Anonymisieren von Transkriptionen

Sandro Hartenstein
sandro.hartenstein@hwr-berlin.de

08.7.2024

TAHAI

Agenda

- Motivation
- Anonymisieren
- Prototyp
- Ergebnisse
- Fazit

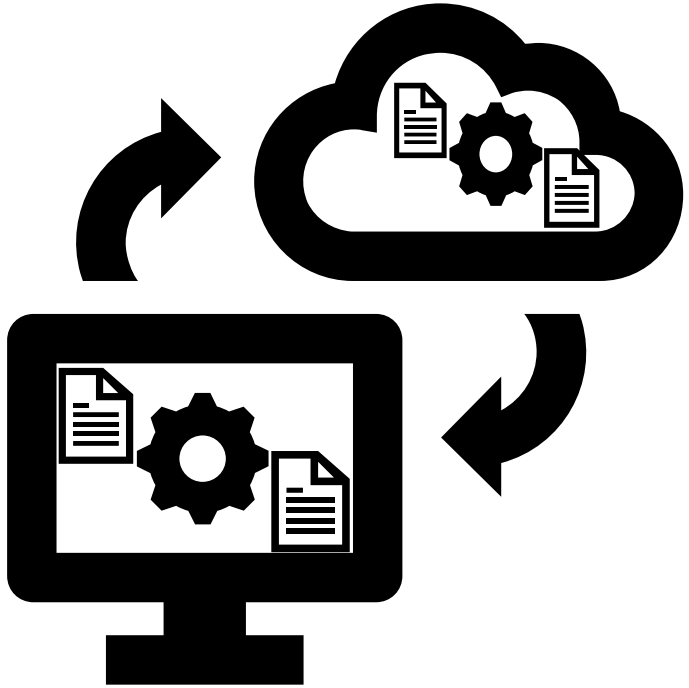
Motivation

Motivation

- Datenschutzkonforme Analyse von Mediationstranskriptionen mit KI-WebAPIs
  - Schutz der personenbezogene Daten der Medianten und Mediatoren
  - Beibehaltung der Kerninhalte

Motivation

- Ermitteln der Möglichkeiten zur automatischen Anonymisierung von Text
- Neueste Entwicklungen prototypisch testen
- Validationsmöglichkeiten



Anonymisieren

Anonymisieren

https://www.johner-institut.de/blog/gesundheitswesen/anonymisierung-und-pseudonymisierung/

Anonymisieren Möglichkeiten

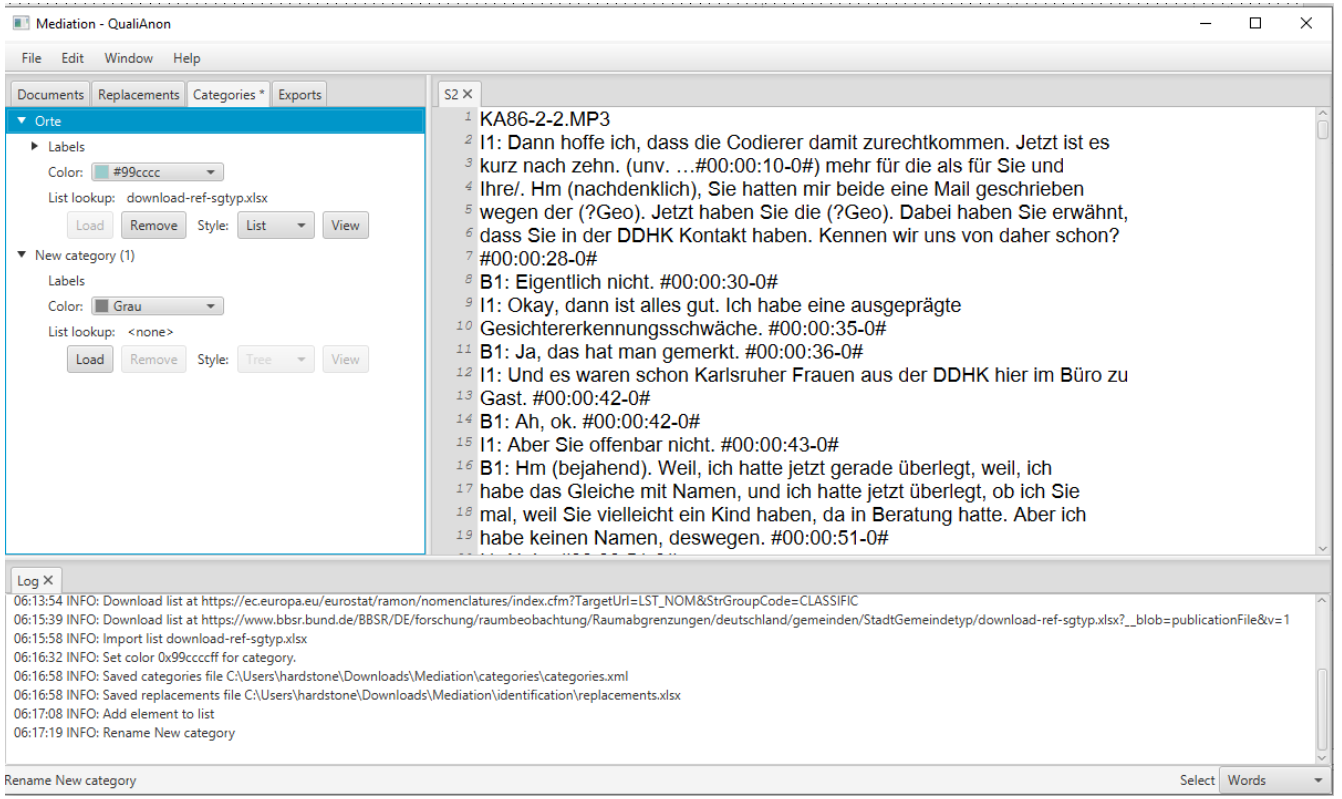
Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

