

A Structured Overview of Use Cases for Natural Language Processing in the Legal Domain

Martina Preis

03.07.2023, Master Thesis Final Presentation

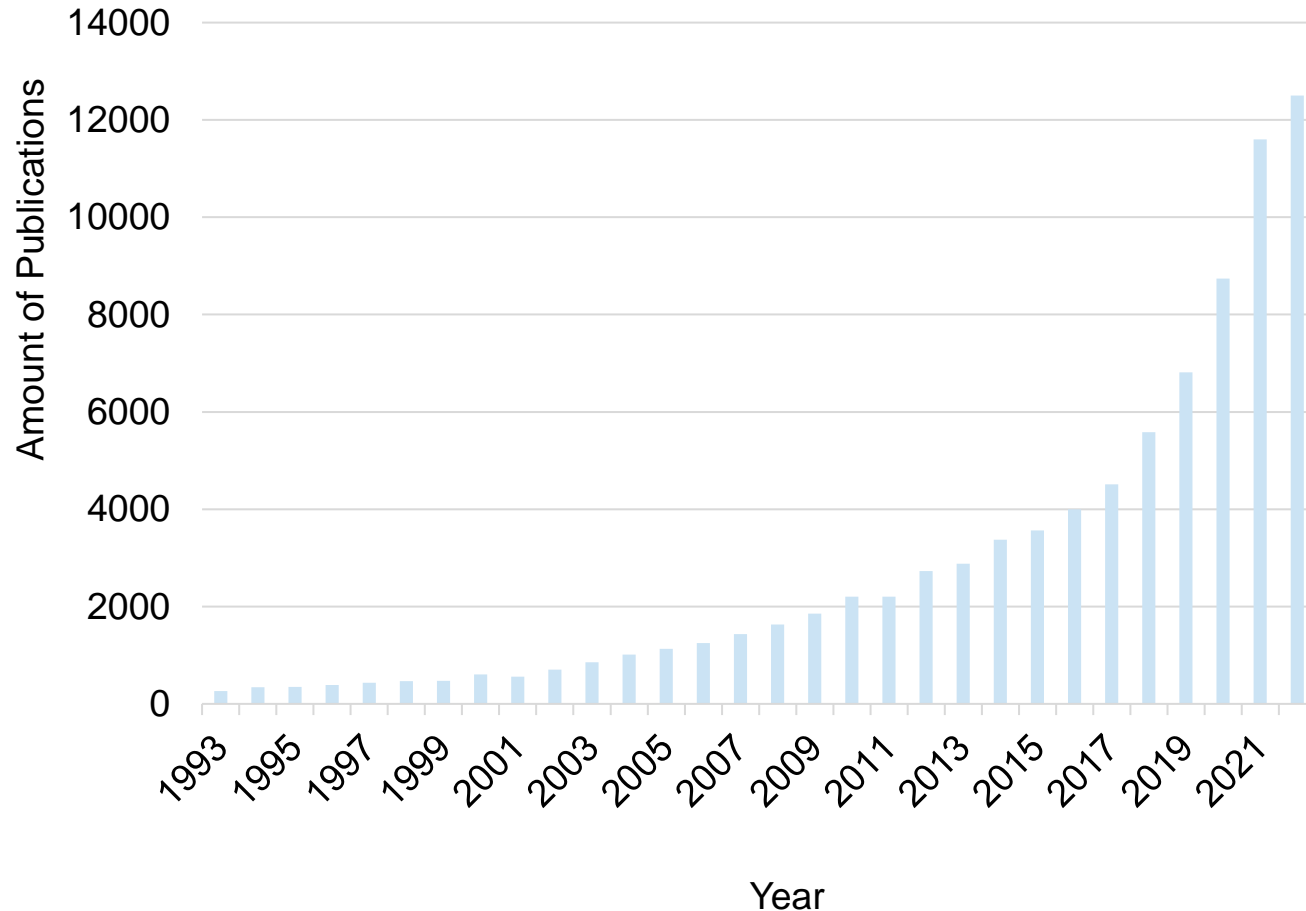
Chair of Software Engineering for Business Information Systems (sebis)
Department of Computer Science
School of Computation, Information and Technology (CIT)
Technical University of Munich (TUM)
www.matthes.in.tum.de

Outline



1. Motivation
2. Research Questions
3. Methodology
4. Results and Interesting Findings
5. Limitations and Future Work

Motivation - Boost of Legal Tech and NLP



Project NLawP – Natural Language Processing and Legal Tech

How AI technologies can impact the legal sector

Partners

- SEBIS (Prof. Dr. Matthes)
- Professorship of Law, Science and Technology (Prof. Dr. Djeflal)

