SCHOOL OF COMPUTATION, INFORMATION AND TECHNOLOGY - INFORMATICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Master's Thesis in Informatics

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

Martina Preis

SCHOOL OF COMPUTATION, INFORMATION AND TECHNOLOGY - INFORMATICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Master's Thesis in Informatics

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

Strukturierter Überblick der Anwendungsfälle von Natural Language Processing im juristischen Bereich

Author:	Martina Preis
Supervisor:	Prof. Dr. Florian Matthes
Advisor:	Juraj Vladika, M.Sc.
Submission Date:	15.06.2023

I confirm that this master's thesis in informatics is my own work and I have documented all sources and material used.

Munich, 15.06.2023 Martina Preis

Acknowledgments

First and foremost, I would like to express my gratitude to Prof. Dr. Florian Matthes for providing me with the opportunity to conduct my master's thesis at the chair for Software Engineering for Buisiness Information Systems (SEBIS). I would like to thank my advisor Juraj Vladika and the two other researchers at the SEBIS-chair Stephen Meisenbacher and Oliver Wardas for the weekly meetings, where they provided support, guidance and constructive feedback. Additionally, I am extremely grateful for the 18 interview participants, who supported my thesis. I am especially thankful for Tom Brägelmann and Dr. Benedikt Quarch, despite their very full schedules, to take the time and take part in the interviews. Last but not least, I want to thank my family and friends - how truly fortunate I am!

Abstract

Context: Natural Language Processing (NLP) has emerged as a powerful technology in the field of Legal Tech, with the potential to revolutionize and reshape the whole legal industry. In order to go a step further into this direction, many things have to be researched on. **Aim:** The aim of this thesis is to investigate the predominant NLP-Technologies, which are applied in the legal domain and to what extend Ethical, Legal and Social Aspects (ELSA) are covered in NLP-Academia and among legal practitioners. Furthermore, legal use cases, in which the identified NLP-Technologies are utilized should be researched on. As a last step, the base for a joint knowledge base between NLP-Researchers and legal practitioners is set. **Approach:** We conducted a Systematic Literature Review (SLR) in order to find out, which predominant NLP-Technologies are present in academia, which legal use cases are included in publications and whether Ethical, Legal and Social Aspects (ELSA) are considered in those publications. After applying our search term to four different databases, 122 results were extracted. After applying the inclusion and exclusion criteria, 49 publications could be identified for further analysis. As the SLR only includes the perspective of the NLP-Academia, Semi-Structured Interview (SSI)s were performed, where 18 legal practitioners contributed. Their professional background is very diverse, as some are working as judges, notaries, lawyers or researchers. They serve in the private sector or in the ministry of Justice and law students were also considered. With the SSIs the legal perspective could be included into the research. **Results and Conclusion:** Regarding the predominant NLP-Technologies applied in the legal domain, we identified in total 18 different NLP-Technologies, which were grouped together into five NLP-Categories. Furthermore, combining the results of the SLR and the SSIs, a total of 31 different legal use cases could be identified. Concerning the ELSA in the SLR, 36.73% of the paper include ELSA. During the SSIs, many *Ethical, Legal and Social Aspects* were mentioned, however 55.56% added, that they only have little concerns regarding ELSA. Lastly, a list of requirements for a *joint knowledge base* between NLP-Researchers and legal practitioners, based on findings from the SLR and mainly the SSIs are given.

Keywords: Natural Language Processing (NLP), Ethical, Legal and Social Aspects (ELSA), Legal Tech, Legal Technology, Legal Use Cases

Contents

1. Introduction

In this chapter, we first present the *Context and Motivation* for this thesis. In a second section the *Integration Into the Project NLawP* is emphasized and in a third section, the *Outline* of the whole thesis is given.

1.1. Context and Motivation

Artificial Intelligence (AI) is currently experiencing a significant rise. There is barely a realm of life or a sector of the work-world, that is not yet affected by it. Whether it is in the context of banks, which use AI systems to detect fraud [1], in the healthcare sector, which includes AI in the context of medical imaging [2] or when shopping online and items are recommended to the user [3]. AI has found a place in our society and is irreplaceable.
One important pillar in terms of AI is Natural Language Processing (NLP). While AI encompasses a broad range of technologies, NLP holds a particularly important place due to its ability to bridge the gap between humans and machines through the power of language. NLP focuses on the interaction between computers and human language. This enables machines to understand, interpret and generate human-like language. This technology has paved the way for numerous advancements and applications across various sectors. NLP-Research is a rapidly evolving field. Generative language models like *GPT-3*, which is a language model with 175 billion parameters, which is a tenfold increase compared to previous non-sparse language models [4], "promote a change of paradigm in NLP" [5]. One such sector, where NLP can play a crucial role and bring transformative changes is the legal domain. NLP-Technologies can aid legal practitioners for example in automating legal processes or it can enhance legal decision-making.
When discussing the application of NLP in the highly sensitive legal sector, it is crucial to consider its ethical implications. NLP-Technologies have the ability to have a very positive impact on legal practitioners and make their lives easier by for example automating repetitive tasks. However, NLP-Technologies also raise potential ethical concerns, which have to be addressed. One way of doing this is by investigating the Ethical, Legal and Social Aspects (ELSA) concerning NLP in the legal domain. As of now, the prerequisites for widespread adoption and the extensive ELSA accompanying NLP remain largely unexplored [6].

1.2. Integration Into the Project NLawP

This thesis is conducted in the scope of the research project *NLawP*, *Natural Language Processing and Legal Tech*[1], which is sponsored by the *Institute for Ethics in AI* of the Technical University of Munich. The *NLawP* project aims to assess the transformative potential of AI technologies in the legal sector. By evaluating the impact of these technologies, the project seeks to understand how AI can revolutionize and reshape the legal industry. The whole project is set up for a time-span of four years and several milestones are included. This thesis marks the kick-off of the *NLawP* project and one aim of this thesis is to contribute to *NLawP* by researching about which predominant NLP-Technologies the NLP-Research works on and to what extent ELSA play a role in the NLP-Academia. Another contribution, which this thesis makes to the *NLawP* project is the investigation in existing and future legal use cases, which utilize NLP and how legal practitioners value ELSA. In the scope of this thesis, a Systematic Literature Review (SLR) and Semi-Structured Interview (SSI)s are conducted to answer those questions.

1.3. Outline

This thesis is structured by the following chapters. First, we set the fundamentals, which are needed in our thesis in Chapter 2. Second, we give an overview of scientific work related to the topics of Natural Language Processing (NLP), Legal Tech and ELSA in Chapter 3. Next, we set the methodology, which will be used in this thesis in Chapter 4. In a next step, we list the results of this thesis in Chapter 5. Those results are then discussed in Chapter 6. Last, we conclude our thesis with a summary and an outlook for future research directions in Chapter 7.

[1] `https://wwwmatthes.in.tum.de/pages/ztm206o67g3q/`

2. Fundamentals

In this chapter the fundamentals, which are needed for this thesis are set. We start of with a section about *Natural Language Processing (NLP)*. The second section in this chapter addresses *Legal Tech*. Lastly, a section about *Ethical, Legal and Social Aspects (ELSA)* is given.

2.1. Natural Language Processing

In this section, a *Definition of Natural Language Processing (NLP)* is given, *Practical Examples* of NLP are stated and *Related Fields of NLP* are addressed.

2.1.1. Definition of NLP

Hapke, Lane, and Howard [7] define Natural Language Processing (NLP) as

> "an area of research in computer science and AI concerned with processing natural languages such as English or Mandarin. This processing generally involves translating natural language into data (numbers) that a computer can use to learn about the world. And this understanding of the world is sometimes used to generate natural language text that reflects that understanding."

NLP consists of a set of methods, which help to make human language comprehensible to computers [8].

2.1.2. Practical Examples

NLP is a very wide-spread field, which has many touchpoints to our everyday lives. Table 2.1 includes some examples, structured by the categories *Search*, *Editing*, *Dialog*, *Email* and *Creative Writing*. When conducting a *Search* for example, one can use NLP to efficiently look through dozens of files to search for specified data or when using search engines, they are more sophisticated nowadays, than just using string matching. NLP can also be encountered in form of dialogues: may it be a chatbot, who helps out in customer service, or a virtual assistant, who helps to plan meetings. The application of NLP is very broad, and Table 2.1 just gives some examples.

Search	Web	Documents	Autocomplete
Editing	Spelling	Grammar	Style
Dialog	Chatbot	Assistant	Scheduling
Email	Spam Filter	Classification	Prioritization
Creative Writing	Movie Scripts	Poetry	Song Lyrics

Table 2.1.: Examples of NLP applications according to categories, based on [7]

2.1.3. Related Fields of NLP

In this section, several areas are listed, which have touchpoints with NLP, namely Computational Linguistics, Machine Learning and Ethics. However, many more disciplines are related to NLP, such as *Speech Processing*, *Text Mining* or *Human-Computer Interaction*.

Computational Linguistics

According to Eisenstein [8], the terms NLP and *Computational Linguistics* could be thought of as being synonyms. Indeed, those two fields have an overlap, but they are far from being synonyms. In *Computational Linguistics*, the focus is on the human language, whereas computational methods only bear a supporting role [8]. In contrast to this, the focus of NLP is on computational algorithms and on the representation for processing human language.

Machine Learning

Numerous techniques in NLP depend extensively on Machine Learning. Machine Learning is a very powerful technology, which provides many general techniques. Eisenstein compares contemporary approaches of NLP to "applied machine learning" [8]. Eisenstein [8] states some fundamental differences between Machine Learning and NLP, for example:

- **Discrete Data**
  Text data is discrete at its core, therefore it is not possible to gradually approach a solution path.
- **Unseen Observations**
  Despite the discrete nature of word sets, new words are continually being created. As a result, the presence of observations that are not included in the training data necessitates the development of highly robust algorithms.

Ethics

As NLP-Technologies become more widespread, its impact on the lives of people all around the world intensifies. Therefore, ethical considerations have to be taken into account, when applying NLP. One question to be answered in this context is the questions, how ethical a project, technology, etc. is. This includes two steps: How ethical is the *process* and how ethical is the *outcome* [9]. Therefore, researchers, industry and the government have responsibility to critically assess the effects of their research [10]. They should take proactive steps to mitigate any potential adverse outcomes. One approach is to already include ethics into the curriculum of NLP lectures, and not only have it as an elective [10]. However, there are also other approaches, which educate researchers, to improve authoring and reviewing of NLP-related publications [11].

2.2. Legal Tech

In this section the focus is on Legal Tech. First, *Different Terminology*, which is used for Legal Tech is stated and in a next step a *Definition of Legal Tech* is given.

2.2.1. Different Terminology

The terms **Legal Tech**, **LegalTech** and **Legal Technology** all share the same meaning. They are used based on different geographic locations. We performed a search on Google Trends [12] to visualize the predominant search term by country, which could be *Legal Tech*, *LegalTech* or *Legal Technology* over the last 10 years. This is visualized in Figure 2.1. The predominantly used search-term decides on the colouring of the country. Figure 2.1 gives an indication, that there is no homogeneous term used worldwide. However, as in Germany the predominant term is *Legal Tech*, this thesis will stick to this terminology.

2.2.2. Definition of Legal Tech

A commonly used impact-based classification of Legal Tech comes from Goodenough [13]. He introduces three different Legal Tech stages, which differentiate to what extend Legal Tech applications support, or even replace legal practitioners.

Legal Technology 1.0

In the first stage, technology supports the human legal practitioner within the current system [13]. An example for this are legal online databases like Juris[1] and Beck-Online[2].

[1] `https://www.juris.de/`
[2] `https://beck-online.beck.de/`

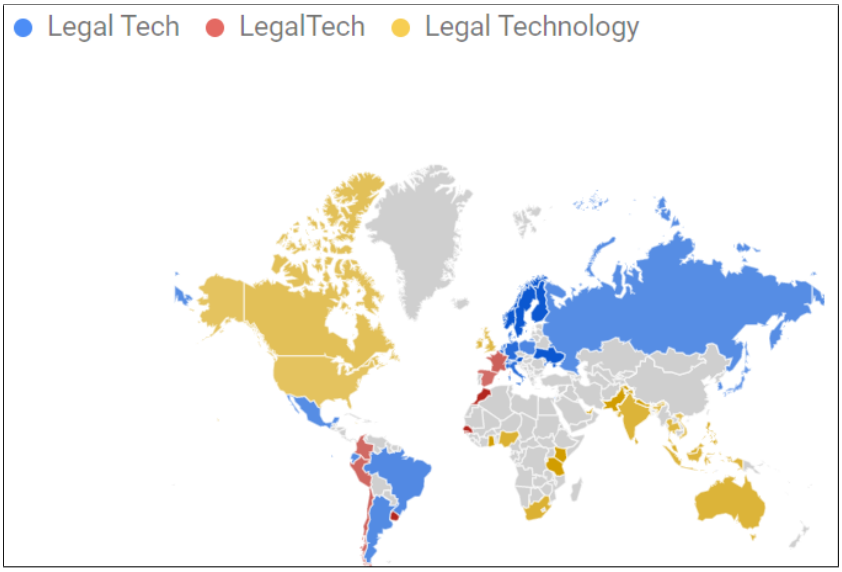

Figure 2.1.: Use of Legal Tech, LegalTech and Legal Technology

Legal Technology 2.0
In the second stage, technology has a greater impact and replaces an increasing number of human law practitioners within the current system [13]. Examples for this are automated contract generators or online compensation portals like RightNow[3].

Legal Technology 3.0
In the third stage, inventions in technology lead to a radical redesign or replacement of the current system itself [13]. This stage questions the human being as the central figure in the provisioning of legal services [14]. An example for this stage could be Smart Contracts.

2.3. Ethical, Legal and Social Aspects

Ethical, Legal and Social Aspects can also be abbreviated as ELSA [15][16]. According to Hullmann, ELSA offer significant insights to the interested public by aiding in the identification of expectations and concerns. Additionally, they hold great importance for policymakers as they help address these needs in terms of promoting good governance of research, which includes effective risk governance [16]. According to Gransche and Manzeschke, the role of ELSA and ELSI, which means Ethical, Legal and Social Implications in research developed over the years from formerly supportive research to integrated research [17].

[3] `https://www.rightnow.de/`

3. Related Work

In this chapter, we give an overview of scientific work related to our thesis' topic. The first section considers the *Boost of Legal Tech and NLP*, followed by a section about *NLP Technologies for Legal Tech*. The third section covers *Legal Use Cases* and lastly *Ethical, Legal and Social Aspects* are considered.

3.1. Boost of Legal Tech and NLP

The interest in the intersection between NLP and the legal domain is increasing. Figure 3.1 plots the amount of papers, which could be browsed for the search string *NLP AND Legal OR NLP AND law* from 1993 until 2022 in Google Scholar [18]. Searching for the year 2022, a total of 12500 results were shown, compared to the amount of search results for the year 2019, where 6810 results were the output. From 2019 on up until 2022, the amount of papers nearly doubled. This shows the increasing interest in the intersection of Legal Tech and NLP.

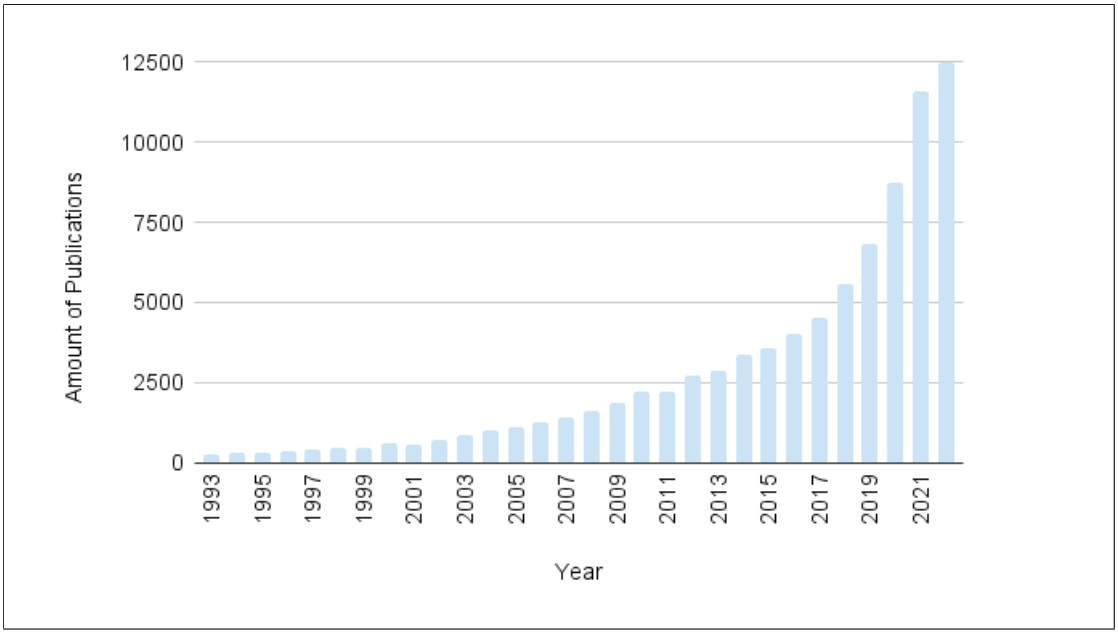

Figure 3.1.: Overview of NLP in Legal Tech over the years

3.2. NLP Technologies for Legal Tech

A variety of publications exist, which summarize NLP-Technologies, which are particularly interesting for the legal domain. Frankenreiter and Nyarko [19] for example provide a non-technical introduction into several NLP-Techniques, which are connected to the domain of Legal Tech. Additionally, their "promises and pitfalls" [19] are explained. The main emphasis of this publication is to deliver comprehensive explanations regarding NLP. The focus is not on exhaustiveness and only a limited level of technical depth is provided.

Another example of a publication, which is summarizing the legal field in regard to NLP, is the publication by Zhong, Xiao, Tu, et al. [20]. Here, the history, current state and future of research in the intersection of NLP and Legal Tech is stated. Also in this publication, the amount of mentioned NLP-Technologies is quite narrow and superficial.

In our research, we want to perform a SLR in the intersecting field between NLP and Legal Tech in order to derive the predominant NLP-Technologies applied in the legal domain.

3.3. Legal Use Cases

A wide range of different legal use cases, which utilize NLP can be found. As an example, the legal use case of *Legal Judgement Prediction* (for example in [21], [22], [23], [24], [25]) or the legal use case of *Legal Document Summarization* (for example addressed in [26], [27], [28]) are both widely discussed in academia. All of those publications include only one specific legal use case, which they elaborate on.

However, also publications exist, which include several legal use cases. Dale [29] for example classified the following use cases: *Legal Research*, *Electronic Discovery*, *Contract Review*, *Document Automation* and *Legal Advice*. Another example of a publication, which includes legal use cases in the publication by Zhong, Xiao, Tu, et al. [20]. Here, the following legal use cases are introduced: *Legal Judgment Prediction*, *Similar Case Matching* and *Legal Question Answering*.

Publications, which include several legal use cases exist, however the amount of legal use cases per paper is rather small. In our research, we want to derive legal use cases from publications in the intersecting field of Natural Language Processing (NLP) and Legal Tech and by conducting Semi-Structured Interview (SSI)s with legal professionals.

3.4. Ethical, Legal and Social Aspects

The concept of Ethical, Legal and Social Aspects (ELSA) is applicable in a variety of different research areas. The objective of the paper written by Berger, Gevers, Siep, and Weltring [30] is to enhance awareness regarding the ELSA associated with the implementation of *nanotechnology in brain implants.* The ELSA are applied to the following three phases: 1) short-term challenges of testing and clinical trials within the current regulatory frameworks, 2) short and medium-term concerns regarding the risks involved in device application and 3) long-term implications associated with enhancement issues.

In the publication by Ikkatai, Hartwig, Takanashi, and Yokoyama [31], the concerns of Ethical, Legal and Social Implications in regard to *AI are compared between different countries*, namely Japan, the United States and Germany. Differences between countries could be detected, especially in the field of *the use of AI for autonomous weapons*.

In the paper of Kapeller, Felzmann, Fosch-Villaronga, et al. [32], ELSI are considered in the context of *Wearable Robot Design*. But not only the implications are stated, but recommendations are given concerning the implementation of ELSI in Wearable Robot design, development and use [32].

In this thesis, the ELSA will be applied to Legal Tech use cases, which are built using NLP-Technologies.

4. Methodology

In this chapter the methodology of this thesis is discussed. In the first section, the *Research Questions* are introduced, followed by a section about *Methodology Design*. The last section in this chapter includes the *The Interview Makeup*.

4.1. Research Questions

This thesis is structured by three research questions, which are introduced in this section.

> **RQ1**: *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?*

This first research question consists of two parts. Firstly, we want to investigate, what the main NLP-Techniques are, which are used in the legal domain. The second aspect of RQ1 focuses on finding out, to what extend ELSA aspects are covered.

> **RQ2**: *What are the use cases in which the identified NLP techniques can be utilized?*

The second research question has its focus rather on use cases, which exist and which include NLP.

> **RQ3:** *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?*

The third and last research question already includes the techniques, which will be applied in this thesis, namely Systematic Literature Review and Semi-Structured Interviews. The goal of this question is to include the findings of the Systematic Literature Review and Semi-Structured Interviews into some sort of shared knowledge base.

4.2. Methodology Design

In order to answer the mentioned Research Questions, a Systematic Literature Review and Semi-Structured Interviews are conducted.

4.2.1. Systematic Literature Review

In this section, a Systematic Literature Review (SLR) according to Kitchenham, Budgen, and Brereton [33] and Wohlin, Runeson, Höst, et al. [34] is performed. The aim of the SLR is to "identify, analyse and interpret all available evidence related to a specific research question" [35].

Definition of the Search Strategy

According to B. A. Kitchenham [35] the most crucial aspect of conducting a systematic review is to specify the research questions that will guide the entire process. Based on those, the search strategy can be set up. We have already defined the research questions in Section 4.1.

