Dec 11th, 12:00 AM

Design and Evaluation of an AI-based Learning System to Foster Students' Structural and Persuasive Writing in Law Courses

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

Thiemo Wambsganss
EPFL, thiemo.wambsganss@bfh.ch

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

Design and Evaluation of an AI-based Learning System to Foster Students' Structural and Persuasive Writing in Law Courses

Completed Research Paper

Florian Weber
University of Kassel
Pfannkuchstraße 1,
34121 Kassel, GER
weber@uni-kassel.de

Thiemo Wambsganss
Bern University of Applied Sciences
Brückenstrasse 73,
3005 Bern, CH
thiemo.wambsganss@bfh.ch

Matthias Söllner
University of Kassel
Pfannkuchstraße 1,
34121 Kassel, GER
soellner@uni-kassel.de

Abstract

Structured and persuasive writing is essential for effective communication, convincing readers of argument validity, and inspiring action. However, studies indicate a decline in students' proficiency in this area. This decline poses challenges in disciplines like law, where success relies on structured and persuasive writing skills. To address these issues, we present the results of our design science research project to develop an AI-based learning system that helps students learn legal writing. Our results from two different experiments with 104 students demonstrate the usefulness of our AI-based learning system to support law students independent of a human tutor, location, and time. Apart from furnishing our integrated software artifact, we also document our assessed design knowledge in the form of a design theory. This marks the first step toward a nascent design theory for the development of AI-based learning systems for legal writing.

Keywords: Design Science Research, Legal Education, Learning from Errors, AI-based Learning System

Introduction

Students' ability to write in a structured and persuasive manner has declined in recent decades (Carter and Harper 2013). This is mainly driven by the increasing use of digital media, which fosters the writing of short, unstructured, and informal texts (Akram and Kumar 2017). The lack of ability to write structured texts leads to inefficient communication, lowered persuasiveness of texts, and difficult comprehensibility of content (Kendeou and Van Den Broek 2007). Students in the legal domain are especially challenged to write structurally and persuasively to present legal doctrine's complex requirements and specific problems. The legal writing instructions varies across countries and is influenced by the specific area of law being studied. Consequently, students employ different methods to learn legal writing based on regional and legal contexts. While in America the "*IRAC*[1] *formula*" is used to teach students legal writing, countries such as China use the "*appraisal style*", while France, for example, uses the "*cas pratique*". Among the most

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

important concepts of German legal writing are the appraisal style and the judgment style, with the appraisal style being especially important for legal education (Urchs et al. 2020). Given that the term "appraisal style" is unique to the German legal context, it lacks a direct English equivalent. To provide a succinct definition, the appraisal style refers to "the form and writing style of a legal opinion" (Stuckenberg 2020). The complexities associated with composing a legal opinion mainly stem from the stringent formalities inherent in the field of legal writing, as well as the challenge of constructing a coherent and precise argument rooted in facts and laws (Pinkwart et al. 2009).

Gupta and Bostrom (2009) have stimulated work in the field of education in information systems research with their research on technology-mediated learning (e. g., Huang et al. 2013, Kabudi 2021, Schlegel et al. 2023). Based on this guidance, researchers in the fields of educational technology and information systems (IS) have developed IS to support students in persuasive writing (Osborne et al. 2016; Schmitt et al. 2021; Wambsganss, Söllner, et al. 2020). Nevertheless, these systems are of limited interest to law students since the writing style in law differs from general writing due to strict formalisms. Researchers and educators claim that the targeted use of IS in law education falls short of expectations mainly due to the missing availability of IS design for the particular domain, such as CATO[2] (Aleven 2003) or ArguMed, a template-based argument mediation system (Verheij 2003). However, these systems face the challenge of not being able to provide students with individual feedback on the errors in their written legal texts. In fact, individual feedback on student errors in texts has been proven to effectively support students in their learning and writing processes (Metcalfe 2017). One way to support students in writing persuasive case solutions is to develop an AI-based learning system based on machine learning (ML) models. Even though there are several algorithms for analyzing text in the legal field, e. g., for classifying judgments (Urchs et al. 2020), summarizing legal texts (Hachey and Grover 2005) and assessing jury verdicts (Poudyal et al. 2019), among many others, there are no algorithms that are specifically trained to help students in writing persuasive case solutions. Furthermore, the literature is scarce in suggesting how such complex systems should be designed to support learners in writing and skill development. It is also difficult to transfer existing design knowledge to other fields, such as legal writing. We aim to address the gaps of limited learning support for students in law courses by designing and evaluating a new form of AI-based learning system for structured and persuasive writing. By leveraging advances in natural language processing (NLP) and machine learning (ML), we aim to generate design knowledge for an AI-based learning system that provides personalized support to students in writing structured and persuasive legal texts and developing their legal writing skills. To achieve our goal, we aim to answer the following research questions (RQ):

RQ1: How should an AI-based learning system be designed to improve law students' structured and persuasive writing skills in large-scale law courses?

