LegalWriter: An Intelligent Writing Support System for Structured and Persuasive Legal Case Writing for Novice Law Students

Florian Weber
weber@uni-kassel.de
University of Kassel
Kassel, Germany

Thiemo Wambsganss
thiemo.wambsganss@bfh.ch
Bern University of Applied Sciences
Bern, Switzerland

Seyed Parsa Neshaei
seyed.neshaei@epfl.ch
EPFL
Lausanne, Switzerland

Matthias Söllner
soellner@uni-kassel.de
University of Kassel
Kassel, Germany

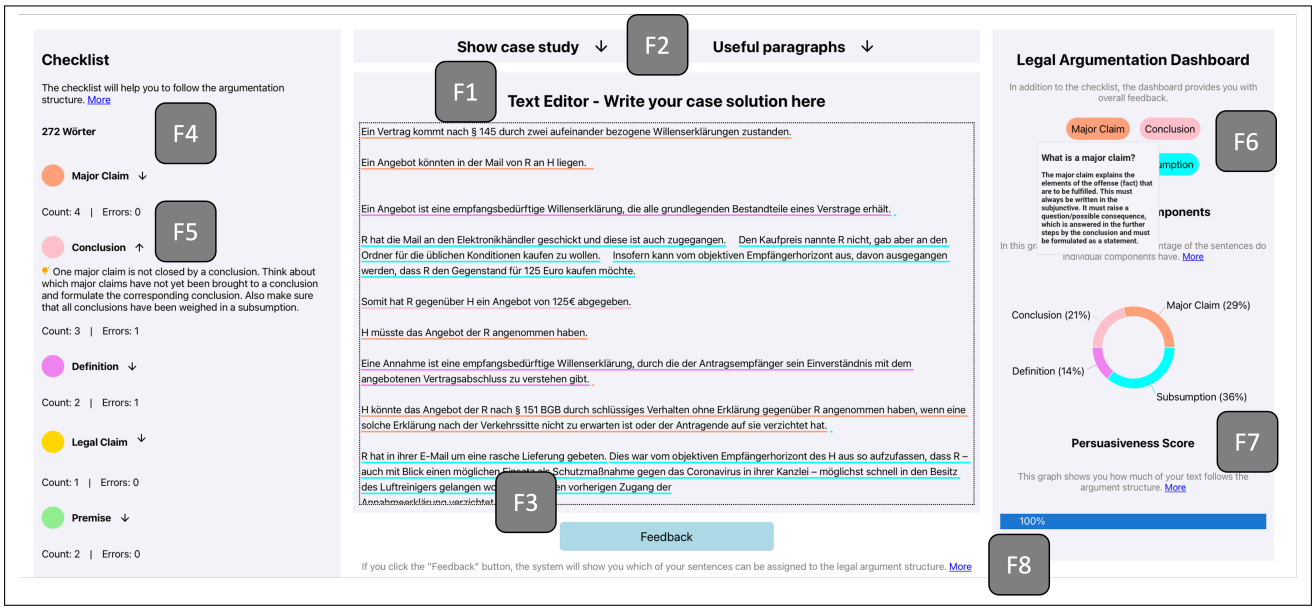

Figure 1: Screenshot of our intelligent writing support system: students write case solutions to legal problems and receive intelligent writing support based on three trained transformer models. The error-based learning helps students improve their skills in writing persuasive and structured case solutions in the appraisal style.

ABSTRACT

Novice students in law courses or students who encounter legal education face the challenge of acquiring specialized and highly concept-oriented knowledge. Structured and persuasive writing combined with the necessary domain knowledge is challenging for many learners. Recent advances in machine learning (ML) have shown the potential to support learners in complex writing tasks. To test the effects of ML-based support on students' legal writing skills, we developed the intelligent writing support system *LegalWriter*. We evaluated the system's effectiveness with 62 students. We showed that students who received intelligent writing support based on their errors wrote more structured and persuasive case solutions with a better quality of legal writing than the current benchmark. At the same time, our results demonstrated the positive effects on the students' writing processes.

CHI '24, May 11–16, 2024, Honolulu, HI, USA

ACM ISBN 979-8-4007-0330-0/24/05
https://doi.org/10.1145/3613904.3642743

CCS CONCEPTS

• **Applied computing** → *Law*; • **Computing methodologies** → **Machine learning**.

KEYWORDS

Writing support systems, Learning systems, Adaptive learning, Learning from errors, Legal education

ACM Reference Format:

Florian Weber, Thiemo Wambsganss, Seyed Parsa Neshaei, and Matthias Söllner. 2024. *LegalWriter*: An Intelligent Writing Support System for Structured and Persuasive Legal Case Writing for Novice Law Students. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI, USA*. ACM, New York, NY, USA, 23 pages. https://doi.org/10.1145/3613904.3642743

1 INTRODUCTION

Within the intricate landscape of our legal system, the imperative to grasp and agree on a cohesive legal interpretation is crucial. This foundational process depends on the adeptness of judges and lawyers to effectively communicate their decisions. Whether articulated through the spoken word in courtrooms or meticulously penned in legal opinions, briefs, or judgments, the art of persuasive communication within the legal realm assumes unparalleled significance. Specifically, legal opinion writing plays a pivotal role in the skill set of legal professionals, serving as the foundation upon which the entire structure of legal dispute resolution relies. As such, how legal experts articulate and defend their perspectives takes center stage, manifesting as a dynamic interplay of diverse approaches. While in America, the "*IRAC formula*"[1] is a reference to teach legal opinion writing to students [48], countries like China and Germany use the "*appraisal style*" [44], whereas, for instance, France uses the "*cas pratique*" [5].

Learning to write legal opinions poses significant challenges not only for novices in law school but also for students from other disciplines who must learn the fundamentals of law, such as business students. As part of their training, students are typically challenged to solve legal problems or case studies as persuasive and structured case solutions [5, 21]. Out of the initial group of first-semester law students at a prestigious university in Germany, 70% reported difficulties when tackling a legal case solution and grasping legal terminology [64]. To write a persuasive and structured legal case solution, a student must follow the structural requirements of the style and justify one's derived results argumentatively via legal claims and premises (see Section 2.1). Researchers, particularly in educational technology, have developed systems to support students with persuasive and structured writing [50, 76, 78]. However, these systems are of limited interest to law students because the writing and argumentation style in law differs from the general argumentation style due to domain-specific nuances. Legal argumentation adheres to the appraisal style [67]. At the same time, it is essential to argue based on the facts that arise from the legal problem. Both requirements, the clearly defined structure and the closeness to the facts of the legal issue, are peculiar to the argumentation in jurisprudence and differ from the generally accepted requirements of a qualitative argumentation (e.g., [69]).

Researchers and educators claim that the targeted use of IT solutions in law education falls short of expectations [8]. The literature concerning legal writing in law courses supports this assertion. Only a few systems, such as CATO [2] [4] and a template-based argument mediation system (ArguMed) [73] are presented. The existing systems train law students' argumentation and writing skills through examples [4] or by presenting exemplary valid arguments [73]. However, none of the currently available systems assist students in writing persuasive and structured case solutions. One way to support students in writing persuasive and structured case solutions in appraisal style is to develop an intelligent writing support system based on machine learning (ML) models. For many tasks, such intelligent writing support systems are already used successfully [85], such as improving students' grammar [33], writing more emphatic peer reviews [81], or supporting creative story writing [14]. Writing support systems are also becoming increasingly important for tasks like argumentation due to continual advancements in ML and natural language processing (NLP) [75]. Researchers in NLP and ML have already developed algorithms for identifying argumentation structures in unstructured texts and providing feedback for improvement [41, 77]. Even though there are several algorithms for analyzing text in the legal field for classifying judgments [70], summarizing legal texts [29], and assessing jury verdicts [56], among many others, only a few algorithms specifically help students write persuasive and structured case solutions [83]. Recent research has also shown that large language models (LLMs) have performed remarkably in many NLP tasks, including commonsense reasoning [35], text comprehension [68], and text translation [57]. However, the current LLMs still fail at several procedures (e.g., [25]) in the legal domain. For example, LLMs still fail to summarize court decisions [17] exactly. Furthermore, Choi et al. [13] show that ChatGPT, a conversational user interface for the GPT-3.5 LLM from OpenAI, consistently exhibited errors when composing legal texts. According to them, ChatGPT's legal writing tends to be slightly lower quality than the legal writing produced by the average law student, based on grade comparisons [13]. LLMs have not demonstrated impressive performance overall in legal education and may benefit from a more rigorous evaluation process when handling legal texts.

Given the potential of applying NLP and ML for identifying structural components in texts, we designed and built *LegalWriter*. *LegalWriter* is an ML-based writing support system that provides students with intelligent support for writing persuasive and structured case solutions. Our objective was to investigate whether a system that provides students with individual support based on their writing errors helps them write more persuasive and structured case solutions in the legal domain. To implement the system, we followed two development approaches to create a user-centered design for *LegalWriter* for learners in law courses. First, we took a theory-driven approach and systematically reviewed the literature on educational technology and pedagogical theories to derive requirements for an initial design of *LegalWriter* [19, 74]. Second, we followed a user-centered design approach where we interviewed ten law students from several universities to derive user requirements for an adaptive learning system for law education. To create an advanced machine learning system, we underwent a process of training and refining three Transformer-based BERT models. We used these models to assess the conformity to appraisal styles and

[1] IRAC is an acronym that stands for issue, rule, application, and conclusion. It serves as a methodology for solving and analyzing legal problems.

[2] CATO is a learning system for case-based argumentation tasks.

the utilization of arguments, along with their interconnectedness within case solutions. We conducted this assessment using the corpus of student-written legal case solutions by Weber et al. [83]. These models served as *LegalWriter*'s underlying feedback algorithm. Based on the feedback algorithm, *LegalWriter* offers users individual writing support based on natural errors[3], providing an adaptive learning dashboard for producing persuasive and structured case solutions and directing students toward more persuasive and structured case solutions through personalized recommendations (see Figures 3 and 1).

To assess the impact of *LegalWriter* on students' ability to develop persuasive and structured case solutions in the appraisal style, we raised and answered two research questions:

RQ1: *How effective is an intelligent version of LegalWriter in supporting novice students in writing persuasive and structured case studies compared to a non-ML-based version of LegalWriter?*

RQ2: *How do novice students perceive the user experience and the learning process during the interaction with an intelligent version of LegalWriter compared to a non-ML-based version of LegalWriter?*

To answer our research questions, we assessed the impact of *LegalWriter* on students' ability to develop persuasive and structured case solutions in the appraisal style. We evaluated our learning system in a scenario where students wrote case solutions to a legal problem. We compared *LegalWriter* to a benchmark system that reflects state-of-the-art legal writing instructions and standard IT support [4]. In an online experiment with 62 students, we observed that the participating students using the intelligent version of *LegalWriter* wrote their legal opinions with a higher quality of legal argumentation and a more accurate structure than the students in the control group. Furthermore, the results indicated a positive user interaction measure in the perceived *feedback accuracy* of the system and *enjoyment*, as well as the *intrinsic motivation* of students using *LegalWriter* in the treatment group compared to students using *LegalWriter* in the control group.

Our work makes five significant contributions. First, we develop a conceptualization for ML-based writing support for law. Second, we provide new design insights for the HCI community, legal tech, and computer-assisted learning in law. Third, we contribute to pedagogy by showing that incorporating learning from errors in conjunction with ML-based feedback is effective [46]. Fourth, we demonstrate the effectiveness of ML-based feedback from an empirical and qualitative side by conducting a field experiment. Fifth, we contribute to legal education and other fields by showing that ML-based learning systems can be trained at scale for typical pedagogical scenarios.