Selection of Databases

An important step in conducting a SLR is the identification of relevant venues and databases to be searched. Table 4.1 includes the four databases, which were selected in the scope of this thesis. Those databases were chosen, as they are widely used in academic research. In terms of *ACL Anthology*, we specified further and focused on searching the Workshop NLLP[1], Natural Legal Language Processing. In a SLR the objective is to comprehensively identify and include as many relevant primary studies as possible, focusing on the research questions, while employing an impartial and rigorous search strategy [35]. Therefore, the use of several databases widens the amount of search results and leads to a greater amount of primary studies.

Database	Website
ACM Digital Library	`https://dl.acm.org/`
Scopus	`https://www.scopus.com/`
IEEEXplore	`https://ieeexplore.ieee.org/`
ACL Anthology	`https://aclanthology.org/`

Table 4.1.: Overview of the databases used in our SLR

Inclusion and Exclusion Criteria

The purpose of study selection criteria is to identify primary studies that directly address the research questions, aiming to minimize bias. It is important to establish these criteria already before the search process [35]. The Inclusion and Exclusion

[1] `https://nllpw.org/`

Criteria should ensure that the research is aligned to the research questions. Table 4.2 lists all the inclusion and exclusion criteria, which we defined for our SLR.

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.

Table 4.2.: Overview of inclusion and exclusion criteria

Definition of Search Terms

In order to search the databases stated in Table 4.1, the following search string is defined:
"NLP" AND "Legal" OR "NLP" AND "Law" OR "Natural Language Processing" AND "Law" OR "Natural Language Processing" AND "Legal " OR "LegalTech" AND "Use Case" OR "Legal Tech" AND "Use Case"
This search string includes the Search Operators *AND* and *OR*. On the one hand the search string includes the technical side (Terms: *Natural Language Processing* and its abbreviation *NLP*), but also the legal side (Terms: *Law*, *Legal*). In order to answer RQ2, which was defined in Section 4.1, also the terms *Use Case* and *Legal Tech*/*LegalTech* are added.
According to Zhang, Babar, and Tell [36] it is not only important where to search for publications, but also in which part of the article the search string should be applied. Depending on the databases, the areas *Article Title*, *Abstract* and *Keywords* were looked through.

Execution of the Search and Filtering

The execution of the process is displayed in Figure 4.1. As input to the four databases we take the search string, which we have defined in Section 4.2.1. This results in 122 findings. In a next step, the access to the publications is checked. After this phase, 119 publications are left. This leads to the next step, which is abstract screening. In this phase a total of 66 papers get excluded. For the rejection, we use the previously defined exclusion criteria stated in Table 4.2. The distribution of exclusion causes regarding those 66 excluded papers can be seen in Table 4.3. As the four databases, which we use (see Table 4.1) are common among Software Engineering, it is no surprise, that the main exclusion criteria in this context is the missing legal-relation. After abstract screening the remaining 53 paper are taken into full text screening. In this phase, four papers get rejected, as depicted in Table 4.4. In the case of the language exclusion, only the abstract was written in English, the rest of the paper was written in Turkish. This results in 49 papers, which qualify for the SLR.

Data Extraction

The data extraction is based on the 49 publications, which have passed abstract and full text screening, as seen in Figure 4.1. The data extraction process covers five specific criteria, which are described in Table 4.5. All of the 49 publications are covered and categorized according to the five categories *SLR-Cat-1 Year*, *SLR-Cat-2 NLP-Technology*, *SLR-Cat-3 Ethical, Legal and Social Aspects (ELSA)*, *SLR-Cat-4 Legal Use Cases* and *SLR-Cat-5 Language of Data*. After that, the results of the categorizations were checked again on another day to make sure all categories are correctly assigned.

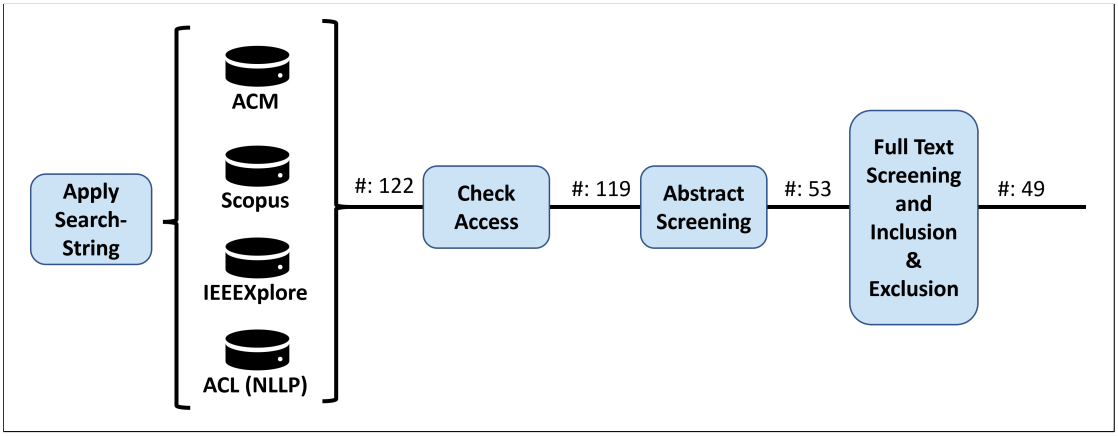

Figure 4.1.: Execution of the search including amount of search-outcomes

Exclusion Criteria	Rejected Papers [#]
No Legal-Relation	45
No Legal Use Case	13
Wrong Publication Type	3
Poor Quality	4
Duplicate	1
Total	**66**

Table 4.3.: Exclusion of publications in abstract screening

Exclusion Criteria	Rejected Papers [#]
Language	1
Publication Type	3
Total	**4**

Table 4.4.: Exclusion of publications in full text screening

ID	Category	Description
SLR-Cat-1	Year	The year, when the publication was published. This is extracted by the BibTeX taken from Google Scholar.
SLR-Cat-2	NLP-Technology	For each publication, the used NLP-related-Technology is extracted and classified.
SLR-Cat-3	Ethical, Legal and Social Aspects	If the publication mentions , those are extracted and classified.
SLR-Cat-4	Legal Use Cases	The legal use cases, which are mentioned in the publications are extracted and grouped together.
SLR-Cat-5	Language of Data	The underlying language, on which the developed NLP-Technologies are applied to.

Table 4.5.: Categories, which are used for data extraction

4.2.2. Semi-Structured Interviews

One core benefit of SSIs, compared to structured interviews, is flexibility. SSIs provides the opportunity to uncover information, that participants deem significant. This information might not have been initially considered as relevant by the research team [37].

Construction of Interview Guide

An interview guide was developed, according to Kallio, Pietilä, Johnson, and Kangasniemi [38], which includes five phases:

Phase 1: Identifying the prerequisites for using semi-structured interviews
The goal of the first phase is to check, whether semi-structured interviews are a suitable technique to collect data, which is needed to answer the defined research questions. Horton, Macve, and Struyven [39] recommend to use SSIs when the views of important protagonists want to be obtained, which is what we want to achieve in our research. They also recommend the use of SSIs, if it is not totally clear beforehand, which questions will be the most important ones to ask, due to the novelty of a topic. This also applies to the whole area of NLP in Legal Tech.

Phase 2: Retrieving and using previous knowledge
The objective of this phase is to gain sufficient comprehension of the subject matter. This entails critically evaluating existing knowledge and considering the potential necessity for additional empirical information to complement the existing understanding. We conducted an extensive SLR to gain a broad understanding of the topic.

Phase 3: Formulating the preliminary semi-structured interview guide
The objective of this phase is to create a preliminary semi-structured interview guide that would serve as an effective tool for collecting data during interviews. This interview guide helps the interviewee to stick to the scope of the interview. As stated by McIntosh and Morse [40], those questions should be open-ended. This prompts unstructured responses by the interviewees and encourages discussion.

Phase 4: Pilot testing the guide
The objective of this phase is to validate the extent and significance of the content included in the initial guide, ensuring its comprehensive coverage. Additionally, the aim of this phase is to identify any potential necessity to reformulate questions. We iterated several times on the interview guide. We discussed about the interview guidelines with our thesis supervisor and two other participants of the SEBIS-chair of the Technical University of Munich. This fits into the category of *internal testing*. We also conducted *field-testing* to finish off phase 4, where we tested our interview guide with a potential study participant.

Phase 5: Presenting the complete semi-structured interview guide
The fifth and last phase has the complete semi-structured interview guide as a result, as depicted in Figure 4.2. This interview guide is a translation of the original guideline, which is in German language and can be seen in Appendix A.1.

The Interview Analysis

Braun and Clarke [41] provide a step-by-step guide on how qualitative data can be analyzed. We follow those six steps of the Thematic Analysis Scheme:

1. **Familiarization with Data**
   In this initial phase, it is important to familiarize oneself with the data, in order to know the "depth and breadth of the content" [41]. All 18 interviewees were taking part in an online setup and the online sessions were recorded. We used the tool Whisper[2] for transcribing those recordings. Whisper is an automatic speech recognition system developed by OpenAI, which is able to transcribe text

[2] `https://openai.com/research/whisper`

Lehrstuhl für Software Engineering für betriebliche Informationssysteme (sebis)
TUM School of Computation, Information and Technology
Technische Universität München

Interview Guideline

Thank you for supporting my master thesis with this interview!

The following is a rough schedule for the interview. The questions listed are intended as a guide - you may not be asked all of them/other questions may be asked. You do not need to explicitly prepare for the interview.

Introduction	• **Formalities:** The interview will be recorded and subsequently transcribed. The transcription itself and any findings contained therein will be used for research purposes and for possible publication in a research paper and/or dissertation. All personally identifiable information will be anonymized. • Mutual brief introduction • Brief outline of the research context in the course of which the interview will take place.
Interview	• What is your understanding of legal tech? • Do you use digital aids to complete your daily tasks? Which ones? • Which repetitive tasks are part of your everyday work? • Where do you see future use cases for legal tech? • What media do you use to stay informed about news in the legal sector? • Where do you think legal tech will develop in the next few years? • Do you have ethical concerns about the use of technology in the legal sector? Which ones?
Outlook	• If we are one step further in our research - can my colleagues contact you again? • Do you know of others in the legal context that I should speak with? • Thank you very much!

Figure 4.2.: Interview guide translated to English

in German, English and many other languages. In a next step, we listened to the interview and refined the previously transcribed version. In yet another step, the transcriptions are read again and some ideas for coding, as well as some general notes, were taken.

2. **Generation of Initial Codes**
   Initial codes are generated by reading through all the transcripts and highlighting information, which might be interesting. For this, we used different colours, to highlight similarities among different transcriptions. We also created an excel sheet, which can be used as a legend to keep track of the different colours and interesting findings.

3. **Searching for Themes**
   Baseline for this phase is the excel-file, which contains all the identified codes across the different data set. In this step, those codes are further analyzed and grouped together in themes. We used excel to structurize this.

4. **Reviewing Themes**
   In this step, the previously defined themes are applied to the data set in order to check, if adjustments need to be taken.

5. **Defining and Naming Themes**
   In the fifth phase, it is verified, that the chosen themes are actually accurately depicting the transcript.

6. **Producing the Report**
   This last step involves the final analysis and write-up of the report. The goal of the write-up "is to tell the complicated story of your data in a way which convinces the reader of the merit and validity of your analysis" [41]. We do this in the scope of this thesis.

4.3. The Interview Makeup

This section consists of three main parts. Firstly, the interview participants are identified, secondly the demographics of the interviewees are displayed and thirdly, a summary of all the interview participants is given.

4.3.1. Identifying Participants

When conducting SSIs, a crucial part is to identify interview participants. Therefore, we are firstly defining the requirements of potential interviewees and secondly the channels, how potential interviewees can be reached.

Requirements of Interview Participants

In order to answer our research questions, we focus on targeting people from the legal sphere. Our goal is to cover a diverse range of legal professions in order to include broad parts of the legal sector.

Channels for Reaching Interview Participants

To reach out to possible interview participants, the following channels were used:

- **Personal Connections**
  This section covers people, which could be immediately connected.
- **Personal Introduction**
  This section covers people, where the first contact was established by a personal connection. We then reached out to them via LinkedIn[3] or E-Mail.
- **Search Results**
  In order to find people from the legal field, we used LinkedIn and entered search terms like *"Notary"* and we also used Google to find interviewees or their contact details. In a next step, we then reached out to the identified people via LinkedIn messages or via E-Mail.
- **Snowball** After the conduction of an interview, the interviewees are then asked to provide recommendations or referrals to other individuals who might also be knowledgeable on the subject. This process continues, with each new interviewee suggesting additional people to interview, creating a "snowball effect" as the number of interviews grows.

Table 4.6 gives an overview of the aforementioned categories, filled with the number of outreach. In total, 60 people were contacted, of whom 18 did conduct an interview. 30% of the contacted people accepted the interview request. The highest acceptance rate in a descending order is among the categories *Personal Connections* (80.00%), *Personal Introduction* (60.00%), *Snowball* (25.00%) and *Search Results* (13.70%). Two people initially scheduled an interview, but one person canceled beforehand and the other person did neither cancel, nor show up. Those two people are not included in this table. When reaching out to the participants via E-Mail or LinkedIn, we included a Link from Calendly[4], where the participants could immediately book a time-slot for the interview. Calendly really made the scheduling of interviews easier, as we stated time-slots, when we are available and participants can book a slot, which suits them. We then used Google Meets[5] to conduct the interviews. Once an interviewee books an appointment,

[3] `https://www.linkedin.com/`
[4] `https://calendly.com/`
[5] `https://meet.google.com/`

Calendly automatically sends out a calendar event via E-Mail, including the access-link for the video conferencing tool.

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%**

Table 4.6.: Channels for reaching interview participants in numbers

4.3.2. Demographics of Participants

Current Position
The purpose of Figure 4.3 is to provide comprehensive and detailed information regarding the current positions in which the interviewees are employed.

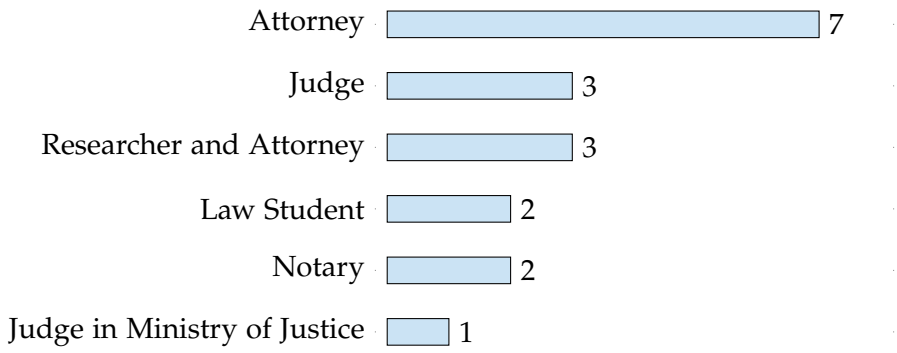

Figure 4.3.: Profession of interviewees

Size of Current Company
In this section the interviewees are categorized based on the company size, for which they are currently working. To cluster this, we include the European Union recommendation 2003/361 [42], which defines micro-sized, small-sized, medium-sized and large-sized companies based on factors like their number of employees and annual

turnover. We extend those categories by two other ones. We introduce the category *State Institutions*, which covers all participants, who are working for the government and the category *Student*, which include students, which are currently not externally employed. The full overview is given in Table 4.7.

Category	Organizations (#)	Organizations (%)	Interviewee ID
Micro-sized	2	11.11%	I-8, I-10
Small-sized	2	11.11%	I-11, I-12
Medium-sized	2	11.11%	I-2, I-9
Large-sized	6	33.33%	I-1, I-3, I-5, I-14, I-15, I-16
State Institutions	4	22.22%	I-6, I-13, I-17, I-18
Student	2	11.11%	I-4, I-7

Table 4.7.: Size of companies, where interviewees are currently employed

Gender
As visualized in Figure 4.4 from the 18 interviewees, 5 identified as female and 13 as male.

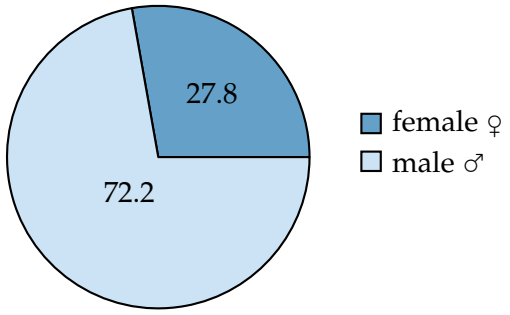

Figure 4.4.: Distribution of gender amongst interviewees

4.3.3. Summary of the Interviewees

Table 4.8 gives an overview of all the interview participants in an anonymized manner. The codes, which are given to each interviewee will be used to reference to the specific interviewees throughout this whole thesis. The *Experience* column of the table refers to

the years of experience among the interviewees, starting from the beginning of their studies until today, as also students are among the interview participants.

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**

Table 4.8.: Detailed overview of the interviewees

5. Results

In this chapter the results of this thesis are discussed. This chapter is structured by three sections. Firstly, the results of the *SLR* are stated, secondly, the results of the *SSIs* are presented. Lastly, the combined results of the *SLR and SSIs* are presented.

5.1. Systematic Literature Review

As defined in Table 4.5, the results of the SLR are structured among the five categories *SLR-Cat-1 Year*, *SLR-Cat-2 NLP-Technology*, *SLR-Cat-3 Ethical, Legal and Social Aspects*, *SLR-Cat-4 Legal Use Cases* and *SLR-Cat-5 Language of Data*.

5.1.1. SLR-Cat-1 Year

In this section, the previously identified 49 publications are structured by their publication year. This is displayed in Figure 5.1

5.1.2. SLR-Cat-2 NLP-Technology

In this section, the NLP-Technologies, which were addressed in the publications are categorized into five categories. Table 5.1 gives an overview of the five main NLP-Categories, which are coded as NLP-Cat-X. Also the number of occurrences of the corresponding categories is stated in this table. In the following, each of the categories will be explained in more detail. For each category, the associated NLP-Technologies (NLP-X) are stated. The category, which occurred most, with 40.5% is *NLP-Cat-4 Text Generation*, followed by *NLP-Cat-3 Text Extraction* with 31.1%. Next, the two categories *NLP-Cat-1 Document Analysis and Processing* and *NLP-Cat-2 Natural Language Understanding and Applications* are occurring equally often with 12.2% each. The least occurring category is *NLP-Cat-5 Other* with 4.0%.

It should be noted that some of the assigned NLP-Technologies could be assigned additionally to other categories (NLP-Cat-X), as the categories are not disjoint and given that NLP-Tasks are very interconnected. Additionally, as NLP is a technology that operates across various disciplines, the mentioned techniques are not exclusively limited to NLP but can also be derived from neighbouring fields.

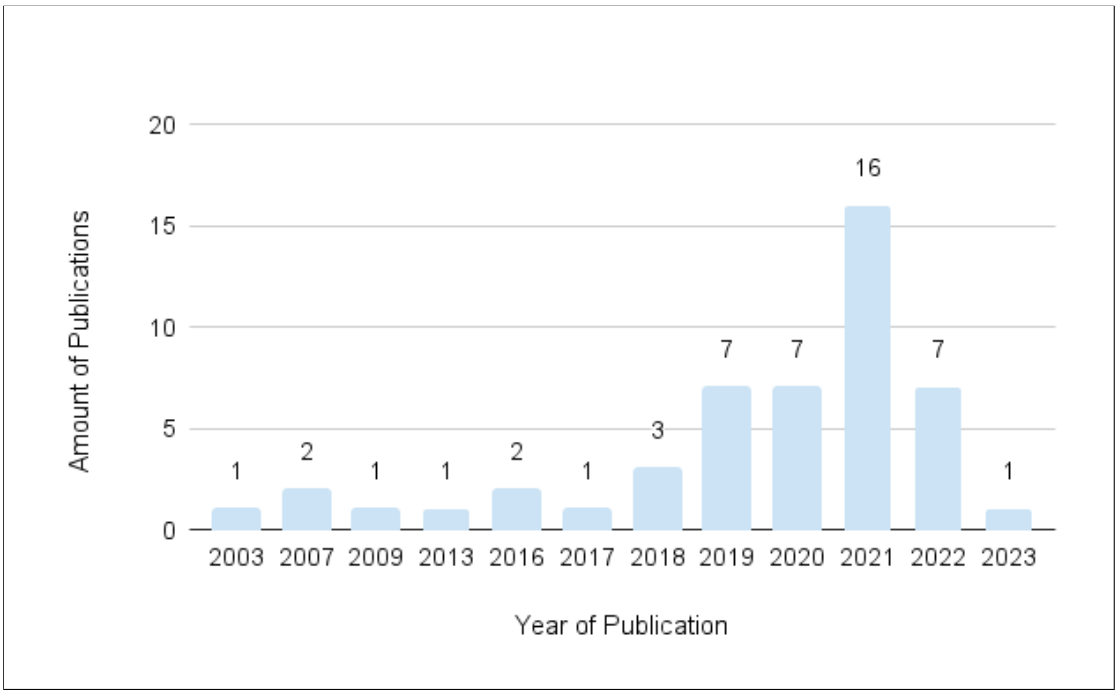

Figure 5.1.: Distribution of search results over the years

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%**

Table 5.1.: Overview of the NLP-Technologies addressed in the SLR

NLP-Cat-1 Document Analysis and Processing

The first of the five categories about NLP-Technologies is *Document Analysis and Processing*, which involves understanding and analyzing documents. Table 5.2 gives an overview of the two technologies, which form this category, namely *NLP-1 Dependency Parsing*, which occurred in 7 publications, and *NLP-2 Document Similarity Analysis*, which occurred twice. In the following, the technology used in each of those categories will be explained.

ID	Subcategory of NLP-Cat-1	Amount
NLP-1	Dependency Parsing	7
NLP-2	Document Similarity Analysis	2
Total		**9**

Table 5.2.: NLP-Technologies, which form NLP-Cat-1 Document Analysis and Processing

NLP-1 Dependency Parsing
Dependency Parsing is the process of analyzing the grammatical structure of a sentence by identifying dependency relations between words [43]. It can help boost the efficiency of a wide range of NLP-Applications such as information extraction or question answering [44].