RQ2: To what extent does an AI-based learning system help law students improve their structural and persuasive writing skills?

To achieve this, we adopt the design science research (DSR) approach by Hevner (2007) and intend to iteratively design and evaluate an AI-based learning system with 104 law students in a lab and a field experiment. In the following, we explain the theoretical background and how learning from errors serves as our guiding kernel theory for designing and evaluating an AI-based learning system (Metcalfe 2017). Next, we outline our research design and describe the eight specific steps we followed in our DSR process. In designing our AI-based learning system, we draw on scientific literature and insights from the field. As we describe our design and evaluation process in detail, we document the generated design knowledge as design principles, following the proposal by Gregor and Hevner (2013). Finally, we summarize our results, discuss limitations, and suggest areas for future research.

Theoretical Background

Legal Writing

Traditionally, students are tasked with resolving legal issues or case studies by writing persuasive case solutions in form of a legal opinion (Enqvist-Jensen et al. 2017). To simplify comprehension, we will use

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

the terms "legal text" or "legal case solution" throughout the paper, both indicating a legal opinion. In these legal case solutions, students are required to employ specialized and deeply concept-driven knowledge. The theoretical knowledge mainly pertains to the accurate application of paragraphs and the establishment of priorities within the case solution. Conversely, concept-driven knowledge predominantly involves the principles of structuring case solutions in a methodical manner. To achieve this, students need to adhere to established legal concepts (Weber et al. 2023). Different methods exist worldwide to teach law students how to write structurally and persuasively. In German jurisprudence, two of the most significant concepts are the appraisal style and the judgment style, with the former being particularly emphasized in legal education (Urchs et al. 2020). The appraisal style is employed for tackling complex legal problems and comprises four distinct components: major claim, definition, subsumption, and conclusion (Urchs et al. 2020). A case solution following the appraisal style invariably commences with a question, known as the major claim. This element outlines the factual details necessary to address a legal problem and is phrased in the subjunctive. Definitions are used to articulate the specifics of the required facts. They are formulated in relation to the points of view raised in the major claim, to be able to assess the legal problems against the background of the law. In the subsumption, the facts of the case are weighed against the definitions and the conditions argumentatively. This weighing follows established models in argumentation theory (Toulmin 2003). These theories show a simple and basic structure of an argument. Accordingly, an argument consists of several components: a claim and at least one premise that supports or refutes it. This simple logic can also be found in jurisdiction. Here, an argument in a subsumption consists of a legal claim and one or more premises. A premise supports the claim's validity in jurisprudence through a statement of fact, a judgment or majority opinions of legal scholars. It is a legitimization that makes a legal claim comprehensible. The characteristic of legal argumentation lies in the fact that the conclusion is derived from the premises. The conclusion is therefore the logical result of the previously mentioned premises. The conclusion is the answer to the posed major claim. Thus, the case solution here comes to a final conclusion. Table 1 provides a succinct explanation of the four components of the appraisal style.

Elements	Explanation	Example
Major claim	The major claim explains the elements of the offense that are to be fulfilled. It raises a question or possible consequence.	D could have the right of compensation for damages against H according to § 280 I 1 BGB.
Definition	The definition establishes the essential conditions to be present in the legal issue for the case solution to be concluded. These elements are contingent upon the question posed in the major claim.	For this right of compensation, there must be an contractual relationship between the parties and a breach of duty on the part of H in accordance with § 241 II. A contractual obligation describes the individual performance relationship between creditor and debtor.
Subsumption (premise and legal claim)	During the subsumption phase, an assessment is conducted to determine the extent to which the conditions of the definition are satisfied. In this process, the case's particulars are measured against the prerequisites outlined in the definitions and underlying premises. Legal consequences can draw from the premises, so-called legal claims.	*Since there are no indications of an invalid contractual relationship from the facts of the case, a valid contract for work and services between D and H is to be assumed. Thus, there is a valid contract for work and services pursuant to § 631 and consequently a valid contractual relationship between the parties.*
Conclusion	A conclusion serves as the response to the major claim. The case solution reaches a final result here.	D therefore has a right of compensation for damages against H in accordance with § 280 I 1 of the German Civil Code (BGB).

Table 1. Elements of a legal opinion in the appraisal style based on Weber et al. 2023.