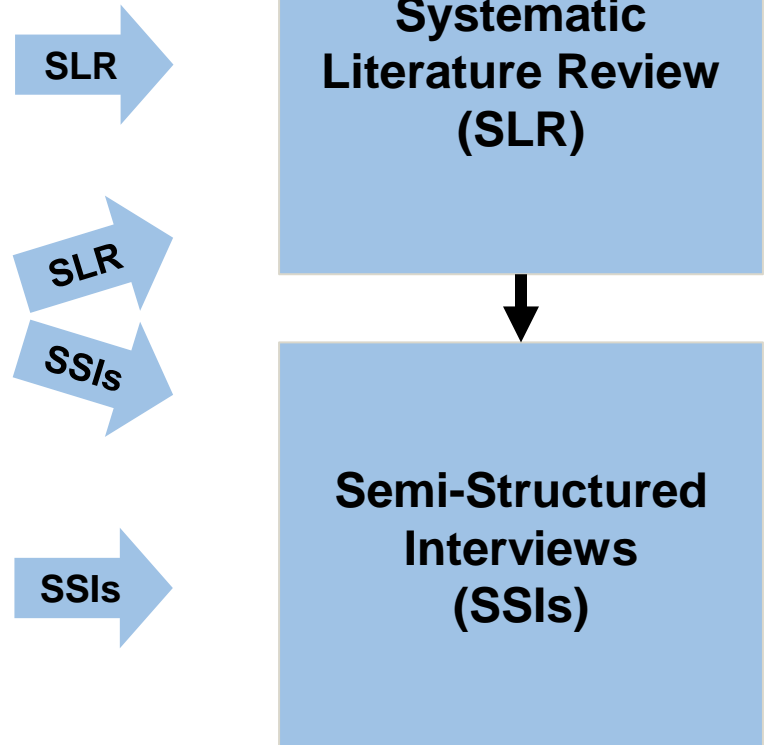
→ Mapping state of the art applications and their implications concerning responsible AI

[1]: Google Scholar: Browse for "NLP AND Legal OR NLP AND law" from 1993 until 2022

Title

Towards a Structured Overview of Use Cases of Natural Language Processing in the Legal Domain

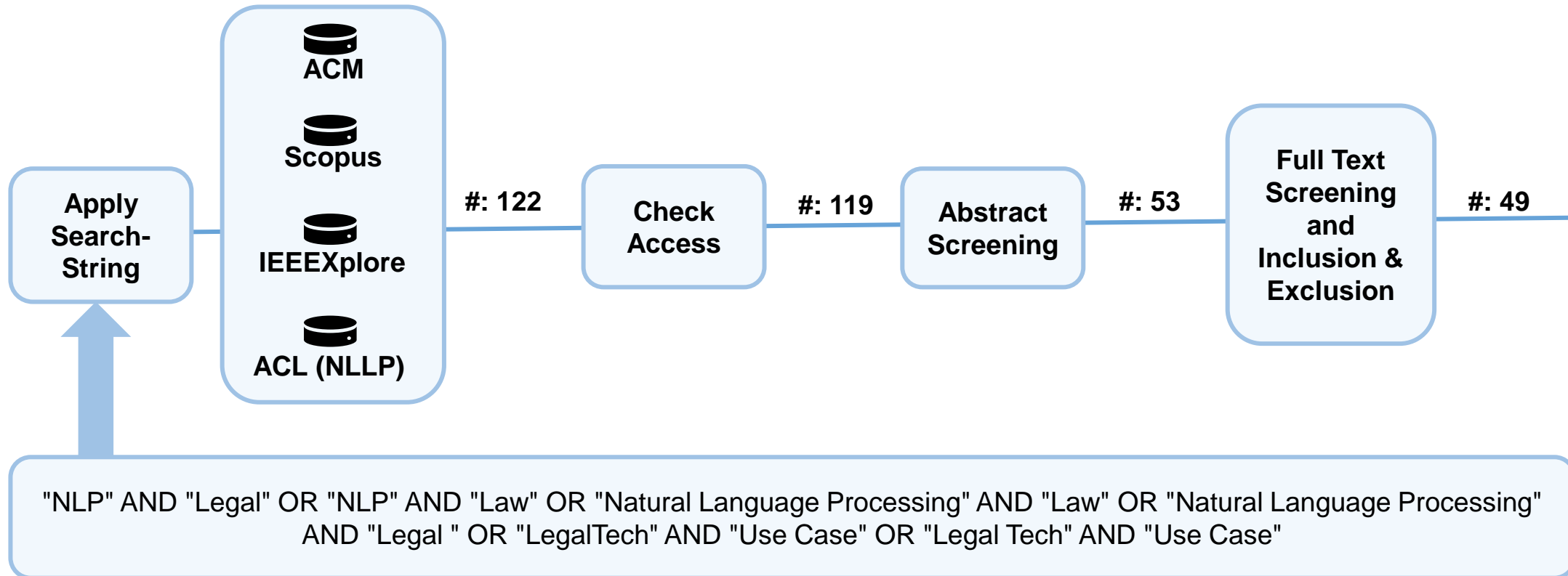
- 1 From a technical perspective, what are the predominant Natural Language Processing techniques being applied in the legal domain, and to what extent are ethical, legal, and social aspects covered?
- 2 What are the use cases in which the identified NLP techniques can be utilized?
- 3 Together with semi-structured interviews, how can the results of the systematic literature review be synthesized with legal expertise to form the basis of a joint knowledge base?



SLR: Inclusion and Exclusion Criteria

Criteria	Inclusion Criteria	Exclusion Criteria
NLP-Relation	Papers, that include NLP are included.	Publications, that do not include NLP are excluded
Legal-Relation	Publications, which include a legal-relation are included.	Publications, which do not include a legal-relation are excluded.
Legal Use Case	Publications, which include a legal use case are included.	Publications, which do not include a legal use case are excluded.
Language	Papers, written in English or in German are included.	Papers written in a language different than English or German are excluded.
Publication Type	Conference papers, journal publications and workshop proceedings are included.	Any other kind of publication, like books or presentations are excluded.
Publication Date	Papers, published between January 1980 and January 2023 are included.	Papers, that have been published before January 1980 or after January 2023 are excluded.
Access	Papers, that are accessible in full text with the rights granted by the Technical University of Munich are included	Papers, that are not accessible in full text with the rights granted by the Technical University of Munich are excluded.
Quality	Publications with correct grammar and vocabulary are included.	Publications, which lack grammar and vocabulary are excluded.
Duplicates	Publications, that are not yet part of the selection process are included.	Publications, which are already part of the selection process are excluded.