Tools	**Scripts**	Services

• QualiAnon
  • Nur docx imput
  • Listenbasierte Suche
• **Openredac**
  • semi-automatisch
  • deutschsprachige Dokumente!
	• **NLP (**NER)
	• Klassifizieren von
	  • Namen
	  • Orten
	  • Zeitangaben
	• Ersetzen mit generierten Entitäten
		• Out of Scope

Anonymisieren
Möglichkeiten
QualiAnon

- Klassifikation nach Listen
  - müssen importiert werden
  - XLS auch individuelle Listen
- Für Forschungsdaten entwickelt
- Unterschiedliche Anonymisierungsgrade (Exports) möglich
- Nur Docx
- https://www.qualiservice.org/de/helpdesk/webinar/tools.html

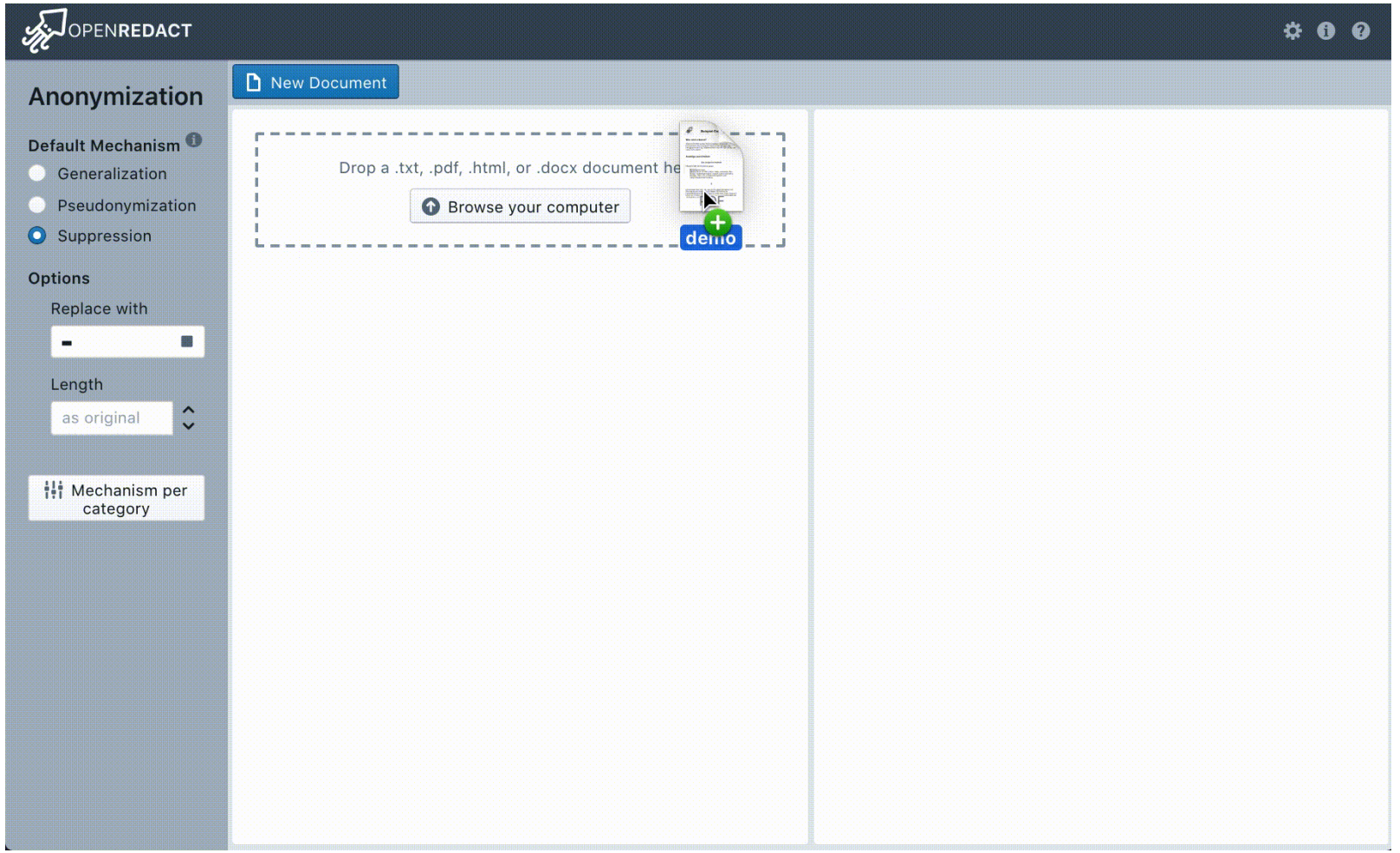


Anonymisieren Möglichkeiten OpenRedact

- Klassifikation (NER)
  - NERwhal
    - regulären Ausdrücken
    - Deep Learning
    - spaCy
    - FlashText