Learning Systems for Legal Writing

Universities and other educational institutions encounter the task of imparting legal writing and reasoning skills. This is partially attributed to instructors' limited pedagogical expertise in teaching persuasive writing within university programs. Additionally, the demand to cover the essential curriculum often leaves minimal room for practicing persuasive writing (Jonassen and Kim 2010). This is true even for topics where persuasive writing is mandated by the curriculum, like law or logic, where teachers' ability to teach persuasive writing is limited by time and availability constraints. As a result, researchers and educators are advocating for an increased emphasis on persuasive and structured writing within the education system (Driver et al. 2000). Consequently, research groups have developed systems to support students in persuasive writing and writing systematics. These systems have been applied in various fields, such as science (Osborne et al. 2016) or business reporting (Wambsganss, Niklaus, et al. 2020). However, researchers and legal educators note that the use of IT systems in legal education falls short of expectations (Beurskens 2016), and this is also true for teaching argument structure and writing persuasive case solutions. Nevertheless, some systems are designed to help students learn persuasive legal writing and structured argumentation. Most of these systems employ methods of argument diagramming (representational guidance approaches) (Pinkwart et al. 2009; Reed et al. 2007). Students are supported by providing them with representations of their reasoning structures, with the goal of supporting their reasoning. A typical example is helping students to represent their reasoning structure in terms of node and link graphs (Pinkwart et al. 2008; Reed et al. 2007). Pioneering work in the legal field has shown that argument diagramming can improve students' ability to make high-quality arguments and can improve the coherence of law students' persuasive writing (Carr 2003; Gordon et al. 2007). Pinkwart et al. 2008 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 (Aleven 2003). This system assists students in argumentation with cases by teaching them to compare their arguments with given cases and offer existing arguments so that the own solution can be improved (discussion scripting approach) (Aleven 2003). A system from the field of e-learning aims to use gamification elements to introduce students to the IRAC formula (Bouki et al. 2014). To sum up, besides some representational guidance approaches or discussion scripting approaches, there seem to be no suitable systems that adaptively support students in ML-based law courses in writing structured and persuasive case solutions. Hence, past literature falls short of a rigorous design study on how to design an AI-based learning system for structured and persuasive law case solutions and lacks rigorous empirical investigations of the effects of adaptive learning support on students' writing style and use experiences. Therefore, we aim to address this literature gap by designing and evaluating an AI-based learning system that helps students learn how to write in a structured and persuasive appraisal style and gives feedback based on the individual errors of the students.

Natural Language Processing and Machine Learning on Law Texts

To develop an AI-based learning system, we aim to build on the literature of NLP and ML to train a novel model that can identify students' legal argumentative components and structures. Given its strict logical structure, law presents a promising field for annotating arguments (Moens et al. 2007; Urchs et al. 2020), the availability of evaluated open-access corpora for training models on legal texts is scarce (Palau and Moens 2009; Reed 2006; Reed et al. 2007; Weber et al. 2023). There are, however, some publicly accessible corpora and models. Hachey and Grover (2005) present a model based on a corpus of 188 annotated English court opinions to construct a system for the automatic summarizing of court judgments. Also, recent advances in NLP are being used to build predictive models that can reveal patterns for judicial decisions (Virtucio et al. 2018). This intelligent support is designed to help lawyers and judges quickly identify cases and recognize patterns to make faster and more accurate decisions (Aletras et al. 2016). Several research groups have employed machine learning models to analyze cases and make predictions regarding their potential outcomes, providing explanations for their predictions (Alarie et al. 2016; Ashley and Brüninghaus 2009). These researchers hope that this will improve computerized legal research. There are also several German corpora in addition to the mostly English-language corpora for recognizing decisions and legal cases (Houy et al. 2013; Urchs et al. 2020). In our literature research, we also included generative models such as GPT3. During the study, an examination of existing literature failed to uncover any evidence supporting the effectiveness of generative models in accurately classifying law texts written by students.

Learning from Errors

To guide the design and development of our AI-based learning system, we rely on the literature on learning from errors, since error-based learning seems to be a suitable underlying concept for helping students to foster their legal writing skills (Ericsson et al. 1993; Ohlsson 1996). In learning and acquiring skills, practice and application play a crucial role (Ericsson et al. 1993). Typically, law students need to solve various cases to internalize the specific type of legal argumentation and the appraisal style. In theoretical lectures, students learn the basic skills for solving cases and get examples from the lecturer. Due to the extensive number of lectures and the significant time demands involved, students seldom receive personalized feedback from their lecturers regarding their cases. Research shows that practicing a specific skill through repeated attempts improves a skill and eventually leads to mastery (Ohlsson 1996). However, errors are bound to occur during these repeated exercises, especially for less experienced students (novices). Current research shows how errors followed by corrective feedback can be effective learning support (Metcalfe 2017). Wong and Lim (2019) distinguished between the approaches of prevention, permission, and promotion of errors. The error-allowance approach permits learners to make mistakes naturally, which are improved through corrective feedback (Lorenzet et al. 2005; Potts and Shanks 2014). We want to create an environment that allows law students to improve their skills to solve legal problems and write persuasive case solutions. In this learning environment, we would like to allow natural mistakes from which students can learn (Metcalfe and Xu 2018). 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 (Kornell et al. 2009). In addition, errors generate enhanced attention because errors reinforce the encoding of subsequent corrective feedback. Additionally, learners might show increased interest in receiving corrective feedback after making an error, driven by their curiosity to obtain the correct answer (Potts and Shanks 2014).