SLR: Process in Numbers



SSIs: Acquisition of Interviewees

Category	Contacted (#)	Accepted (#)	Accepted (%)
Personal Connections	5	4	80.00%
Personal Introduction	10	6	60.00%
Search Results	29	4	13.79%
Snowball	16	4	25.00%
Total	60	18	30.00%

- Snowballing is a great source of acquiring further interviewees
- Personalized first messages really make a difference

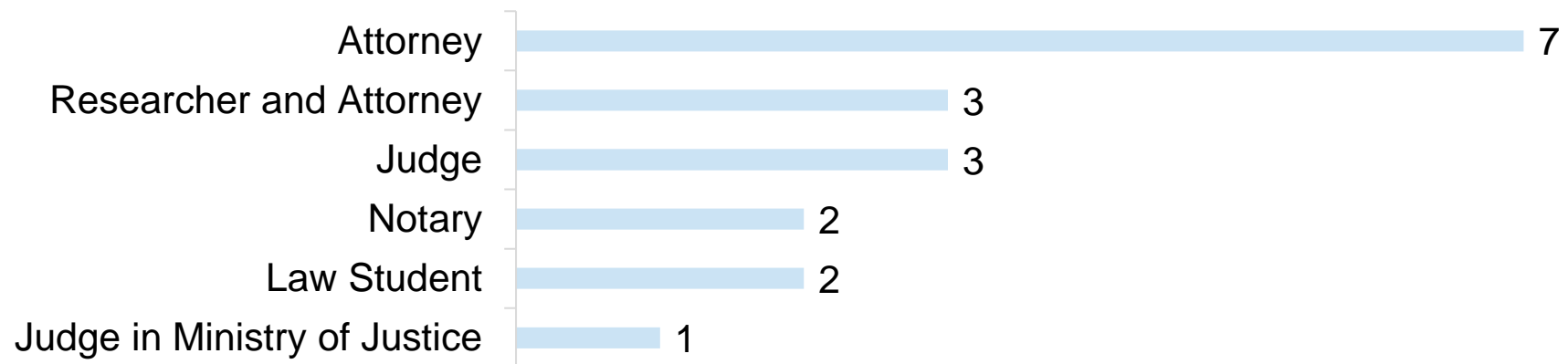
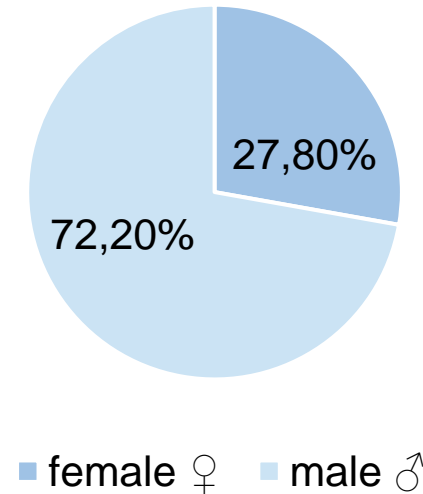
SSIs: Overview of Interviewees

Code	Position	Organization	Experience (years)	Duration (min)
I-1	Researcher and Attorney	Large-sized	9	31
I-2	Attorney	Medium-sized	11	58
I-3	Attorney	Large-sized	9	31
I-4	Law Student	Student	5	35
I-5	Attorney	Large-sized	29	19
I-6	Judge in Ministry of Justice	State Institution	16	61
I-7	Law Student	Student	6	38
I-8	Attorney	Micro-sized	14	55
I-9	Attorney	Medium-sized	27	55
I-10	Notary	Micro-sized	9	36
I-11	Notary	Small-sized	29	42
I-12	Attorney	Small-sized	12	30
I-13	Judge	State Institution	10	58
I-14	Researcher and Attorney	Large-sized	7	36
I-15	Researcher and Attorney	Large-sized	10	23
I-16	Attorney	Large-sized	9	36
I-17	Judge	State Institution	13	47
I-18	Judge	State Institution	18	49
Average			12.3	43.5

SSIs: Closer Look at Interviewees

Category	Organizations (#)	Organizations (%)
Micro-sized	2	11.11%
Small-sized	2	11.11%
Medium-sized	2	11.11%
Large-sized	6	33.33%
State Institution	4	22.22%
Student	2	11.11%

* Company Size according to European Union Recommendation 2003/361



ID	Category	#	%
NLP-Cat-1	Document Analysis and Processing	9	12.16%
NLP-Cat-2	Natural Language Understanding and Applications	9	12.16%
NLP-Cat-3	Text Extraction	23	31.08%
NLP-Cat-4	Text Generation	30	40.54%
NLP-Cat-5	Other	3	4.05%
Total		74	100%

ID	Category
NLP-Cat-1	Document Analysis and Processing
NLP-Cat-2	Natural Language Understanding and Applications
NLP-Cat-3	Text Extraction
NLP-Cat-4	Text Generation
NLP-Cat-5	Other

- Dependency Parsing
- Document Similarity Analysis

ID	Category
NLP-Cat-1	Document Analysis and Processing
NLP-Cat-2	Natural Language Understanding and Applications
NLP-Cat-3	Text Extraction
NLP-Cat-4	Text Generation
NLP-Cat-5	Other

- Chatbot Development
- Concept Models
- Part-of-Speech Tagging
- Question Answering
- Text Classification