2 RELATED WORK AND CONCEPTUAL BACKGROUND

2.1 Legal Appraisal Style

Legal education has specialized characteristics, and students must process the specialized knowledge of jurisprudence and concept-driven knowledge [21]. Conceptual knowledge includes various argumentation schemes or particular styles to write a case solution. In the English-speaking areas, the thinking and writing styles include the IRAC formula [48] and the HIRAC formula[4] [31]. Among the most critical concepts of German jurisprudence are the appraisal and judgment styles [67, 70]. Since the term appraisal style is peculiar to the German legal language, there is no direct equivalent in English. The appraisal style is *"the form and writing style of a legal opinion"* [67]. The appraisal style helps solve complex legal problems and always starts with a question, the so-called major claim [67]. Thus, the clarity of the legal issue is the most significant difference between the two styles. Complex problems use the appraisal style, while unambiguous problems use the judgment style. Students must solve these legal problems or case studies as persuasive and structured case solutions [21]. In a case solution, the correct application of the appraisal style is the basis for writing a structured and persuasive legal opinion. The appraisal style, unlike the judgment style, consists of four components: *major claim*, *definition*, *subsumption*, and *conclusion* [60, 82]. Table 1 briefly explains the four elements of the appraisal style. Table 13 in the Appendix provides more examples of each component.

2.2 Writing Support and Learning Systems for Legal Writing

Educational institutions like universities face the challenge of teaching structural writing. This fact is partially due to external pressures to complete the core curriculum, so there are limited opportunities to practice persuasive and structured writing [32]. This lack is true even for topics where the curriculum mandates persuasive and structured writing, like law or logic, where time and availability constraints limit the teachers' ability to teach persuasive and structured writing. As a result, researchers and educators are calling on the education system to increase the role of developing persuasive and structured writing and argumentation [20]. Consequently, research groups have developed systems to support students in persuasive and structured writing. These systems have appeared in various fields, such as science [50], conversational argumentation [16, 80], business reporting [78], and online debates [89]. However, researchers and legal educators note that the use of IT systems in legal education falls short of expectations [8]. Nevertheless, some systems help students learn persuasive and structured legal writing. Most of these systems employ methods of argument diagramming (representational guidance approach). It provides students with representations of their reasoning structures to support their reasoning. A typical example is helping students represent their reasoning structure with node and link graphs [53, 58]. Pioneering work in the legal field has shown that argument diagramming can improve students' ability to make high-quality arguments and the coherence of law students' persuasive and structured writing [11, 27, 61, 62]. Pinkwart et al. [53] have developed the LARGO system (Legal Argument Graph Observer), which allows law students to display examples of legal interpretations with hypothetical arguments graphically. Besides the diagram argumentation systems, there are a few other systems, such as CATO[5] [6]. This system assists students in argumentation with cases by teaching them to

[3]Natural errors occur during the application of complex skills without being prevented or promoted [88].

[4]HIRAC is an acronym for heading, issue, rule, application, and conclusion. It serves as a methodology for solving and analyzing legal problems.

[5]CATO is a learning system for case-based argumentation tasks.

Table 1: Essential elements of a legal opinion in appraisal style [44, 67, 83]. Explanation of the various elements and an example from a student-written case solution. To facilitate communication, we translated the case solution from German into English.

Components	Example
Major Claim: The major claim explains the elements of the offense that are to be fulfilled. It raises a question or possible consequence. The question is answered in the further steps and concludes in the final step.	*H could be entitled to payment for the air purifier if a purchase agreement between H and R was valid.*
Definition: The definition determines the constituent elements that must occur in the legal problem so that the case solution can conclude. The elements always depend on the question raised in the major claim.	*An offer could lie in the mail from R to H. An offer is a declaration of intent that must be received and contains all the elements of a contract.*
Subsumption (*premise* and *legal claim*): The subsumption examines the extent of the conditions (elements) of the definition. It weighs the facts of the case against the preconditions from the definitions and the premises. Legal consequences are drawn from the premises, so-called legal claims.	*R has sent the mail to the electronics retailer, who has received it. R did not state the purchase price but said he wanted to buy the folder for the usual conditions. In this respect, it can be assumed from the objective recipient's horizon that R would like to purchase the item for $125.*
Conclusion: The conclusion is the answer to the major claim. Thus, the case solution reaches a final result here.	*Thus, R has made an offer of $125 to H.*

compare their arguments with given cases and offer existing arguments to improve their own solution (discussion scripting approach) [4, 6].

2.3 Individual Error-Based Learning

We rely on the literature on error-based learning to guide the design and development of our intelligent writing support system [23, 47, 88]. This foundational concept facilitates enhancing legal writing skills among novice law students [22, 49]. Practice and application play a crucial role in learning and acquiring skills [22]. Typically, law students need to solve various cases to internalize legal argumentation and appraisal style. In lectures, students learn the basic skills for solving cases and get examples from the lecturer. Due to the time commitment and the abundance of content that they must master, students rarely receive individual feedback on their cases from lecturers. Research shows that practicing a skill through repeated attempts improves it and eventually leads to mastery [49]. However, errors are bound to occur during these repeated exercises, especially for less experienced novice students. Current research shows how errors followed by corrective feedback can support learning [46, 88]. Wong and Lim [88] distinguished between the approaches of prevention, permission, and promotion of errors. The error-allowance approach permits learners to make mistakes naturally, which corrective feedback can improve [43, 55]. Corrective feedback is particularly efficient when given promptly and based on individual errors [23, 30]. For students in law courses and other students interested in writing persuasive and structured case solutions, we would like to create an environment that allows them to improve their skills to solve legal problems and write persuasive and structured case solutions. In this learning environment, we permit natural mistakes from which students can learn [47]. Error learning theory, or allowing errors to occur, assumes that errors have an activating effect and endow an alternative path to reaching the correct solution [37]. In addition, errors generate increased attention, enhancing the effectiveness of subsequent corrective feedback. Also, learners pay more attention to corrective feedback after making an error since they are curious to get the correct answer [55].

3 DESIGN OF *LEGALWRITER*

To investigate how an intelligent writing support system can influence persuasive and structured case solution writing, we developed *LegalWriter*. *LegalWriter*[6] relies on the theory of error-based learning on one side and scaffolding approaches on the other. An ML-based adaptive feedback algorithm provides the underlying back-end to help students with feedback at their individual skill levels. An overview of the basic concept of *LegalWriter* appears in Figure 2.

3.1 Literature Requirements and User Requirements

3.1.1 ***Literature Requirements****:* We developed a theory-driven and user-centered writing support system using two strategies: a rigorous theory-driven approach and a user-centered approach. For the theory-driven approach, we used the Webster and Watson [86] and Vom Brocke et al. [74] methods to derive specific theoretical requirements for the design of an intelligent writing support system (e.g., [19]). To achieve this design, we followed two literature streams. First, we analyzed the literature on writing support systems to find out how they can be effective. In the second step, we analyzed the literature on technology-based learning to understand how to accomplish effective learning. To create an effective learning and support system, we finally explored the constructivist theory of error-based learning (see Section 2.3). Further literature related to the theoretical consideration of our systems is in the fields of immediate [9, 38] and adaptive feedback [30]. Furthermore, we

[6] *LegalWriter*'s interaction appears in a video: https://www.youtube.com/watch?v=xqennT2AFqU.

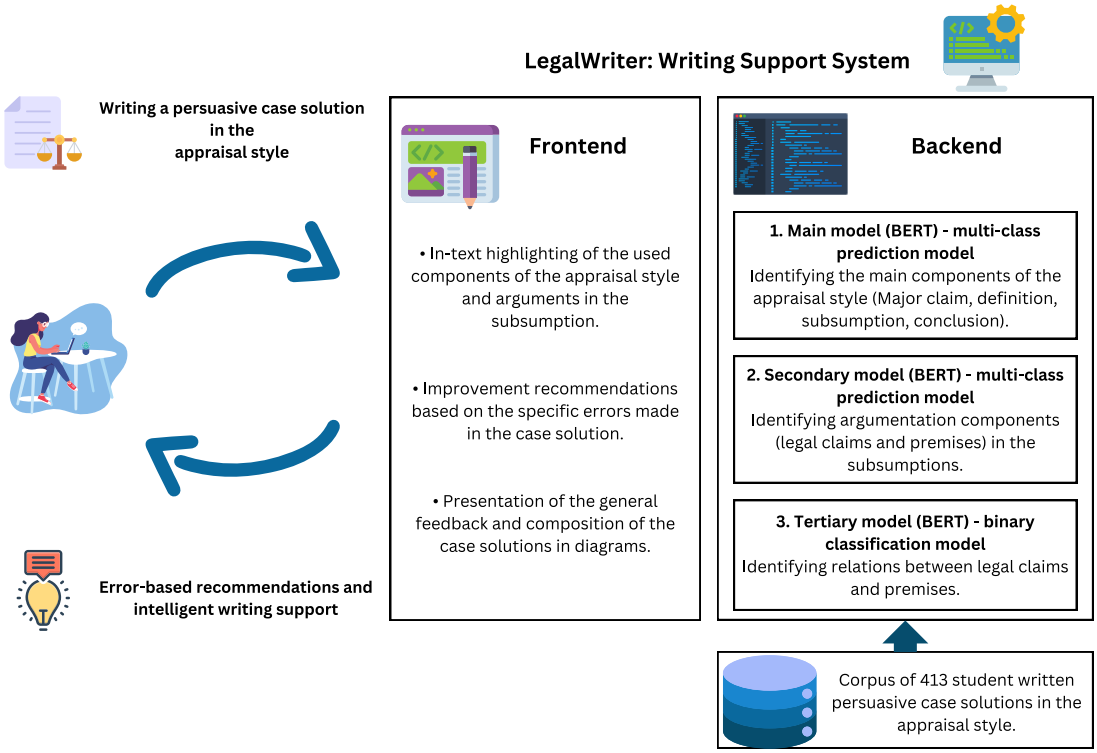

Figure 2: Basic concept of *LegalWriter*: Students complete a case solution writing exercise and are guided to write more persuasively through intelligent feedback and recommendations. For the feedback and the recommendations, we trained various BERT models based on the corpus of student-written case solutions from [83].

incorporated the theory of scaffolding to provide students with a clearly structured learning process and to help them improve their solutions in a self-directed way [10, 88]. The scaffolding approach is especially critical for novices, as they need more guidance for writing than more experienced students. In addition, we followed established learning methods from legal theory [48, 67] and case-based learning [21]. We aligned the system with an experience-based learning approach to provide an environment where students can practice their skills continuously and learn individually from their errors [12, 46]. In the next step, we translate all the requirements from the literature into meta-requirements (MR) (see Table 2). These MRs help us translate the knowledge from the theory into the descriptions for our artifact [19, 28].

3.1.2 ***User Requirements****:* In addition to the theory-driven approach, we applied a user-centered approach to customizing the system to the preferences of the future target group. We conducted ten user interviews with law students to collect the user requirements (UR) from students for a writing support system for persuasive and structured case solutions. The user interviews followed the methodology of Rubin and Chisnell [59]. The interviews utilized a semi-structured questionnaire based on three sets of questions (30 questions in total). The questions concerned law students' learning requirements, the considerations for implementing technology-enhanced education systems in law courses, and the design requirements for a writing support system. First, we asked the students about their learning requirements in law courses, seeking to understand the diverse aspects of their educational needs. Second, a significant focus of the interviews concerned the considerations inherent in implementing technology-enhanced education systems within legal education. We asked participants to share their perspectives on integrating technology, addressing potential challenges, benefits, and the overall impact on the learning outcomes of law students. This line of questioning aimed to uncover insights that could inform the development and implementation of effective and user-centric learning systems tailored to the unique demands of legal education. Third, we asked questions about the intricate design requirements for a writing support system catering to the needs of law students. We asked participants for their preferences and expectations concerning features, functionalities, and user interface design that would optimize their writing and learning experiences. The interviews lasted between 16 and 95 minutes (mean = 36.60, SD = 20.89).