NLP-2 Document Similarity Analysis
According to Gahman and Elangovan [45], Document Similarity Analysis is a task, where the similarity between two or more documents is analyzed, based on their content. It calculates a similarity score that quantifies the degree of resemblance or relatedness between documents. Frequently employed methods for this, such as cosine similarity or Euclidean distance are used to assess the similarity between documents based on their textual characteristics. These techniques assign a similarity score ranging from 0 to 1, where a score of 1 represents complete similarity and a score of 0 indicates no similarity.

NLP-Cat-2 Natural Language Understanding and Applications

In the second category **Natural Language Understanding and Applications**, the focus is on the comprehension and interpretations of human natural language by computers. This category is further divided into five technologies, as stated in Table 5.3.

NLP-3 Chatbot Development
Chatbots are "artificially intelligent creatures which can converse with humans"

ID	Subcategory of NLP-Cat-2	Amount
NLP-3	Chatbot Development	1
NLP-4	Concept Models	1
NLP-5	Part-of-Speech Tagging	1
NLP-6	Question Answering	2
NLP-7	Text Classification	4
Total		**9**

Table 5.3.: NLP-Technologies, which form NLP-Cat-2 Natural Language Understanding and Applications

[46]. This could be in a written form or in a spoken manner. Chatbot development involves creating conversational agents or virtual assistants that can interact with users through natural language [46]. Chatbots are designed to understand user queries, provide automated responses, and assist users with various tasks.

NLP-4 Concept Models
Concept Models are models that represent concepts or ideas and capture their relationships and hierarchies [47]. They are used to organize and categorize knowledge in a structured manner.

NLP-5 Part-of-Speech Tagging
Part-of-speech Tagging is the process of assigning grammatical tags or labels to words in a sentence, indicating their syntactic role, such as noun, verb, etc. [48]. This helps in analyzing the structure of a sentence and enables further processing tasks such as parsing.

NLP-6 Question Answering
Question Answering systems can understand questions posed in natural language and provide accurate answers based on available information like a database or a collection of natural language documents [49]. Question Answering involves analyzing the question, retrieving relevant information, and generating a concise and accurate response.

NLP-7 Text Classification
Text Classification is a task that assigns a given document to a set of pre-defined categories based on its content, topic and extracted features [50]. Text Classification has many applications, such as product review analysis or spam filtering [50].

NLP-Cat-3 Text Extraction

The third category is *Text Extraction*, which mainly focuses on extracting specific information from text data. Table 5.4 shows that this category can be further divided into five NLP-Technologies, to allow a more nuanced classification, namely *NLP-8 Entity Linking*, *NLP-9 Keyword Extraction*, *NLP-10 Lexical Normalization*, *NLP-11 Named Entity Recognition* and *NLP-12 Tokenization*.

ID	Subcategory of NLP-Cat-3	Amount
NLP-8	Entity Linking	1
NLP-9	Keyword Extraction	2
NLP-10	Lexical Normalization	4
NLP-11	Named Entity Recognition	10
NLP-12	Tokenization	6
Total		**23**

Table 5.4.: NLP-Technologies, which form NLP-Cat-3 Text Extraction

NLP-8 Entity Linking
Entity linking refers to the process of connecting a mentioned entity in text with its corresponding real-world entity in an existing knowledge base [51].

NLP-9 Keyword Extraction
Keyword Extraction involves automatically identifying and extracting important keywords or terms from a text document. Those keywords represent the main topics and concepts contained in the document [52].

NLP-10 Lexical Normalization
Lexical Normalization is the process of standardizing or normalizing text by converting different word forms or spellings to a canonical form, consistent with dictionaries [53]. Lexical Normalization helps to reduce variations and inconsistencies in textual data, and makes it easier to process and analyze in a next step.

NLP-11 Named Entity Recognition
According to J. Li, Sun, Han, and C. Li [54] Named Entity Recognition (NER) is the process of identifying and categorizing mentions of rigid designators within textual data. Rigid designators refer to specific entities such as persons, organizations or locations. The objective of NER is to accurately detect and

classify these mentions into predefined semantic types. NER is often the base for NLP-Applications.

NLP-12 Tokenization
Tokenization is the process of splitting text into smaller segments called tokens (mostly words) [7]. It is a fundamental step in NLP as it breaks down the text into smaller units for further analysis.

NLP-Cat-4 Text Generation

The fourth category is Text Generation, which focuses on generating human-like text. As stated in Table 5.5, this category can be further divided into five technologies, namely *NLP-13 Language Modeling*, *NLP-14 Machine Translation*, *NLP-15 Text Summarization*, *NLP-16 Topic Modeling* and *NLP-17 Word Embedding*.

ID	Subcategory of NLP-Cat-4	Amount
NLP-13	Language Modeling	15
NLP-14	Machine Translation	1
NLP-15	Text Summarization	4
NLP-16	Topic Modeling	1
NLP-17	Word Embedding	9
Total		**30**

Table 5.5.: NLP-Technologies, which form NLP-Cat-4 Text Generation

NLP-13 Language Modeling
In Language Models, the model is trained to predict the next word, based on preceding words [55]. This involves building statistical models that capture the probability of word sequences in a language.

NLP-14 Machine Translation
Machine Translation is the automatic translation of text or speech from a source language to a target language.

NLP-15 Text Summarization
Text Summarization involves generating a summary of a longer text while preserving its key information and main points. It aims to condense the content of the text, making it easier for users to bear. By reading a summary, the users can extract the essential information without reading the entire document.

NLP-16 Topic Modeling
Topic Modeling is a statistical technique used to analyze a collection of documents and uncover the latent topics or themes present within them [7]. It helps in identifying patterns and relationships between words to determine the main topics that emerge from the corpus.

NLP-17 Word Embedding
Word Embedding is a technique used to represent words or phrases as dense, low-dimensional vectors in a high-dimensional space. These vectors capture semantic relationships and similarities between words [56].

NLP-Cat-5 Other

Table 5.6 forms the last category named Other. This category covers *NLP-18 NLP Overview*. Three papers did not go into detail about one specific NLP-Technology, but rather gave a shallow overview of several NLP-Technologies.

ID	Subcategory of NLP-Cat-5	Amount
NLP-18	NLP Overview	3

Table 5.6.: NLP-Cat-5 Other

5.1.3. SLR-Cat-3 Ethical, Legal and Social Aspects

In this section, the 49 publications are examined for ELSA. Figure 5.2 visualizes that 18 publications, which are 36.73% of the publications, include ELSA. The remaining 31 publications, which are 63.27%, do not mention it. Those 18 papers, which mention ELSA, include in total 30 different ELS-Facets. Figure 5.3 shows the distribution among those 30 *Ethical*, *Legal* and *Social* aspects.

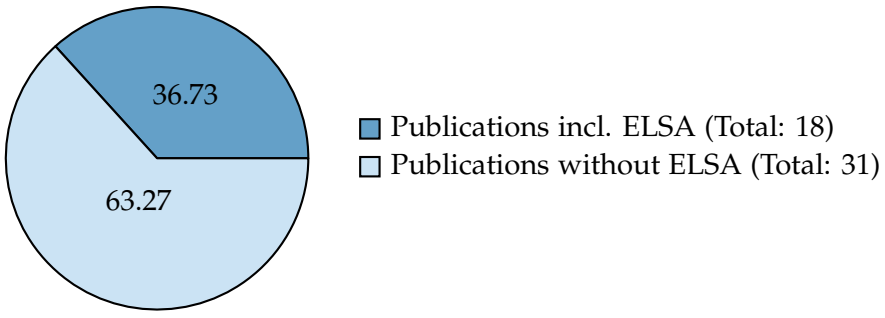

Figure 5.2.: Amount of papers, which include ELSA

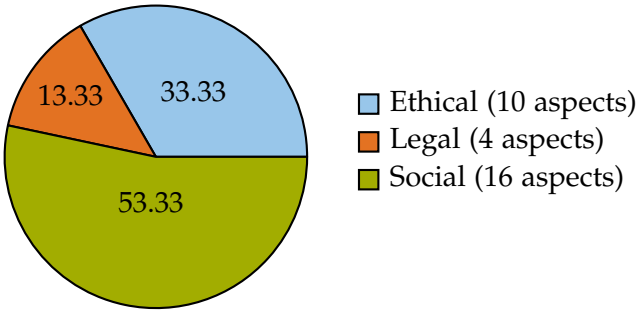

Figure 5.3.: Distribution of ELSA among SLR

Ethical

Table 5.7 lists the ten *ethical aspects*, which can be further divided into four categories, namely *Eth-1 Black-Box Principle*, *Eth-2 Empowering Support, Not Final Verdict*, *Eth-3 Threat to Discrimination* and *Eth-4 Transparency*. In the following section, a description to each of those categories is given.

ID	Ethical Aspect	Amount
Eth-1	Black-Box Principle	1
Eth-2	Empowering Support, Not Final Verdict	3
Eth-3	Threat to Discrimination	3
Eth-4	Transparency	3
Total		**10**

Table 5.7.: Overview of the ethical part of the ELSA in SLR

Eth-1 Black-Box Principle
The black-box principle, within the legal context, refers to the concept of treating certain aspects of a system or process as opaque or hidden. It suggests that legal professionals can make use of the outputs or results of a particular system or process without needing to understand the detailed internal workings or algorithms that generate those outputs. Questions covered in this category are: How much does a legal practitioner need to know about the depths of a Legal Tech solution to be able to properly use it, like which technology is used to construct it or which algorithms are included? To what extent is a Black-Box-Approach ethical?

Eth-2 Empowering Support, Not Final Verdict
In this category, the appropriate role of decision-making in the legal domain is addressed. Eth-2 emphasizes the idea that Legal Tech should be seen as a supportive tool rather than a definitive authority in legal matters. It suggests that while technology can assist legal professionals by providing data analysis, research support, or predictions, it should not replace the human judgment and decision-making that is inherent to the legal profession.

Eth-3 Threat to Discrimination
This third ethical category covers the discrimination potential of Legal Tech solutions. The worry is that if technical solutions are not properly designed, trained, or monitored, they may inadvertently perpetuate or amplify existing biases and discrimination that are prevalent in society or introduce new ones.

Eth-4 Transparency
The last category in the ethical domain covers the importance of openness and accessibility in technology used within legal systems, such as public source code or open data. By advocating for transparency, especially through open-source code, the ethical concern aims to ensure accountability and mitigate potential risks associated with hidden biases, errors, or unfairness within technological systems. It allows for independent scrutiny, expert review, and identification of any unintended consequences or shortcomings.

Legal

This section focuses on the legal aspects concerning ELSA. Table 5.8 lists the two sub-groups *Leg-1 Data Protection* and *Leg-2 Legal Compliance*, into which this category can be further divided.

ID	Legal Aspect	Amount
Leg-1	Data Protection	3
Leg-2	Legal Compliance	1
Total		**4**

Table 5.8.: Overview of the legal part of the ELSA in SLR

Leg-1 Data Protection
This legal concern emphasizes the importance of safeguarding the personal data of clients of legal practitioners. Legal professionals have a duty to adhere to data protection laws and regulations governing the collection, storage, and processing

of customer data. This includes mechanisms like implementing robust security measures to prevent data breaches or the transparent handling of customer information.

Leg-2 Legal Compliance
This second legal category includes the concern of legal professionals to be legally compliant. This revolves around their obligation to comply with the laws and regulations applicable to their practice.

Social

In the third pillar of the ELSA-Framework, the focus is on *social* aspects. Table 5.9 gives an overview of the two sub-categories, which form this section, namely *Soc-1 Accessibility of Jurisprudence for Public* and *Transformation of Work*.

ID	Social Aspect	Amount
Soc-1	Accessibility of Jurisprudence for Public	7
Soc-2	Transformation of Work	9
Total		**16**

Table 5.9.: Overview of the social part of the ELSA in SLR

Soc-1 Accessibility of Jurisprudence for Public
This consideration highlights the importance of making jurisprudence accessible to the public. Ensuring that legal information is easily available and understandable allows individuals to be aware of their rights, obligations, and the legal processes they may encounter. By promoting accessibility, society can foster the empowerment in legal fields, enabling people to make informed decisions and participate actively in the legal system.

Soc-2 Transformation of Work
This category refers to the changes in nature and structure of the employment for legal practitioners. With the rise of technology and automation, the tasks of legal professionals are evolving. This category includes both, the impact of those transformations on individuals, but also on the whole society.

5.1.4. SLR-Cat-4 Legal Use Cases

In this section, the 49 publications are studied for legal use cases, where NLP-Technologies can be used to help with legal tasks. Table 5.10 gives an overview of the 51 derived use cases, which are structured in eight different use case categories.

ID	Category	#	%
UC-Cat-1	Compliance and Risk Management	4	7.84%
UC-Cat-2	Document Analysis and Management	9	17.65 %
UC-Cat-3	Document Generation and Management	8	15.69 %
UC-Cat-4	Information Processing and Extraction	14	27.45 %
UC-Cat-5	Legal Decision Making and Dispute Resolution	8	15.69%
UC-Cat-6	Legal Information Retrieval and Support	5	9.80 %
UC-Cat-7	Legal Research and Information Management	2	3.92 %
UC-Cat-8	Other	1	1.96 %
Total		**51**	**100%**

Table 5.10.: Overview of the legal use cases addressed in the SLR

UC-Cat-1 Compliance and Risk Management

As illustrated in Table 5.11, this use case category includes three specific use cases, namely *UC-1 Automation of Auditing*, *UC-2 GDPR Compliance* and *UC-3 Risk Assessment*. In the following, a short description of each use case is given.

ID	Use Case of UC-Cat-1	Amount
UC-1	Automation of Auditing	1
UC-2	GDPR Compliance	2
UC-3	Risk Assessment	1
Total		**4**

Table 5.11.: Legal use cases from SLR, which form UC-Cat-1 Compliance and Risk Management

UC-1 Automation of Auditing
The first use case is about streamlining the auditing process. With the help of automation legal documents, contracts, or financial records can be efficiently reviewed for errors, discrepancies, or compliance issues. The whole auditing process could be automated.

UC-2 GDPR Compliance
The General Data Protection Regulation (GDPR) sets guidelines for the protection of personal data. Achieving GDPR compliance involves a variety of sub-tasks. Legal tech solutions can assist in this to make sure organizations adhere to GDPR requirements.

UC-3 Risk Assessment
This use case category involves the evaluation of potential risks associated with legal matters or business activities. Legal Tech software can aid in this process by analyzing historical data, identifying risk factors, and generate risk profiles.

UC-Cat-2 Document Analysis and Management

Table 5.12 shows that UC-Cat-2 can be further refined into *UC-4 Automatic File Difference Tracking*, *UC-5 Document Classification*, *UC-6 Document Management* and *UC-7 Error Detection*.

ID	Use Case of UC-Cat-2	Amount
UC-4	Automatic File Difference Tracking	1
UC-5	Document Classification	6
UC-6	Document Management	1
UC-7	Error Detection	1
Total		**9**

Table 5.12.: Legal use cases from SLR, which form UC-Cat-2 Document Analysis and Management

UC-4 Automatic File Difference Tracking
Automating the tracking of differences in files simplifies the process of comparing and identifying changes between different versions.

UC-5 Document Classification
This use case involves the automatic categorization of legal documents based on their content and characteristics. It includes efficiently managing and accessing documents.

UC-6 Document Management
This use case category includes the effective organization, storage, retrieval, and collaboration of legal documents throughout their life span.

UC-7 Error Detection
Error detection can be handled through Legal Tech. It involves utilizing algorithms and software tools to identify errors, or potential issues within documents, such as contracts.

UC-Cat-3 Document Generation and Management

The third use case category consists of *UC-8 Automatic Contract Generation*, *UC-9 Enrichment of Legal Documents* and *UC-10 Summarization*, as shown in Table 5.13

ID	Use Case of UC-Cat-3	Amount
UC-8	Automatic Contract Generation	1
UC-9	Enrichment of Legal Documents	2
UC-10	Summarization	5
Total		**8**

Table 5.13.: Legal use cases from SLR, which form UC-Cat-3 Document Generation and Management

UC-8 Automatic Contract Generation
This use case includes the automatic generation of various kinds of legal contracts.

UC-9 Enrichment of Legal Documents
This use case includes Legal Tech tools, which enrich legal documents by for example adding additional information, annotations, or references. This additionally added information can enhance the comprehensiveness of legal documents by for example adding additional context, cross-references, or relevant legal citations.

UC-10 Summarization
This use case category includes the summarization of lengthy legal texts, such as court rulings or contracts.

UC-Cat-4 Information Processing and Extraction

The fourth use case category can be further classified into three specific use cases, namely *Anonymisation*, *Automatic Information Extraction and Insertion* and *Patent Retrieval*. This is depicted in Table 5.14.

UC-11 Anonymisation
Anonymisation refers to the process of removing sensitive information in a document, or masking it, that the sensitive information is not identifiable anymore.

ID	Use Case of UC-Cat-4	Amount
UC-11	Anonymisation	1
UC-12	Automatic Information Extraction and Insertion	12
UC-13	Patent Retrieval	1
Total		**14**

Table 5.14.: Legal use cases from SLR, which form UC-Cat-4 Information Processing and Extraction

UC-12 Automatic Information Extraction and Insertion
This use case includes the extraction of specified information from documents and also the automatic fill-in of information.

UC-13 Patent Retrieval
This use case includes the search and retrieval of relevant patents.

UC-Cat-5 Legal Decision Making and Dispute Resolution

Use case category number five can be seen in Table 5.15. It is shaped by the three specific use cases *UC-14 Legal Decision Making*, *UC-15 Legal Reasoning* and *UC-16 Recommendations Based on Previous Court Rulings*.

ID	Use Case of UC-Cat-5	Amount
UC-14	Legal Decision Making	1
UC-15	Legal Reasoning	5
UC-16	Recommendations Based on Previous Court Rulings	2
Total		**8**

Table 5.15.: Legal use cases from SLR, which form UC-Cat-5 Legal Decision Making and Dispute Resolution

UC-14 Legal Decision Making
This use case includes the automatic decision-making process in the legal domain.

UC-15 Legal Reasoning
The use case *Legal Reasoning* covers the assistance of legal professionals by Legal

Tech tools in terms of legal reasoning and argumentation. This can be done for example by analyzing and organizing legal principles or other legal resources.

UC-16 Recommendations Based on Previous Court Rulings
Legal Tech can help legal professionals develop effective legal strategies and recommendations by analyzing and extracting insights from previous decisions and court rulings.

UC-Cat-6 Legal Information Retrieval and Support

The sixth use case category is shown in Table 5.16. It includes the three use cases *UC-17 Chatbot*, *UC-18 Question Answering* and *UC-19 Ranking of Lawyers*.

ID	Use Case of UC-Cat-6	Amount
UC-17	Chatbot	1
UC-18	Question Answering	3
UC-19	Ranking of Lawyers	1
Total		**5**

Table 5.16.: Legal use cases from SLR, which form UC-Cat-6 Legal Information Retrieval and Support

UC-17 Chatbot
Chatbots in the legal domain can occur in many ways. They can engage with the public interactively for example by answering basic legal questions. However, they can also occur and take over legal tasks like client-intake in a law firm.

UC-18 Question Answering
Question answering systems are designed to answer specific legal questions accurately. The answer is generated based on a give context or knowledge base.

UC-19 Ranking of Lawyers
In this category, Legal Tech tools are used to rank or evaluate lawyers based on experience, such as success rate, expertise, etc.

UC-Cat-7 Legal Research and Information Management

Table 5.17 shows use case category number seven. It includes the two specific use cases *UC-20 Changes in Law* and *UC-21 Database for Court Decisions*.

ID	Use Case of UC-Cat-7	Amount
UC-20	Changes in Law	1
UC-21	Database for Court Decisions	1
Total		**2**

Table 5.17.: Legal use cases from SLR, which form UC-Cat-7 Legal Research and Information Management

UC-20 Changes in Law
This use case includes the automated tracking and monitoring of changes in law, so that legal professionals are always up-to-date with the current jurisdiction.

UC-21 Database for Court Decisions
In this use case a comprehensive database exists, which includes many court decisions. This database can be used by legal professionals for searching, analyzing or referencing.

UC-Cat-8 Other

This section contains the legal use case of *UC-22 Law Firm Management-Software*, see Table 5.18. This includes the whole management software within a law-firm. This is a wholistic approach, which could include many of the afore mentioned use cases.

ID	Use Case of UC-Cat-8	Amount
UC-22	Law Firm Management-Software	**1**

Table 5.18.: Overview of UC-Cat-8 Other with insights from SLR

Law Firm Management-Software
This use case includes the main software within a law firm. This can include centralized platforms for case management, billing, time-keeping, collaborating-opportunities and much more.

5.1.5. SLR-Cat-5 Language of Data

All of the 49 publications are written in English language. Since NLP depends a lot on the language, which is processed, Figure 5.4 gives an overview of the languages lying the base in each of the 49 paper. The English language is predominant, as 31 papers developed new NLP-Technologies for the English language. Four of the publications

stated, that they used law texts, which have an origin in languages such as Dutch, Japanese or French and were translated into English before the application of NLP-Technologies. Those four publications are included in the *English* category.