Research Methodology

Our research project follows the DSR approach proposed by Hevner (2007). DSR is highly effective in addressing practical issues and enhancing the current body of knowledge by designing and evaluating a novel research artifact. Figure 1 illustrates the eight-step process we followed to ensure the creation of design knowledge based on insights from the application domain and the knowledge base.

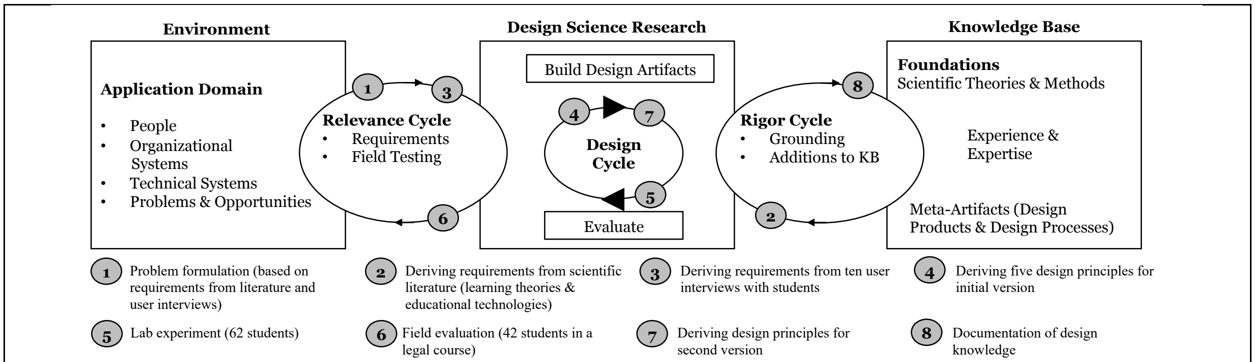

Figure 1. Three cycle design science process (DSR) according to Hevner (2007).

We began by analyzing the requirements for an AI-supported learning system in the legal domain, considering the legal learning context (environment) and existing knowledge in the scientific literature. Building upon these requirements, we formulated design principles, which were then implemented and evaluated in our AI-based learning system in a series of design cycles. Our project contributes to research by providing new design knowledge in the form of design principles that prescribe how to design artifacts in this class (Gregor et al. 2020). We followed a theory-driven design approach by grounding our research on the theory of learning from errors by Metcalfe 2017, which motivated the overall design and evaluation of our DSR approach. In applying the DSR cycle, we followed the first step by formulating the problem and describing its meaning in the introduction and theoretical background. We then derived meta-requirements from the scientific literature for the design of AI-based learning systems in the field of education and gathered user stories and requirements from semi-structured interviews with ten law students. We used these inputs to derive design principles and implemented our AI-based learning system called *LegalWriter*. Within the system's development, we initially defined eight design features as a concrete manifestation of our design principles (e. g., Meth et al. 2015). Therefore, in this paper we use the term "design feature" to refer to a collection of features that a legal learning system could potentially provide. To gauge its performance, we employ the FEDS framework introduced by Venable et al. in 2016 for evaluation. Hence, in the fifth step, we performed a proof-of-concept artificial evaluation with 62 students to measure the attitudes of the target groups toward the system and to make initial statements about the short-term effectiveness of the system (Venable et al. 2016). Based on these findings, we refined our system and worked mainly on the front-end and usability. We then evaluated *LegalWriter* in a field experiment with 42 students in a law course (Venable et al. 2016). The field experiment took four weeks and was designed to provide insight into the long-term effectiveness of the system. We conducted the experiment in two groups, the control group learned in a classical law course and the treatment group learned in a course that works with our system. In step 7, we revised the design principles. In the last step, we document our design knowledge as a nascent design theory (Gregor and Hevner 2013). Overall, our research project contributes to the design of AI-based learning systems for education that can be embedded to enable students to receive intelligent feedback based on their individual errors. The evaluated design knowledge from our research project is summarized in a novel design theory for AI-based legal learning systems based on the theory of learning from errors to support learning of structured and persuasive writing.