The interviewees were students from various German universities, all potential users of our system. Therefore, we interviewed students from law and business law programs. The interviews with the business law students helped us broaden the system's scope. The mean age of the students was 23.00 years (SD = 2.03). Seven

women and three men participated in the interviews. We tape-recorded or videotaped all interviews and transcribed them. For transcription, we followed the approach of Mayring [45]. Based on the transcription, we derived categories found in all interviews and thus identified user stories from the interviews. We used open coding to form a unified coding system during the analysis [45]. Based on these results, we collected user stories and summarized the most frequent ones following Cohn [15]. The analysis of the interviews reveals the significant requirements of the students. They favored a user-friendly interface that is easy to understand, the ability to receive training in various legal domains (UR1), and the inclusion of in-text highlighting for elements of the appraisal style (UR2). Additionally, they emphasized the importance of clear explanations regarding the appraisal style and the explicit utilization of the system to ensure that students focused on the writing (UR3). More information on the requirements appears in Table 2.

3.2 User Interaction of *LegalWriter*

Following our requirements, we developed *LegalWriter* as a web-based application compatible with various devices. Figure 1 showcases a screenshot of *LegalWriter* with its primary functionalities (F1 - F8), while a detailed examination of individual features appears in Figure 3. *LegalWriter* comprises three key components: a text editor, a checklist, and a dashboard. It focuses on reviewing during writing, emphasizing the analysis and enhancement of written content, aligning with cognitive theory [24], and learning from errors theory [46]. It facilitates continuous feedback and recommendations for law students by allowing users to input their cases into the text editor (F1). Positioned directly above the text editor are two essential buttons, "*show case study*" and "*useful paragraphs*" (F2). These buttons enable students to access predefined case studies from specific legal domains and obtain content-related tips via the "*useful paragraphs*" button, preventing cognitive overload in novice students. Students receive personalized feedback by pressing the feedback button, stimulating text analysis, and generating highlights within their text (F3). These highlights denote the components of the appraisal style used, leaving unmarked sections that don't adhere to this style. Moreover, the "*feedback*" button also displays overall feedback in the dashboard and provides recommendations within the checklist.

On the left side, a checklist includes a word count and a summary of critical components related to the appraisal style (F4). The checklist also features a counting function to track used components and identify text errors (F5). The system offers 24 recommendations tailored to specific errors made by students (refer to Figure 3 and Tables 9, 10, and 11). Errors may occur when students forget components, fail to use precise legal terminology, or when components do not align with each other. For instance, a conclusion must follow each major claim. The system alerts the user if a major claim lacks a corresponding conclusion, though the user must independently assess the content alignment between the two. Upon encountering an error in the checklist, students receive tailored recommendations for addressing the issue (F5). The right side provides a legal argumentation dashboard. Here, students can access general feedback (see Table 11) and gain insights into the specific elements of their appraisal style (F6). Two charts offer information on text composition, with a ring diagram providing an overview of text composition based on sentences and their alignment with appraisal style components (refer to Figure 3). This text composition analysis aids students in assessing whether they have emphasized the correct elements in their case solution. Additionally, a bar chart gauges the persuasiveness and structure of the whole text. The system computes the persuasiveness score by comparing recognized appraisal style components, legal claims, and premises with unassigned text sections (F7). This score is determined by dividing the number of sentences not categorized within lawful components by the total number of sentences in the text. To further assist students, the "*more*" button provides detailed information about the system's functionalities and offers relevant suggestions for effective usage (F8).

3.3 Feedback Algorithm of LegalWriter

To develop an intelligent writing support system that promotes the writing of persuasive and structured case solutions grounded on error-based learning, we trained and tuned different models that allow students to learn from their errors.

For the model training process, we referred back to the legal corpus of Weber et al. [83]. This corpus comprises 413 case studies written by students in class. These texts were annotated by two annotators educated in law. The annotation showed satisfying results regarding Kripp. α and Fleiss' Kappa. Hence, a consistent observation is that annotating structured and persuasive elements in student case solutions is reliably achievable [83]. The dataset adheres to a rigorous annotation guideline[7], exhibiting moderate agreement, and has previously delivered adaptive feedback to law students focused on enhancing appraisal style evaluations. Example entries from the corpus for each legal component incorporated in the corpus appear in Table 13 in the Appendix. We developed the back end of *LegalWriter* using the Flask framework[8]. The back end consists of three ML models, two of which receive the output of the other model as their input data. All three models are transformer-based [71] and use BERT architecture [18]. We selected BERT models because they achieve the best accuracy, e.g., compared to RoBERTA models [42] in preliminary experiments. We obtained the pre-trained model from HuggingFace [87] and then trained it on the training dataset. BERT can transfer what it has learned from the initial dataset to the domain of lawful texts. The corpora were pre-processed with Spacy[9] to prepare suitable inputs for the models. Training the model happened in batches of size 8. The BERT models had a learning rate of $4e^{-5}$ and a warm-up ratio of 0.06.

Descriptions of the number of sentences in each class from the corpora of Weber et al. [83] and the performance and accuracy of the models appear below. For training the models, We randomly selected and used 20% of the original dataset for evaluation (i.e., test set) and the rest (80%) for the training set.

A) **Lawful components classifier:** This model is a five-class classification model, performing one sentence at a time. It classifies each of the sentences in the legal text as a *major claim* (3514

[7] Available in github.com
[8] https://flask.palletsprojects.com/
[9] https://spacy.io/

Table 2: Collected design requirements for developing an ML-based legal writing support system that supports writing persuasive and structured case solutions. Meta-requirements (MR) are collected through a literature search and act as a theory lens in system design. The user requirements (UR) were collected with the help of user interviews.

Name	Requirements
MR1: Provide the system with ...	... case studies that describe realistic legal problems and represent an established legal education method.
MR2: Provide the system with ...	... a feedback system, which evaluates adaptive students' errors in legal writing assessments.
MR3: Provide the system with ...	... an analysis system that checks for consistent adherence to the appraisal style and its components (major claim, definition, subsumption, and conclusion) and the connections between the individual components.
MR4: Provide the system with ...	... conceptual scaffolds in the form of recommendations that help students improve their case solutions.
UR1: Provide the system with ...	... cases from different areas of law.
UR2: Provide the system with ...	... highlighting that indicates the appraisal style and errors in the appraisal style.
UR3: Provide the system with ...	... explanations about the appraisal style and the explicit use of the system so that the system is user-friendly.

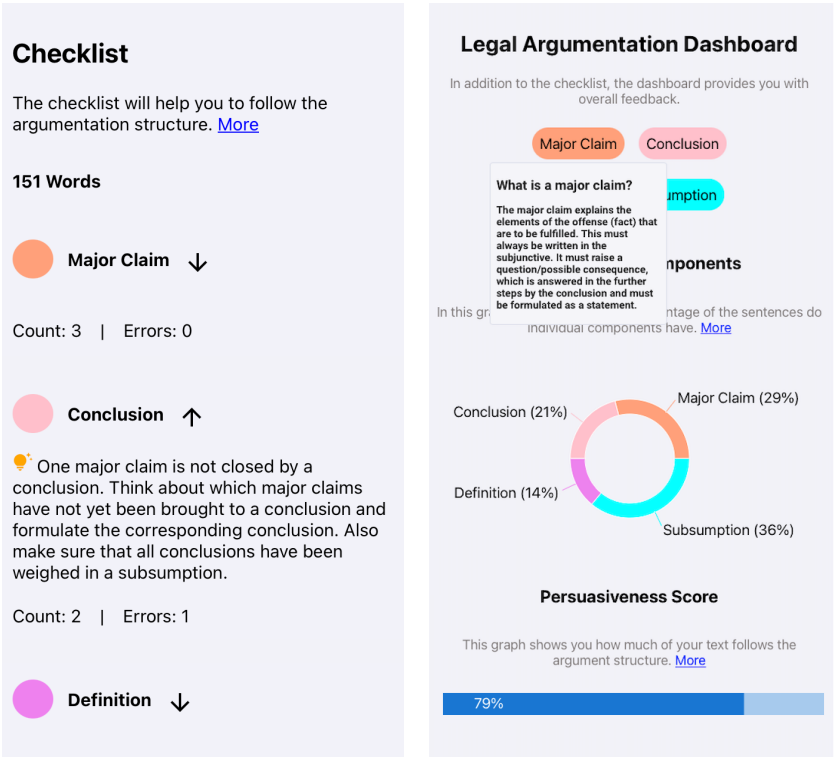

Figure 3: Screenshot of the two main functions of *LegalWriter*. Left: Checklist with guidelines for writing in the appraisal style and recommendations on how to improve errors based on three trained transformer models, Right: Dashboard which gives an overview of the use of the components of the appraisal style and provides students with a persuasiveness score, which gives an assessment of the extent to which structured and persuasive legal writing has been adhered to. Additional recommendations for improvement can also accessed in the dashboard.

sentences), *conclusion* (3531 sentences), *definition* (2288 sentences), *subsumption* (8090 sentences), or *none* (5320 sentences). With the help of this model, various lawful components are highlighted in *LegalWriter*'s interface. In addition, the subsumption sentences are used in model B to identify premises and legal claims.

B) **Subsumption types classifier:** This model is a three-class classification model, performing one sentence at a time. It classifies each of the sentences in the list of subsumption sentences obtained by model A as a *legal claim* (1949 sentences), *premise* (3304 sentences), or *None* (2837 sentences). With the help of this model, legal claims and premises are highlighted in *LegalWriter*'s interface. In addition, the legal claims and premises are used in model C to identify the relations between the premises and the legal claims.

C) **Legal Claim-Premise relation classifier:** This model is a two-class classification model, performing on each pair of sentences at a time. It classifies each of the legal claim-premise pairs in the list obtained by model B as having an argumentative relation or not. With the help of this model, *LegalWriter*'s user interface presents errors regarding backing the legal claims with evidence.

Additional pre-processing removed duplicate sentences or ones with five characters or less before using them as training and testing data for the models. The results obtained after the training and the evaluation of the models appear in Table 3.

4 EXPERIMENTAL EVALUATION

Our work tests whether intelligent writing support for individual errors helps novice students write higher-quality legal case solutions. We assessed case solution quality by evaluating the legal writing quality exhibited within the texts. The exact description of the text measurements is in Section 4.4. To test our hypothesis, we designed an experiment that asked participants to solve a case study on a legal problem. This task is a typical learning scenario across law curricula worldwide and a wide variety of courses that deal with legal content [21, 44, 48].

4.1 Participants

We recruited 96 students in an online experiment via Prolific. We selected Prolific as an experimental platform since past research on behavioral research platforms has reported it has the highest response quality and sample diversity [51]. Our objective was to assess the *LegalWriter* system using participants who are law novices. To ensure that we tested our design with a valid target group, we ensured that students participating in our study represented potential end users of a legal writing support system. Furthermore, we restricted it to students with adequate German skills (the current version of the system is in German). We also restricted the study to students who declared a self-interest in writing persuasive and structured legal case solutions. They could include students of business administration, business law, law, business informatics students, or any other who have to take a lecture on the basics of law. We identified Prolific as a suitable platform for our study, as we could generate a broad spectrum of participants and thus test our system on a diverse target group. In addition, we asked participants about their experience with the appraisal style (1- 7 Likert scale). We deliberately excluded experienced participants, defined as those who reported a substantial familiarity with the appraisal style (rated 3 or higher on the Likert scale) from the study. This decision aligns with the study's primary focus on novices in the field of law, ensuring that insights gleaned are particularly relevant to those with limited prior experience in the appraisal style.

After randomization and the experience test, we counted 32 completed valid outcomes in the treatment group and 30 in the control group. The average age of the control group was 29.4, and for the treatment group, it was 28.5. The participants in the control group were 17 males and 13 females. In the treatment group, ten were male, 21 were female, and one was non-binary. To examine the subjects' experiences and to ensure that all students had the same experiences, we asked about students' experiences with appraisal styles in the pre-survey (see Table 8). In the control group, we found a mean of 2.11, and in the treatment group, a mean of 1.84 (1- 7 Likert scale). We detected no significant difference between the groups, so we assumed both groups had the same prior knowledge. All participants received 9£ of compensation per hour of participation in the experiment. Participants took an average of 59.37 minutes to complete the experiment.