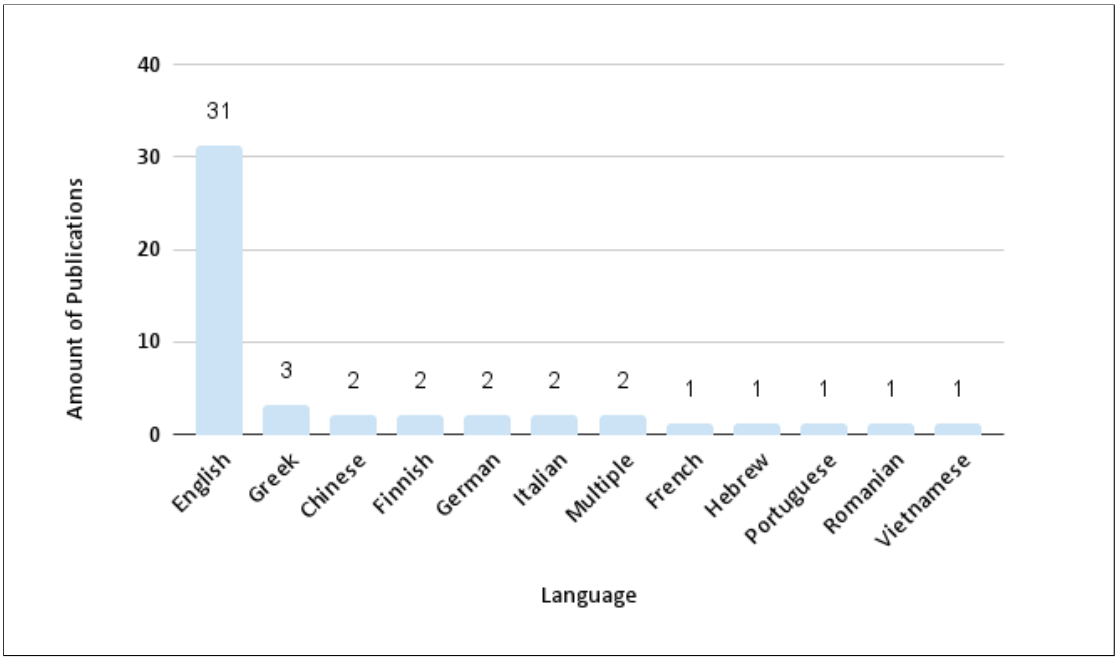

Figure 5.4.: Languages of the underlying dataset for each publication

5.2. Semi-Structured Interviews

In this section, all the results derived from the Semi-Structured Interviews are presented. They are structured by the following criteria *Understanding of Legal Tech*, *Ethical, Legal and Social Aspects*, *Legal Use Cases*, *Further Classification of Legal Use Cases*, *Requirements for Legal Tech Solutions*, *Source of Information* and *Interest in Underlying Technology*. As the interviews were all conducted in the German language, all direct quotes of interviewees in this thesis were translated into English by the author. The original German quotes are listed in the Appendix Section A.2.

5.2.1. Understanding of Legal Tech

One of the first question asked to 17 of the 18 participants was, what their understanding of Legal Tech is. This question is also included in the interview guide in Figure 4.2. As a next step, we used the definition of Legal Tech from Section 2.2, specifically the three

stages of Legal Tech defined by Goodenough [13]. Table 5.19 gives an overview of the mapping of the interviewees answers to the mentioned three criteria. Four participants stated a definition of Legal Tech, where the model of Goodenough was not applicable. Therefore, we introduced a fourth category named *Other*. For example I-6 put the focus in their answers on the differentiation between *E-Justice, E-Government, Legal Tech* and I-12 differentiated Legal Tech in terms of *B2B Legal Tech, B2C Legal Tech and State Legal Tech*. Two candidates I-5 and I-15 gave answers, which could not be applied. For example I-5 stated: "Actually, for me, there is no such thing as Legal Tech. There is only technology that works and technology that doesn't work." (I-5).

Category	Amount	Code
Legal Technology 1.0	12	I-1, I-2, I-3, I-4, I-7, I-8, I-10, I-11, I-13, I-16, I-17, I-18
Legal Technology 2.0	5	I-1, I-2, I-8, I-14, I-17
Legal Technology 3.0	5	I-2, I-3, I-13, I-14, I-16
Other	4	I-5, I-6, I-12, I-15

Table 5.19.: Classification of the Legal Tech understanding of interviewees

5.2.2. Ethical, Legal and Social Aspects

All 18 of the interviewees were asked, whether they have ethical concerns regarding the application of Legal Tech in the legal domain and what they are. Without further asking, ten people, 55.56% stated right away that they have no big ethical concerns. Whether the remaining 44.44% also have no big ethical concerns was not explicitly stated. This is illustrated in Figure Table 5.5. Figure 5.6 gives an overview of the distribution among Ethical, Legal and Social Aspects within the SSIs. The 18 interviewees stated in total 79 Ethical, Legal and Social Aspects.

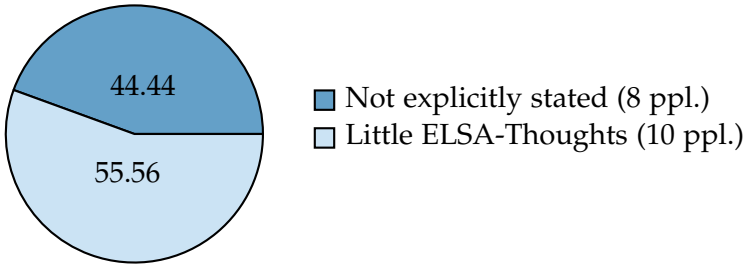

Figure 5.5.: Amount of interviewees, which include ELSA during the SSIs

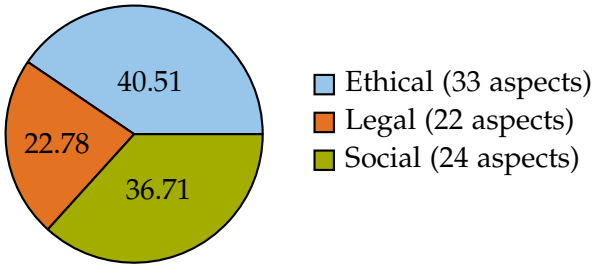

Figure 5.6.: Distribution of ELSA among SSIs

Ethical

Table 5.20 includes the ethical findings within the SSIs. Depicted in gray are the four already introduced categories from the SLR. Three new ethical categories could be discovered, namely *Eth-5 Artificially Created Opinion*, *Eth-6 Exploitable Use of Technology* and *Eth-7 Monopoly*. For the previously introduced categories in gray, only a quote from the SSIs will be given. For the new categories, additionally an explanation will be given.

ID	Ethical Aspect	Amount
Eth-1	Black-Box Principle	4
Eth-2	Empowering Support, Not Final Verdict	13
Eth-3	Threat to Discrimination	5
Eth-4	Transparency	3
Eth-5	Artificially Created Opinion	6
Eth-6	Exploitable Use of Technology	1
Eth-7	Monopoly	1
Total		**33**

Table 5.20.: Overview of the ethical part of the ELSA in the SSIs

Eth-1 Black-Box Principle
I-6, which is a judge working in the Ministry of Justice stated that "it is unethical to say that I provide a tool to judges that they do not understand and then they might have to use it in the end." (I-6). I-6 is responsible in the Ministry of Justice for suggesting technical solutions, which might go into a pilot-phase in specified courts and later on could be rolled-out in whole Bavaria. For the ministry, ethical

considerations play a very important role when selecting new software solutions. Another aspect that comes into play in the scope of *Eth-1 Black-Box Principle* is the impact, which understanding a software solution has on its usability. Interviewee I-7 stated that "I think one works better with it or perhaps (...) already a step further, one can only work effectively with the technology when one understands it at least fundamentally." (I-7). According to I-7, thinking of a Legal Tech program fully as a black-box would harm its usability.

Eth-2 Empowering Support, Not Final Verdict
For interviewee I-18, a judge, it is very important that the actual decision-making is done by a human being. Considering Legal Tech tools, which take decisions, "That would be terrible." I-18 states. I-18 goes even further "It would be bad if, after studying for years, a computational model suddenly comes along and says, "What you're doing here is wrong." It might even be true, but I still wouldn't want to hear it." (I-18). I-18 is not the only one who thinks like this. In total 13 of the 18 participants, which is 72.22% have this ethical concern.

Eth-3 Threat to Discrimination
In total 5 of the 18 participants, which is 27.78% of the participants have concerns regarding Legal Tech tools to be a threat to discrimination. One of them is I-14 for example, who states that "Bias does concern me, especially when it comes to integrating [Legal Tech solution] into judicial proceedings, where it could play a role in court decisions." (I-14). Interviewee I-14 is particularly afraid of using Legal Tech in the scope of courts, where it could have significant consequences.
Interviewee I-7 has concerns regarding the dataset, which is used for the AI Systems and that this could introduce bias to the algorithm. I-7 also states that "Currently, we lack the legal framework, which is necessary to effectively utilize data or create high-quality datasets." (I-7). Even though we do not have the needed datasets, our current legal framework is quite strict in terms of data protection. I-7 even goes one step further and elaborates that it would be negative to use datasets from other countries, as "the data used must represent our values and our population." (I-7). The integration of datasets from other countries could introduce some kind of bias.
A different view on this topic has interviewee I-5, who sees bias in the data as a "very serious challenge, but one that can be addressed." (I-5). I-5 is very optimistic concerning the topic of bias in data, as he points out that there exist AI techniques, which can be used to identify and mitigate this bias.

Eth-4 Transparency
The fourth ethical aspects, is the one about transparency. In total 3 of the 18 interviewees mentioned this concern, which is 16.67%.
Interviewee I-12 for example says that "(...) if we could manage to widely publish all court decisions, that would be great." (I-12). This plays into having

transparency for example concerning which techniques are used to derive this data. I-12 even goes one step further by saying about the publication of court decisions: "These are things that would bring the entire industry forward." (I-12). If a great variety of datasets from the own native language exist, Legal Tech solutions could make use of them and sharpen their results tremendously. I-12 also compares the transparency in the legal domain with open access in other domains, such as for example Computer Science and concludes that the legal system in Germany has to change at some point and make more things accessible.

Eth-5 Artificially Created Opinion
This fifth ethical aspect refers to AI systems, which generate opinions. The fear is that the generated opinions are not representative for the values of a nation. In total 6 of the 18 participants mentioned this concern, which is about 33.33%, every third person. This is an aspect, which also concerns interviewee I-4, who wonders about "how prevailing opinions are formed, how disputes of opinion arise." (I-4). Even without the use of Legal Tech solutions, how can the opinion of a people be captured? I-4 further arises concerns regarding once a Legal Tech solution would capture the opinion of a nation in a wrong way, how this could boost a minority-opinion to become the prevailing opinion. I-4 thinks that this would be possible, because in law everyone cites and references other cases. I-4 thinks that "that is where problems are most likely to arise". (I-4).
Interviewee I-9 points out to the difficulty of grasping the public opinion of a people. Especially in the field of law, I-9 states that "the same text of the law can be used for different consequences, that is the dangerous thing about law." (I-9). A law can be interpreted in different directions, it is always up to the zeitgeist of the people, of the judges. I-9 continues that "law is inherently unstable. This is also due to the fact that societies are unstable." (I-9). 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.

Eth-6 Exploitable Use of Technology
This category refers to the potential misuse or abuse of technology for unethical or malicious purposes. In total only one out of the 18 asked interviewees stated this concern, which is 5.56%.
In this context, interviewee I-9 mentioned an example, where a company introduced a bot some years ago, but had to shut it down again after two days, as it "just produced horrible stuff." (I-9). This is of course a problem, which arises from technology, as algorithms are trained with data from human beings. I-9 states that he is "not surprised by that. We humans are also cruel and horrible to some extent and if you learn from human texts or human nature, there would be a whole lot of horrible stuff." (I-9). Measurements have to be found to limit the exploitable use of technology.

Eth-7 Monopoly
This ethical aspect refers to the concentration of power and control over Legal Tech solutions in the hands of a single dominant entity or a small group of companies. It raises concerns about the potential negative effects of such monopolistic control on for example access to justice, and the overall fairness and diversity of the legal landscape. In total one of the 18 interviewees is concerned about this aspect, which marks 5.56%.
Interviewee I-2 states that as a lawyer it is important to always stay informed about current jurisprudence, in fact it is even legally required. Therefore, for lawyers it is very important to always stay up-to-date. I-2 was wondering how this can be done most easily and concluded that "if it is centralized, it is always the easiest. The problem with centralization is, of course, that if someone centralizes it, they gain a monopoly position." (I-2).

Legal

In this section, the legal part of the ELSA of the SSIs is presented. Table 5.21 gives an overview of the four legal aspects, that shape this category. Compared to the grayed rows in the table, which are categories, that were already introduced in the SLR, two new legal aspects are introduced, namely *Leg-3 Compliant Creation* and *Leg-4 Faulty Results*. In the following, quotes from the SSIs is given to each of the four categories.

ID	Legal Aspect	Amount
Leg-1	Data Protection	5
Leg-2	Legal Compliance	11
Leg-3	Compliant Creation	2
Leg-4	Faulty Results	4
Total		**22**

Table 5.21.: Overview of the legal part of the ELSA in the SSIs

Leg-1 Data Protection
The legal category of *Data Protection* was mentioned by 5 out of the 18 interviewees, which marks 27.78%.
Interviewee I-11, who is a notary states that he places great importance on software, which he used in his firm to not be cloud based, as this is "basically prohibited, because no data should leave the premises." (I-11). He even went one step further and brought up the example of demand-ordering for his copier,

where his copier would automatically report via the cloud to his supplier that the toner is empty, and I-11 would get a new toner automatically. But this is prohibited, as this would result in "data floating out of my office that I can't control and that I can't prevent." (I-11). I-11 would wish for a software design, where data only leaves the premises, if authorized beforehand.
Interviewee I-8 sets the data protection in the context of how data protection was handled in the past. I-8 explained that even in the past, he had uncertainties in the law firm, where he was working and which was located in a central and urban area. He stated that "we were not immune to surveillance through directional antennas. If someone wanted to eavesdrop on me, they could do so even back then." (I-8). He expressed that the means have changed, but "whether it is actually being done more frequently now, I don't know for sure." (I-8). This is an interesting perspective on data protection.

Leg-2 Legal Compliance
This legal aspect was mentioned by 11 of 18 interviewees, which accounts for 61.11% of the interviewees mentioned *Leg-2 Legal Compliance*. 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?" (I-7). This is an important question, which concerns many of the interviewees.
Interviewee I-9 stated about Legal Tech software that for lawyers "it is sufficient if it is legally correct, even if it is not completely secure." (I-9). This means that if the legal situation states that something is within the legal frame, lawyers would rely on this.

Leg-3 Compliant Creation
This legal aspect was mentioned by two out of the 18 interviewees, which marks 11.11%. This legal aspect focuses on the development of Legal Tech solutions in a way, that adheres to legal and regulatory requirements.
Interviewee I-9 expressed an example about a software company, which has acted unethically during the production of its AI algorithms, as employees were exploited. I-9 states that "I wish it wasn't like this. It is certainly a problem." (I-9).
Interviewee I-9 approached this topic also from another angle, namely the handling of copyright, when it comes to the training of AI algorithms. If algorithms train with copyright-related or creative content, "the rights of the authors are ignored." (I-9).

Leg-4 Faulty Results
The fourth legal aspect points out the potential of Legal Tech solutions to proceed inaccurate results. This can have significant consequences for for example legal practitioners or for the rights of individuals. This aspect was mentioned by 4 of the 18 participants, which accounts for 22.22%.
Interviewee I-9 points out that lawyers are very sensitive in terms of errors, as

they are personally liable for errors. In I-9's regard, a Legal Tech software would need to have "an accuracy of well over 98%. Otherwise, people would say it's not precise enough." (I-9). This means that if a legal practitioner has to invest a lot of personal effort to verify the output of a software and manually adjust it, they would not want the Legal Tech software.
Interviewee I-11 approaches this topic from a very interesting angle. When it comes to AI software, I-11 would like to have "some form of control or control-mechanism to assess how much the software is guessing." (I-11). This is in the context of hallucination of AI systems, where an AI systems confidently executes or answers a task, even if this is not justified by its training data [57]. Interviewee I-11 would wish for a program, which highlights or warns, once a certain probability threshold is reached.

Social

In this section, the social aspects of the SSIs are covered. Table 5.22 gives an overview of the six social aspects, which could be derived from the interviews. In gray are two categories, which have already been introduced in the scope of SLR. Four new social aspects were addressed by the interviewees, namely *Soc-3 Dependability on Technology*, *Soc-4 Empathy* , *Soc-5 Peer Pressure* and *Soc-6 Readiness of Society*.

ID	Social Aspect	Amount
Soc-1	Accessibility of Jurisprudence for Public	3
Soc-2	Transformation of Work	10
Soc-3	Dependability on Technology	1
Soc-4	Empathy	7
Soc-5	Peer Pressure	1
Soc-6	Readiness of Society	2
Total		**24**

Table 5.22.: Overview of the social part of the ELSA in the SSIs

Soc-1 Accessibility of Jurisprudence for Public
This first social section was mentioned by 3 of the 18 participants, which marks 16.67%. Interviewee I-13 sees great potential in Legal Tech in order to make law more accessible "especially for the normal population." (I-13). I-13 gets asked a lot by those around him about cases in rental agreement, claims from flights,

etc. I-13 thinks that for the average population, the legal system is incredibly opaque if they have no knowledge about it or don't even know where to begin looking. "How are they ever really supposed to get that?" (I-13). Therefore, Legal Tech software has great potential to make the law more accessible to people, for example by giving a rough estimate about a legal situation, which is more or less reliable.

Soc-2 Transformation of Work
The second social aspect was mentioned by 10 of the 18 participants, which marks a quote of 55.56%. Regarding this aspects, different opinions have been stated, for example interviewee I-14 has "of course a bit of concern about one's own economic existence." (I-14). This aspect is rooted in the fear that technology is quickly expanding and then eventually workflows could be largely carried out without human participation.
Interviewee I-3 has a slightly different opinion concerning this topic. I-3 sees that the role of the lawyer could change over time to the role of a reviewer. I-3 also raised thoughts not just about the role of a single lawyer, but about a whole law-firm. 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-3).
Interviewee I-5 is not concerned about technology replacing people or making them surplus. I-5 elaborates that "it will only replace those who resist the change, those that say that they want to work oldschool, traditional, without the use of state of the art technology solutions for all eternity." (I-5). In the eyes of I-5, there is only little concern regarding technology taking over. Similar views on this topic has interviewee I-13, who thinks that "machines will never replace humans. I don't feel threatened in any way regarding my existence." (I-13). To sum it up, *Soc-2 Transformation of Work* is a very broad field with many different opinions among the interviewees.

Soc-3 Dependability on Technology
This third social aspect was mentioned by one of the 18 interviewees, marking 5.56%. This aspect includes the reliability of legal practitioners on technology. This could lead to a loss of skills for legal practitioners, if they are used to performing their tasks with the help of machinery, which once again strengthens their *Dependability on Technology*. Interviewee I-1 stated regarding the application of Legal Tech software that "they need to function 24 hours a day, every day, so there should be a service available in case of emergencies." (I-1). In this context, the service, which would fix the software in case of emergency gives the legal practitioners security and increases the reliability of the software.

Soc-4 Empathy
The fourth social aspect was mentioned by 7 of the 18 interviewees, which accounts for 38.89%. As in legal matters humans are involved, emotions often play a role. Therefore, a challenge for a Legal Tech solution lies in incorporating technology in a way that enhances empathy, by providing tools that support human connection and emotional intelligence.
Interviewee I-11, who works as a notary hears many personal stories of its clients, for example when they create a last will. In the scope of this, I-11 gave an example, where a client has some hidden money on a bank account in Switzerland, and I-11 is bound to secrecy, once the client tells this to I-11. The question to be asked in this context is, whether the client would have the same trust to a machine and tell this story to a machine, too. I-11 is wondering: "Would you tell a machine that when you are married, that you still have a premarital child that the spouse doesn't know about?" (I-11). This is a very complex field, how the trust to a machine can be established, how a machine can be empathetic.

Soc-5 Peer Pressure
The fifth social category was mentioned by 1 of the 18 interviewees, which represents 5.56%. With the introduction of Legal Tech solutions, there may be a concern that legal professionals feel pressured to adopt certain technologies, even if they might not be aligned with their values or expertise. It is essential to foster an environment where legal professionals have the freedom to critically evaluate and choose technologies based on their merits and ethical considerations, and not because of peer pressure.
Interviewee I-9 compares this with the situation "when the pocket calculator came along, you could wish it away, but it's still there". (I-9). This is the same with Legal Tech solutions. If a competitors is using Legal Tech tools, they might be able to finish their processes faster and to a cheaper rate. Oneself can still be very good at the job, but if the competition is way cheaper, the own customers might switch to the competition.

Soc-6 Readiness of Society
This sixth and last social category was mentioned by 2 of the 18 participants, marking a coverage of 11.11%. The adoption and acceptance of Legal Tech solutions require a societal readiness of the people to adapt to technological changes. Interviewee I-2 for example sees great potential in the automation of law enforcement. However, I-2 is also a bit disillusioned, as I-2 states that "But as long as society doesn't want it, it won't happen." (I-2). When changes in technology are affecting the lives of the people, it is important to involve and bring along society.

5.2.3. Legal Use Cases

In this section, all use cases, which were mentioned by the 18 interviewees are categorized. Within the interviews, 95 use cases were identified. The previously defined eight categories of Table 5.10, which were used for the categorization of the use cases of the SLR, are now used to categorize the legal use cases from the SSIs. Table 5.23 give an overview of the concrete amounts of mentions and the percentages of each use case category. In the following, we will go through each of the eight categories and also introduce some new use cases.

ID	Category	#	%
UC-Cat-1	Compliance and Risk Management	1	1.05%
UC-Cat-2	Document Analysis and Management	6	6.32%
UC-Cat-3	Document Generation and Management	23	24.21%
UC-Cat-4	Information Processing and Extraction	16	16.84%
UC-Cat-5	Legal Decision Making and Dispute Resolution	14	14.74%
UC-Cat-6	Legal Information Retrieval and Support	9	9.47%
UC-Cat-7	Legal Research and Information Management	14	14.74%
UC-Cat-8	Other	12	12.63%
Total		**95**	**100%**

Table 5.23.: Overview of the legal use cases addressed in the SLR

UC-Cat-1 Compliance and Risk Management

The *Compliance and Risk Management* use case category encompasses various use cases, which aim to streamline and enhance processes related to regulatory compliance and risk assessment. Table 5.24 gives an overview of the three use cases, which shape this category. The use cases are marked in gray, as they were already introduced in the SLR. For each of the following categories, which were mentioned in the SSIs, a quote from the SSIs will be stated, in this category this is the use case *UC-3 Risk Assessment*.