ID	Category
NLP-Cat-1	Document Analysis and Processing
NLP-Cat-2	Natural Language Understanding and Applications
NLP-Cat-3	Text Extraction
NLP-Cat-4	Text Generation
NLP-Cat-5	Other

- Entity Linking
- Keyword Extraction
- Lexical Normalization
- Named Entity Recognition
- Tokenization

ID	Category
NLP-Cat-1	Document Analysis and Processing
NLP-Cat-2	Natural Language Understanding and Applications
NLP-Cat-3	Text Extraction
NLP-Cat-4	Text Generation
NLP-Cat-5	Other

- Language Modeling
- Machine Translation
- Text Summarization
- Topic Modeling
- Word Embedding

ID	Category
NLP-Cat-1	Document Analysis and Processing
NLP-Cat-2	Natural Language Understanding and Applications
NLP-Cat-3	Text Extraction
NLP-Cat-4	Text Generation
NLP-Cat-5	Other

- NLP Overview

- **SLR:** 51 Use Cases
- **SSIs:** 95 Use Cases

→ **31 disjoint use cases could be identified, grouped together in 8 use case categories**



ID	Category
UC-Cat-1	Compliance and Risk Management
UC-Cat-2	Document Analysis and Management
UC-Cat-3	Document Generation and Management
UC-Cat-4	Information Processing and Extraction
UC-Cat-5	Legal Decision Making and Dispute Resolution
UC-Cat-6	Legal Information Retrieval and Support
UC-Cat-7	Legal Research and Information Management
UC-Cat-8	Other

ID	Category
UC-Cat-1	Compliance and Risk Management
UC-Cat-2	Document Analysis and Management
UC-Cat-3	Document Generation and Management
UC-Cat-4	Information Processing and Extraction
UC-Cat-5	Legal Decision Making and Dispute Resolution
UC-Cat-6	Legal Information Retrieval and Support
UC-Cat-7	Legal Research and Information Management
UC-Cat-8	Other

Example: UC-3 Risk Assessment

Interviewee I-14 mentioned a tool:

- Currently being built
- Used in context of due diligence

Example: due diligence, where hundreds of rental contracts had to be manually searched through by legal practitioners in order to find potential issue with open-ended leases.

To overcome manual work: Legal Tech solution, which develops *“long-term risk profiles, which would then be assessed by a lawyer in the second step to determine how to factor them into the purchase price.”*

ID	Category
UC-Cat-1	Compliance and Risk Management
UC-Cat-2	Document Analysis and Management
UC-Cat-3	Document Generation and Management
UC-Cat-4	Information Processing and Extraction
UC-Cat-5	Legal Decision Making and Dispute Resolution
UC-Cat-6	Legal Information Retrieval and Support
UC-Cat-7	Legal Research and Information Management
UC-Cat-8	Other

Example: UC-4 Automatic File Difference Tracking

Interviewee I-8 mentioned a tool:

- Being already used by I-8 (attorney)
- Used in context of Transaction Management

Example:

Use of Legal Tech Software in order to track the change process during negotiations.

“These transaction documents are extensive documents, usually 50 to 60 pages long. And there are an insane number of attachments in there.” Manually keeping track of all these changes is a very time-consuming and laborious process. I-8 has now *“solved this using software, where an algorithm in the background automatically compares the documents for us.”*

ID	Category
UC-Cat-1	Compliance and Risk Management
UC-Cat-2	Document Analysis and Management
UC-Cat-3	Document Generation and Management
UC-Cat-4	Information Processing and Extraction
UC-Cat-5	Legal Decision Making and Dispute Resolution
UC-Cat-6	Legal Information Retrieval and Support
UC-Cat-7	Legal Research and Information Management
UC-Cat-8	Other

Example: UC-8 Automatic Contract Generation

Interviewee I-2 mentioned a tool:

- Which I-2 envisions for the future
- Used in context of Labor Law

Example:

Automatic Contract Generation could be use for the creation and issuing of employment contracts or for the preparation of termination documents.

Creating those kind of contracts are very repetitive tasks, as only nuances are changing. In this context, the wish of I-2 would be to have software, which would be able to *“extract relevant data from application documents or any other source, and generates my employment contract with just a click of a button.”*

ID	Category
UC-Cat-1	Compliance and Risk Management
UC-Cat-2	Document Analysis and Management
UC-Cat-3	Document Generation and Management
UC-Cat-4	Information Processing and Extraction
UC-Cat-5	Legal Decision Making and Dispute Resolution
UC-Cat-6	Legal Information Retrieval and Support
UC-Cat-7	Legal Research and Information Management
UC-Cat-8	Other

Example: UC-11 Anonymisation

Interviewee I-1 mentioned a tool:

- Which I-1 already used at previous occupation
- Used at a large law firm

Example:

Interviewee I-1 used to work in a law firm, where they had a Legal Tech tool, *“where you could blacken parts in documents, such as personal data and so on.”*

ID	Category
UC-Cat-1	Compliance and Risk Management
UC-Cat-2	Document Analysis and Management
UC-Cat-3	Document Generation and Management
UC-Cat-4	Information Processing and Extraction
UC-Cat-5	Legal Decision Making and Dispute Resolution
UC-Cat-6	Legal Information Retrieval and Support
UC-Cat-7	Legal Research and Information Management
UC-Cat-8	Other

Example: UC-15 Legal Reasoning

Interviewee I-9 mentioned a tool, which I-9 envisions for the future

Example:

I-9 envisions that the Legal Tech solution could assist in the effective articulation of an own opinion, or it could be used to verify the arguments of the opposing side and to check for inconsistencies.