Für Dokumente entwickelt

- Unterschiedliche Anonymisierungsgrade möglich

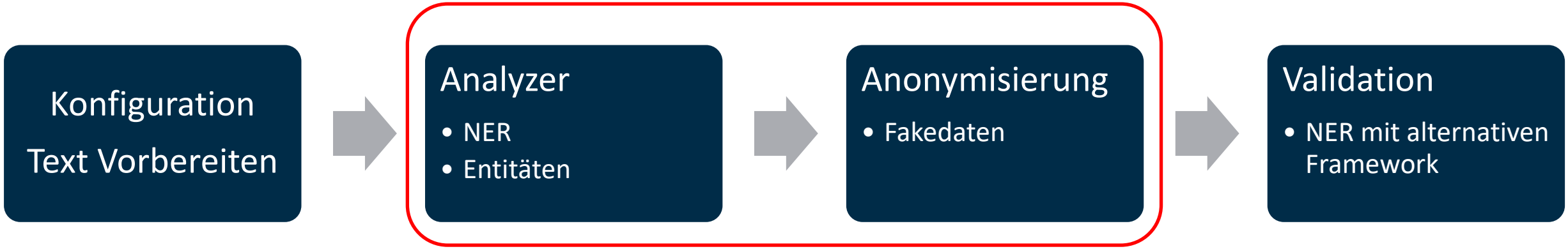


https://openredact.org/prototypefund

Anonymisieren Frameworks

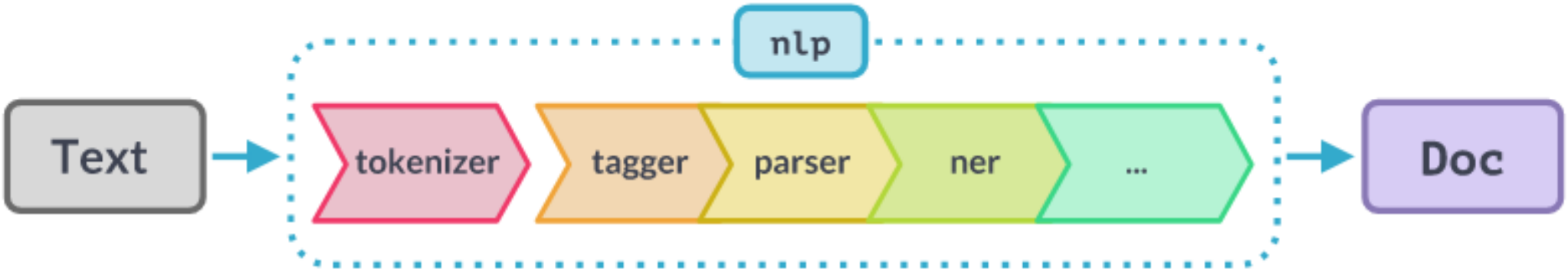
Basic NLPs	Huginface / Transformer	**MS Presidio**
• **spaCy** • NLTK • Flair • SciKit-Learn	• Vortrainierte Modelle • Eigenes Training • NER via LLM • Generate Fakes	• Auf PII vortrainierte NLP Modelle • Analyze (Spacy default) Tokenbased • Anonymizer

Anonymisieren / Pseudonymisieren
Prozess / Pipeline



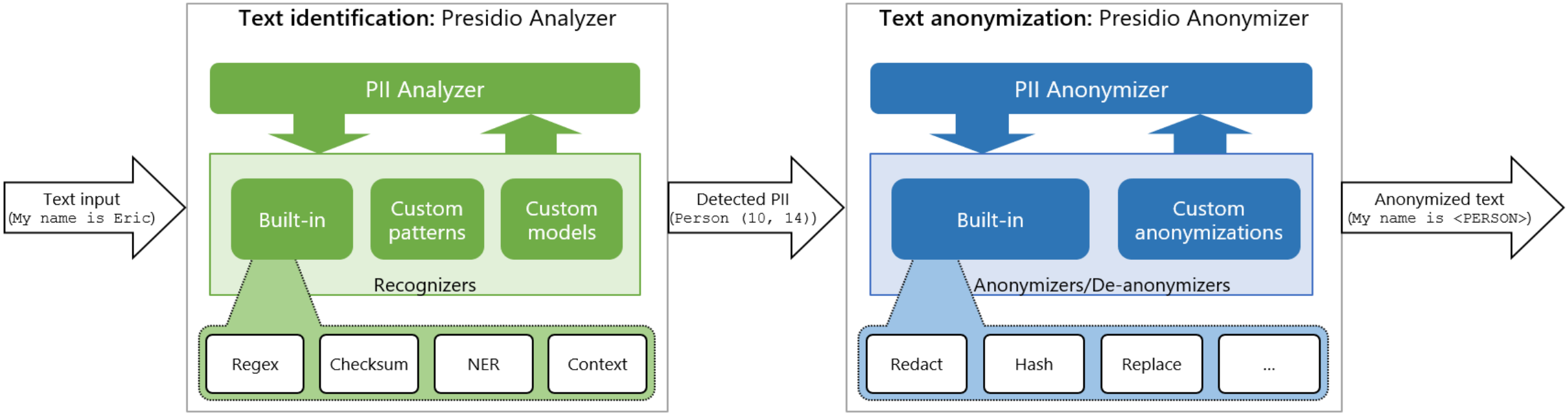
https://arunprasad86.medium.com/nlp-pipeline-a-primer-d9f2a5129e94

Anonymisieren
Spacy



https://spacy.io/usage/processing-pipelines

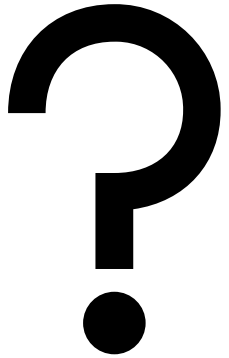
Anonymisieren
Presidio



Prototype

Prototypen
Herausforderungen

- Vorbereiten der Transkripte
  - Format (rtf -> txt)
  - Codepage (utf-8)
  - Aggregation (S1-S6 -> full.txt)
- Ausschließlich lokale Verarbeitung
  - Limits (Spacy z.B. 1mio Zeichen)
  - Beachtung der Anforderung (Sinninhalt muss erhalten bleiben) -> Fakedaten statt Schwärzen
- Validation
  - Stichproben (bzw. Findingstellen anzeigen. Und Gegenanalysen)