UC-3 Risk Assessment
The use case of *Risk Assessment* was addressed once during the SSIs. Interviewee I-14 mentioned a tool, which is under construction at the law firm, where I-14 is working. When conducting a due diligence, many things have to be taken into consideration. I-14 mentioned an example, where hundreds of rental contracts

ID	Use Case of UC-Cat-1	Amount
UC-1	Automation of Auditing	0
UC-2	GDPR Compliance	0
UC-3	Risk Assessment	1
Total		**1**

Table 5.24.: Legal use cases from SSIs, which form UC-Cat-1 Compliance and Risk Management

had to be manually searched through by legal practitioners in order to find potential issue with open-ended leases. To overcome this manual work the idea is to have a 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." (I-14).

UC-Cat-2 Document Analysis and Management

The *Document Analysis and Management* use case category comprises a range of use cases that focus on enhancing the efficiency and effectiveness of document-related processing within the legal domain. Table 5.25 shows that this category is formed by four use cases, which have already been introduced within the SLR and are therefore marked in gray. The two use cases, which occurred in both, the SLR and in the SSIs are *UC-4 Automatic File Difference Tracking* and *UC-6 Document Management*. For those two categories, quotes from the SSIs will be presented in the following.

ID	Use Case of UC-Cat-2	Amount
UC-4	Automatic File Difference Tracking	2
UC-5	Document Classification	0
UC-6	Document Management	4
UC-7	Error Detection	0
Total		**6**

Table 5.25.: Legal use cases from SSIs, which form UC-Cat-2 Document Analysis and Management

UC-4 Automatic File Difference Tracking
The use case of *Automatic File Difference Tracking* was addressed twice during the SSIs. Interviewee I-18 has heard of this use case from his colleague, who works as a prosecutor. When performing corruption investigations, one has a lot of documents to be browsed through. In this scope, some kind of Legal Tech software is used from I-18's colleague in order to find information, which does not match or prices, which are improbable. "Then you realize: someone made this up." (I-18).
Interviewee I-8 personally uses a Legal Tech solution for *Automatic File Difference Tracking*. I-8 utilizes this software in the field of Transaction Management in order to track the change process during negotiations. In those kind of contexts, many stakeholders like investors, other lawyers, etc. are involved and it is hard to keep track of all the mails going back and forth, each mail including changes of contracts, additional clauses, etc. This is a lot of manual work, as "These transaction documents are extensive documents, usually 50 to 60 pages long. And there are an insane number of attachments in there." (I-8). 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." (I-8).

UC-6 Document Management
The use case of *Document Management* was mentioned four times during the SSIs. Interviewee I-2 is one of them and explained in detail, how at his workplace, they "have a file management system where documents are assigned to a client number and a specific task." (I-2). Additionally, each of the files can be assigned to different employees, can have a subject, can have a status, etc. This way, the documents are always organized, which reduces the susceptibility to errors.

UC-Cat-3 Document Generation and Management

The *Document Generation and Management* use case category encompasses several use cases, which focus on the generation, automation and optimization of document-related processes in the legal field. This section is formed by six use cases, as depicted in Table 5.26. Three of those use cases have been previously introduced during the SLR, therefore they are depicted gray in the table. Three new use cases have been introduced, namely *UC-23 Deadline Management*, *UC-24, E-Mail Communication* and *UC-25 Mass-Trials*. All of the six use cases have been mentioned by at least one interviewee. In the following, a quote from the SSIs will be given for all of the six categories and additionally for the three newly introduced use cases, an explanation about them will be given.

UC-8 Automatic Contract Generation
The use case category of *Automatic Contract Generation* was mentioned ten times

ID	Use Case of UC-Cat-3	Amount
UC-8	Automatic Contract Generation	10
UC-9	Enrichment of Legal Documents	1
UC-10	Summarization	3
UC-23	Deadline Management	4
UC-24	E-Mail Communication	1
UC-25	Mass-Trials	4
Total		**23**

Table 5.26.: Legal use cases from SSIs, which form UC-Cat-3 Document Generation and Management

during the SSIs. Interviewee I-2 could imagine the automation of contract generation very well for the context of a labor-law firm, which among other things creates and issues employment contracts or handles the terminations of employees. 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." (I-2).

UC-9 Enrichment of Legal Documents
The use case of *Enrichment of Legal Documents* was mentioned once during the SSIs. Interviewee I-9 knows of a Legal Tech software, which already exists in the US. It works as following: "you upload a brief and then it gets checked." (I-9). The brief gets checked, whether something else could be added as citation and whether the footnotes are accurate. According to I-9 this software also checks: "Does what I cite accurately reflect the content?" (I-9). After utilizing this Legal Tech solution, it is still recommended by I-9 to double-check the results.

UC-10 Summarization
The use case of *Summarization* was mentioned three times within the SSIs. Interviewee I-11 elaborated on summarization in the context of legal research. Whenever I-11 is conducting a search in one of the legal databases, this results in getting articles, court decisions, and so on in the search results. "And then I have to click through it again for about 20 minutes. It would be nicer if I could get it all presented as a summarized text right away." (I-11).

UC-23 Deadline Management
The use case of *Deadline Management* was mentioned in total four times during the SSIs. This use case refers to the efficient and effective management of legal deadlines and timelines associated with a variety of legal processes, such as probation. It includes tracking, monitoring and also the compliance with critical legal deadlines. Interviewee I-17, shared that if a suspended-sentence is given, the judge is responsible for probation-supervision, which ensures that individuals comply with their probation conditions. This causes a lot of work. I-17 thinks that this enforcement process could be streamlined, for example if a "computer system only notifies the judge when someone fails to comply with their probation conditions." (I-17).

UC-24 E-Mail Communication
The use case of *E-Mail Communication* was brought up once during the SSIs. It refers to the utilization of Legal Tech to facilitate efficient and effective E-Mail communication between legal professionals, clients, and other stakeholders within the legal ecosystem.
Interviewee I-3 stated that when communicating with the clients via E-Mail, it would be desirable to have technical support to prepare automated E-Mails, which are already well formulated. I-3 brought up the example, when I-3's client sends in documents, which were previously requested by the lawyer, it would be nice to have an easy solution, that the tool would politely respond to the client. I-3 further explains that "a significant amount of time is currently invested in crafting the perfect response for the client. If this process can be made easier or taken over by AI, even better." (I-3).

UC-25 Mass-Trials
The use case of *Mass-Trials* was mentioned in total four times during the SSIs. This use case includes mass-trials, where a great amount of similar cases are heard in court. The goal is to improve the efficiency and organization of those mass-trials by exploiting the similarity of the single cases.
Interviewee I-13 states that "Legal Tech is being used more and more in mass-proceedings." (I-13). I-13 further gave examples for this, namely the Diesel-Emissions-Scandal in the automotive industry, but also for example flight compensations, which are handled as mass-trials.

UC-Cat-4 Information Processing and Extraction

The *Information Processing and Extraction* use case category covers a wide range of use cases that revolve around the extraction, analysis and processing of information from legal documents and sources. As depicted in Table 5.27, this category consists of four use cases, of whom three were mentioned by interviewees. In the following, for the use

cases *Anonymisation* and *Automatic Information Extraction and Insertion*, which have been introduced already during the SLR, quotes from the interviewees will be mentioned. For the newly introduced *UC-26 Transcription*, a short explanation and a quote of an interviewee will be given.

ID	Use Case of UC-Cat-4	Amount
UC-11	Anonymisation	2
UC-12	Automatic Information Extraction and Insertion	8
UC-13	Patent Retrieval	0
UC-26	Transcription	6
Total		**16**

Table 5.27.: Legal use cases from SSIs, which form UC-Cat-4 Information Processing and Extraction

UC-11 Anonymisation
The use case of *Anonymisation* was mentioned twice during the SSIs.
Interviewee I-1 for example 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." (I-1).

UC-12 Automatic Information Extraction and Insertion
The use case of *Automatic Information Extraction and Insertion* was brought up eight times during the SSIs.
Interviewee I-18, who works as a judge now, brought up an example, where I-18 was working as a prosecutor in the traffic department, where the amount of standard-cases is very high. Back then, I-18 received about 200 new cases per month. Those cases were rather small, but still. I-18 resumes about this time: "If it weren't all standardized, you couldn't handle it." (I-18). I-18 describes the work in this context as following: "It's really like this: you spend 90% of your day entering data into text fields." (I-18). Of course, legal intuition is needed to determine, whether something is a serious case or not, but this repetitive work was not "particularly exciting in terms of content." (I-18).

UC-26 Transcription
The use case of *Transcription* was mentioned six times during the SSIs. Legal Proceedings require precise documentation at every stage. Therefore, transcription plays a crucial role in the legal domain. Transcription is the conversion of audio or

video recordings to written text. Interviewee I-17, who works as a judge, explains that they have two options regarding the topic of transcription. On the one hand, the judges could use a dictation machine and then a typist has to transcribe the spoken words manually. "Older colleagues use this very, very regularly", I-17 states. The second option is to use software solutions.

UC-Cat-5 Legal Decision Making and Dispute Resolution

The *Legal Decision Making and Dispute Resolution* use case category includes various use cases, which focus on improving the efficiency and effectiveness of decision-making processes and resolving legal disputes. This use case category is formed by four use cases, which were all mentioned within the SSIs. Table 5.28 gives an overview of those use cases. In gray are the use cases, which have been previously defined during the SLR. In the following, quotes, which have been derived from the SSIs will be given for those use cases, namely *UC-14 Legal Decision Making*, *UC-15 Legal Reasoning* and *UC-16 Recommendations Based on Previous Court Rulings*. The use case *UC-27 Dispute Resolution Mechanism* was identified during the SSIs and will be explained and provided with a quote of an interviewee in this section.

ID	Use Case of UC-Cat-5	Amount
UC-14	Legal Decision Making	7
UC-15	Legal Reasoning	2
UC-16	Recommendations Based on Previous Court Rulings	2
UC-27	Dispute Resolution Mechanism	3
Total		**14**

Table 5.28.: Legal use cases from SSIs, which form UC-Cat-5 Legal Decision Making and Dispute Resolution

UC-14 Legal Decision Making

The use case of *Legal Decision Making* was mentioned seven times during the SSIs. Interviewee I-2 stated regarding automation of legal enforcement that "this is an area that I find highly interesting." (I-2). In the example of I-2, an AI-based system would provide a draft decision, including the current situation and the legal decision of the underlying matter. This would be a cost-effective decision-making option. As a result, only complex cases would end up with the state judiciary in the form of judges.

UC-15 Legal Reasoning
The use case of *Legal Reasoning* was mentioned twice among the SSIs.
According to I-9, Legal Tech tools could be used in the future in the context of legal reasoning. 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." (I-9). And the Legal Tech solution could assist in this.

UC-16 Recommendations Based on Previous Court Rulings
The use case of *Recommendations Based on Previous Court Rulings* was mentioned twice during the SSIs.
Interviewee I-13 would support recommendations based on previous court rulings, but I-13 is concerned regarding the data-collection of court rulings in order to be able to come up with recommendations. Many hearings are public hearings, "but who sits in every public hearing? Nobody does that." (I-13). What also pays into the difficulty of the collection of data are data collection laws, as "not everyone is granted access to review the case files." (I-13).

UC-27 Dispute Resolution Mechanism
The use case of *Dispute Resolution Mechanism* was mentioned three times during the SSIs. It refers to the process of resolving a conflict or dispute between parties in an automated manner. This provides an alternative to court proceedings, as it gives the parties the opportunity to resolve their disagreement without litigation. Interviewee I-6 brought up the example about dispute resolution mechanisms from big platforms like PayPal. I-6 is fine with this procedure, as I-6 states "if there is an electronic proposal and the parties like it and accept it, all fair enough, but not for the state judiciary system." (I-6). I-6 agrees with dispute resolution mechanisms in the private sector, but would not agree to them as a public solution.

UC-Cat-6 Legal Information Retrieval and Support

The *Legal Information Retrieval and Support* use case category encompasses a range of use cases that focus on providing efficient access to legal information and supporting legal professionals in their decision-making process. This use case category consist of five use categories, as depicted in Table 5.29. Four of those categories were mentioned during the SSIs and in the following, quotes of the interviewees will be given for those use cases, namely *UC-17 Chatbot*, *UC-18 Question Answering*, *UC-28 Credibility of Witnesses* and *UC-29 Translation*. The latter two are introduced in this section and therefore, additionally an explanation will be given.

ID	Use Case of UC-Cat-6	Amount
UC-17	Chatbot	5
UC-18	Question Answering	1
UC-19	Ranking of Lawyers	0
UC-28	Credibility of Witnesses	1
UC-29	Translation	2
Total		**9**

Table 5.29.: Legal use cases from SSIs, which form UC-Cat-6 Legal Information Retrieval and Support

UC-17 Chatbot
The use case of including chatbots in the legal domain was mentioned five times during the SSIs. Interviewee I-8 envisions the use of chatbots for the "partial automation of the consultation process." (I-8). This includes for example chatbots, which can do initial interviews with clients. In this case, the chatbot would ask these typical questions, that otherwise attorneys would ask. Interviewee I-8 sees two main advantages when using chatbots. First, they reduce the susceptibility to errors. "When conducting interviews, we attorneys are often inclined to rely on our memory rather than using checklists." (I-8). The second benefit concerns synchronization. I-8 states "I am no longer dependent on being emotionally, temporally, and physically available at the same time as my clients to conduct such an interview." (I-8).

UC-18 Question Answering
The use case of *Question Answering* was mentioned once during the SSIs. I-11 sees great potential in the use of question answering software in the context of legal research. Additionally, I-11 states about question answering in the legal research procedure that "if it had voice control, it would be much better." (I-11).

UC-28 Credibility of Witnesses The use case of *Credibility of Witnesses* was mentioned once during the SSIs. It involves the application of technology to evaluate the credibility of witnesses in legal proceedings. It is a tool or system designed to assist legal professionals in making decisions regarding the trustworthiness of witnesses.
Interviewee I-17, who is a judge, is especially excited about this use case, as in his opinion "There is no place where more lies are told than in court." (I-17). 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.” (I-17).

UC-29 Translation
The use case of *Translation* was mentioned twice during the SSIs. This use case includes the translation of legal documents, such as contracts, from one language to another. Interviewee I-8 mentioned, that he is using software that provides translation at his workplace. In financing rounds with international investors, many things have to be captured in a bilingual format. Regarding the use of translation process in this context I-8 states that “it saves a lot of time in the process and manually doing that would be a nightmare.” (I-8).

UC-Cat-7 Legal Research and Information Management

The *Legal Research and Information Management* use case category consists of four use cases, as Table 5.30 shows. All four of them were mentioned by interviewees during the SSIs. Therefore for each of those four use cases, namely *UC-20 Changes in Law*, *UC-21 Database for Court Decisions* , *UC-30 Divergence Between Law Systems* and *UC-31 Research Tool*, a quote from the SSIs will be provided. The last two use cases are introduced during the SSIs, therefore a short explanation of them will be given as well.

ID	Use Case of UC-Cat-7	Amount
UC-20	Changes in Law	1
UC-21	Database for Court Decisions	1
UC-30	Divergence Between Law Systems	1
UC-31	Research Tool	11
Total		**14**

Table 5.30.: Legal use cases from SSIs, which form UC-Cat-7 Legal Research and Information Management

UC-20 Changes in Law
The use case *Changes in Law* was addressed once during the SSIs. Interviewee I-3 mentioned that in the law firm, where I-3 is employed, they are currently building up a Legal Tech software regarding staying up-to-date about changes in law. However, I-3 said that “I can’t really say anything about that.” (I-3). It appears that this Legal Tech software under development is secretive, and it is not meant to be communicated, yet.

UC-21 Database for Court Decisions
The use case regarding *Database for Court Decisions* was mentioned once during the SSIs. Interviewee I-12 would support it, if it would be possible to widely publish all court decisions. "These are things that would bring the entire industry forward." (I-12). I-12 also compared the closed field of law with other sciences, where everything is open access and I-12 concludes "that [law] has to change." (I-12).

UC-30 Divergence Between Law Systems
The use case of *Divergence Between Law Systems* was mentioned once during the SSIs. This use case addresses the challenges and the complexity that arises due to different legal systems across different countries.

Interviewee I-3 knows that in the company, where I-3 is working a tool exists, which monitors the divergence between British law and EU law and shows the changes. The United Kingdom initially incorporated a lot of European law into its own legal system, which was the starting point of this Legal Tech software. Since then, the laws have been diverging. This tool "keeps track of this divergence and shows the differences between EU law and UK law in various areas." (I-3).

UC-31 Research Tool
The use case category *Research Tool* was mentioned 11 times within the SSIs. This use case includes the legal research process. Legal Tech solutions can be used to streamline this process by for example supporting 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." (I-2).

UC-Cat-8 Other

This last use case category named *Other* consists of one single use cases, as shown in Table 5.31. This use case was already introduced in the SLR, but in this section a quote from the SSIs will be added to make this use case more tangible.

ID	Use Case of UC-Cat-8	Amount
UC-22	Law Firm Management-Software	12

Table 5.31.: Overview of UC-Cat-8 Other with insights from SSIs

UC-22 Law Firm Management-Software
Interviewee I-10 stated that a wide range of notary-management programs are available. I-10 then explained in detail, how the law firm management software helps at his work 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." (I-10). This is just one example of a law firm management software. Other interviewees named other highly specified software solutions, for example for attorneys or for judges.

5.2.4. Further Classification of Legal Use Cases

In this section, the 95 identified use cases from the SSIs are categorized by the previously defined use case categories and additionally by the following three categories:

Already Used Use Case This category includes all the use cases, which the interviewees personally used.

Knows-Existing Use Case Part of this category are all the use cases, of which the interviewees heard of. Those existing use cases have not yet been tried out by the interviewees.

Future Use Case This category covers all the use cases, which the interviewees could imagine for the future and which to the best of their knowledge do not exist yet.

Table 5.32 gives a summary of the use cases in the SSIs. For each use case category, also the interview-codes of the interviewees, which mentioned them are stated.

ID	Use Case	#	Cat. (%)	Total (%)	Code
UC-3	Risk Assessment	1		1,05%	
	Already Used UC	0	0%	0%	
	Knows-Existing UC	0	0%	0%	
	Future UC	1	100%	1.05%	I-14
UC-4	Automatic File Difference Tracking	2		2.11%	
	Already Used UC	1	50.00%	1.05%	I-8
	Knows-Existing UC	1	50.00%	1.05%	I-18
	Future UC	0	0%	0%	

ID	Use Case	#	Cat. (%)	Total (%)	Code
UC-6	Document Management	4		4.21%	
	Already Used UC	4	100%	4.21%	I-2, I-3, I-5, I-8
	Knows-Existing UC	0	0%	0%	
	Future UC	0	0%	0%	
UC-8	Automatic Contract Generation	10		10.53%	
	Already Used UC	0	0%	0%	
	Knows-Existing UC	3	30.00%	3.16%	I-2, I-7, I-15
	Future UC	7	70.00%	7.37%	I-2, I-2, I-4, I-9, I-10, I-11, I-18
UC-9	Enrichment of Legal Documents	1		1.05%	
	Already Used UC	0	0%	0%	
	Knows-Existing UC	1	100%	1.05%	I-9
	Future UC	0	0%	0%	
UC-10	Summarization	3		3.16%	
	Already Used UC	2	66.67%	2.11%	I-9, I-15
	Knows-Existing UC	0	0%	0%	
	Future UC	1	33.33%	1.05%	I-11
UC-11	Anonymisation	2		2,11%	
	Already Used UC	1	50.00%	1.05%	I-1
	Knows-Existing UC	0	0%	0%	
	Future UC	1	50.00%	1.05%	I-6
UC-12	Automatic Information Extraction and Insertion	8		8.42%	
	Already Used UC	5	62.50%	5.26%	I-3, I-3, I-3, I-14, I-18
	Knows-Existing UC	2	25.00%	2,11%	I-9, I-9
	Future UC	1	12.5%	1.05%	I-18

ID	Use Case	#	Cat. (%)	Total (%)	Code
UC-14	Legal Decision Making	7		7.37%	
	Already Used UC	0	0%	0%	
	Knows-Existing UC	0	0%	0%	
	Future UC	7	100%	7.37%	I-6, I-9, I-11, I-12, I-14, I-17, I-18
UC-15	Legal Reasoning	2		2.11%	
	Already Used UC	0	0%	0%	
	Knows-Existing UC	0	0%	0%	
	Future UC	2	100%	2.11%	I-4, I-9
UC-16	Recommendations Based on Previous Court Rulings	2		2.11%	
	Already Used UC	0	0%	0%	
	Knows-Existing UC	0	0%	0%	
	Future UC	2	100%	2.11%	I-9, I-13
UC-17	Chatbot	5		5.26%	
	Already Used UC	0	0%	0%	
	Knows-Existing UC	1	20.00%	1.05%	I-9
	Future UC	4	80.00%	4.21%	I-4, I-8, I-8, I-10
UC-18	Question Answering	1		1.05%	
	Already Used UC	0	0%	0%	
	Knows-Existing UC	0	0%	0%	
	Future UC	1	100%	1.05%	I-11
UC-20	Changes in Law	1		1.05%	
	Already Used UC	0	0%	0%	
	Knows-Existing UC	0	0%	0%	
	Future UC	1	100%	1.05%	I-3
UC-21	Database for Court Decisions	1		1.05%	
	Already Used UC	0	0%	0%	
	Knows-Existing UC	0	0%	0%	
	Future UC	1	100%	1.05%	I-12

ID	Use Case	#	Cat. (%)	Total (%)	Code
UC-22	Law Firm Management-Software	12		12.63%	
	Already Used UC	11	91.67%	11.58%	I-2, I-5, I-8, I-10, I-11, I-13, I-13, I-14, I-16, I-17, I-18
	Knows-Existing UC	1	8.33%	1.05%	I-6
	Future UC	0	0%	0%	
UC-23	Deadline Management	4		4.21%	
	Already Used UC	0	0%	0%	
	Knows-Existing UC	0	0%	0%	
	Future UC	4	100%	4.21%	I-2, I-13, I-17, I-18
UC-24	E-Mail Communication	1		1.05%	
	Already Used UC	0	0%	0%	
	Knows-Existing UC	0	0%	0%	
	Future UC	1	100%	1.05%	I-3
UC-25	Mass-Trials	4		4.21%	
	Already Used UC	1	25.00%	1.05%	I-12
	Knows-Existing UC	3	75.00%	3.16%	I-1, I-2, I-13
	Future UC	0	0%	0%	
UC-26	Transcription	6		6.32%	
	Already Used UC	4	66.67%	4.21%	I-9, I-11, I-13, I-17
	Knows-Existing UC	1	16.67%	1.05%	I-4
	Future UC	1	16.67%	1.05%	I-18
UC-27	Dispute Resolution Mechanism	3		3.16%	
	Already Used UC	0	0%	0%	
	Knows-Existing UC	1	33.33%	1.05%	I-6
	Future UC	2	66.67%	2.11%	I-2, I-14

ID	Use Case	#	Cat. (%)	Total (%)	Code
UC-28	Credibility of Witnesses	1		1.05%	
	Already Used UC	0	0%	0%	
	Knows-Existing UC	0	0%	0%	
	Future UC	1	100%	1.05%	I-17
UC-29	Translation	2		2.11%	
	Already Used UC	2	100%	2.11%	I-8, I-9
	Knows-Existing UC	0	0%	0%	
	Future UC	0	0%	0%	
UC-30	Divergence Between Law Systems	1		1.05%	
	Already Used UC	0	0%	0%	
	Knows-Existing UC	1	100%	1.05%	I-3
	Future UC	0	0%	0%	
UC-31	Research Tool	11		11.58%	
	Already Used UC	8	72.73%	8.42%	I-2, I-2, I-4, I-10, I-11, I-12, I-13, I-16
	Knows-Existing UC	0	0	0%	
	Future UC	3	27.27%	3.16%	I-2, I-4, I-18
Total		**95**		**100**%	

Table 5.32.: Summary of all the use cases of the SSIs

5.2.5. Requirements for Legal Tech Solutions

During the SSIs, 17 of the the 18 participants were asked, what requirements they have for a Legal Tech solution. Table 5.33 gives an overview of the answers. In total 53 requirements were stated, which are categorized into eleven categories. In the following, each of the eleven categories will be explained. Furthermore, the table includes the percentage representation of each requirement, indicating its proportion in relation to the total number of 53 listed requirements.