Another application within this use case could be: *“As a lawyer, I may need to conceal my own weaknesses or those of my client, avoiding certain areas where we could be vulnerable.”* And the Legal Tech solution could assist in this.

ID	Category
UC-Cat-1	Compliance and Risk Management
UC-Cat-2	Document Analysis and Management
UC-Cat-3	Document Generation and Management
UC-Cat-4	Information Processing and Extraction
UC-Cat-5	Legal Decision Making and Dispute Resolution
UC-Cat-6	Legal Information Retrieval and Support
UC-Cat-7	Legal Research and Information Management
UC-Cat-8	Other

Example: UC-28 Credibility of Witnesses

Interviewee I-17 mentioned a tool:

- Which I-17 envisions for the future
- Used in the context of legal proceedings

Example:

It is a tool or system designed to assist legal professionals in making decisions regarding the trustworthiness of witnesses.

“There is no place where more lies are told than in court.” As a judge, one is always required to listen to everything, but whether one believes statements to be true is up to the judge. *“Sometimes you would be grateful if you had some kind of assistance, that you could ask whether the witness is lying or not.”*

ID	Category
UC-Cat-1	Compliance and Risk Management
UC-Cat-2	Document Analysis and Management
UC-Cat-3	Document Generation and Management
UC-Cat-4	Information Processing and Extraction
UC-Cat-5	Legal Decision Making and Dispute Resolution
UC-Cat-6	Legal Information Retrieval and Support
UC-Cat-7	Legal Research and Information Management
UC-Cat-8	Other

Example: UC-31 Research Tool

Interviewee I-2 mentioned a tool:

- Which I-2 envisions for the future
- Used in the context of legal research

Example:

Legal Tech solutions could be used to support legal practitioners with their research or by completely automating the legal research.

Interviewee I-2 shows the great range within the category of research tool. On the one hand, it could be something like Juris and Beck-Online, which is the querying of databases, or it could be something *“that takes over research work completely, where I don’t have to go and conduct the research myself.”*

ID	Category
UC-Cat-1	Compliance and Risk Management
UC-Cat-2	Document Analysis and Management
UC-Cat-3	Document Generation and Management
UC-Cat-4	Information Processing and Extraction
UC-Cat-5	Legal Decision Making and Dispute Resolution
UC-Cat-6	Legal Information Retrieval and Support
UC-Cat-7	Legal Research and Information Management
UC-Cat-8	Other

Example: UC-22 Law Firm Management-Software

Interviewee I-10 mentioned a tool:

- Which I-10 uses on a daily basis
- Used in the context of notary management-software

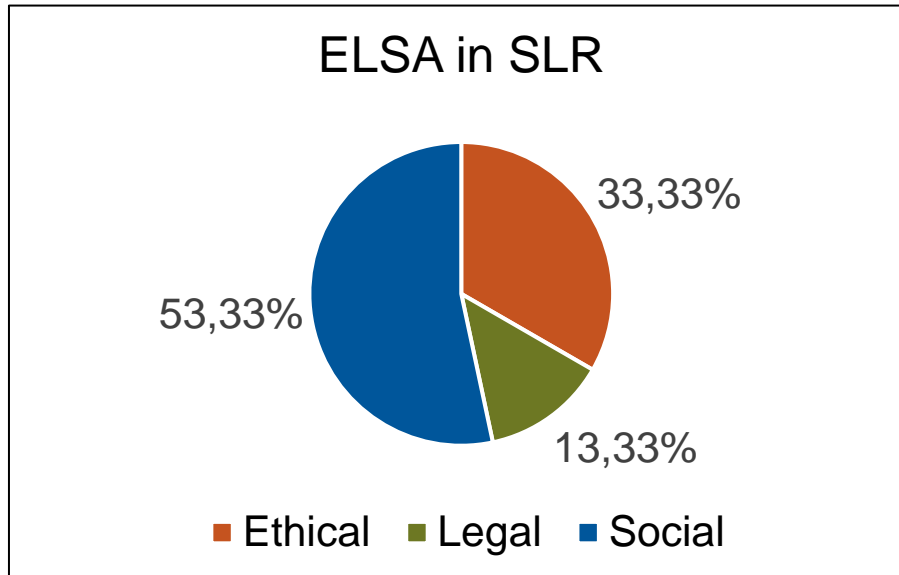
Example:

I-10 utilizes the law firm management-software for example to keep track of the customer database, to interface with official state institutions, to write contracts. For the latter, *“there are templates available for this purpose. In a regular home purchase, there are always the same building blocks, which are already included in the software.”*

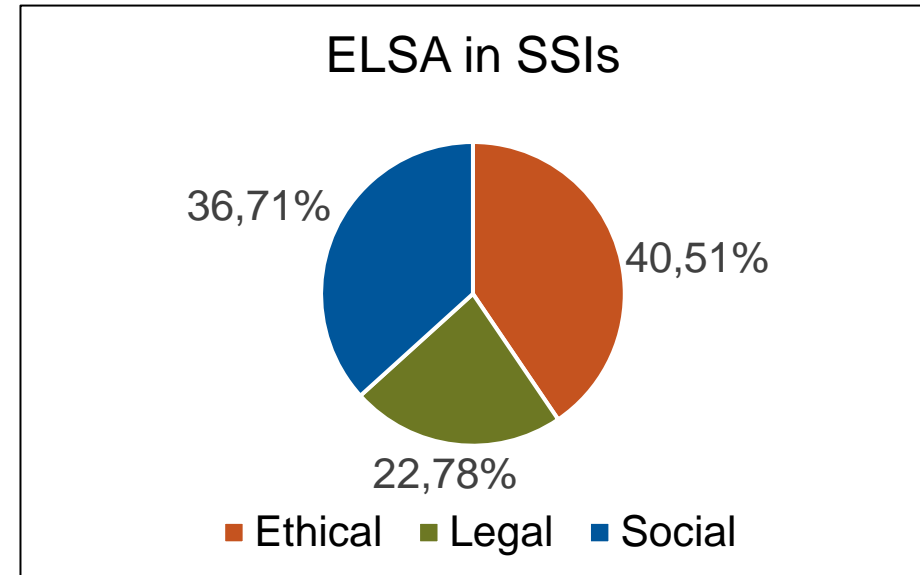
This is just one example, other interviewees named other highly specified software solutions, for example for attorneys or for judges.