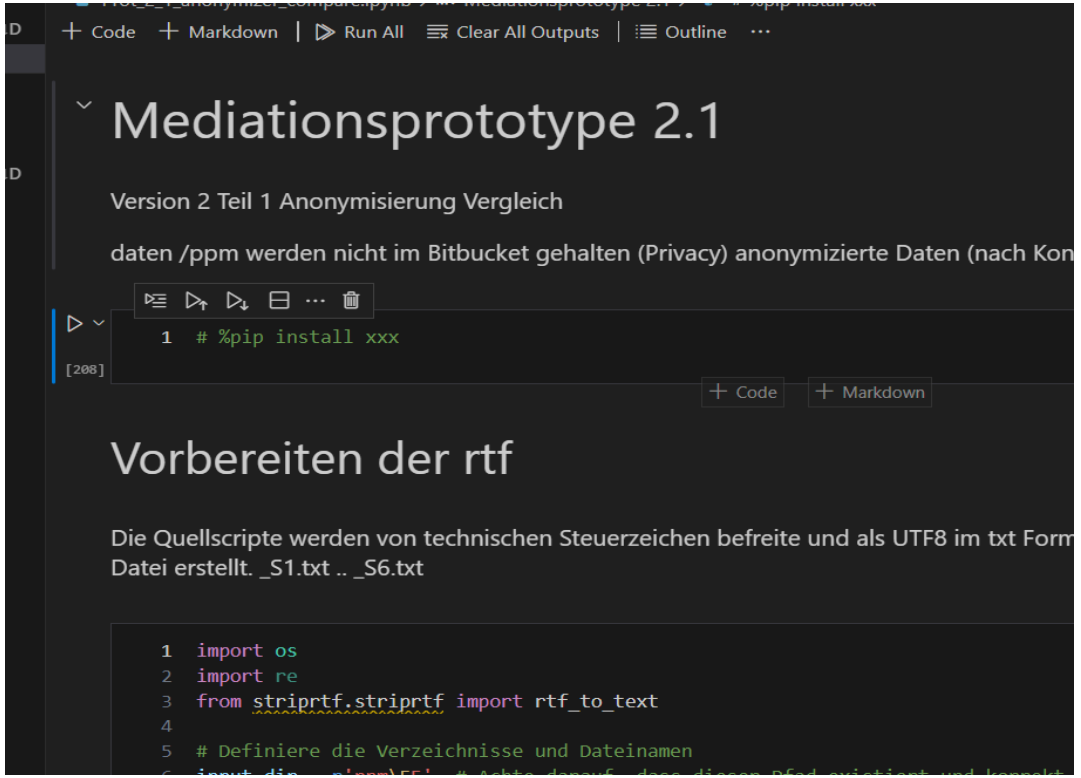


Prototype
Versuchsaufbau Vergleich

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

Python (Jupyter Notebook)

1. Konfiguration (Dateinamen, Modelle, Ausnahmen)
2. Laden des Transkript
3. Festlegen der Entities (Analyse) -> NER / NER via LLM / Transformer
4. Festlegen der Operators (Anonymisierung) → random / Faker
5. Schreiben des Anonymen Transkript
6. Manuelle Validierung