Design and Evaluation of *LegalWriter*

In the following chapter, we explain the design and evaluation of our AI-based learning system *LegalWriter*, following eight steps of the DSR approach according to Hevner (2007) (see Figure 1).

Step 1: Problem Formulation

We initiated the process by formulating the problem. The problem is derived from the introduction, and the theoretical background further establishes the foundation for our system's requirements. Additionally, the problem scope was expanded through student interviews, as detailed in Steps 2 and 3.

Step 2 & 3: Deriving Requirements from Scientific Literature and User Interviews

To identify the requirements for an AI-based learning system for structured and persuasive legal writing, we conducted a systematic literature search using the methodological approaches of Cooper (1988) and Vom Brocke et al. (2015). The process involved four steps: defining the review scope, conceptualizing the topic, searching the literature, and analyzing the findings related to requirements. In the first step, we defined the review scope, primarily focusing on studies that demonstrate the successful implementation of AI-based learning systems in different writing scenarios. In the second step, we conceptualized the topic and identified two broad areas for deriving requirements: *Human Computer Interaction*, *Machine Learning, and Information Systems*. Since creating such a system is a complex project, we needed knowledge from different research areas, such as *psychology*, *computer science*, and *writing studies*. In the third step, we conducted a literature search on several databases, including *ArXiv*, *Science Direct*, *ACM Digital Library*, *ProQuest*, and *IEE Xplore*. We used the following search terms: *"Writing System", "Writing Assistance", "Writing Tool", „Writing Process", „Learning from Errors", „Learning Theory", "Legal Learning",* and "*Learning Theories*". We established criteria for inclusion and exclusion, subsequently evaluating the titles and abstracts of our search outcomes, ultimately selecting 64 papers for more intensive analysis. We summarized similar topics of these contributions as meta-requirements for the design of an AI-based learning system for structured and persuasive legal writing. The meta-requirements encompass the integration of a suitable legal pedagogy (Cannon 1955; Xu and Yan 2008), implementing a feedback system that detects individual errors (Kornell et al. 2009; Metcalfe 2017), and integrating guidance through the writing process (Cagiltay 2006; Flower and Hayes 1981). In the fourth step, we conducted ten user interviews with law students to derive requirements from the future user group. Following the methodology of Rubin and Chisnell (2008), we used a semi-structured questionnaire with three sets of questions related to the students' learning requirements, considerations for implementing technology-enhanced education systems, and design requirements for an AI-based learning system. The interviews lasted between 16 and 95 minutes (mean = 36.6; SD = 20.89), and the interviewees were students from different German universities, including both law and business law programs. 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 all interviews and transcribed them using the approach of Mayring (2010). Based on the transcription, we derived categories that were found in all interviews and identified user stories from the interviews. We used open coding to form a unified coding system during the analysis and collected user stories to create a first clickable prototype and visualize the design ideas. The analysis of the interviews revealed the students' first important requirements, including a comprehensible user experience, the possibility to train in different areas of law (**DP1**), and in-text highlighting of the components of the appraisal style (**DP2**).

Design Principles (DP)	Theory Lens / Requirements (UR/MR)
For educational designers to effectively design AI-based legal learning systems for students …	Provide the system with …
DP1: … they should embed realistic legal case studies into the system so that students can solve legal problems from different areas of law individually and thus digitally apply an established learning technique in law.	**MR1:** … case studies that describe realistic legal problems and represent an established legal education method (Cannon 1955; Xu and Yan 2008).
	UR1: … different cases from different areas of law.
DP2: … they should provide individualized feedback on errors in adherence to the structure of legal writing in the appraisal style so that students can intuitively identify errors in adherence to the structure of legal writing.	**MR2:** … a feedback system which evaluates the students' errors (Metcalfe 2017).
	MR3: … an analysis system that checks for consistent adherence to structured legal writing and its components (major claim, definition, subsumption, and conclusion) as well as the connections between the individual components.
	UR2: … a direct and individual feedback of student errors.
DP3: … they should integrate support that gives students recommendations on how to adhere to the structure of the appraisal style, how to formulate the components of the appraisal style and how to build a stringent legal argumentation so that students can improve their legal texts in a self-directed manner.	**MR4:** … conceptual scaffolds in the form of recommendations that help students improve their persuasive and structured writing (Cagiltay 2006).

Table 2. Derived design principles according to Gregor et al. 2020.