ID	Category	#	%
Req-1	Added Value	6	11.32%
Req-2	Correct Output	6	11.32%
Req-3	Cost-Sensitivity	5	9.43%
Req-4	Cross-Platform Usability	2	3.77%
Req-5	Data Protection	12	22.64%
Req-6	Easy Usability	9	16.98%
Req-7	Flexibility	2	3.77%
Req-8	Reliability	2	3.77%
Req-9	Seamless Integration	5	9.43%
Req-10	Service for Emergency	2	3.77%
Req-11	Understandability of Technology	2	3.77%
Total		**53**	**100%**

Table 5.33.: Overview of the requirements for a Legal Tech solution stated by the interviewees

Req-1 Added Value

This requirement refers to the software's ability to provide additional benefits or features that simplify the overall legal process for the user.
Interviewee I-5 expressed: "The software solution must also provide a very tangible and measurable added value." (I-5). In order to be tangible, this must be something that really speeds up processes, improves them, and generates higher quality results compared to not using the solution.

Req-2 Correct Output

This requirement focuses on the software's capability to generate accurate and precise results.
Interviewee I-15 describes legal practitioners as "risk-averse, which means their requirements for precision are very high. Exceptionally high, when it comes to AI systems." (I-15).

Req-3 Cost-Sensitivity

This requirement includes the prevision of cost-effective Legal Tech solutions without compromising essential functionalities or quality.
Legal Tech solutions do come with a certain cost. Interviewee I-1 explained a situation at I-1's workplace, where some functionality of a Legal Tech Tool was not included and they had to buy it separately. I-1 states that within the law firm "it became a topic of discussion whether the price was justified, (...) or whether it would be more cost-effective to pay the lawyer or interns to handle the tasks." (I-1).

Req-4 Cross-Platform Usability

This requirement includes the Legal Tech software's ability to function seamlessly across different platforms, such as desktops, laptops, or mobile devices.
Interviewee I-10 has a clear vision about this topic. I-10 demands for Legal Tech solutions, which are usable across different platforms. I-10 even goes one step further and states: "I would like to have the ability to access everything on my mobile phone as well." (I-10). In the eyes of I-10, an optimal solution for this are browser-based solutions, as they are cross-platform compatible.

Req-5 Data Protection

This requirement involves the implementation of security measures to safeguard sensitive and confidential information.
Interviewee I-1 knows from the experience of working in big law firms that "data protection is always a concern that law firms pay great attention to." (I-1).

Req-6 Easy Usability

This requirement refers to a user-friendly interface of the software, which is intuitive to understand.
This is a concern, which is also important for interviewee I-10. Legal Tech software must be designed in a clear and intuitive way so that it is understandable and usable for everyone. Expressed in the words of I-10: "It doesn't need to be rocket science. Essentially, it should be user-friendly and easy to operate." (I-10).

Req-7 Flexibility

This requirement includes customization or configuration to accommodate varying workflows, document templates, or jurisdiction-specific rules. Flexible software enables users to tailor the system to their specific preferences and needs.
This is a requirement, which is particularly important for Interviewee I-18. A Legal Tech software should give its user the opportunity to have some degree of freedom.

I-18 gives the example of automatically generated documents: “Because I might not want to write "This fact is established after the conducted main examination." Maybe I don’t like that sentence, and I want to phrase it differently.” (I-18). Therefore, the software should have some space for flexibility and adaptability.

Req-8 Reliability

This requirement ensures that the Legal Tech software performs consistently and accurately without downtime.
Interviewee I-17 stated that at work, they have IT partners, who take care of the utilized software through remote maintenance. However, it is still quite common for I-17 to experience software disruptions. Usually, this can be resolved quite fast, but not always. I-17 states that those disruptions “result in a significant loss of working time, that’s for sure.” (I-17).

Req-9 Seamless Integration

Seamless integration involves the software’s ability to integrate smoothly with other systems or tools commonly used in the legal industry.
Interviewee I-8 is a big fan of interfaces, as the landscape of software providers in the legal field becomes more and more fragmented. Therefore, it is necessary, that different software solutions can communicate with each other, meaning that the tools can work together. I-8 stated an example, where a software itself is working well, but the interface with another program is not fully developed. I-8 even goes as far as, if this “software doesn’t fix the issue by the end of the year, I will replace it.” (I-8).

Req-10 Service for Emergency

This requirement includes the software provider’s ability to offer prompt and reliable support or assistance in emergency situations.
Interviewee I-1 supports this requirement with the statement that “the Legal Tech solutions must function 24 hours a day, every day. It is important to have a service available in case of emergencies.” (I-1).

Req-11 Understandability of Technology

This requirement focuses on ensuring that the software’s underlying technology or algorithms are understandable to legal professionals. This can involve clear documentation, explanations, or training materials that enable users to comprehend how the software works and trust its results.
Interviewee I-7 states “I would say that a good user manual is definitely necessary.”

(I-7). According to I-7, this manual should include explanations, that are understandable even for non-technical individuals. It should be clear what is happening within the software so that users can trust the process.

5.2.6. Source of Information

The interviewees were asked, what media they use to stay informed about news in the legal sector. Figure 5.7 includes the answers of the 18 interviewees. The following nine categories could be derived:

- **Company-Internal Source**
  Some of the interviewees work in big law firms. For example Interviewee I-3 stated that its company employs *Knowledge Lawyers*, whose job it is to inform themselves about news in the legal domain, also changes of law, and communicating important news to the rest of the lawyers. This can be done for example with company-internal newsletters or chat-groups.

- **Mainstream Print Media**
  Some of the interviewees inform themselves via *Print Media*. For example Interviewee I-17 mentioned that they get the most important legal news from regular print media, which I-17 reads daily.

- **Newsletter**
  Newsletters are common amongst legal practitioners. Interviewee I-4 for example regularly receives newsletters from the *Federal Court of Justice* or the *Federal Constitutional Court*.

- **Official Institutions**
  Interviewee I-13 for example gets informed via the website of the Press-Release-Offices of the Federal Court of Justice and the Federal Constitutional Court.

- **Online Databases**
  Interviewee I-16 for example likes to get informed via the legal databases Juris and Beck-Online.

- **Social Media**
  The specific Social Media channels, which form this category are *LinkedIn*, *Twitter*, *Instagram* and *Podcasts*. However, within this category, *LinkedIn* is used most among the interviewees.

- **Special Legal Literature**
  Interviewee I-2 for example regularly reads the legal magazine *Neue Juristische Wochenschrift* to stay informed about news in the legal field.

- **Websites**
  Interviewee I-15 is very interested in programming. Therefore, I-15 for examples regularly gets information from *https://artifact.news/* to stay informed about news in terms of Legal Tech.

- **Workshops**
  Interviewee I-11 regularly participates in advanced trainings organized by the *Chamber of Notaries*.

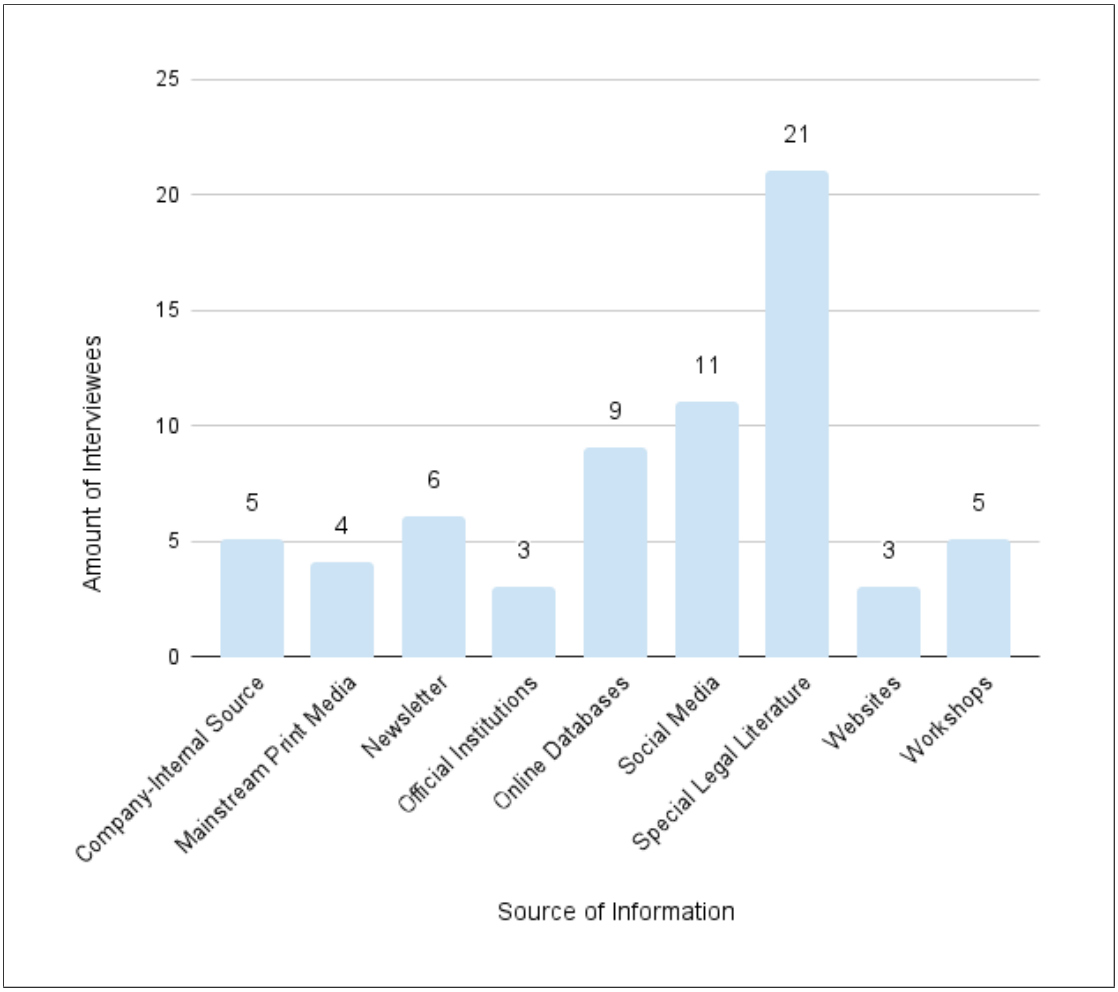

Figure 5.7.: Overview of the sources, which the interviewees use to stay up-to-date

5.2.7. Interest in Underlying Technology

Of the 18 interviewees, 15 were asked how much technical depth they want to know about a Legal Tech software, which they are using. To categorize the answers, the following categories are introduced:

- **Interested**
  This category includes answers, where the interviewees clearly stated, that they are interested to know which underlying technology is used to build the Legal Tech software.
- **Interested in Basics**
  This category includes answers, where the interviewee does not want to go into deep technical depth, but rather wants to get an overview.
- **Not Interested**
  This sections includes interviewees, which are not interested to know which technologies are used to build the software. Requirement for this category is that the software is generating correct results and is legally allowed to use.

Table 5.34 gives an overview of the three categories including the amount of occurrence per category and the interviewees-Codes for each category.

Category	#	Code
Interested	5	I-3, I-6, I-7, I-13, I-14
Interested in Basics	1	I-8
Not Interested	9	I-2, I-4, I-10, I-11, I-12, I-15, I-16, I-17, I-18
Total	**15**	

Table 5.34.: Distribution of the 15 interviewees regarding their interest in the underlying technology

5.3. Systematic Literature Review and Semi-Structured Interviews in Context

In this section, the findings from the SLR are combined with the findings of the SSIs. A combined version containing both, the use cases of the SLR and the SSIs is generated and the found use cases are compared with the NLP-Technologies from the SLR.

5.3.1. Legal Use Cases

In Table 5.35 the use cases from the SLR and from the SSIs are included.

ID	Category	SLR(#)	SSI(#)
UC-Cat-1	**Compliance and Risk Management**	**4**	**1**
UC-1	Automation of Auditing	1	0
UC-2	GDPR Compliance	2	0
UC-3	Risk Assessment	1	1
UC-Cat-2	**Document Analysis and Management**	**9**	**6**
UC-4	Automatic File Difference Tracking	1	2
UC-5	Document Classification	6	0
UC-6	Document Management	1	4
UC-7	Error Detection	1	0
UC-Cat-3	**Document Generation and Management**	**8**	**23**
UC-8	Automatic Contract Generation	1	10
UC-9	Enrichment of Legal Documents	2	1
UC-10	Summarization	5	3
UC-23	Deadline Management	0	4
UC-24	E-Mail Communication	0	1
UC-25	Mass-Trials	0	4
UC-Cat-4	**Information Processing and Extraction**	**14**	**16**
UC-11	Anonymisation	1	2
UC-12	Automatic Information Extraction and Insertion	12	8
UC-13	Patent Retrieval	1	0
UC-26	Transcription	0	6

ID	Category	SLR(#)	SSI(#)
UC-Cat-5	**Legal Decision Making and Dispute Resolution**	**8**	**14**
UC-14	Legal Decision Making	1	7
UC-15	Legal Reasoning	5	2
UC-16	Recommendations Based on Previous Court Rulings	2	2
UC-27	Dispute Resolution Mechanism	0	3
UC-Cat-6	**Legal Information Retrieval and Support**	**5**	**9**
UC-17	Chatbot	1	5
UC-18	Question Answering	3	1
UC-19	Ranking of Lawyers	1	0
UC-28	Credibility of Witnesses	0	1
UC-29	Translation	0	2
UC-Cat-7	**Legal Research and Information Management**	**2**	**14**
UC-20	Changes in Law	1	1
UC-21	Database for Court Decisions	1	1
UC-30	Divergence Between Law Systems	0	1
UC-31	Research Tool	0	11
UC-Cat-8	**Other**	**1**	**12**
UC-22	Law Firm Management-Software	1	12
Total		**51**	**95**

Table 5.35.: Overview of all the legal use cases addressed in the SLR and in the SSIs

5.3.2. Mapping of Use Case Category to NLP-Technology

In this section, the identified NLP-Technologies from the SLR, summarized in Table 5.1 and the identified use case categories, which are depicted in Table 5.35 are mapped. The result of this mapping can be seen in Table 5.36. This table includes the percentages

of the use case categories among the derived use cases from both, the SLR and the SSIs. Additionally, the percentages of each NLP Category among all the reviewed 49 papers is stated. However, a mapping by category cannot cover all individual cases, it rather provides a direction. It is possible that individual use cases may also be realized by a combination of several NLP-Technologies from different categories.

Use Case Category	**UC-Cat. (%)**	**NLP Category**	**NLP-Cat. (%)**
UC-Cat-1 Compliance and Risk Management	3.42%	NLP-Cat-1 Document Analysis and Processing	12.16%
UC-Cat-2 Document Analysis and Management	10.27%	NLP-Cat-1 Document Analysis and Processing	12.16%
UC-Cat-3 Document Generation and Management	21.23%	NLP-Cat-4 Text Generation	40.54%
UC-Cat-4 Information Processing and Extraction	20.54%	NLP-Cat-3 Text Extraction	31.08%
UC-Cat-5 Legal Decision Making and Dispute Resolution	15.07%	NLP-Cat-2 Natural Language Understanding and Applications	12.16%
UC-Cat-6 Legal Information Retrieval and Support	9.59%	NLP-Cat-2 Natural Language Understanding and Applications	12.16%
UC-Cat-7 Legal Research and Information Management	10.96%	NLP-Cat-1 Document Analysis and Processing	12.16 %
UC-Cat-8 Other	8.90%	NLP-Cat-5 Other	4.05%

Table 5.36.: Mapping of use case categories to NLP-Technologies

6. Discussion

This chapter consists of two main sections. Firstly, the *Key Findings* of this thesis are presented. Secondly, also its *Limitations* are pointed out.

6.1. Key Findings

In this section, important findings from the SLR and from the SSIs are pointed out.

Increasing Amount of Publications
Figure 5.1 indicated, that specifically in 2021 a great amount of papers was published. This could be the case, because due to the COVID-19 pandemic, an "increase in use of new technologies" [58] was registered. And therefore the incentive for academia to investigate and publish application-related papers was there.
Another reason for the rise in the amount of published papers in 2020/2021 are the innovations in terms of technology. For example in terms of Large Language Models, in 2020 OpenAI introduced GPT-3. This was a significant milestone in the development of Large Language Models [59].

Understanding of Legal Tech
When the participants were asked what their understanding of Legal Tech is, 12 of the 17 asked interviewees immediately answered with supporting aspects of Legal Tech, which can be categorized as *Legal Technology 1.0* according to Goodenough [13]. Only five of the interviewees, which marks 29.41% included elements regarding *Legal Technology 3.0*.
This shows that in the majority of the minds of legal practitioners, they do not even consider *Legal Technology 3.0* as a possibility or they simply are not well enough informed.
I-4, who is a law student stated:

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

The German law university program could also be the cause of the lack of knowledge about Legal Tech. Interviewee I-7, who is still a law student, also mentioned that Legal Tech does not play a role in the law studies. This is something, which should

be changed. Legal Tech has to be a mandatory part in legal studies, in order for legal practitioners to take informed decisions about the use of Legal Tech.

ELSA Over the Years

The concept of Ethical, Legal and Social Aspects was only included in 18 of the 49 investigated papers. This marks only 36.73%. In total, 30 different ELSA could be derived from those papers. And those papers, which indeed included ELSA on average included several aspects, to be precise 1.67 aspects.
In the scope of this thesis, a trend regarding the percentage of ELSA-mentions over the years could not be identified, as seen in Table 6.1. The dataset of papers, which include ELSA is too small to derive significant hypothesis. This is up for future work.

Year	ELSA per Year	Total per Year	% of ELSA
2007	1	2	50.00%
2016	1	2	50.00%
2019	3	7	42.86%
2020	3	7	42.86%
2021	7	16	43.75%
2022	2	7	28.57%
2023	1	1	100.00%

Table 6.1.: Overview of occurrence of ELSA in the examined publications

Comparison between SLR and SSIs concerning ELSA

Table 6.2 summarizes the distribution between the three parts *ethical*, *legal* and *social*. The focus of the interviewees is less on the social implications, but rather on ethical and legal ones, compared to the SLR. Especially the higher percentage in terms of *legal* is to highlight here. Under no circumstances do legal practitioners want to violate the law. As I-9 states:

> "And lawyers are very sensitive to this (...). So 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 solution would increase the usage of Legal Tech software amongst practitioners.

Aspect	SLR (%)	SSIs (%)
Ethical	33.33%	40.51%
Legal	13.33%	22.78%
Social	53.33%	36.71%

Table 6.2.: Comparison of the occurrence of ELSA regarding SLR and SSIs

Tangible Use Cases
Those use cases, which are very close to the legal practitioners every-day routines, like *UC-22-Law Firm Management Software*, *UC-8 Automatic Contract Generation*, *UC-12 Automatic Information Extraction and Insertion* or *UC-31 Research Tool* were mentioned most by the interviewees. This observation can be underlined by the *Understanding of Legal Tech* of the interviewees, where only 29.41% could be categorized into the stage *Legal Technology 3.0*. Interviewee I-6 for example stated in the context of the use of AI: "(...) but sometimes the low-hanging fruits are not bad either." (I-6). This attitude reflects the mindset of many legal practitioners, who feel most comfortable when dealing with systems, which are visualizable for them and that can be easily integrated into their existing workflows. All of this speaks very much for the stage *Legal Technology 1.0*.

Mapping NLP-Research with Derived Use Cases
Table 5.36 gives an overview of the most occurring NLP categories and the most occurring use case categories from both, SLR and SSIs. Also the mapping between the use case categories and the NLP categories, which are needed to implement the use case categories, are stated. As an example: The most occurring use case category, namely *UC-Cat-3 Document Generation and Management*, which covered 21.23% of all use cases is mapped to the biggest NLP category, which is *NLP-Cat-4 with 40.54%*. In general, the directions of the NLP-Research and the defined use cases coincide.

Interest in Underlying Technology
When asking the interview participants about their interest in the underlying technology of a software, which they use on a regular basis, 9 out of 15 answered, that they are not particularly interested. For those 60% of the interview participants, it would be sufficient for them to know that the software is allowed to be used in the legal context and that the provided results are correct. I-11 for example stated:

> " 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).