4.2 Experimental Setup

Both groups received a different version of the writing support system *LegalWriter* for the same writing task. The treatment group worked with an ML-based version of *LegalWriter* that provided intelligent writing support for correcting the students' errors and analyzing their written text. The support for improvement was based on the annotated corpus and the trained feedback algorithm. Based on this algorithm, the system highlights the user's text. This highlighting shows the user which components of the appraisal style are used and which ones might be missing. In addition, users can select the *"error"* button (on the left side of the system) to receive individual recommendations for improvement (see Figure 3). The text analysis dashboard on the right side of *LegalWriter* shows the students the distribution of the individual components, the persuasiveness of the text, and recommendations for improvement (see Tables 9, 10, and 11).

In the control group, participants utilized a baseline edition of *LegalWriter*. This version lacked the machine learning-based feedback algorithm and text analysis featured in the ML-based version of *LegalWriter*. The system comprised static recommendations aligned with the contemporary state of legal education systems [4]. Consequently, the checklist guided students in structuring their discussions, diminishing cognitive demands, and facilitating the creation of more structured texts. Additionally, students received general feedback at the conclusion to enhance their written work (see Table 12).

We implemented several functions in both versions to ensure consistency between the two iterations of *LegalWriter* and uphold the alternative learning approach. Both versions incorporated a case study and a practical paragraph checklist, albeit the static version's checklist did not highlight individual errors. Moreover, the explanation buttons and user interaction mechanisms remained identical in both.

Table 3: Recall, precision, and F1 score for the BERT models used in *LegalWriter*. A model comparison to other transformer models appears in [83].

Model Name	Model Type	Precision	Recall	F1 Score	Class
Lawful Components Classifier	BERT, 5-class	0.92	0.95	0.93	Major Claim
		0.87	0.92	0.89	Conclusion
		0.78	0.86	0.82	Definition
		0.69	0.73	0.71	Subsumption
		0.91	0.86	0.88	None
Subsumption Types Classifier	BERT, 3-class	0.78	0.58	0.66	Legal Claim
		0.62	0.79	0.69	Premise
		0.83	0.74	0.78	None
Legal Claim-Premise Relation Classifier	BERT, 2-class	0.90	0.90	0.90	-

Figure 4: Overview of the procedural steps implemented in our online experiment, wherein participants were tasked with writing a legal opinion. The participants were randomly divided into two groups. The treatment group created a case solution with ML support, while the control group worked only with a static version of the system [4, 73].

4.3 Experimental Procedure

To evaluate *LegalWriter*, we conducted a three-stage experiment, divided into the **pre-survey**, **writing exercise**, and **post-survey**. The **pre-survey** (randomization and experience test) and **post-survey** phases were the same for all participants. In the **writing exercise**, the two groups used different versions of *LegalWriter*. We randomly assigned the participants to a control or treatment group. The treatment group used an ML-based version of *LegalWriter*, and the control group used a non-ML-based version, considered the benchmark system (see Section 4.2).

1. Pre-survey: The experiment started with a questionnaire asking the participants how familiar they were with the appraisal style (see Section 4.1). We then tested the technology acceptance constructs to determine whether the randomization was successful [3, 65, 72]. We used a 1- to 7-point Likert scale for the pre-survey constructs. All questions and constructs appear in the Appendix.

2. Writing task: Before the writing task started, the participants received a two-page introduction PDF that described the appraisal style and instructions on how to solve a legal problem. During the reading, a countdown that shows the time remaining to complete the task begins. We asked four questions about the components of the appraisal style afterward to motivate the participants to read the introduction carefully (see Table 8). In the writing phase of the experiment, participants had to solve a case study involving a legal problem. We gave the participants in the control group and those in the treatment group the same legal problem. Before the task started, both groups received an introductory video on their respective version of the system. We told participants they would need at least 30 minutes to complete the writing exercise. They

could not complete the writing exercise until a countdown expired. The treatment group used the ML-based version of *LegalWriter* to write a case solution. In contrast, the control group used the non-ML-based version of *LegalWriter*. The participants who used the ML-based version of *LegalWriter* received individualized feedback based on the feedback algorithm we developed and recommendations for improvement. The participants using the non-ML-based version received static recommendations and a checklist to control the appraisal style in the case solution.

3. Post-survey: In the post-survey, we tested the impact of *LegalWriter* on student's experience with the learning process. Therefore, we measured the constructs of *feedback accuracy* [54], *intrinsic motivation* [40], and *enjoyment* [36] of the user while interacting with the system. We included two control questions to test whether the participants completed the survey conscientiously. All constructs appear in the Appendix. For all the constructs in the post-survey, we used a 1- to 7-point Likert scale. Following the constructs, we used four questions for a qualitative evaluation of the system and collected suggestions for improvement (see Table 8).

4.4 Measuring the Quality of Legal Writing

We aimed to investigate and measure the impact on students' writing quality by *LegalWriter*. Consequently, we assessed the quality of the legal writings based on the appraisal style. In the following, we briefly outline how we measured the legal text quality using the appraisal style in texts [44, 67]. To measure the correct application of the appraisal style, we analyzed if it used all the essential components of the appraisal style. Additionally, we conducted an assessment to determine the logical connection between the components. We established a quantifiable measure of evaluation quality with a 5-point scale modeled after the grading system used for law examinations at our university (refer to Table 7). This scale aligns with the customary evaluation criteria employed in German legal education, as outlined in Zekoll and Wagner [90]. We abstracted the grading system to enhance its measurability and objectivity. To ensure impartiality and prevent any identification of text origins, we analyzed all texts from the participants in a randomized sequence. A third-party legal scholar, unaffiliated with our project, assessed the students' texts based on the scores in Table 7.

5 RESULTS

To test the hypothesis of whether a system that provides students with individualized feedback based on their errors and recommendations during the writing process helps them write more persuasive and structured case solutions [90], we raised two research questions in the introduction. To answer the first research question, we analyzed the texts written by the experiment participants in the two versions of *LegalWriter* concerning the quality of legal writing.

We answered the second research question by analyzing constructs and comparing whether there are significant differences between the two groups. To assess the influence of *LegalWriter* on users' learning and writing processes, we employed the concepts of *enjoyment* and *intrinsic motivation*. These concepts helped us evaluate whether the user experience was pleasant and whether participants felt a higher inherent drive to engage in learning through feedback derived from machine learning techniques [36, 40]. We used the *feedback accuracy* construct to determine whether participants perceived ML-based feedback as more accurate [54]. Moreover, we performed a comprehensive qualitative assessment to elucidate the intricate ways *LegalWriter* impacts the student learning process (see Section 5.3).

We evaluated the differences between the treatment and control groups by performing two-tailed t-tests. We tested data relevant to RQ1 and RQ2 for normal distribution, for which we created graphical histograms for the control and the treatment group. The analysis revealed that each group's data were distributed normally. In addition to a normal distribution, we tested for equality of variances between groups using a Levene test [26]. As Levene's test did not reveal equal significance for the construct *enjoyment*, we conducted a Welch test to assess differences in the *enjoyment* construct between the two groups. We assumed that randomization was successful because we randomly assigned participants to the groups during the online experiment. Additionally, there was no noteworthy distinction between the control group (mean = 5.21) and the treatment group (mean = 5.77) when assessing attitudes toward technology ($p = 0.077$). Furthermore, we detected no significant variance in participants' prior experience in legal writing, as outlined in Section 4.1.

5.1 Impact on Students' Legal Text Quality

We analyzed the quality of the legal writing to get insights into the texts and determine if they were written persuasively. We compared the quality of the appraisal style of the control group with that of the treatment group. As a basis for the comparison, we used the assigned scores on the quality of the appraisal style. Figure 5 presents the data distribution with boxplots. The control group is on the left, while the treatment group is on the right. The interquartile ranges notably highlight a more extensive score distribution in the treatment group. Also, mean values appear as red points for enhanced visualization. The median for both groups is 3 (see Table 4). As Table 4 shows, the mean score (mean = 3.41) of the treatment group is higher than that of the control group (mean = 2.60). The difference between the two groups shows statistical significance, which underlines the effectiveness of ML-based feedback.

5.2 Impact on Students' Learning Process

To determine the impact on the students' learning process, we measured *intrinsic motivation* and the level of *enjoyment* during the interaction with the system. Figure 6 presents the data distribution of the two constructs with boxplots. *Intrinsic motivation* is on the left, while *enjoyment* is on the right. We compare the control (CG) and treatment group (TG) for both constructs. The median in the treatment group (TG) is 5 for *enjoyment* and 5.08 for *intrinsic motivation* (see Table 5) for the medians of the control group (CG)). Notably, the interquartile ranges highlight more extensive distributions in the treatment group for both constructs (see Figure 6). Also, mean values appear as red points for enhanced visualization. *Enjoyment* averaged a rating of 4.64 (SD = 1.16), and *intrinsic motivation* averaged 4.911 (SD = 1.047). Both constructs are significantly higher than the control group (see Table 5). Both *enjoyment* and *intrinsic motivation* are above the neutral mean of 4. *Enjoyment* increases acceptance and learning success among students [39].

Table 4: Results of the analysis of the *quality of legal writing*. We show the mean, the standard deviation, the median, and the degrees of freedom (df) of the control group and the treatment group, as well as the results of a double-tailed t-test with equal variances. We set the significance level at alpha 0.05: p<=0.01.**

Quality of legal writing (appraisal style)	Control group	Treatment group
Mean	2.600	3.406**
Standard Deviation (SD)	1.220	1.012
Median	3	3
p value	0.006	
t value	2.839	
Degrees of Freedom (df)	60	

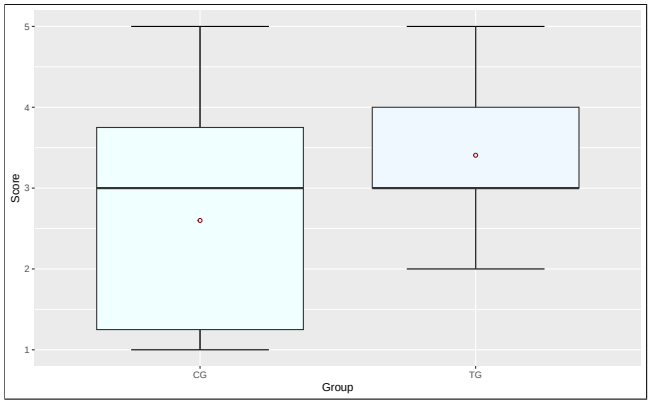

Figure 5: Data distribution of scores for legal text quality. The treatment group (TG) receives intelligent feedback and recommendations, while the control group (CG) receives static recommendations. Mean values are highlighted with red points.

Empirical research supports the view that activating positive emotions can improve academic performance. Specifically, *enjoyment* in a learning process fosters student's achievements [52]. Our data analysis revealed that students who tested the ML-based version of *LegalWriter* displayed significantly higher *intrinsic motivation* than students in the control group (see Table 5). *Intrinsic motivation* refers to the actions done for an end unto themselves or because they seem inherently pleasurable. Individuals motivated by intrinsic factors engage in novel acts without expecting a reward. As a result, *intrinsic motivation* positively affects learning processes and skills acquisition [7]. The literature on error-based learning discusses the effects of errors on motivation and enjoyment of learning [34] since too many errors can quickly demotivate learners and lead to frustration [12]. Against this background, the results that the participants rated *motivation* and *enjoyment* above the neutral value of 4 are interesting.