Tangible Use Cases were mentioned most by the interviewees, like:

- UC-8 Automatic Contract Generation (mentioned 20 times)
- UC-12 Automatic Information Extraction and Insertion (mentioned 20 times)
- UC-22 Law Firm Management Software (mentioned 13 times)
- UC-31 Research Tool (mentioned 11 times)



- 18 of the 49 publications include ELSA (36.73%)
- 30 aspects were mentioned



- 79 aspects were mentioned in 18 interviews
- At least 10 of 18 ppl. have little ELSA thoughts (55.56%)

In the SSIs, stronger legal concerns: *“Liability errors do not occur so often, but when they do, they can also become existential.” (I-9)*

→ **Therefore, having some kind of certification for a Legal Tech solutions would increase the usage of Legal Tech software amongst practitioners.**

ID	Ethical Aspect
Eth-1	Black-Box Principle
Eth-2	Empowering Support, Not Final Verdict
Eth-3	Threat to Discrimination
Eth-4	Transparency
Eth-5	Artificially Created Opinion
Eth-6	Exploitable Use of Technology
Eth-7	Monopoly

ID	Ethical Aspect
Eth-1	Black-Box Principle
Eth-2	Empowering Support, Not Final Verdict
Eth-3	Threat to Discrimination
Eth-4	Transparency
Eth-5	Artificially Created Opinion
Eth-6	Exploitable Use of Technology
Eth-7	Monopoly

Eth-5 Artificially Created Opinion

This aspect includes the fear that artificially generated opinions are not representative for the values of a nation.

- I-4 wonders *“how prevailing opinions are formed, how disputes of opinion arise.”*
- I-9 points out to the difficulty of grasping the public opinion of a people. *“The same text of the law can be used for different consequences, that is the dangerous thing about law.”* A law can be interpreted in different directions, it is always up to the zeitgeist of the people, of the judges.
- Law is not like a binary system, which is either 1 or 0, but rather like quantum physics, that is how I-11 describes the influence of the people on legal decisions.

ID	Legal Aspect
Leg-1	Data Protection
Leg-2	Legal Compliance
Leg-3	Compliant Creation
Leg-4	Faulty Results

ID	Legal Aspect
Leg-1	Data Protection
Leg-2	Legal Compliance
Leg-3	Compliant Creation
Leg-4	Faulty Results

Leg-2 Legal Compliance

This aspect includes the compliance of a Legal Tech solution with the law.

- Interviewee I-7 poses the following questions to itself: *“is it even permitted to use it? Does it meet any legal requirements for its use to be allowed at all?”*
- Interviewee I-9: *“it is sufficient if it is legally correct, even if it is not completely secure.”* This means that if the legal situation states that something is within the legal frame, lawyers would rely on this.

ID	Social Aspect
Soc-1	Accessibility of Jurisprudence for Public
Soc-2	Transformation of Work
Soc-3	Dependability on Technology
Soc-4	Empathy
Soc-5	Peer Pressure
Soc-6	Readiness of Society

ID	Social Aspect
Soc-1	Accessibility of Jurisprudence for Public
Soc-2	Transformation of Work
Soc-3	Dependability on Technology
Soc-4	Empathy
Soc-5	Peer Pressure
Soc-6	Readiness of Society

Soc-2 Transformation of Work

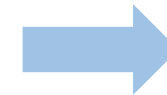
This aspects includes the impact that the rise of Legal Tech could have on the workforce.

- I-14 has *“of course a bit of concern about one’s own economic existence.”*
- I-3 sees that the role of the lawyer could change over time to the role of a reviewer.
- I-3 thinks that if the clients of a law firm would use Legal Tech software themselves and only hand the documents in to the law firm, then the main benefit of the law firm would be to *“take responsibility. And then it would practically become an insurance company rather than a law firm.”*
- I-13 thinks that *“machines will never replace humans. I don’t feel threatened in any way regarding my existence.”*

- **Legal Technology 1.0**
 - Technology supports human legal practitioner with current system
 - *Example:* Legal Databases like Juris or Beck-Online
- **Legal Technology 2.0**
 - Technology has a greater impact
 - Replacement of human law practitioners with current system
 - *Example:* Online compensation tools like RightNow
- **Legal Technology 3.0**
 - Radical redesign or replacement of the current system
 - Questioning of the human being as central figure
 - *Example:* Smart Contracts

Interesting Findings: Differentiation of Legal Tech according to Goodenough [2]

- **Legal Technology 1.0**
 - Technology supports human legal practitioner with current system
 - *Example:* Legal Databases like Juris or Beck-Online
- **Legal Technology 2.0**
 - Technology has a greater impact
 - Replacement of human law practitioners with current system
 - *Example:* Online compensation tools like RightNow
- **Legal Technology 3.0**
 - Radical redesign or replacement of the current system
 - Questioning of the human being as central figure
 - *Example:* Smart Contracts