Step 4: Deriving Design Knowledge and Implementation of LegalWriter

Based on these findings in step 2 & 3, we derived three design principles for an AI-based learning system for structured and persuasive legal writing. Our design principles are based on the requirements derived from the literature and from the interviews with the law students. The design principles are illustrated in Table 2. For formulating the design principles, we relied on the conceptual schema of Gregor et al. 2020. Following the scheme, we have formulated our DP with the aim, the context, the mechanism, and the rationale (Gregor et al. 2020), so we assume that the DP are self-explanatory. To instantiate our design knowledge, we built the writing system *LegalWriter,* which was developed as a web-based application. A screenshot of *LegalWriter* and its core design features (**DF1 - DF8**) can be seen in Figure 2.

LegalWriter consists of three components: a text editor, a checklist, and a legal argumentation dashboard. The structure of *LegalWriter* is based on a typical writing task, which consists of the components planning, translating, and reviewing (Flower and Hayes 1981). *LegalWriter* focuses on the category of reviewing in the writing process and places the analysis and improvement of the written text in the foreground. To achieve this goal the system is fundamentally aligned with the theory of learning from errors (Metcalfe 2017). The feedback mechanism is specifically designed to enable the system to provide continuous feedback and recommendations to law students regarding their individual case solutions. Thus, *LegalWriter* follows the classical advantages of intelligent IT-based learning and can enable time and location-independent skill training (Arkorful and Abaidoo 2015). To receive feedback, students can enter a case in a text editor or copy an already solved case into the text editor (**DF1**). Directly above the text editor are the buttons "*show case study*" and "*useful paragraphs*". The buttons allow students to call up a predefined case from a specific field of law and train with it. At the same time, the learner can call up content tips via the "*useful paragraphs*" button (**DF2**), so that students are not cognitively overwhelmed when writing. The "*useful paragraphs*" and the corresponding case studies must be stored in the current version of *LegalWriter*, in a separate page by the system administrators to ensure that students can retrieve the appropriate paragraphs. However, it is also possible for students to use the system without predefined cases. In this scenario, students would have to refer to the code of law to find the relevant paragraphs. For students to receive individual feedback, they can press the feedback button. This stimulates the text analysis; thus, students receive highlighting's in their text (**DF3**) (Afrin et al. 2021). The text highlighting marks the used components of the appraisal style. Text fragments that do not fit into the appraisal style remain unmarked. In addition, the "*feedback*" button also allows the overall feedback to be visible in the dashboard and the recommendations in the checklist (both functions are explained in more detail below). On the left side, we integrated a checklist. This contains a word count and an overview of the most important components of the appraisal style (**DF4**). The checklist also integrates a counting function that tracks the components used and the errors that occurred in the text (**DF5**). The recommendations result from a matching of the specified components (by the ML-models) and the use of certain heuristics. Errors can occur when students either forget components, do not use the exact legal wording or components used do not refer to other components. For example, each major claim must be answered by a conclusion. The system shows the user if a major claim is not closed by a conclusion, but the user must check by himself if the conclusion has a content that fits the major claim. When a student calls up an error in the checklist he or she receives the individual recommendations (see Figure 2) for improving the error (**DF5**) (Cagiltay 2006). On the right side, we integrated a dashboard for legal argumentation. In the dashboard, students can access general feedback and receive explanations of the individual elements of their appraisal style (**DF6**). Two charts provide students with information about the composition of their texts. Additionally, a ring diagram gives an overview of the composition of the text based on sentences and their assignment to the components of the appraisal style (see Figure 2). The composition of the text based on the components of the appraisal style offers the advantage for the students that they can recognize whether they have set the correct emphasis in their case solution. At the same time, explanations and recommendations for improvement can also be called up here: "*It is important that the subsumption takes up one of the largest parts, i. e. that the argumentative weighing of the facts and the condition of the case takes place here*". A bar chart shows how persuasive the text is overall. The persuasiveness score is calculated by relating the recognized components of the appraisal style as well as the claims and premises used, with the unassigned text sections (**DF7**). By using the "*more*" button, students get information about the functionalities (**DF8**).