Since *LegalWriter* uses an intelligent ML-based feedback algorithm, we aimed to analyze how accurately the groups perceived the feedback and recommendations in the ML-based version of *LegalWriter* compared to the non-ML-based version of *LegalWriter*. Figure 7 presents the data distribution of *feedback accuracy* through boxplots. The control group (CG) (median = 2.88) is on the left, while the treatment group (TG) is on the right (median = 4.5). The interquartile ranges notably highlight a more extensive score distribution in the treatment group (TG). Additionally, mean values appear as red points for enhanced visualization. We evaluated the *feedback accuracy* construct in the treatment group with a mean value of 4.5 (SD = 1.20). However, the control group only revealed an average score of 3.158 (SD = 0.99) in the evaluation. The data analysis shows that the treatment group rates the accuracy of the feedback and the recommendations significantly higher than the control group (see Table 5). Thus, we can show that ML-based feedback is perceived as more accurate than static feedback in our skill-learning scenario.

5.3 Qualitative User Statements

At the end of our survey, we asked the experiment participants some open questions to explain what they liked about the system and what could be improved. Positive sentiments for *LegalWriter* were evident: "*Very clean design and intuitive user interface. Easy to understand and color-highlighted where errors lie*". Participants noted that they liked the "*direct feedback and the colored background of the sentences*." They also praised the clean design and user-friendliness. They perceived the dashboard with the overall feedback as positive. Despite the positive assessment of the system, the participants still had suggestions for improvement. Sometimes, the students desired minor technical improvements to make the interaction even more intuitive. For example: "*It would be advantageous to be able to open the tabs 'Case study' and 'Useful paragraphs' in parallel to increase efficiency.*" Another important point mentioned by the participants was that the system should explain the errors more precisely and should indicate in the text where the individual errors lie more clearly.

To gain deeper insights into the qualitative statements of the participants, we examined their feedback with a focus on their perceptions of their learning process. We discerned three primary themes across the treatment group (TG), encompassing 1) perceived enjoyment, 2) learner motivation, and 3) feedback accuracy. We provide a concise overview of the key points (see Table 6), with a more detailed discussion in the following.

1) Perceived enjoyment: The feedback from the treatment group contained positive remarks indicating the students' perceived enjoyment during the learning process. For example, 76% of the treatment group students stressed that they found the interaction

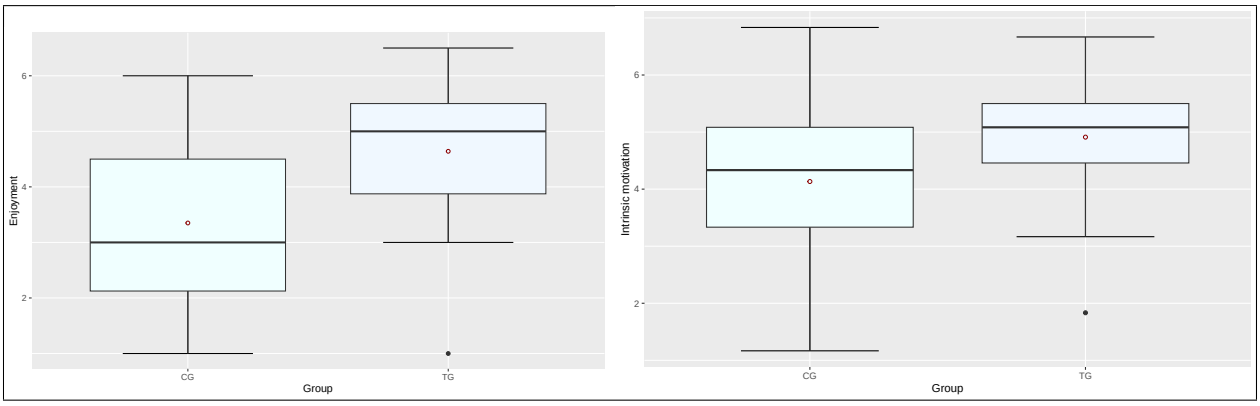

Figure 6: Data distribution of enjoyment and intrinsic motivation. The treatment group (TG) receives ML-based intelligent feedback and recommendations, while the control group (CG) receives static recommendations. Mean values are highlighted with red points.

Table 5: Results of the analysis of the perceived *enjoyment*, the *intrinsic motivation*, and *feedback accuracy*. We show the mean, the standard derivation, and the degrees of freedom (df) of the control and treatment group, as well as the results of a double-tailed t-test with equal variances. We used a Welch test for the *enjoyment* construct since a Levene test could not detect equal significance. We set the significance level at alpha 0.05: p<=0.05*, p<=0.001*.**

Group	Enjoyment	Intrinsic motivation	Feedback accuracy
Mean (TG)	4.640***	4.911*	4.500***
Mean (CG)	3.350	4.133	3.158
SD (TG)	1.166	1.047	1.204
SD (CG)	1.515	1.453	0.986
Median (TG)	5	5.08	4.5
Median (CG)	3	4.32	2.88
p value	> 0.001	0.018	> 0.001
t value	3.742	2.430	4.474
df	54.432	60	60

with the system enjoyable. Some students in this group mentioned the highlights as particularly thrilling (30% of the students). One stated, "*I appreciated that my sentences were highlighted to maintain an overview of my structure.*" These statements suggest that the highlighting feature, in particular, contributed to the students' perceptions that writing with *LegalWriter* is an enjoyable experience. Additionally, students rated the interaction as easy and comfortable.

2) Learner motivation: The examination of the results regarding the effects of the intelligent writing support system showed that the students in the treatment group achieved better results in legal writing than the students in the control group (see Section 5.1). Statements from students in the treatment group concerning their motivation supported this assertion, as increased motivation leads to increased learning success [7, 63]. Some students expressed that their interaction with the system inspired them to revise their texts more frequently. They identified ML-based feedback as a significant factor of heightened motivation: "*The feedback helped me adhere to the writing style and motivated me to achieve a better quality rating.*"

3) Feedback accuracy: Comparing the statements from the two groups, it is evident that students in the treatment group perceived the system as valuable since 95% of all participants stated that result in the treatment group. In contrast, the control group had significantly fewer comments reflecting this sentiment (just 28% of all participants in the CG). Most positive comments in the treatment group are related to feedback. These comments illustrate that they regarded ML-based feedback as precise and accurate. For instance, students mentioned, "*Feedback provided a sense of security, making it easier to continue writing, especially after receiving positive feedback.*" Moreover, students in the treatment group attributed to the *LegalWriter* system the capability of "*recognizing the basic structure and classifying sentences correctly as far as possible. Thus, it appears that the system genuinely aids in the writing process.*"

6 DISCUSSION OF FINDINGS

In this study, we developed and evaluated a novel student-centered, theory-driven writing support system called *LegalWriter*. *LegalWriter* enables novice students to improve their skills in writing

Table 6: Representative examples of qualitative user topics from the treatment group (TG).

Topic	Example User Responses
Enjoyment in the learning process	**On the ML-based feedback, e.g.,** "*I found the use enjoyable and eventually took advantage of the feedback.*" **On the highlighting [2], e.g.,** "*I appreciated that my sentences were highlighted to maintain an overview of my structure.*" **On the whole system perception, e.g.,** "*While writing in the system, I experienced ease and comfort.*"
Motivation in the learning process	**On the ML-based feedback, e.g.,** "*The feedback helped me adhere to the writing style and motivated me to achieve a better quality rating.*" **On the whole system perception, e.g.,** "*The system prompted me to reconsider the text multiple times.*"
Accuracy of the ML-based feedback	**On the whole system perception, e.g.,** "*I think it's good that LegalWriter recognizes the basic structure and can classify sentences correctly as far as possible. So, apparently, it really does help with writing.*" **On the whole system perception, e.g.,** "*LegalWriter demonstrated a commendable ability to discern and categorize each sentence according to its respective component.*" **On the ML-based feedback, e.g.,** "*Feedback provided a sense of security, making it easier to continue writing, especially after receiving positive feedback.*"

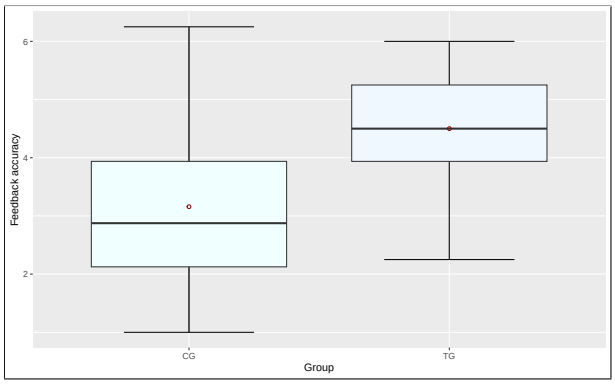

Figure 7: Data distribution of feedback accuracy. The treatment group (TG) receives intelligent ML-based feedback and recommendations, while the control group (CG) receives static recommendations. Mean values are highlighted with red points.

persuasive and structured case solutions. We evaluated our intelligent writing support system in a quantitative study and qualitative user evaluation (see Section 5). We found that students who received ML-based writing support and error-based feedback while writing legal case solutions wrote more persuasive and structured case solutions, as they better adhered to the appraisal style than those who received only general recommendations. Furthermore, we observed that novice students who used the ML-based system rated the feedback from the system as more accurate. At the same time, they felt more *enjoyment* and *intrinsic motivation* than the students who used the alternative system. Past research in computer-supported legal education focused on the design and evaluations of representational guidance approaches or discussion scripting approaches (e.g., [4, 53]). Hence, with our study, we provide not only insights into the design of a novel legal writing support system based on ML but also rich empirical evidence about the effects of intelligent writing support through a controlled experimental study.

6.1 Theoretical and Practical Contributions

Our study has multiple contributions to the literature on HCI, legal tech, and educational technology.

First, our study provides an interdisciplinary conceptualization and overview of legal learning systems based on novel technologies. Our approach combined insights from computer-supported learning, legal writing, and pedagogical theory. We found that insights into the HCI and the design of ML-based writing support systems in law are rare. Most research is based on representational guidance or discussion scripting approaches, which offer limited individual support due to a lack of adaptivity. We provided students individualized feedback on their natural errors by developing an intelligent writing support system that extends the HCI and systems available in legal education [4, 53]. Our novel writing support system differs from existing systems in legal writing by providing individualized feedback on written texts and natural errors. Traditional writing support systems usually only support general argumentation approaches [50], which are not sufficient when writing legal case solutions due to the strict formalism and the precisely organized manner of presenting arguments (see Section 2.1).

Second, with our study, we offer design knowledge for the HCI community, legal tech, and computer-supported learning in law. We followed a rigorous and transparent methodology to derive requirements from users and theory to design a new writing support system based on ML and NLP, especially for the legal domain. We present seven theory- and user-centered requirements and show their usefulness through a qualitative and quantitative evaluation through *LegalWriter*. We believe that other researchers or educational designers can apply these requirements to develop their own

educational tools in the legal domain. Educational research can follow our requirements (Table 2) to train legal writing skills in other languages or domains. For example, it could combine legal case studies from different areas of law with adaptive feedback. These requirements contribute not only to HCI but also to educational technology research and legal tech by bridging the gap between a user-centered design and rigorous theoretical embedding. By following the approach of learning from errors, we believe that pedagogical designers can build upon our approach to implement and test this theory embedded in our tool.

Third, we also provide a theoretical contribution to pedagogy by illustrating that incorporating the learning from errors approach [46] with ML-based feedback algorithms can aid students in enhancing their legal writing skills and capability to address legal issues persuasively. The approach of "error permission" [88] has proven to be especially effective in our work, as we have given less experienced students (novices) in the learning environment the opportunity to practice their skills and improve them for natural errors [23]. Thus, we contribute to the research on learning from errors as we demonstrate the "error permission" approach — a specific method within the framework of learning from errors, as discussed by Wong and Lim [88] — in digital environments. Furthermore, our work supports the view that direct feedback or recommendations following an individual error support the student learning process [43, 55].