Category	Amount
Legal Technology 1.0	12
Legal Technology 2.0	5
Legal Technology 3.0	5
Other	4

* 17 interviewees asked

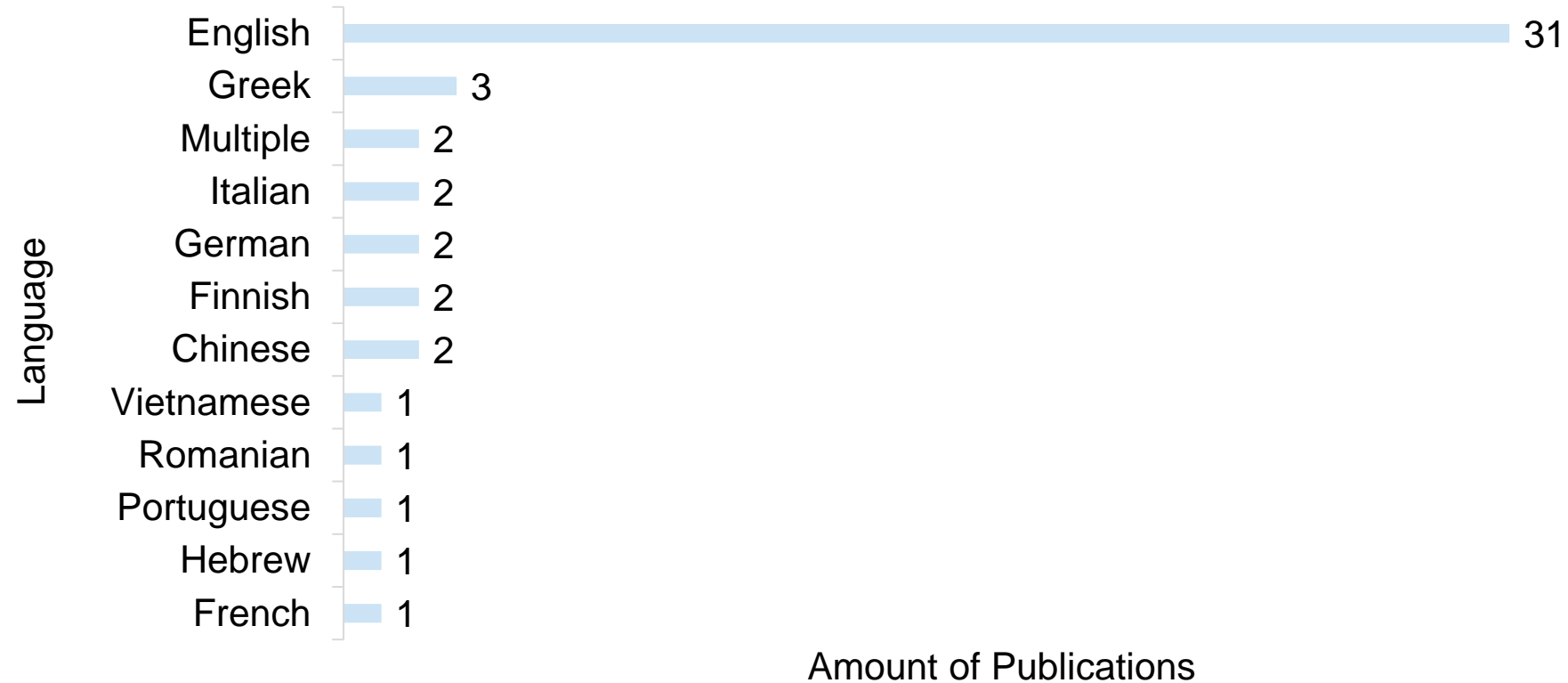
→ Only 29.41% include Legal Technology 3.0

→ **Information-gap regarding Legal Tech:**

“ I have to warn you in advance, Legal Tech doesn't play a big role in my studies.” (1-4).

[2] O. R. Goodenough. “Getting to Computational Jurisprudence 3.0”. In: The Challenge of Innovation in Law: The Impact of Technology and Science on Legal Studies and Practice (2015), pp. 3–17.