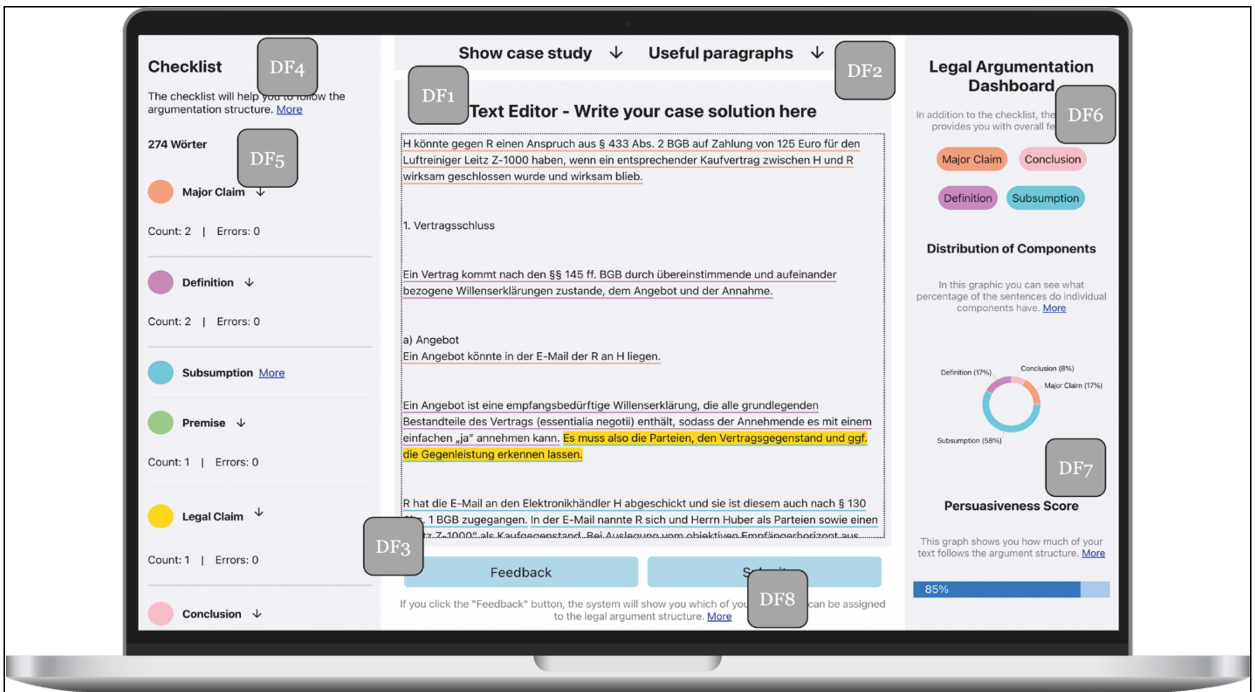

Figure 2. Screenshot of the interface of *LegalWriter* with design features. To facilitate understanding we translated the user interface to English.

Constructing a Corpus of Annotated Student Written Legal Case Solutions

Developing an AI-based learning system that can provide students with personalized feedback on their writings requires a significant amount of annotated text data, known as corpora. Unfortunately, there are currently no adequate annotated corpora available in German that provide legal case solutions for students[3]. As a result, we decided to create our own corpus and collected a total of 413 case solutions from

[3] We used generative models like for student feedback, but no suitable LLM was available for satisfactory results on German law texts during our research.

students in two law courses at our university in 2021 and 2022 (Weber et al. 2023). During these courses, the students were given various cases to solve from two different areas of law. To ensure the quality of our annotations, we followed established methodologies to create a data set for student writing support (Persing and Ng 2016; Stab and Gurevych 2014). The collected case solutions were annotated by two German-speaking experts in law. These annotators followed a comprehensive thirteen-page guide, which outlined the appraisal style, argument details, and the interconnections of arguments within the subsumption. Our guide describes in detail the four main components of the appraisal style, namely, major claim, definition, subsumption, and conclusion. We provide specific instructions on identifying and annotating each of these components, ensuring that our annotators clearly understand what they are looking for. In addition to the four components of the appraisal style, our guideline also covers the arguments and relationships within the subsumption. We provide detailed instructions on how to identify and annotate the various arguments and their relationships, such as premises and legal claims. The first step in the annotation process involved marking the four components of the appraisal style. In the second step, the annotators took a closer look at the subsumption and assembled legal claims and premises. Finally, the relationships between the legal claims and premises were annotated in the third step. To maintain uniformity in the annotation process, we organized multiple training sessions aimed at resolving any discrepancies among annotators and fostering a shared comprehension of the annotation guidelines. We used the tool tagtog[4] for annotation. The tool offers the advantages of a graphical interface for marking up text units and allows for the monitoring of Inter-Annotator Agreements (IAA) through a dashboard of metrics. After 100 annotated texts, we calculated the IAAs using Krippendorff's α (Krippendorf 1980). Our analysis showed that we achieved a minimum of substantial agreement for all components and fair agreement for relations (Landis and Koch 1977). These levels of agreement are an indicator that our annotation process was dependable and consistent, which means that our annotated corpus is a valuable resource for developing AI-based learning systems that can offer students personalized feedback on their writing (Weber et al. 2023).