Fourth, we show empirical evidence for the effectiveness of ML-based writing support systems for law. Our work shows that ML-based writing support systems are successful, especially in support of structured writing, which is indispensable for persuasive and structured case solutions in law courses. We demonstrate that novice students using the ML-based version of *LegalWriter* write case solutions with a qualitatively higher appraisal style than students using the alternative version of the system (see Table 4). Past work on interdisciplinary research between HCI and legal education mainly utilized methods of argument diagramming (representational guidance approach). Students were supported through visualization of their legal argumentation chances in node and linked graphs [53, 58]. Our work extends this work with new design features to legal learning systems and combines them with ML-based feedback. By evaluating their impact in an online experiment, we also indicate the behavioral and perceptual effects of our new design. Students who received ML-based writing support and error-based feedback while writing legal case solutions wrote more persuasive and structured case solutions, as they better adhered to the appraisal style than those who received only general recommendations. Moreover, the students indicated a better user experience measured by a higher *enjoyment* and *intrinsic motivation* than those who used the alternative system. Furthermore, it became evident that patients consistently rated the accuracy of the feedback higher than static feedback. This observation supports the hypothesis that students perceive ML-based feedback as accurate and valuable. These insights might shed light on how to utilize AI-based tools for adaptive education. We extended these quantitative findings with qualitative insights. Overall, we saw that students rated both versions of our system as useful, especially with the ML-based version. Students stated that the ML-based version motivated them to revise their texts several times, and they felt supported while writing because the system gave them more confidence (see Table 6). They also recognized feedback's utility in addressing individual errors, underscoring its precision. Students found the text highlights beneficial [2], implying that they complement the feedback-based learning approach.

Fifth, our findings offer valuable insights not only for HCI researchers and educational designers but also for students and teachers within the legal realm. The application of our system can prove advantageous, as it has the potential to enrich traditional learning setups in legal courses by enabling novice students to receive personalized feedback throughout their learning journey. Previously constrained by organizational and financial limitations, personalized feedback was limited, especially when a few lecturers had to cater to a large student population. Moreover, our system supports students during self-directed learning, fostering a more active and engaging educational experience. Consequently, we present a practical use case illustrating skills training on a large scale within a standard pedagogical setting. The insights derived from this scenario extend beyond writing support and pose practical implications for diverse academic applications. By doing so, we advocate for the widespread adoption and implementation of intelligent writing support systems in educational settings and institutions.

7 LIMITATIONS AND FUTURE WORK

In respect to our work, some limitations must be mentioned. Although our model shows good values for the prediction of the components of the appraisal style (precision between 69% and 92%), the values for the determination of the legal claims and premises are lower compared to the other values (62% and 78%) (see Table 3). However, compared to other studies in argumentation mining, they show acceptable values. Since values for accuracy in argumentation mining are rare for the legal field, we can only draw a comparison with other fields. For example, Wambsganss et al. [79] showed good results in supporting the argumentation skills of students with an accuracy of 65.4%. As our values are higher, they seem to be reasonable, specifically since, at the same time, our results in perceived feedback accuracy also suggest that our models can provide useful feedback for students (see Table 5). Our empirical and qualitative evaluations showed positive effects on the user experience, the learning process, and the adherence to the appraisal style through the interaction with *LegalWriter*. Accordingly, the influence of error-based learning combined with ML-based feedback improved writing persuasive and structured case studies [46].

Although our system shows positive results in student text quality, we assume that a single application of the system is insufficient to teach a complex skill such as persuasive and structured writing of legal case solutions. We can determine true learning outcomes and long-term value only through repeated use over an extended period [66]. We plan a field experiment in civil law courses at various universities to demonstrate the long-term learning effects. The field experiment will allow us to evaluate the system in parallel with the courses in the longer term and possibly measure learning progress. To this end, we would like to schedule three interventions of the system followed by an interaction without the system to see if students still achieve success without the system. We want to

compare these results with a control group with no intervention with the system. The field experiment will follow the approach of Abbasi et al. [1].

Furthermore, our evaluation lacks a control group that receives no system support and instead receives feedback following the legal gold standard by discussing a sample solution with a tutor. In previous studies, we have demonstrated the effectiveness of our system compared to a control group without system feedback and discussed some implications of why ML-based feedback improves the quality of students' texts [84]. The results showed that students who used our systems (mean = 11.08, p = 0.002) performed significantly better on a 15-grade scale than students who learned according to the legal gold standard (mean = 8.84) Weber et al. [84]. However, in our CHI paper, we aimed to focus on the HCI contribution and hence compared two different design versions of *LegalWriter*.

As a final limitation, the current version of *LegalWriter* is limited to applying the appraisal style in German. In the future, further efforts must investigate the transferability of our system to other countries with different legal systems and languages. However, we assume this task is possible in principle since some countries like China now also use the appraisal style in law teaching [44], and countries like the USA use similar approaches like learning with case studies using the IRAC formula [48]. Nevertheless, it will require some adaptation of the system to transfer it to other legal systems, as legal logic and language in each country have their own specificities.

8 CONCLUSION

Novice students in law courses must acquire specialized and highly concept-oriented knowledge to solve legal problems. Structured and persuasive writing combined with the required specialized knowledge is challenging for many learners. At the same time, we see that the traditional and computer-based support for students to write persuasive and structured case solutions falls short of expectations. Therefore, we developed and evaluated a novel student-centered and theory-driven system called *LegalWriter*. *LegalWriter* is an intelligent writing support system that provides students feedback on their individual errors. To develop *LegalWriter*, we derived new design knowledge in seven requirements. We tested *LegalWriter* in an evaluation with 62 novice law students. The results show that students who used the ML-based version of *LegalWriter* produced more persuasive and structured case solutions with a higher quality of the appraisal style than students who used an alternative version of the system. Furthermore, we showed that students perceived the interaction with *LegalWriter* as fun and had a higher intrinsic motivation to use it than a comparable system. Additionally, students rated the ML-based feedback as accurate and useful. The results suggest that our student-centered system, which relies on the error-based learning theory and uses ML-based feedback, helps novice students write more persuasive and structured case solutions and improves the user experience and the student's learning processes.

REFERENCES

[1] Ahmed Abbasi, Zhu Zhang, David Zimbra, Hsinchun Chen, and Jay F Nunamaker Jr. 2010. Detecting fake websites: The contribution of statistical learning theory. *Mis Quarterly* (2010), 435–461.

[2] Tazin Afrin, Omid Kashefi, Christopher Olshefski, Diane Litman, Rebecca Hwa, and Amanda Godley. 2021. Effective Interfaces for Student-Driven Revision Sessions for Argumentative Writing. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–13. https://doi.org/10.1145/3411764.3445683

[3] Ritu Agarwal and Elena Karahanna. 2000. Time flies when you're having fun: Cognitive absorption and beliefs about information technology usage. *MIS quarterly* (2000), 665–694.

[4] Vincent Aleven. 2003. Using background knowledge in case-based legal reasoning: a computational model and an intelligent learning environment. *Artificial Intelligence* 150, 1-2 (2003), 183–237.

[5] Pascal Ancel, Olivier Gout, and Ingrid Maria. 2012. *Travaux dirigés: introduction au droit et droit civil*. Vol. 3. LexisNexis.

[6] Kevin D Ashley and Vincent Aleven. 1991. Toward an intelligent tutoring system for teaching law students to argue with cases. In *Proceedings of the 3rd international conference on Artificial intelligence and law*. 42–52.

[7] Andrew G Barto, Satinder Singh, Nuttapong Chentanez, et al. 2004. Intrinsically motivated learning of hierarchical collections of skills. In *Proceedings of the 3rd International Conference on Development and Learning*. Piscataway, NJ, 112–19.

[8] Michael Beurskens. 2016. Neue Spielräume durch Digitalisierung? E-Learning in der deutschen Rechtslehre. *ZDRW Zeitschrift für Didaktik der Rechtswissenschaft* 3, 1 (2016), 1–17.

[9] Paul Black and Dylan Wiliam. 2009. Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability (formerly: Journal of Personnel Evaluation in Education)* 21, 1 (2009), 5–31.

[10] Kursat Cagiltay. 2006. Scaffolding strategies in electronic performance support systems: Types and challenges. *Innovations in education and Teaching International* 43, 1 (2006), 93–103.

[11] Chad S Carr. 2003. Using computer supported argument visualization to teach legal argumentation. In *Visualizing argumentation*. Springer, 75–96.

[12] Kelly A Chillarege, Cynthia R Nordstrom, and Karen B Williams. 2003. Learning from our mistakes: Error management training for mature learners. *Journal of business and psychology* 17, 3 (2003), 369–385.

[13] Jonathan H Choi, Kristin E Hickman, Amy Monahan, and Daniel Schwarcz. 2023. Chatgpt goes to law school. *Journal of Legal Education (Forthcoming), Available at SSRN: https://ssrn.com/abstract=4335905 or http://dx.doi.org/10.2139/ssrn.4335905* (2023).

[14] Elizabeth Clark, Anne Spencer Ross, Chenhao Tan, Yangfeng Ji, and Noah A Smith. 2018. Creative writing with a machine in the loop: Case studies on slogans and stories. In *23rd International Conference on Intelligent User Interfaces*. 329–340.

[15] Mike Cohn. 2004. *User stories applied: For agile software development* (1 ed.). Addison-Wesley Professional.

[16] Reuma De Groot, Raul Drachman, Rakheli Hever, Baruch B Schwarz, Ulrich Hoppe, Andreas Harrer, Maarten De Laat, Rupert Wegerif, Bruce M McLaren, et al. 2007. Computer supported moderation of e-discussions: The ARGUNAUT approach. *Mice, minds, and society—The computer supported collaborative learning (CSCL)* (2007), 165–167.

[17] Aniket Deroy, Kripabandhu Ghosh, and Saptarshi Ghosh. 2023. How Ready are Pre-trained Abstractive Models and LLMs for Legal Case Judgement Summarization? *Proceedings of the Third International Workshop on Artificial Intelligence and Intelligent Assistance for Legal Professionals in the Digital Workplace* (2023), 8–19.

[18] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL* (2018), 4171–4186.

[19] Ernestine Dickhaut, Andreas Janson, Matthias Söllner, and Jan Marco Leimeister. 2023. Lawfulness by design–development and evaluation of lawful design patterns to consider legal requirements. *European Journal of Information Systems* (2023), 1–28.

[20] Rosalind Driver, Paul Newton, and Jonathan Osborne. 2000. Establishing the norms of scientific argumentation in classrooms. *Science education* 84, 3 (2000), 287–312.

[21] Cecilie Enqvist-Jensen, Monika Nerland, and Ingvill Rasmussen. 2017. Maintaining doubt to keep problems open for exploration: An analysis of law students' collaborative work with case assignments. *Learning, culture and social interaction* 13 (2017), 38–49.

[22] K Anders Ericsson, Ralf T Krampe, and Clemens Tesch-Römer. 1993. The role of deliberate practice in the acquisition of expert performance. *Psychological review* 100, 3 (1993), 363–406.

[23] Lisa K Fazio and Elizabeth J Marsh. 2009. Surprising feedback improves later memory. *Psychonomic Bulletin & Review* 16, 1 (2009), 88–92.

[24] Linda Flower and John R Hayes. 1981. A cognitive process theory of writing. *College composition and communication* 32, 4 (1981), 365–387.

[25] Gaël Gendron, Qiming Bao, Michael Witbrock, and Gillian Dobbie. 2023. Large Language Models Are Not Abstract Reasoners. *arXiv preprint arXiv:2305.19555* (2023).

[26] Gene V Glass. 1966. Testing homogeneity of variances. *American Educational Research Journal* 3, 3 (1966), 187–190.

[27] Thomas F Gordon, Henry Prakken, and Douglas Walton. 2007. The Carneades model of argument and burden of proof. *Artificial Intelligence* 171, 10-15 (2007), 875–896.