Interesting Findings: Language of Dataset



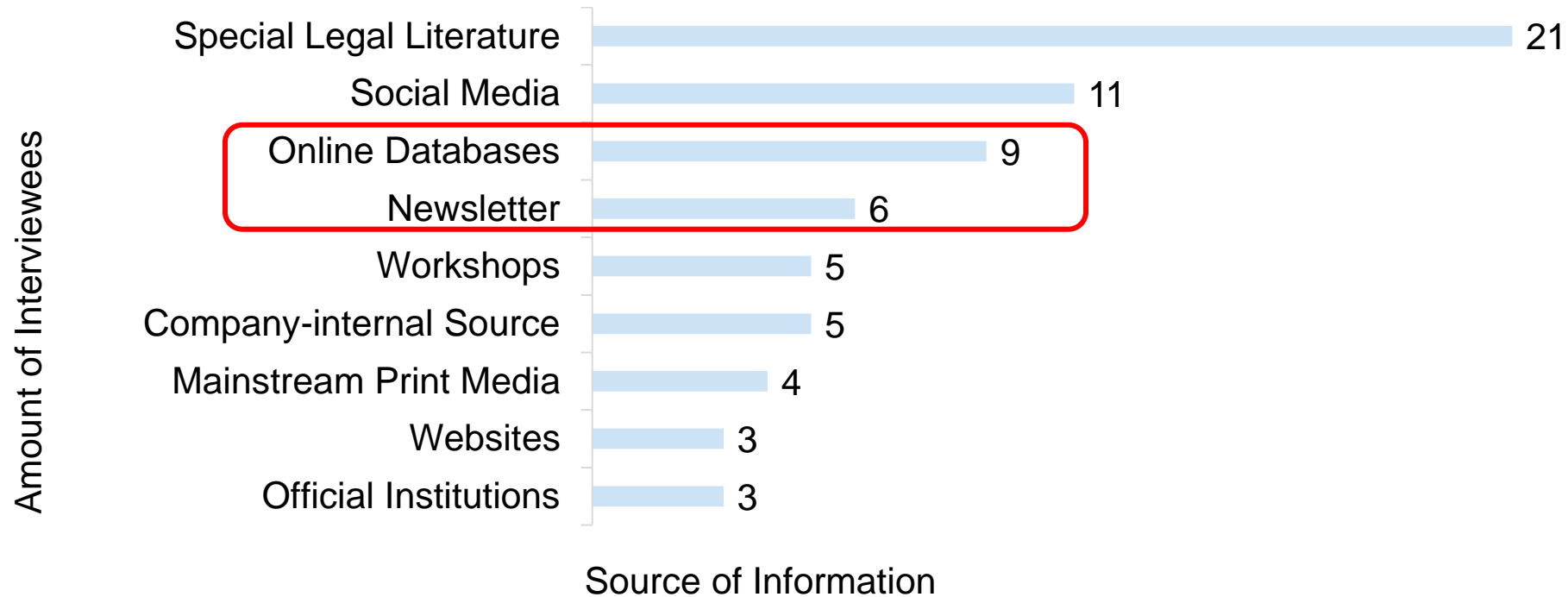
- The opinion created by a NLP algorithm should reflect the opinion of the people of a country.
- Each country should foster their own research in the field of NLP and invest in building up own datasets for their native language.

Interesting Findings: Requirement of Legal Tech Solution

Nr.	Category	#
1	Data Protection	12
2	Easy Usability	9
3	Added Value	6
4	Correct Output	6
5	Cost Sensitivity	5
6	Seamless Integration	5
7	Cross-Platform Usability	2
8	Flexibility	2
9	Reliability	2
10	Service for Emergency	2
11	Understandability of Technology	2
Total		53

* 17 interviewees asked

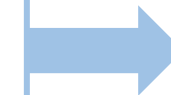
Interesting Findings: Source of Information



* 18 interviewees asked

- **Newsletter are common in the legal sphere – every third interviewee regularly reads newsletters**
- **Even though high dissatisfaction with legal databases, everybody uses them: “very powerful in terms of content, technically very weak and completely outdated.” (I-9).**

- **Interested**
 - Interested to know about the underlying technology in depth
- **Interested in Basics**
 - Interviewee does not want to go into technical depth
 - Interviewee wants to get overview of technology used
- **Not Interested**
 - Interviewees are not interested in underlying technology
 - *Requirement 1*: Solution generates correct results
 - *Requirement 2*: legally allowed to use solution



Category	Amount
Interested	5
Interested in Basics	1
Not Interested	9

* 15 interviewees asked

→ For 60%, it would be sufficient that the solution is creating correct results and that it is allowed to use:
“Ultimately, I would need a qualified or reliable (...) assurance that this program complies with the legal framework that I myself must adhere to.” (I-11).

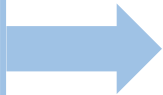
Research Questions

1

From a technical perspective, what are the predominant Natural Language Processing techniques being applied in the legal domain, and to what extent are ethical, legal, and social aspects covered?

2

What are the use cases in which the identified NLP techniques can be utilized?



Use Case Category	NLP Category
UC-Cat-1 Compliance and Risk Management	NLP-Cat-1 Document Analysis and Processing
UC-Cat-2 Document Analysis and Management	NLP-Cat-1 Document Analysis and Processing
UC-Cat-3 Document Generation and Management	NLP-Cat-4 Text Generation
UC-Cat-4 Information Processing and Extraction	NLP-Cat-3 Text Extraction
UC-Cat-5 Legal Decision Making and Dispute Resolution	NLP-Cat-2 Natural Language Understanding and Applications
UC-Cat-6 Legal Information Retrieval and Support	NLP-Cat-2 Natural Language Understanding and Applications
UC-Cat-7 Legal Research and Information Management	NLP-Cat-1 Document Analysis and Processing
UC-Cat-8 Other	NLP-Cat-5 Other

3

Together with semi-structured interviews, how can the results of the systematic literature review be synthesized with legal expertise to form the basis of a joint knowledge base?



When putting up a joint knowledge base between NLP-Academia and legal practitioners, the following points should be considered:

- Abstraction from technology
- Educational section for legal practitioners
- Newsletter
- Marketing of the joint knowledge base
- Intuitive design
- Added value
- Network
- Keeping information up-to-date

Limitations

Research Bias

Interviewee Pool

Generalizability

Future Work

Joint Knowledge Base

ELSA over Time

Certification



Martina Preis

martina.preis@tum.de

Technical University of Munich (TUM)
TUM School of CIT
Department of Computer Science (CS)
Chair of Software Engineering for Business
Information Systems (sebis)

Boltzmannstraße 3
85748 Garching bei München

+49.89.289.17132
matthes@in.tum.de
www.matthes.in.tum.de