This leads to the need for some kind of certification, which gives the lawyers some guidance and reassurance when introducing Legal Tech software to their day-to-day tasks.

Requirements of Legal Tech Solution
The top five requirements for Legal Tech solutions, according to the interviewees are:

1. **Req-5 Data Protection**
   12 out of 17 interviewees mentioned this requirement, which is 70.59%
2. **Req-6 Easy Usability**
   9 out of 17 interviewees mentioned this requirement, which is 52.94%
3. **Req-1 Added Value**
   6 out of 17 interviewees mentioned this requirement, which is 35.29%
4. **Req-2 Correct Output**
   6 out of 17 interviewees mentioned this requirement, which is 35.29%
5. **Req-3 Cost Sensitivity**
   5 out of 17 interviewees mentioned this requirement, which is 29.41%

Having a legal requirement as number one is not surprising, as legal practitioners pay great attention to adhering to law. The second, third and fourth requirement can be considered together. The majority of legal practitioners do not want to get in depth about underlying technology, they want an easy user interface, which outputs the correct values and therefore adds real value to their work. Legal practitioners are also business-people, for example attorneys, which have their own law firms or notaries.

Dissatisfaction with Legal Databases
In Germany there are legal databases, which have some kind of monopoly-position. Already whilst studying, law students get in touch with those databases to prepare for their exam and they keep using it during their career. Many of the interviewees were not satisfied with the usability of those databases. For example the search is very inconvenient and the corresponding search results are not precise. A lot of fine-tuning by the user, e.g. with search filters is needed to properly use the database. Interviewee I-9 describes those databases as "very powerful in terms of content, technically very weak and completely outdated." (I-9). The ethical concerns *Eth-4 Transparency* and *Eth-7 Monopoly* play in here as well. This whole situation leaves room for improvement and the law practitioners, which were interviewed generally make a very open impression, that they would be open to also include new player in the market.

Newsletters are Common
This finding is concerned about the source of information, how legal practitioners inform themselves about news in the legal field. Six out of the 18, which is every third person asked, likes to get informed about news in the legal domain via newsletters.

Language Barrier
Figure 5.4 gives an overview of the languages of the datasets, which were used to train NLP algorithms. The English language is predominant in this field, as 63% of the publications were built up on the English language. The other 11 identified languages mark each either 2.05% or 4.08%. Some languages, such as Dutch or French were translated into English before applying NLP-Algorithms. However, this correlates with the ELSA *Eth-5 Artificially Created Opinion*. The opinion created by a NLP algorithm should reflect the opinion of the people of a country. People from other parts of the world could have different cultural values as a people, which play a crucial role in the opinion of a people. Therefore, each country should foster their own research in the field of NLP and invest in building up own datasets in their native language.

6.2. Limitations

In order to discuss the validity and value of the SLR and SSIs, important limitations are derived in this section.

6.2.1. Researcher Bias

A factor, which could have impacted the comprehensiveness of the SLR is the potential presence of *Researcher Bias*. During the SLR, 49 publications were screened for the five criteria introduced in Table 4.5. As this was performed by the author alone, this could introduce errors and misinterpretations and therefore influence the accuracy and completeness of the results. To mitigate this, the author iterated on the process of data extraction.
Also the SSIs were performed and analysed only by the author. To mitigate this, the author performed the analysis on different days, which reduces the bias. Also in this case the experiences and interpretations of the researcher might have an impact on the outcome of the SSIs.

6.2.2. Interviewee Pool

In this section, the interviewees, which participated in the SSIs are inspected. The amount of 18 legal practitioners is quite decent and extended the initially planned amount of interviews. When breaking it down to the exact professions like *notary* or *judge* however, the amount per profession is quite narrow. Also the gender-ratio

among the interviewees is not balanced, as only 27.78% identify as female and the rest is male. Additionally, as the legal landscape is very diverse and scattered, other participant-groups like prosecutors, lawyers for mass proceedings or employees of legal protection insurance companies would still be of interest and were not interviewed. In any case it would be beneficial for future research to include even more interview participants, specifically also other legal target groups.

6.2.3. Generalizability

As this whole thesis investigates within the intersection of the legal field and NLP in terms of Legal Tech, in general results are only applicable in this domain. Additionally, the SSIs were only conducted among legal practitioners, which makes those results not generalizable for example among NLP-Researchers. Conducting SSIs with NLP-Researchers would strengthen the generalizability of the results of this thesis and are recommended for future work.

7. Conclusion

This chapter consists of two main sections. First, a *Summary* is given and second, recommendations for *Future Work* are given.

7.1. Summary

This thesis provides an overview of the intersection between NLP and Legal Tech. By performing a SLR, 122 publications from four databases were analyzed. After abstract and full-text screening, 49 publications were analyzed in detail regarding year of publication, used NLP-Technology, ELSA, included legal use cases and language of data. In a second step, SSIs were performed. In total, 18 legal practitioners were interviewed. The corresponding interview transcripts were categorized regarding the interviewees understanding of Legal Tech, ELSA, legal use cases, sources of information and their interest in Legal Tech's underlying technology. In this section, the answers to the research questions provided in Section 4.1 are summarized.

> **RQ1**: *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?*

In total, 74 NLP-Technologies were discovered in the 49 publications. Those are grouped together to the following five categories:

- **NLP-Cat-1 Document Analysis and Processing**
  This category covers 12.2% of the occurring NLP-Technologies.
- **NLP-Cat-2 Natural Language Understanding and Applications**
  Similar to NLP-Cat-2, this category also includes 12.2% of the occurring instances.
- **NLP-Cat-3 Text Extraction**
  This category covers31.1 % of the NLP-Technologies and therefore marks the second biggest category.
- **NLP-Cat-4 Text Generation**
  This category includes 40.5% of the NLP-Technologies and it therefore the largest among the five categories.

- **NLP-Cat-5 Other 3**
  This category only marks 4.0%, and therefore is the least-frequent one.

The extent, to which ELSA are covered in the SLR is quite small, as only 18 of the examined 49 papers include ELSA. Those 18 however include on average 1.67 aspects.

> **RQ2**: *What are the use cases in which the identified NLP techniques can be utilized?*

When combining the use cases from the SLR and the SSIs, a total of 31 different use cases could be identified. In total, 51 use cases were found in the SLR and 95 in the SSI. Some of the use cases within the SLR and SSIr were overlapping, which results in a total of 31 disjoint use cases. A full list of all use cases is included in Table 5.35. The derived use cases can be grouped together into eight use case categories:

- **UC-Cat-1 Compliance and Risk Management**
  This category accounts for 3.42% of the total use case occurrences.
- **UC-Cat-2 Document Analysis and Management**
  This category represents 10.27% of the total use case occurrences.
- **UC-Cat-3 Document Generation and Management**
  This category constitutes 21.23% of the total use case occurrences.
- **UC-Cat-4 Information Processing and Extraction**
  This category contributes with 20.54% of the total use case occurrences.
- **UC-Cat-5 Legal Decision Making and Dispute Resolution**
  This category comprises 15.07% of the total use case occurrences.
- **UC-Cat-6 Legal Information Retrieval and Support**
  This category accounts for 9.59% of the total use case occurrences.
- **UC-Cat-7 Legal Research and Information Management**
  This category accounts for 10.96% of the total use case occurrences.
- **UC-Cat-8 Other**
  This category represents 8.90% of the total use case occurrences.

Table 5.36 includes a mapping of the eight identified use case categories to the five identified NLP-Categories.

> **RQ3**: *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**
  When introducing legal use cases, abstract from the underlying technology, as this could scare-off legal practitioners.

- **Educational section for legal practitioners**
  Include an educational section, where the basic NLP-Technologies are explained very intuitively. However, this section should not be mandatory to understand the rest of the knowledge base. This is rather for interested people to widen their horizon.

- **Newsletter**
  As newsletter are widely spread among legal practitioners, this communication technique could be used to keep the readers engaged about news from the Legal Tech domain.

- **Marketing of the joint knowledge base**
  An important factor when introducing a joint knowledge-base is how the target groups of NLP-Researches and legal practitioners become aware of it. Among legal practitioners, social media, specifically LinkedIn is widely spread and could be used to introduce the knowledge base. Among NLP-Researchers however, an E-Mail could be sent to different NLP-Chairs to spread its popularity.

- **Intuitive design**
  The focus should be on an easy and intuitive use interface, so that the users of the knowledge base know how the software works. This was one of the major requirements of the interviewees for a Legal Tech solution.

- **Added value**
  The knowledge base needs to add a specific value to its user. The value for the legal practitioners can vary from the value of the NLP-Researchers. But still, an added value has to be clarified. One point is of course bringing legal practitioners and NLP-Researchers on the same table

- **Network**
  During the SSIs, the impression arose that legal practitioners are a big fan of connecting themselves with other people and discussing about Legal Tech solutions. The knowledge base should be a place where interaction between different stakeholders is possible and fostered.

- **Keep information up-to-date**
  Another point, which still has to be investigated more on is keeping the knowledge base up to date. Who inserts new technologies or updates existing ones?

7.2. Future Work

In this section, suggestions for different directions for future research are stated, which could be derived during the conduction of this thesis, namely *Joint Knowledge Base*, *ELSA* and *Certification*.

Joint Knowledge Base
Regarding the implementation of a joint knowledge base, which can be used by NLP-Researchers and legal practitioners, several things are still open, such as: Which value can the joint knowledge base add to NLP-Researchers and which values to legal practitioners? Design Thinking workshops could be conducted with both, NLP-Researchers and legal practitioners to find out, which intrinsic need this knowledge base could solve.
In the scope of this thesis, interviews with legal professionals were conducted to find out their needs. To obtain a holistic view of the topic, we recommend a study including SSIs with researchers from the NLP-Field.

ELSA
In the scope of ELSA in SLR over time, we were not able to generate significant results. Therefore, a SLR in a broader field, for example Legal Tech in general would make sense, in order to derive trends regarding the presence of ELSA in academia over time.

Certification
During the SLR, many interviewees were very cautious regarding the use of Legal Tech, as they are not able to understand all the underlying technology in detail. It would help them to have some kind of *certification*, where they could have confidence that the Legal Tech solution, which they want to use is *allowed* and not violating any law. Therefore, further research in the field of certification for Legal Tech software is recommended. Some questions that could be addressed: How should a certification look like? Who takes the legal responsibility? Who carries financial responsibility? What requirements are needed for a certification? How are ELSA integrated into certification?

A. General Addenda

This section consists of Section A.1, which includes the questionnaire sent out to the interviewees of the SSIs prior to their interviews and Section A.2, which includes the translations of the direct quotes of the interviewees.

A.1. Interview Questionnaire

Figure A.1 includes the interview guide in its original language, which is German.

A.2. Translations

Table A.1 includes the direct quotes of the interviewees, which are mentioned in this thesis. As the interviews were conducted in German, for the scope of this thesis, the quotes were translated into English. The corresponding codes of the interviewees, as defined in Table 4.8 are also stated and structure the table.

Code	Original Quote [German]	Translation [English]
I-1	(...) die müssen ja 24 Stunden am Tag, jeden Tag funktionieren, es ist wichtig, dass es da eben auch einen Service gibt in Notfällen.	(...) they need to function 24 hours a day, every day, so there should be a service available in case of emergencies.
I-1	(...) wo man zum Beispiel Dokumente schwärzen kann, also personenbezogene Daten und so weiter.	(...) where you could blacken parts in documents, such as personal data and so on.
I-1	Die [Legal Tech Lösungen] müssen ja 24 Stunden am Tag, jeden Tag funktionieren. Dass es da eben auch einen Service gibt in Notfällen ist wichtig.	The [Legal Tech Solutions] must function 24 hours a day, every day. It is important to have a service available in case of emergencies.

Code	Original Quote [German]	Translation [English]
I-1	Datenschutz ist immer so ein Thema, auf das Kanzleien sehr acht geben.	Data protection is always a concern that law firms pay great attention to.
I-1	Das war dann auch so ein Thema, ob sich der Preis da rentiert, (...) oder ob man den Anwalt oder Praktikanten eher zahlen will, die das dann machen.	It became a topic of discussion whether the price was justified, (...) or whether it would be more cost-effective to pay the lawyer or interns to handle the tasks.
I-2	(...) wenn das zentralisiert ist, ist es immer am einfachsten. Problem von Zentralisierung ist natürlich, wenn das einer zentralisiert, erlangt er irgendwo Monopolstellung.	(...) if it is centralized, it is always the easiest. The problem with centralization is, of course, that if someone centralizes it, they gain a monopoly position.
I-2	Aber solange die Gesellschaft nicht will, wird das nicht kommen.	But as long as society doesn't want it, it won't happen.
I-2	(...) haben ein Dateimanagementsystem. Da wird ein Dokument einer Mandantennummer und einem Auftrag zugeordnet.	(...) have a file management system where documents are assigned to a client number and a specific task.
I-2	(...) sich die relevanten Daten rauszieht aus irgendwelchen Bewerbungsunterlagen oder sonst was und mir auf einen Knopfklick meinen Arbeitsvertrag und so weiter erstellt.	(...) extract relevant data from application documents or any other source, and generates my employment contract with just a click of a button.
I-2	(...) das ist ein Bereich, den ich für hochinteressant erachte.	(...) this is an area that I find highly interesting.
I-2	(...) das mir Recherchearbeiten komplett übernimmt, wo ich nicht selber hergehen muss und recherchieren muss.	(...) that takes over research work completely, where I don't have to go and conduct the research myself.

Code	Original Quote [German]	Translation [English]
I-3	(...) Verantwortung übernehmen. Und dann wird es praktisch schnell eine Versicherung, nicht eine Kanzlei.	(...) take responsibility. And then it would practically become an insurance company rather than a law firm.
I-3	(...) dass da schon sehr viel Zeit eben investiert wird, die perfekte Antwort für den Mandanten zu schmieden. Wenn das erleichtert werden kann oder von der KI übernommen werden kann, umso besser.	(...) that a significant amount of time is currently invested in crafting the perfect response for the client. If this process can be made easier or taken over by AI, even better.
I-3	Da darf ich nicht wirklich was sagen.	I can't really say anything about that.
I-3	(...) diese Spaltung in den Augen behält und uns in verschiedenen Themenbereichen eben zeigt, was die Unterschiede zwischen EU-Recht sind und UK-Recht.	(...) keeps track of this divergence and shows us the differences between EU law and UK law in various areas.
I-4	(...) wie entstehen herrschende Meinungen, wie entsteht ein Meinungsstreit. (...) da könnte es halt am ehesten Problem geben.	(...) how prevailing opinions are formed, how disputes of opinion arise. (...) that is where problems are most likely to arise.
I-4	Ich muss dich schon mal vorwarnen, Legal Tech findet im Studium jetzt nicht so die große Rolle.	I have to warn you in advance, Legal Tech doesn't play a big role in my studies.
I-5	(...) sehr ernstzunehmende Herausforderung, der man aber begegnen kann.	(...) very serious challenge, but one that can be addressed.
I-5	Eigentlich gibt es für mich gar nicht Legal Tech. Es gibt nur Tech, die funktioniert und Tech, die nicht funktioniert.	Actually, for me, there is no such thing as Legal Tech. There is only technology that works and technology that doesn't work.

Code	Original Quote [German]	Translation [English]
I-5	Es werden halt nur diejenigen ersetzt, die sich dem Wandel quasi widersetzen, die sagen, dass sie bis auf alle Ewigkeit oldschool, traditionell, ohne den Einsatz von State of the Art Technologielösung arbeiten wollen.	(...) it will only replace those who resist the change, those that say that they want to work oldschool, traditional, without the use of state of the art technology solutions for all eternity.
I-5	Es muss natürlich auch einen sehr, sehr greifbaren und messbaren Mehrwert geben, den die Softwarelösung bietet.	The software solution must also provide a very tangible and measurable added value.
I-6	(...) es ist ja unethisch zu sagen, ich gebe einem Richterinnen und Richter ein Tool, das sie nicht verstehen und dann vielleicht am Ende einsetzen müssen.	(...) it is unethical to say that I provide a tool to judges that they do not understand and then they might have to use it in the.
I-6	(...) aber manchmal sind auch die Low-Hanging-Fruits nicht schlecht.	(...) but sometimes the low-hanging fruits are not bad either.
I-6	(...) wenn das ein elektronischer Vorschlag ist und die Parteien den gut finden und den annehmen, alles fair, aber nicht von einer staatlichen Justiz.	(...) if there is an electronic proposal and the parties like it and accept it, all fair enough, but not for the state judiciary system.
I-7	Ich finde, man arbeitet besser damit oder vielleicht (...) man kann erst sinnvoll mit der Technologie arbeiten, wenn man sie zumindest grundsätzlich versteht.	I think one works better with it or perhaps (...) already a step further, one can only work effectively with the technology when one understands it at least fundamentally.

Code	Original Quote [German]	Translation [English]
I-7	(...) darf man das überhaupt verwenden? Also erfüllt das irgendwelche gesetzlichen Rahmenbedingungen, dass es überhaupt erlaubt ist, sowas einzusetzen?	(...) is it even permitted to use it? Does it meet any legal requirements for its use to be allowed at all?
I-7	Wir haben im Moment noch nicht die rechtlichen Voraussetzungen, um Daten effektiv zu nutzen oder richtig gute Datensätze zu erstellen(...) Die benutzten Daten müssen ja unsere Werte vertreten und es muss ja unsere Bevölkerung dann auch repräsentieren.	Currently, we lack the legal framework, which is necessary to effectively utilize data or create high-quality datasets. (...) The data used must represent our values and our population.
I-7	Ich würde sagen, auf jeden Fall muss eine gute Bedienungsanleitung dabei sein.	I would say that a good user manual is definitely necessary.
I-8	(...) war man auch schon immer gegen Richtantennen nicht immun. Ja, also wenn mich jemand abhören wollte auch früher, dann ging das. (...) aber ob das jetzt wirklich faktisch mehr gemacht wird, weiß ich jetzt nicht.	(...) we were not immune to surveillance through directional antennas. If someone wanted to eavesdrop on me, they could do so even back then. (...) whether it is actually being done more frequently now, I don't know for sure.
I-8	Solche Transaktionsdokumente sind, umfangreiche Dokumente, 50, 60 Seiten lang. Und da sind wahnsinnig viele Anlagen enthalten. (...) Wir haben das jezt sofware basiert gelöst, dass einfach ein Algorithmus im Hintergrund die Dokumente permanent miteinander vergleicht.	These transaction documents are extensive documents, usually 50 to 60 pages long. And there are an insane number of attachments in there. (...) We have now solved this using software, where an algorithm in the background automatically compares the documents for us.

Code	Original Quote [German]	Translation [English]
I-8	(...) Teilautomatisierung des Beratungsprozesses (...) dass wenn wir bei der Befragung einfach notorisch geneigt sind, keine Checklisten zu verwenden, sondern das zu fragen, was uns gerade so einfällt. (...) Ich bin nicht mehr darauf zwingend angewiesen, dass ich zur gleichen Zeit, wie meine Mandanten, emotional breit, ja aber auch ganz schlicht zeitlich und örtlich in der Lage bin, so eine Befragung durchzuführen.	(...) partial automation of the consultation process. (...) When conducting interviews, we attorneys are often inclined to rely on our memory rather than using checklists. (...) I am no longer dependent on being emotionally, temporally, and physically available at the same time as my clients to conduct such an interview.
I-8	(...) das nimmt sehr, sehr viel Zeit aus dem Prozess und wenn man das händisch nachzieht, kommt man in die Hölle.	(...) it saves a lot of time in the process and manually doing that would be a nightmare.
I-8	Ich hätte ganz gerne, dass ich das alles auch auf meinem Handy benutzen kann.	I would like to have the ability to access everything on my mobile phone as well.
I-8	(...) wenn die das nicht beheben bis zum Ende des Jahres, dann werde ich sie ersetzen.	(...) software doesn't fix the issue by the end of the year, I will replace it.
I-9	Als Anwalt muss ich auch vielleicht verschleiern, wo ich selber Schwächen habe, oder mein Mandant, wo man sagt, hier wären wir angreifbar, da wollen wir dann mal nicht so drauf eingehen.	As a lawyer, I may need to conceal my own weaknesses or those of my client, avoiding certain areas where we could be vulnerable.

Code	Original Quote [German]	Translation [English]
I-9	(...) der gleiche Text des Gesetzes kann für unterschiedliche Folgen herangezogen werden, das ist auch das Gefährliche an Jura. (...) Jura ist halt auch instabil. Es liegt auch daran, dass Gesellschaften instabil sind.	(...) the same text of the law can be used for different consequences, that is the dangerous thing about law. (...) law is inherently unstable. This is also due to the fact that societies are unstable.
I-9	(...) einfach nur noch grässliches Zeug produziert hat. (...) mich wundert das nicht. Wir Menschen sind nun auch teilweise grausam und fürchterlich und insofern, wenn man an menschlichen Texten lernt oder an der menschlichen Natur, würde auch eine ganze Menge Furchtbares dabei sein.	(...) just produced horrible stuff. (...) not surprised by that. We humans are also cruel and horrible to some extent and if you learn from human texts or human nature, there would be a whole lot of horrible stuff.
I-9	(...) für uns Juristen reicht das immer, wenn es im Rechtssinn ist, auch wenn es nicht sicher ist.	(...) it is sufficient if it is legally correct, even if it is not completely secure.
I-9	Ich würde mir wünschen, es wäre nicht so. Das ist natürlich ein Problem. (...) die Rechte der Autoren werden ignoriert.	I wish it wasn't like this. It is certainly a problem. (...) the rights of the authors are ignored.
I-9	(...) inhaltlich sehr mächtig sind, technisch sehr schwach und völlig verhaltet.	(...) very powerful in terms of content, technically very weak and completely outdated.
I-9	Und da sind Anwälte sehr empfindlich. (...) Also es kommen Haftungsfehler nicht so häufig vor, aber wenn sie vorkommen, können sie auch existenziell werden.	And lawyers are very sensitive to this. (...) So liability errors do not occur so often, but when they do, they can also become existential.