[28] Shirley Gregor and Alan R Hevner. 2013. Positioning and presenting design science research for maximum impact. *MIS quarterly* (2013), 337–355.

[29] Ben Hachey and Claire Grover. 2005. Automatic legal text summarisation: experiments with summary structuring. In *Proceedings of the 10th International Conference on Artificial intelligence and Law*. 75–84.

[30] John Hattie and Helen Timperley. 2007. The power of feedback. *Review of educational research* 77, 1 (2007), 81–112.

[31] Nick James. 2012. Logical, critical and creative: Teaching'thinking skills' to law students. *Law and Justice Journal* 12, 1 (2012), 66–88.

[32] David H. Jonassen and Bosung Kim. 2010. Arguing to learn and learning to argue: Design justifications and guidelines. *Educational Technology Research and Development* 58, 4 (2010), 439–457. https://doi.org/10.1007/s11423-009-9143-8

[33] Laksnoria Karyuatry. 2018. Grammarly as a tool to improve students' writing quality: Free online-proofreader across the boundaries. *JSSH (Jurnal Sains Sosial dan Humaniora)* 2, 1 (2018), 83–89.

[34] Nina Keith and Michael Frese. 2008. Effectiveness of error management training: a meta-analysis. *Journal of Applied Psychology* 93, 1 (2008), 59–69.

[35] Jacob Devlin Ming-Wei Chang Kenton and Lee Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT*, Vol. 1. 4171–4186.

[36] Soomin Kim, Joonhwan Lee, and Gahgene Gweon. 2019. Comparing data from chatbot and web surveys: Effects of platform and conversational style on survey response quality. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–12.

[37] Nate Kornell, Matthew Jensen Hays, and Robert A Bjork. 2009. Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 35, 4 (2009), 989–998.

[38] James A Kulik and Chen-Lin C Kulik. 1988. Timing of feedback and verbal learning. *Review of educational research* 58, 1 (1988), 79–97.

[39] Matthew Lee, Christy MK Cheung, and Zhaohui Chen. 2005. Acceptance of Internet-based learning medium: the role of extrinsic and intrinsic motivation. *Information & management* 42, 8 (2005), 1095–1104.

[40] Lijia Lin, Robert K Atkinson, Robert M Christopherson, Stacey S Joseph, and Caroline J Harrison. 2013. Animated agents and learning: Does the type of verbal feedback they provide matter? *Computers & Education* 67 (2013), 239–249.

[41] Marco Lippi and Paolo Torroni. 2016. Argumentation mining: State of the art and emerging trends. *ACM Transactions on Internet Technology (TOIT)* 16, 2 (2016), 1–25.

[42] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. , 67 pages. https://doi.org/10.48550/ARXIV.1907.11692

[43] Steven J Lorenzet, Eduardo Salas, and Scott I Tannenbaum. 2005. Benefiting from mistakes: The impact of guided errors on learning, performance, and self-efficacy. *Human Resource Development Quarterly* 16, 3 (2005), 301–322.

[44] JIN Man. 2022. The Appraisal-Based Case Teaching Method in China's Legal Education. *Canadian Social Science* 18, 2 (2022), 1–4.

[45] Philipp Mayring. 2010. Qualitative Inhaltsanalyse. In *Handbuch Qualitative Forschung in der Psychologie*. VS Verlag für Sozialwissenschaften, 601–613. https://doi.org/10.1007/978-3-531-92052-8{_}42

[46] Janet Metcalfe. 2017. Learning from errors. *Annual Reviews Inc.* 68 (2017), 465–489.

[47] Janet Metcalfe and Judy Xu. 2018. Learning from one's own errors and those of others. *Psychonomic Bulletin & Review* 25, 1 (2018), 402–408.

[48] Jeffrey Metzler. 2002. The importance of IRAC and legal writing. *University of Detroit Mercy Law Review* 80 (2002), 501–514.

[49] Stellan Ohlsson. 1996. Learning from performance errors. *Psychological review* 103, 2 (1996), 241–262.

[50] Jonathan F. Osborne, J. Bryan Henderson, Anna MacPherson, Evan Szu, Andrew Wild, and Shi Ying Yao. 2016. The development and validation of a learning progression for argumentation in science. *Journal of Research in Science Teaching* 53, 6 (2016), 821–846. https://doi.org/10.1002/tea.21316

[51] Eyal Peer, Laura Brandimarte, Sonam Samat, and Alessandro Acquisti. 2017. Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology* 70 (2017), 153–163.

[52] Reinhard Pekrun, Thomas Goetz, Wolfram Titz, and Raymond P Perry. 2002. Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational psychologist* 37, 2 (2002), 91–105.

[53] Niels Pinkwart, Kevin D Ashley, Vincent Aleven, and Collin F Lynch. 2008. Graph Grammars: An ITS Technology for Diagram Representations.. In *FLAIRS Conference*. 433–438.

[54] Philip M Podsakoff and Jiing-Lih Farh. 1989. Effects of feedback sign and credibility on goal setting and task performance. *Organizational behavior and human decision processes* 44, 1 (1989), 45–67.

[55] Rosalind Potts and David R Shanks. 2014. The benefit of generating errors during learning. *Journal of Experimental Psychology: General* 143, 2 (2014), 644–667.

[56] Prakash Poudyal, Teresa Gonçalves, and Paulo Quaresma. 2019. Using Clustering Techniques to Identify Arguments in Legal Documents.. In *ASAIL@ ICAIL*.

[57] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, Peter J Liu, et al. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.* 21, 140 (2020), 1–67.

[58] Chris Reed, Douglas Walton, and Fabrizio Macagno. 2007. Argument diagramming in logic, law and artificial intelligence. *The Knowledge Engineering Review* 22, 1 (2007), 87–109.

[59] Jeffrey Rubin and Dana Chisnell. 2008. *Handbook of usability testing: how to plan, design and conduct effective tests*. John Wiley & Sons.

[60] David R Samuelson. 1997. Introducing legal reasoning. *J. Legal Educ.* 47 (1997), 571–598.

[61] Patricia Schank and Michael Ranney. 1995. Improved reasoning with convince me. In *Conference companion on Human factors in computing systems*. 276–277.

[62] Oliver Scheuer, Frank Loll, Niels Pinkwart, and Bruce M McLaren. 2010. Computer-supported argumentation: A review of the state of the art. *International Journal of Computer-supported collaborative learning* 5, 1 (2010), 43–102.

[63] Sofia Marlena Schöbel, Andreas Janson, and Matthias Söllner. 2020. Capturing the complexity of gamification elements: a holistic approach for analysing existing and deriving novel gamification designs. *European Journal of Information Systems* 29, 6 (2020), 641–668.

[64] Daniela Schröder. 2022. Challenges of German first-year law students. *European Journal of Legal Education* 3, 1 (2022), 23–47.

[65] Matthias Söllner, Axel Hoffmann, and Jan Marco Leimeister. 2016. Why different trust relationships matter for information systems users. *European Journal of Information Systems* 25, 3 (2016), 274–287.

[66] Matthias Söllner, Abhay Nath Mishra, Jan-Michael Becker, and Jan Marco Leimeister. 2022. Use IT again? Dynamic roles of habit, intention and their interaction on continued system use by individuals in utilitarian, volitional contexts. *European Journal of Information Systems* (2022), 1–17.

[67] Carl-Friedrich Stuckenberg. 2020. Der juristische Gutachtenstil als cartesische Methode. *ZDRW Zeitschrift für Didaktik der Rechtswissenschaft* 6, 4 (2020), 323–341.

[68] Alon Talmor, Oyvind Tafjord, Peter Clark, Yoav Goldberg, and Jonathan Berant. 2020. Leap-of-thought: Teaching pre-trained models to systematically reason over implicit knowledge. *Advances in Neural Information Processing Systems* 33 (2020), 20227–20237.

[69] Stephen E. Toulmin. 2003. *The uses of argument: Updated edition*. 1–247 pages. https://doi.org/10.1017/CBO9780511840005

[70] Stefanie Urchs, Jelena Mitrović, and Michael Granitzer. 2020. Towards Classifying Parts of German Legal Writing Styles in German Legal Judgments. In *2020 10th International Conference on Advanced Computer Information Technologies (ACIT)*. IEEE, 451–454.

[71] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention Is All You Need. *In Advances in neural information processing systems* Nips (2017), 5998–6008.

[72] Viswanath Venkatesh and Hillol Bala. 2008. Technology acceptance model 3 and a research agenda on interventions. *Decision sciences* 39, 2 (2008), 273–315.

[73] Bart Verheij. 2003. Artificial argument assistants for defeasible argumentation. *Artificial intelligence* 150, 1-2 (2003), 291–324.

[74] Jan Vom Brocke, Alexander Simons, Kai Riemer, Bjoern Niehaves, Ralf Plattfaut, and Anne Cleven. 2015. Standing on the shoulders of giants: Challenges and recommendations of literature search in information systems research. *Communications of the association for information systems* 37, 1 (2015), 205–224.

[75] Thiemo Wambsganss. 2021. Designing Adaptive Argumentation Learning Systems Based on Artificial Intelligence. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–5.

[76] Thiemo Wambsganss, Andreas Janson, and Jan Marco Leimeister. 2022. Enhancing argumentative writing with automated feedback and social comparison nudging. *Computers & Education* 191 (2022), 1–17.

[77] Thiemo Wambsganss, Nikolaos Molyndris, and Matthias Söllner. 2020. Unlocking transfer learning in argumentation mining: a domain-independent modelling approach. In *15th International Conference on Wirtschaftsinformatik*. 1–16.

[78] Thiemo Wambsganss, Christina Niklaus, Matthias Cetto, Matthias Söllner, Siegfried Handschuh, and Jan Marco Leimeister. 2020. AL: an adaptive learning support system for argumentation skills. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14.

[79] Thiemo Wambsganss, Christina Niklaus, Matthias Söllner, Siegfried Handschuh, and Jan Marco Leimeister. 2020. A corpus for argumentative writing support in German. *Proceedings of the 28th International Conference on Computational Linguistics* (2020), 856–869.

[80] Thiemo Wambsganss and Roman Rietsche. 2019. Towards designing an adaptive argumentation learning tool. Proceedings of the International Conference on Information Systems (ICIS) 2019, 1–9.

[81] Thiemo Wambsganss, Matthias Söllner, Kenneth R Koedinger, and Jan Marco Leimeister. 2022. Adaptive Empathy Learning Support in Peer Review Scenarios.

In *CHI Conference on Human Factors in Computing Systems*. 1–17.

[82] Dirk Weber. 2918. *Methodik der Fallbearbeitung*. Number 1. Nomos Verlagsges. MBH+ Co.

[83] Florian Weber, Thiemo Wambsganss, Seyed Parsa Neshaei, and Matthias Soellner. 2023. Structured Persuasive Writing Support in Legal Education: A Model and Tool for German Legal Case Solutions. In *Findings of the Association for Computational Linguistics: ACL 2023*. 2296–2313.

[84] Florian Weber, Thiemo Wambsganss, and Matthias Soellner. 2023. Design and Evaluation of an AI-based Learning System to Foster Students' Structural and Persuasive Writing in Law Courses. In *Proceedings of the Forty-Fourth International Conference on Information Systems (ICIS)*. 1–17.

[85] Florian Weber, Thiemo Wambsganss, and Matthias Soellner. 2023. Supporting Human Cognitive Writing Processes: Towards a Taxonomy of Writing Support Systems. In *Proceedings of the Forty-Fourth International Conference on Information Systems (ICIS)*. 1–17.

[86] Jane Webster and Richard T Watson. 2002. Analyzing the past to prepare for the future: Writing a literature review. *MIS quarterly* (2002), 13–22.

[87] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, et al. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations*. 38–45.