https://github.com/hardstoneed/ai_mediation_analyse

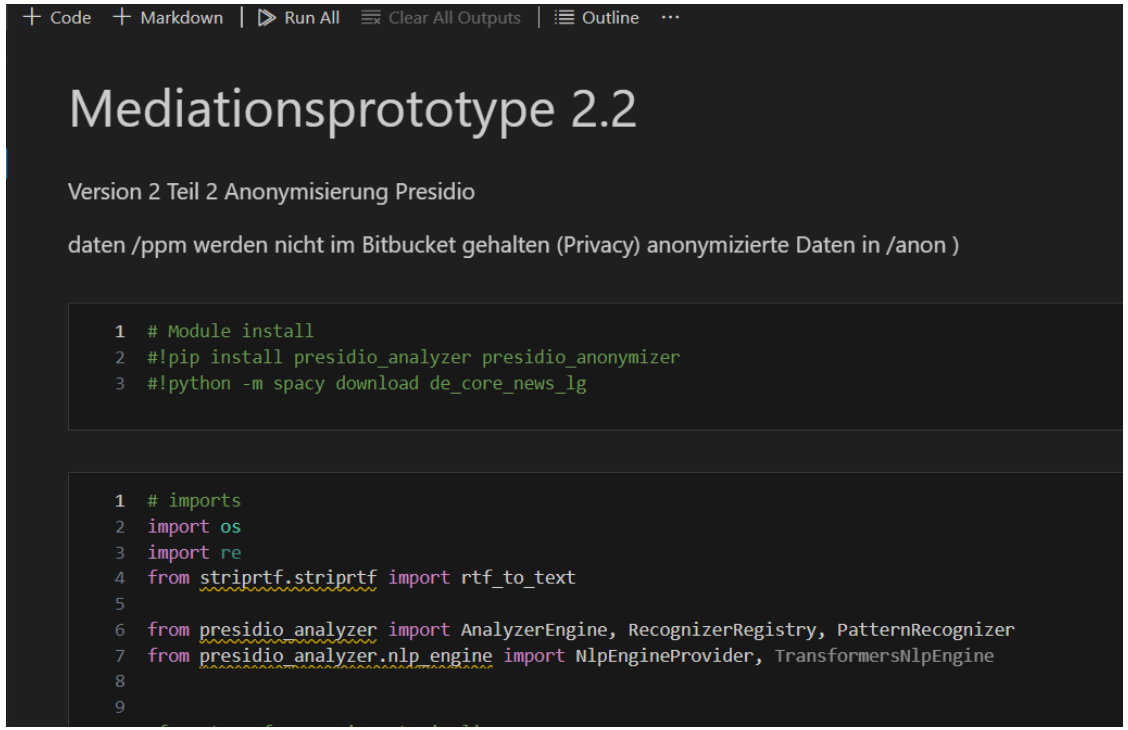
Prototype Experimente

Config / Framework	Transformer	SpaCy	MS Presidio	OpenRedact
Model	Isotonic/distilbert-base-german-cased_finetuned_ai4privacy_v2	de_core_news_lg	Spacy: de_dep_news_trf Transformers: domischwimmbeck/bert-base-german-cased-fine-tuned-ner Regex -> Pattern	Use nerwahl -> SpaCy(Tokenize) und Stanza (NER)
score_threshold	0.8	0.8	0.8	
entities		["PER", "ORG", "GPE", "LOC"]	["CREDIT_CARD", "PERSON_NAME", "PHONE_NUMBER", "LOCATION"]	["COUNTRY", "DATE", "EMAIL", "LOC", "MISC", "MONEY", "NUMBER", "ORG", "PER", "PHONE"]
Demotext	Bernd wohnt in Berlin. Max Mustermann arbeitet bei Bosch und wohnt in (Karlsruhe?). in der Auenstrasse 3. Max hat aber auch Verwandte in Berlin.			

Prototype
Versuchsaufbau
Presidio

Python (Jupyter Notebook)

1. Konfiguration (Dateinamen, Modelle, Ausnahmen)
2. Laden des Transkript
3. Festlegen der Entities (Analyse)
4. Festlegen der Operators (Anonymisierung)
5. Schreiben des Anonymen Transkript
6. Manuelle Validierung