Code	Original Quote [German]	Translation [English]
I-9	(...) eine Genauigkeit von weitaus mehr als 98% haben. Ansonsten würden die Leute sagen, das ist mir nicht genau genug.	(...) but it needs to have an accuracy of well over 98%. Otherwise, people would say it's not precise enough.
I-9	(...) wie der Taschenechner irgendwann mal kam, kann man den auch wegwünschen, aber der ist ja trotzdem da.	(...) when the pocket calculator came along, you could wish it away, but it's still there.
I-9	(...) wo Sie meinetwegen einen Schriftsatz hochladen und dann wird der überprüft (...) Ist das, was ich zitiere korrekt, besagt das dass so ungefähr?	(...) you upload a brief and then it gets checked (...) Does what I cite accurately reflect the content?
I-10	(...) hierfür gibt es Vorlagen. Beim normalen Hauskauf sind es immer dieselben Bausteine. Die sind schon in der Software drin.	(...) 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.
I-10	(...) aber es muss jetzt nicht irgendwie Raketenwissenschaft sein. Ganz pragmatisch, es muss einfach leicht zu bedienen sein.	(...) it doesn't need to be rocket science. Essentially, it should be user-friendly and easy to operate.
I-11	(...) weil hier keine Daten rausgehen dürfen. Das ist schon verboten im Grunde (...) Datenabfluss aus meinem Büro, den ich nicht steuern kann und den ich nicht verhindern kann.	(...) basically prohibited, because no data should leave the premises. (...) data floating out of my office that I can't control and that I can't prevent.
I-11	(...) irgendeine Art von Kontrolle oder Kontrollmechanismus, wie sehr die Software jetzt geraten hat.	(...) some form of control or control-mechanism to assess how much the software is guessing.

Code	Original Quote [German]	Translation [English]
I-11	Würden Sie einer Maschine erzählen, dass, wenn Sie verheiratet sind, dass Sie noch ein vor-eheliches Kind haben, von dem der Ehepartner nichts weiß?	Would you tell a machine that when you are married, that you still have a pre-marital child that the spouse doesn't know about?
I-11	Und dann kann ich mich da wieder durchklicken, 20 Minuten lang. Und schöner wäre, wenn ich das dann gleich alles zusammengefasst als Volltext präsentiert bekomme.	And then I have to click through it again for about 20 minutes. It would be nicer if I could get it all presented as a summarized text right away.
I-11	(...) wenn es eine Sprachsteuerung hätte, wäre das wesentlich schöner.	(...) if it had voice control, it would be much better.
I-11	Letztlich bräuchte ich eine qualifizierte oder belastbare (...) Sicherheit, dass dieses Programm jetzt den rechtlichen Rahmen einhält, den ich selber einhalten muss.	Ultimately, I would need a qualified or reliable (...) assurance that this program complies with the legal framework that I myself must adhere to.
I-12	Wenn man das hinkriegen würde, wirklich großflächig alle Urteile zu veröffentlichen, das fände ich schon toll. (...) Das sind Sachen, was die Gesamtbranche da auch voranbringt. (...) das muss sich ändern.	(...) if we could manage to widely publish all court decisions, that would be great. (...) These are things that would bring the entire industry forward. (...) that has to change.
I-13	(...) gerade auch für die normale Bevölkerung. (...) Wie soll man da jemals wirklich durchsteigen?	(...) especially for the normal population. (...) How are you ever really supposed to get that?
I-13	Ich glaube nicht, dass die Maschine dann jemals den Menschen ersetzen wird. Ich fühle mich da in keinster Weise bedroht in meiner Existenz.	I don't believe that machines will ever replace humans. I don't feel threatened in any way regarding my existence.

Code	Original Quote [German]	Translation [English]
I-13	In Massenverfahren wird auch immer mehr Legal Tech genutzt.	Legal Tech is being used more and more in mass-proceedings.
I-13	Aber wer sitzen in jeder öffentlichen Verhandlung. Das macht ja niemand. (...) Akteneinsicht kriegt auch nicht jeder.	Of course, this is a public hearing. But who sits in every public hearing? Nobody does that. (...) not everyone is granted access to review the case files.
I-14	Der Bias, der würde mir da irgendwie Sorgen machen, insbesondere wenn es dann dazu kommt, dass das [Legal Tech Lösung] in gerichtlichen Verfahren, wo es bei gerichtlichen Entscheidungen dann eine Rolle spielen kann.	I would say that bias does concern me, especially when it comes to integrating [Legal Tech solution] into administrative and potentially even judicial proceedings, where it could play a role in court decisions.
I-14	(...) natürlich ein bisschen die Bedenken um die eigene wirtschaftliche Existenz.	(...) of course a bit of concern about one's own economic existence.
I-14	(...) langfristig Risikoprofile zu entwickeln, dass das dann im zweiten Schritt nochmal von einem Anwalt irgendwie eingeschätzt wird, wie man das im Kaufpreis einpreisen kann.	(...) 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.
I-15	(...) risiko-averse. Das bedeutet, die Requirements an Präzision sind sehr hoch. Ungewöhnlich hoch für künstliches Intelligenzsystem vor allem.	(...) risk-averse, which means their requirements for precision are very high. Exceptionally high, when it comes to artificial intelligence systems.
I-17	(...) Computer meldet dem Richter dann quasi nur, wenn sich jemand nicht an Bewährungsauflagen hält.	(...) computer system only notifies the judge when someone fails to comply with their probation conditions.
I-17	Das nehmen ältere Kollegen sehr, sehr regelmäßig in Beschlag.	Older colleagues use this very, very regularly.

Code	Original Quote [German]	Translation [English]
I-17	Es gibt ja auch keinen Ort, wo mehr gelogen wird als vor Gericht (...) Da wäre man manchmal schon dankbar, wenn man da so eine Assistenz hätten, die man fragen könnte, ob der Zeuge lügt oder nicht.	There is no place where more lies are told than in court (...) Sometimes you would be grateful if you had some kind of assistance, that you could ask whether the witness is lying or not.
I-17	Da geht schon arg viel Arbeitszeit verloren, das muss man einfach sagen.	This result in a significant loss of working time, that's for sure.
I-18	Das könnten Sie quasi, wenn das nicht alles standardisiert wäre, nicht anders wegarbeiten. (...) Es ist wirklich so, Sie verbringen 90% Ihres Tages damit, dass Sie Daten in so Textmasken eingeben. (...) inhaltlich nicht besonders spannend.	If it weren't all standardized, you couldn't handle it (...) It's really like this: you spend 90% of your day entering data into text fields. (...) particularly exciting in terms of content.
I-18	Das wäre ja furchtbar. (...) Wäre ja schlecht, wenn man jahrelang studiert hat und dann kommt da auf einmal das Rechenmodell und sagt, das ist jetzt aber falsch, was du hier machst. Kann ja sogar sein, aber das möchte ich dann trotzdem nicht hören.	That would be terrible. (...) It would be bad if, after studying for years, a computational model suddenly comes along and says, "What you're doing here is wrong." It might even be true, but I still wouldn't want to hear it.
I-18	Dann stellt man fest: das hat sich jemand ausgedacht.	Then you realize: someone made this up.
I-18	Weil ich will vielleicht nicht schreiben "Dieser Sachverhalt steht nach der durchgeführten Hauptbeamtung fest." Vielleicht finde ich diesen Satz blöd und dann möchte ich das anders.	Because I might not want to write "This fact is established after the conducted main examination." Maybe I don't like that sentence, and I want to phrase it differently.

Table A.1.: Quotes translated from German to English

Lehrstuhl für Software Engineering für betriebliche Informationssysteme (sebis)
TUM School of Computation, Information and Technology
Technische Universität München

Interviewleitfaden

Vielen Dank, dass Sie mit diesem Interview meine Masterarbeit unterstützen!

Im Folgenden ist ein grober Ablaufplan für das Interview dargestellt. Die genannten Fragen gelten als Orientierung – ggf. werden nicht alle gestellt/auch andere Fragen gestellt. Sie müssen sich nicht explizit auf das Interview vorbereiten.

Einstieg	• **Formalitäten**: Das Interview wird aufgezeichnet und anschließend transkribiert. Die Transkription selbst und alle darin enthaltenen Ergebnisse werden zu Forschungszwecken und für eine eventuelle Veröffentlichung in einer Forschungsarbeit und/oder Dissertation verwendet. Alle persönlich identifizierbaren Informationen werden anonymisiert. • Kurze Vorstellung • Kurzer Umriss des Forschungskontexts, im Zuge dessen das Interview stattfindet
Interview	• Was ist Ihr Verständnis von Legal Tech? • Nutzen Sie digitale Hilfen bei der Erledigung Ihrer täglichen Aufgaben? Welche? • Welche repetitiven Aufgaben gehören zu Ihrem Arbeitsalltag? • Wo sehen sie zukünftige Anwendungsfälle von Legal Tech? • Welche Medien nutzen Sie, um sich über Neuigkeiten im juristischen Sektor zu informieren? • Wohin wird sich Legal Tech Ihrer Meinung nach in den nächsten Jahren entwickeln? • Haben sie ethische Bedenken hinsichtlich des Einsatzes von Technologie im juristischen Sektor? Welche?
Ausblick	• Wenn wir in unserer Forschung einen Schritt weiter sind – können meine Kollegen Sie nochmals kontaktieren? • Kennen Sie weitere Personen aus dem juristischen Kontext, mit denen ich noch sprechen könnte? • Vielen Dank!

Figure A.1.: Interview guide in original language, German

List of Figures

List of Tables

Acronyms

AI Artificial Intelligence. 1–3, 9, 42, 43, 45, 46, 55, 65, 76

ELSA Ethical, Legal and Social Aspects. iv, 1–3, 6, 7, 9, 10, 13, 15, 23, 29–32, 39–41, 44, 46, 75, 76, 78, 80, 81, 83, 96–98

ELSI Ethical, Legal and Social Implications. 6, 9

GDPR General Data Protection Regulation. 34

NER Named Entity Recognition. 27, 28

NLP Natural Language Processing. iv, 1–5, 7–10, 12, 13, 15, 23–25, 27–29, 32, 38, 39, 70, 72, 73, 76, 78–83, 96–98, 101

RQ1 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?. 10, 80

RQ2 What are the use cases in which the identified NLP techniques can be utilized?. 10, 13, 81

RQ3 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?. 10, 81

SEBIS Software Engineering for Buisiness Information Systems. iii, 16

SLR Systematic Literature Review. iv, 2, 8, 10–13, 16, 23, 24, 30–38, 41, 44, 46, 49–51, 54, 55, 59, 70–76, 78, 80, 81, 83, 96–98

SSI Semi-Structured Interview. iv, 2, 8, 10, 15, 18, 23, 39–41, 44, 46, 49–60, 64, 70–76, 78–84, 96, 98

Bibliography

[1] M. M. Alhaddad. "Artificial Intelligence in Banking Industry: A Review on Fraud Detection, Credit Management, and Document Processing". In: *ResearchBerg Review of Science and Technology* 2.3 (2018), pp. 25–46.

[2] I. Castiglioni, L. Rundo, M. Codari, G. Di Leo, C. Salvatore, M. Interlenghi, F. Gallivanone, A. Cozzi, N. C. D'Amico, and F. Sardanelli. "AI applications to medical images: From machine learning to deep learning". In: *Physica Medica* 83 (2021), pp. 9–24.

[3] R. Zimmermann, D. Mora, D. Cirqueira, M. Helfert, M. Bezbradica, D. Werth, W. J. Weitzl, R. Riedl, and A. Auinger. "Enhancing brick-and-mortar store shopping experience with an augmented reality shopping assistant application using personalized recommendations and explainable artificial intelligence". In: *Journal of Research in Interactive Marketing* 17.2 (2023), pp. 273–298.

[4] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, et al. "Language models are few-shot learners". In: *Advances in neural information processing systems* 33 (2020), pp. 1877–1901.

[5] J. M. Gómez-Pérez. "Towards Language-centric Scientific AI". In: *arXiv preprint arXiv:2210.15327* (2022).

[6] *NLawP - Natural Language Processing and Legal Tech*. URL: `https://wwwmatthes.in.tum.de/pages/ztm206o67g3q/NLawP-Natural-Language-Processing-and-Legal-Tech` (visited on 06/04/2023).

[7] H. M. Hapke, H. Lane, and C. Howard. *Natural language processing in action*. 2019.

[8] J. Eisenstein. *Introduction to natural language processing*. MIT press, 2019.

[9] J. L. Leidner and V. Plachouras. "Ethical by design: Ethics best practices for natural language processing". In: *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing*. 2017, pp. 30–40.

[10] E. M. Bender, D. Hovy, and A. Schofield. "Integrating ethics into the NLP curriculum". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts*. 2020, pp. 6–9.

[11] L. Benotti, K. Fort, M.-Y. Kan, and Y. Tsvetkov. "Understanding Ethics in NLP Authoring and Reviewing". In: *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics: Tutorial Abstracts*. 2023, pp. 19–24.

[12] *Google Trends analysis to visualize the predominant search terms regardgin Legal Technology*. 2023. URL: `https://trends.google.com/` (visited on 05/10/2023).

[13] 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.

[14] J. Wagner. *Legal Tech und Legal Robots: Der Wandel im Rechtswesen durch neue Technologien und Künstliche Intelligenz*. Springer, 2020.

[15] S. Vijayavenkataraman, W. Lu, and J. Y. H. Fuh. "3D bioprinting–an ethical, legal and social aspects (ELSA) framework". In: *Bioprinting* 1 (2016), pp. 11–21.

[16] A. Hullmann. "European activities in the field of ethical, legal and social aspects (ELSA) and governance of nanotechnology". In: *DG Research, Brussels: European Commission* (2008).

[17] B. Gransche and A. Manzeschke. *Das geteilte Ganze: Horizonte Integrierter Forschung für künftige Mensch-Technik-Verhältnisse*. Springer-Verlag, 2020.

[18] *Google Scholar search to see the amount of publications intersecting the NLP and legal domain*. 1993-2022. URL: `https://scholar.google.de/` (visited on 06/10/2023).

[19] J. Frankenreiter and J. Nyarko. "Natural language processing in legal tech". In: *Legal Tech and the Future of Civil Justice (David Engstrom ed.)* (2022).

[20] H. Zhong, C. Xiao, C. Tu, T. Zhang, Z. Liu, and M. Sun. "How does NLP benefit legal system: A summary of legal artificial intelligence". In: *arXiv preprint arXiv:2004.12158* (2020).

[21] M. Masala, R. C. A. Iacob, A. S. Uban, M. Cidota, H. Velicu, T. Rebedea, and M. Popescu. "jurBERT: A Romanian BERT model for legal judgement prediction". In: *Proceedings of the Natural Legal Language Processing Workshop 2021*. 2021, pp. 86–94.

[22] J. Niklaus, I. Chalkidis, and M. Stürmer. "Swiss-judgment-prediction: A multilingual legal judgment prediction benchmark". In: *arXiv preprint arXiv:2110.00806* (2021).

[23] G. Sukanya and J. Priyadarshini. "A meta analysis of attention models on legal judgment prediction system". In: *International Journal of Advanced Computer Science and Applications* 12.2 (2021).

[24] V. G. Pillai and L. R. Chandran. "Verdict Prediction for Indian Courts Using Bag of Words and Convolutional Neural Network". In: *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*. IEEE. 2020, pp. 676–683.

[25] D. Trautmann, A. Petrova, and F. Schilder. "Legal Prompt Engineering for Multilingual Legal Judgement Prediction". In: *arXiv preprint arXiv:2212.02199* (2022).

[26] A. Shukla, P. Bhattacharya, S. Poddar, R. Mukherjee, K. Ghosh, P. Goyal, and S. Ghosh. "Legal case document summarization: Extractive and abstractive methods and their evaluation". In: *arXiv preprint arXiv:2210.07544* (2022).

[27] K. Merchant and Y. Pande. "NLP Based Latent Semantic Analysis for Legal Text Summarization". In: *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. 2018, pp. 1803–1807. DOI: 10.1109/ICACCI.2018.8554831.

[28] V. Pandya. "Automatic text summarization of legal cases: A hybrid approach". In: *arXiv preprint arXiv:1908.09119* (2019).

[29] R. Dale. "Law and word order: NLP in legal tech". In: *Natural Language Engineering* 25.1 (2019), pp. 211–217.

[30] F. Berger, S. Gevers, L. Siep, and K.-M. Weltring. "Ethical, legal and social aspects of brain-implants using nano-scale materials and techniques". In: *NanoEthics* 2 (2008), pp. 241–249.

[31] Y. Ikkatai, T. Hartwig, N. Takanashi, and H. M. Yokoyama. "Segmentation of ethics, legal, and social issues (ELSI) related to AI in Japan, the United States, and Germany". In: *AI and Ethics* (2022), pp. 1–17.

[32] A. Kapeller, H. Felzmann, E. Fosch-Villaronga, K. Nizamis, and A.-M. Hughes. "Implementing ethical, legal, and societal considerations in wearable robot design". In: *Applied Sciences* 11.15 (2021), p. 6705.

[33] B. A. Kitchenham, D. Budgen, and P. Brereton. *Evidence-based software engineering and systematic reviews*. Vol. 4. CRC press, 2015.

[34] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén. *Experimentation in software engineering*. Springer Science & Business Media, 2012.

[35] S. C. B. A. Kitchenham. *Guidelines for performing systematic literature reviews in software engineering*. 2007.

[36] H. Zhang, M. A. Babar, and P. Tell. "Identifying relevant studies in software engineering". In: *Information and Software Technology* 53.6 (2011), pp. 625–637.

[37] P. Gill, K. Stewart, E. Treasure, and B. Chadwick. "Methods of data collection in qualitative research: interviews and focus groups". In: *British dental journal* 204.6 (2008), pp. 291–295.

[38] H. Kallio, A.-M. Pietilä, M. Johnson, and M. Kangasniemi. "Systematic methodological review: developing a framework for a qualitative semi-structured interview guide". In: *Journal of advanced nursing* 72.12 (2016), pp. 2954–2965.

[39] J. Horton, R. Macve, and G. Struyven. "Qualitative research: experiences in using semi-structured interviews". In: *The real life guide to accounting research*. Elsevier, 2004, pp. 339–357.

[40] M. J. McIntosh and J. M. Morse. "Situating and constructing diversity in semi-structured interviews". In: *Global qualitative nursing research* 2 (2015), p. 2333393615597674.

[41] V. Braun and V. Clarke. "Using thematic analysis in psychology". In: *Qualitative research in psychology* 3.2 (2006), pp. 77–101.

[42] E. Union. *Commission Recommendation of 6 May 2003 concerning the definition of micro, small and medium-sized enterprises (notified under document number C(2003) 1422)*. May 2003. URL: https://eur-lex.europa.eu/eli/reco/2003/361/oj/eng (visited on 05/25/2023).

[43] S. Kübler, R. McDonald, and J. Nivre. "Dependency parsing". In: *Synthesis lectures on human language technologies* 1.1 (2009), pp. 1–127.

[44] S. Singkul and K. Woraratpanya. "Thai dependency parsing with character embedding". In: *2019 11th International Conference on Information Technology and Electrical Engineering (ICITEE)*. IEEE. 2019, pp. 1–5.

[45] N. Gahman and V. Elangovan. "A Comparison of Document Similarity Algorithms". In: *arXiv preprint arXiv:2304.01330* (2023).

[46] T. Lalwani, S. Bhalotia, A. Pal, V. Rathod, and S. Bisen. "Implementation of a Chatbot System using AI and NLP". In: *International Journal of Innovative Research in Computer Science & Technology (IJIRCST) Volume-6, Issue-3* (2018).

[47] M. L. Zeng. "Knowledge organization systems (KOS)". In: *KO Knowledge Organization* 35.2-3 (2008), pp. 160–182.

[48] S. Gupta, S. Malik, L. Pollock, and K. Vijay-Shanker. "Part-of-speech tagging of program identifiers for improved text-based software engineering tools". In: *2013 21st International Conference on Program Comprehension (ICPC)*. IEEE. 2013, pp. 3–12.

[49] M. A. C. Soares and F. S. Parreiras. "A literature review on question answering techniques, paradigms and systems". In: *Journal of King Saud University-Computer and Information Sciences* 32.6 (2020), pp. 635–646.

[50] M. K. Dalal and M. A. Zaveri. "Automatic text classification: a technical review". In: *International Journal of Computer Applications* 28.2 (2011), pp. 37–40.

[51] W. Shen, J. Wang, P. Luo, and M. Wang. "Linden: linking named entities with knowledge base via semantic knowledge". In: *Proceedings of the 21st international conference on World Wide Web*. 2012, pp. 449–458.

[52] S. Beliga. "Keyword extraction: a review of methods and approaches". In: *University of Rijeka, Department of Informatics, Rijeka* 1.9 (2014).

[53] B. Ahmed. "Lexical normalisation of twitter data". In: *2015 Science and Information Conference (SAI)*. IEEE. 2015, pp. 326–328.

[54] J. Li, A. Sun, J. Han, and C. Li. "A survey on deep learning for named entity recognition". In: *IEEE Transactions on Knowledge and Data Engineering* 34.1 (2020), pp. 50–70.

[55] B. Min, H. Ross, E. Sulem, A. P. B. Veyseh, T. H. Nguyen, O. Sainz, E. Agirre, I. Heinz, and D. Roth. "Recent advances in natural language processing via large pre-trained language models: A survey". In: *arXiv preprint arXiv:2111.01243* (2021).

[56] J. Hirschle. *Deep Natural Language Processing: Einstieg in Word Embedding, Sequence-to-Sequence-Modelle und Transformer mit Python*. Carl Hanser Verlag GmbH Co KG, 2022.

[57] G. Eysenbach et al. "The role of ChatGPT, generative language models, and artificial intelligence in medical education: a conversation with ChatGPT and a call for papers". In: *JMIR Medical Education* 9.1 (2023), e46885.

[58] T. C. Li. "Privacy in pandemic: law, technology, and public health in the COVID-19 crisis". In: *Loy. U. Chi. LJ* 52 (2020), p. 767.

[59] L. Fan, L. Li, Z. Ma, S. Lee, H. Yu, and L. Hemphill. "A bibliometric review of large language models research from 2017 to 2023". In: *arXiv preprint arXiv:2304.02020* (2023).