[88] Sarah Shi Hui Wong and Stephen Wee Hun Lim. 2019. Prevention-permission-promotion: A review of approaches to errors in learning. *Educational Psychologist* 54, 1 (2019), 1–19.

[89] Meng Xia, Qian Zhu, Xingbo Wang, Fei Nei, Huamin Qu, and Xiaojuan Ma. 2022. Persua: A Visual Interactive System to Enhance the Persuasiveness of Arguments in Online Discussion. *Proceedings of the ACM on Human-Computer Interaction* (2022), 1–30.

[90] Joachim Zekoll and Gerhard Wagner. 2018. *Introduction to German law* (3 ed.). Kluwer Law International BV.

APPENDIX

Evaluation metrics

Table 7: Evaluation metric used to evaluate the case solution to make the quality of the appraisal style comparable.

Scoring	Quality of the appraisal style	Percentage distribution over the text
1	There are nearly no components (*major claim*, *definition*, *subsumption*, and *conclusion*) of the appraisal style in the text, and no connections between the components can be seen (e.g., raised major claims are not closed).	0-10%
2	There are a few components of the appraisal style in the text (*major claim*, *definition*, *subsumption*, and *conclusion*). Some connections between the components become apparent (e.g., raised major claims are not closed).	10-40%
3	In the text, all components of the appraisal style are included, and connections between the components can be seen. However, not all components are connected, or some text passages cannot be assigned to the components.	40-70%
4	Most of the text can be divided into the components of the appraisal style, and many connections between the components can be seen.	70-90%
5	The complete text can be divided into the components of the appraisal style, and all components have logical connections to each other.	90-100%

Survey items

Table 8: Overview of items measured in the study.

Section	Variable	Items	Scale
Post-survey	Demographics	1. Age 2. Gender 3. Language	open
Pre-survey	Previous experience with legal writing	*"Did you learn appraisal style in your degree?"* *"How confident do you feel in writing persuasive and structured case solutions in the appraisal style?"*	Open Question and 1 - 7 Likert scale (7: highest)
Pre-survey	Attitude Towards Technology	*"I like to experiment with new technologies and try them out."* *"Usually, I am hesitant to try new technologies."* *"When I hear about new technologies, I look for a way to experiment with them."* *"Among my friends, I am usually the first to try new digital media / new technologies."*	1 - 7 Likert scale (7: highest)
Introduction	Assignment	*"Before you can test the system, answer the following questions. Explain briefly what is meant by a major claim. Explain briefly what is meant by a definition. Explain briefly what is meant by a subsumption. Explain briefly what is meant by a conclusion."*	Open questions
Writing exercise	Assignment	*"In the following, you can solve your uncle's case. Use the appraisal style (theorem, definition, subsumption, and conclusion). The LegalWriter system will help you write your uncle's case and will also provide you with the exact facts of the case in the form of a case study. Your case solution should be about 300-400 words (the system will show you your word count)."*	Open question
Post-survey	Enjoyment [36]	*"I enjoyed and enjoyed interacting with the system."* *"Interaction with the system was exciting."*	1 - 7 Likert scale (7: highest)
Post-survey	Intrinsic motivation [40]	*"I would describe this writing task as very interesting."* *"I think this activity could be valuable for me."* *"I think writing with the system is a boring activity."* *"I think this activity is useful for understanding how to write persuasive and structured case solutions in the appraisal style."* *"I liked the content on legal argumentation and the appraisal style."* *"I think this activity could be beneficial for me."*	1 - 7 Likert scale (7: highest)
Post-survey	Feedback accuracy [54]	*"LegalWriter's evaluation of my case solution reflects my actual performance."* *"LegalWriter has accurately evaluated my performance."* *"The recommendations I received from LegalWriter was an accurate assessment of my performance."* *"I assume that LegalWriter will help me improve my ability to write persuasive and structured case solutions in the appraisal style."*	1- 7 Likert scale (7: highest)
Post-survey	Control questions	*"Please check "Strongly agree."* *"A certain word was mentioned in the LegalWriter tutorial video. Please write this word in the text box below."*	1 - 7 Likert scale (7: highest)
Post-survey	Qualitative Impression	*"What do you like in the interaction with LegalWriter?"* *"What would you improve in LegalWriter?"* *"Do you have any additional ideas? What else would you like to add to the system?"*	Open Questions

Recommendations used in LegalWriter

Table 9: Recommendations in the ML-based version of LegalWriter.

Component of the appraisal style	Recommendation	Calculation
Major Claim	A conclusion is not opened by a major claim. Formulate a corresponding major claim in the conjunctive.	"if one conclusion is not opened by major claims "
Major Claim	Two conclusions are not opened by major claims. Formulate two corresponding major claims in the conjunctive.	"if two conclusions are not opened by major claims"
Major Claim	Three conclusions are not opened by major claims. Formulate three corresponding major claims in the conjunctive.	"if three conclusions are not opened by major claims"
Major Claim	Several conclusions are not opened by major claims. Formulate three more corresponding major claims in the conjunctive.	"if three or more conclusions are not opened by major claims"
Definition	Check that you have explained all the major claims by a definition. This is important so that all the facts raised in the major claim can be explained. Note: Some subordinate clauses do not need an additional definition, so check carefully if you need to add another definition.	"if one or more major claims are not explained by a definition"
Subsumption	It seems a conclusion is not supported argumentatively by a subsumption; use legal claims and premises to be able to explain your conclusion.	"if a subsumption between definition and conclusion is missing "
Subsumption	It seems that several conclusions are not supported argumentatively by subsumption; use legal claims and premises to be able to explain your conclusion.	"if two or more subsumptions between definitions and conclusions are missing"
Conclusion	One major claim is not closed by a conclusion. Think about which major claims have not yet been brought to a conclusion and formulate the corresponding conclusion.	"if one major claim is not closed by a conclusion"
Conclusion	Two major claims are not closed by conclusions. Think about which major claims have not yet been brought to a conclusion and formulate the corresponding results.	"if two major claims are not closed by a conclusion"
Conclusion	Three major claims are not closed by conclusions. Think about which major claims have not yet been brought to a conclusion and formulate the corresponding results.	"if three major claims are not closed by a conclusion"
Conclusion	Several major claims are not closed by conclusions. Think about which major claims have not yet been brought to a conclusion and formulate the corresponding results.	"if three or more major claims are not closed by a conclusion"
None	Several sentences or one sentence cannot be assigned to the components of the argument structure. Consider how you can rephrase the sentences accordingly or whether the sentences may be superfluous.	"if a sentence is in class none (hover)"

Table 10: Recommendations in the ML-based version of LegalWriter (legal claim and premise).

Component of the appraisal style	Recommendation	Calculation
Legal claim	A legal claim is not backed by a premise. Try to derive one or more premises for the legal claim from the case study. If you cannot find supporting premises in the case study, your legal claim may not be admissible.	"if a claim is not backed by premises"
Legal claim	Two legal claims are not backed by a premise. Try to derive one or more premises for the legal claim from the case study. If you cannot find supporting premises in the case study, your legal claim may not be admissible.	"if two claims are not backed by premises"
Legal claim	Three legal claims are not backed by a premise. Try to derive one or more premises for the legal claim from the case study. If you cannot find supporting premises in the case study, your legal claim may not be admissible.	"if three claims are not backed by premises"
Legal claim	Several legal claims are not backed by a premise. Try to derive one or more premises for the legal claim from the case study. If you cannot find supporting premises in the case study, your legal claim may not be admissible.	"if three or more claims are not backed by premises"
Premise	A premise does not appear to have a connection to a legal claim. Try to draw a legal claim from the premise or consider whether the single premise that does not cover a legal claim is necessary.	"if a premise has no connection to a legal claim"
Premise	Two premises do not appear to have a connection to one or more legal claims. Try to draw a legal claim from the premise or consider whether the two single premises are necessary.	"if two premises have no connection to one or more legal claims"
Premise	Three premises do not appear to have a connection to one or more legal claims. Try to draw a legal claim from the premise or consider whether the three single premises are necessary.	"if three premises have no connection to one or more legal claims"
Premise	Several premises do not appear to have a connection to one or more legal claims. Try to draw a legal claim from the premise or consider whether these premises are necessary.	"if three or more premises have no connection to one or more legal claims"

Table 11: Dashboard recommendations in the ML-based version of LegalWriter.

Diagram chart	Recommendation	Calculation
Distribution of components	The distribution of the individual components in your case solution seems reasonable. The argumentative part in the subsumption takes an acceptable part of your case solution.	"if 60-80% of the text can be classified by the ML-models as component subsumption "
Distribution of components	The distribution of the individual components in your case solution seems partly reasonable. The argumentative part in the subsumption takes an acceptable part of your case solution. Nevertheless, try to formulate more legal claims and premises	"if 50-60% of the text can be classified by the ML-models as component subsumption "
Distribution of components	You failed to make your conclusions argumentative. Try to focus more on the subsumption.	"if >50% of the text can be classified by the ML-models as component subsumption"

Table 12: Recommendations in the static version of LegalWriter.

Components	Recommendation
Major Claim	The major claim explains the elements of the offense (fact) that are to be fulfilled. The major claim must always be written in the subjunctive. It must raise a question/possible consequence, which is answered in the further steps by the conclusion and must be formulated as a statement. Tip for improvement: Remember that every major claim must be closed by a conclusion. Set it up so that it is written in the subjunctive and describes a legal problem that you need to solve.
Definition	The definition defines the constituent elements to be fulfilled, according to which the point of view raised in the major claim is considered in more detail from a legal point of view. Here, the focus should be on the essentials, knowledge without regard to relevance is out of place here. Tip for improvement: Check that you have explained all the major claims by a definition. This is important so that all the facts raised in the major claim can be explained.
Subsumption	In the subsumption, it is examined to what extent the conditions (elements) of the definition are given. Here, the facts of the case are weighed against the preconditions from the definitions and the premises (facts). Tip for improvement: Remember that every claim must be covered by at least one premise. Check if all claims are related to a suitable premise.
Conclusion	The conclusion is the answer to the major claim. Thus, the case solution reaches a final result here. The question formulated in the major claim is answered. The conclusion is always written in the indicative. Reasons are out of place here, they only belong in the definition or subsumption. Tip for improvement: Remember that you have to close every major claim you have opened with a solution.
Distribution of components	Remember that the subsumption must be the main part of your case solution. Check that you have adequately explained all your conclusions with premises and legal claims. Next, the definitions should take up most of your case solution, so that the facts to be fulfilled are defined.

Examples of each component from the corpus

Table 13: Three examples for each of the components of legal opinions in appraisal style, randomly selected from the corpus we used. We translated the examples from German to English for presentation in this paper.

Component	Three Examples for Each Component
Major Claim	1) It is questionable whether there is an offer from R. 2) B would have to have declared the acceptance on behalf of A. 3) For this to happen, an effective purchase contract must have been concluded.
Definition	1) An offer is a declaration of intent that requires a receipt, specifying the essential negotii (purchase price, item purchased, and contractual parties), which is presented in such a way that the recipient only has to agree. 2) The decisive factor is the assessment from the objective perspective of the recipient. 3) A purchase contract is created through an offer by the seller and acceptance by the buyer or by an order from the buyer and an order acceptance by the seller, since the buyer or the seller can take action first.
Legal Claim	1) So, it is just an insignificant motive error. 2) However, the email and therefore the offer was accepted in a timely manner. 3) This means that A is given the power of attorney by V.
Premise	1) The external facts were fulfilled by submitting the offer when sending the email. 2) Since there was also a willingness to purchase, the internal requirement is also fulfilled. 3) Trainee A is 16 years old and therefore has limited legal capacity.
Conclusion	1) In this respect, there are initially two consistent declarations of intent, which is why a purchase contract was concluded. 2) Consequently, the declaration of challenge was sent to the correct opponent. 3) A therefore had the power of representation to sell the vehicle to K at a price of $ 27,000.