https://github.com/hardstoneed/ai_mediation_analyse

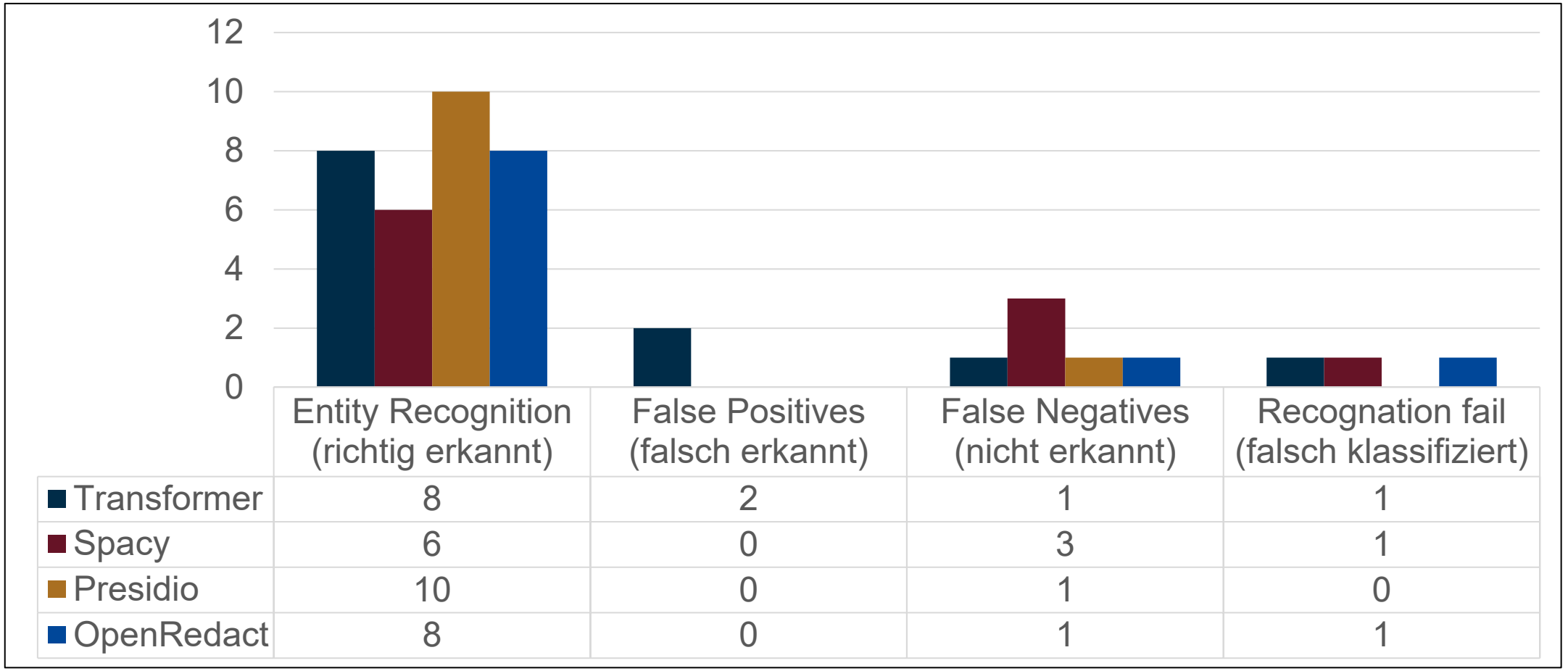
Ergebnisse

Ergebnisse

	Transformer Pipeline	SpaCy	MS Presidio	OpenReduct
Pseudonymisierter Text mit Markierung für Ungenauigkeiten	<PER 1> wohnt in <LOC 1>. <PER 2> <PER 1>er<PER 3><PER 4> arbeitet bei <ORG 1> und wohnt in <LOC 2> in der Auestrasse 3. <PER 3> hat aber auch Verwandte in <LOC 3>	<PER 1> wohnt in <LOC 1> . <PER 2> <PER 3> arbeitet bei <ORG 1> und wohnt in Karlsruhe in der Auestrasse 3. Max hat aber auch Verwandte in <LOC 2>	<PERSON_2> wohnt in <LOCATION_0>. <PERSON_1> arbeitet bei <ORGANIZATION_0> und wohnt in (<LOCATION_2>?). in der <LOCATION_1> 3. <PERSON_0> hat aber auch Verwandte in <LOCATION_0.>	<PERSON 1> wohnt in <LOCATION 1>. <PERSON 2> arbeitet bei <PERSON 3> und wohnt in (<LOCATION 2>). in der <LOCATION 3> 1 <PERSON 4> hat aber auch Verwandte in <LOCATION 1>.
Genauigkeit (Qualität) Demotext	einige False Positives	einige Nicht Erkennung	Sehr genau	Sehr genau
Laufzeit Demotext	2.1s	1.2s	3.4	5s
Laufzeit Mediation (incl. read and write file)	40.3s	4.1s	282.0s	622s
Genauigkeit Mediation ZeroShot (qualitativen Fehleranalyse)	einige False Positives	einige False Negatives	Sehr genau	Sehr genau
Fall 5			1044 ersetzungen in 18m	Fail (timeout)
Fall 8			569 Ersetzungen in 11m	Fail (timeout)

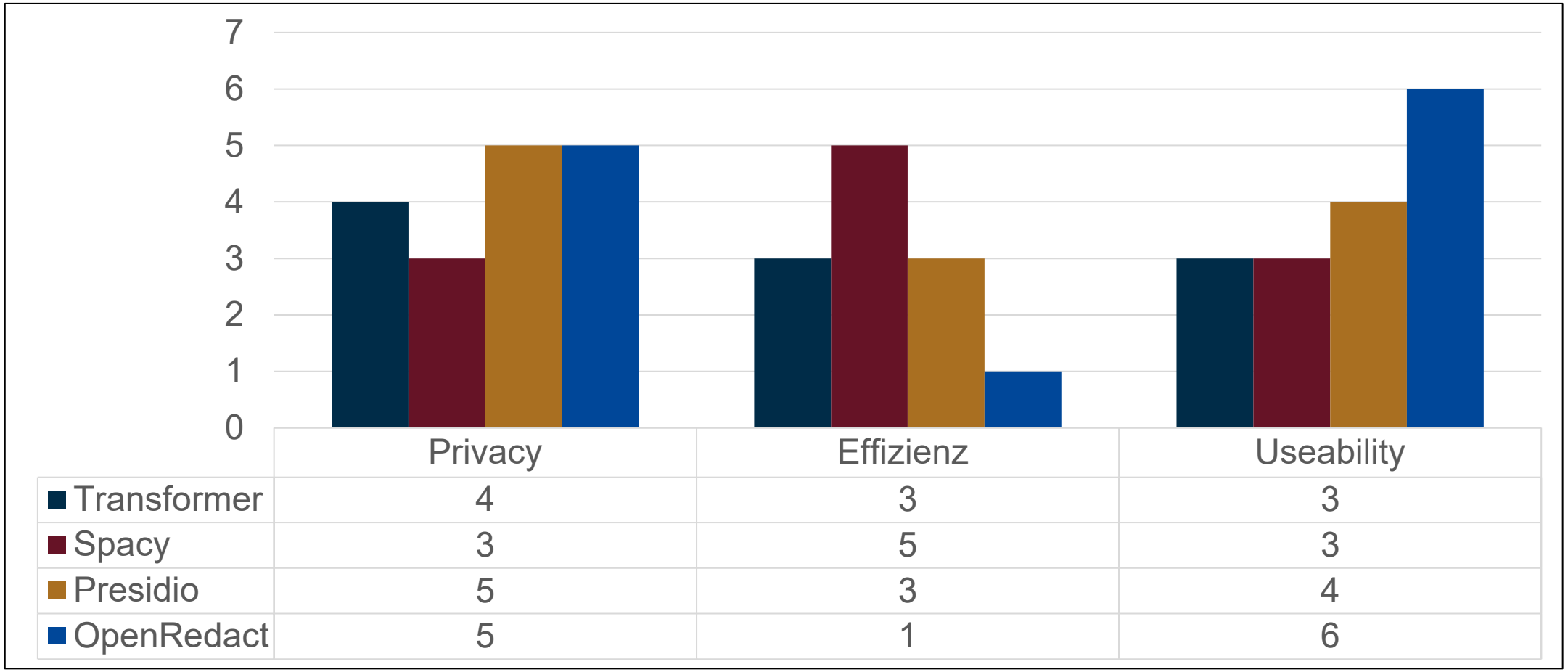
Ergebnisse – Genauigkeitsmessung

Hochschule für
Wirtschaft und Recht Berlin
Berlin School of Economics and Law



	Entity Recognition (richtig erkannt)	False Positives (falsch erkannt)	False Negatives (nicht erkannt)	Recognation fail (falsch klassifiziert)
Transformer	8	2	1	1
Spacy	6	0	3	1
Presidio	10	0	1	0
OpenRedact	8	0	1	1

Ergebnisse – Leistungsbewertung



	Privacy	Effizienz	Useability
Transformer	4	3	3
Spacy	3	5	3
Presidio	5	3	4
OpenRedact	5	1	6

Fazit

Fazit
Lessons-Learned

- openreduct oder ms presidio liefern gute Ergebnisse
  - Bei Standardtexten und Dokumente
  - wenig Konfiguration
- Transkriptionen sind für die automatisierte NER herausfordernd
  - Teilweise besondere Syntax (Pausen, Zwischenreden, Klammern)
  - Unterschiedliche Grade der Anonymisierung für den Sinnerhalt notwendig
    - z.B. Volkwagen (Produkt oder Organisation)
  - Manuelle Unterstützung zielführend
    - Listen für bekannte Entitäten

Fazit
Ausblick

- Bereitstellung und Erläuterung des Prototypen für Fachexperten
- Dokumentation des Prototypen
  - Technische Anforderungen
  - Best Practices
- Dokumentation der Experimente publizieren

OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG
FAKULTÄT FÜR INFORMATIK

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

Vielen Dank!
Fragen?

Sandro Hartenstein
sandro.hartenstein@hwr-berlin.de

TAHAI