Building the Feedback Algorithm to Provide AI-based Legal Writing Feedback

The backend of *LegalWriter* was developed using the Flask framework and consists of three ML models. Two of these models use the output of the first model as input data. All three models are based on the Transformer architecture and employ BERT, which we obtained as a pre-trained model from HuggingFace[5]. We then trained the BERT model on a training set. The model was trained in batches of size 8, with a learning rate of 4e−5 and a warmup ratio of 0.06. Overall, *LegalWrites* backend is a robust and well-designed system that utilizes cutting-edge ML techniques to analyze and provide feedback on legal texts. To implement individualized feedback on errors in adherence to the structure of legal writing (**DP2**), our model was exported and implemented in the *LegalWriter* backend. When a student enters a case solution in the text editor and clicks the feedback button, the text is sent to our three trained models. The first model classifies the four components of the appraisal style, while the second one identifies the legal claims and premises in the subsumption. The third model determines the relationships between the legal claims and premises. Once the text is classified, it is sent back to the front-end of *LegalWriter*, which provides individual feedback by highlighting the text's components in different colors (Afrin et al. 2021). Non-persuasive sentences or sentences that don't meet the requirements of the appraisal style are not highlighted. Additionally, *LegalWriter* suggests recommendations for improving the text (**DP3**), based on the classified components (see step 4). Additionally, students can use a dashboard to view the distribution of components across the text. Depending on the distribution, students receive further suggestions for improvement to achieve a balanced distribution of the components.

Step 5: Proof of Concept Evaluation in a Lab Experiment

The evaluation of *LegalWriter* involved a three-stage lab experiment. In the first stage participants were assessed for their familiarity with legal appraisal and their background in law. The success of randomization was tested through two constructs measured on a 1-7 Likert scale (Agarwal and Prasad 1998; Ashford et al. 2003). During the subsequent stage, participants carried out a writing assignment using one of two distinct version of *LegalWriter*, depending in which of the two randomly assigned groups (control and treatment) they perform the writing assignment. They were introduced to the appraisal style and instructions on

4 https://tagtog.net

5 https://huggingface.co/docs/hub/index

solving a legal problem, followed by a writing task to solve a specific case study. The control and treatment groups received the same problem but used different versions of *LegalWriter*. The treatment group used an ML-based version with personalized feedback, while the control group used a non-ML-based version with static recommendations (**DP2** and **DP3** were not ML-based). The static feedback aims to closely emulate the comprehensive guidance of a tutor, drawing inspiration from an example solution. This form of feedback, rooted in a sample solution, closely approximates the esteemed benchmark in legal education in Germany. The third stage, the post-survey, measured the impact of *LegalWriter* on user experience and support effectiveness. Constructs such as *intention to use, perceived usefulness (Agarwal and Prasad 1998)* and *perceived ease of us*e (Bala and Venkatesh 2007) were assessed. Qualitative feedback was collected to evaluate the system and gather suggestions for improvement (see Step 7). The experiment was conducted based on our university's ethical rules and the rules of the platform operator. Incomplete surveys and surveys in which certain control questions were not correctly answered were not included. After randomization, we counted 32 completed valid outcomes in the treatment group and 30 in the control group. The average age of the control group was 27.8 (SD = 3.66). In the treatment group at 27.7 (SD = 3.25). 17 participants in the control group were male and 13 were female. In the treatment group, 11 were male and 21 female. The final sample consisted of 62 students.

Results

To determine the participants' user experience, we compare the results from the ML-based version of *LegalWriter* with the static version of *LegalWriter*. For data analysis, we performed a double-tailed t-test with equal variances to assess whether differences between both groups are statistically significant. To assess the normal distribution of the data, we employed the Shapiro-Wilk test in addition to conducting a graphical analysis. To verify the homogeneity of variances, we utilized the Levene's test. For the construct of *perceived ease of use* we used a Welch test, since a Levene test could not detect equal significances. The construct *intention to use* was rated with an average value of 5.14^{**} (SD= 1.21) and the *perceived usefulness* with an average value of 5.00^{*} (SD= 1.04). Notably, the *intention to use value* significantly surpasses that of the control group, and the *perceived usefulness* was also rated significantly higher compared to the control group. *Perceived ease of use* was also rated higher in the treatment group compared to the control group and showed a statistical difference between both groups (see Table 3). The results show that the participants of our experiment positively evaluated the technology acceptance towards the ML-based version of *LegalWriter* compared to the use of the alternative version. Moreover, the mean scores of *LegalWriter* show outstanding initial results. All treatment group results are higher than the neutral value of 4. Particularly, the significantly high value of *intention to use* indicates that the participants are receptive to the system and are inclined towards its prospective adoption. Also, the values in *perceived usefulness* and *perceived ease of use* for writing structured and persuasive case solutions with the system *LegalWriter* provide promising results, as the constructs significantly impact the influence on the acceptance of IT systems (Agarwal and Prasad 1998).

Group (n = 62)	Intention to Use	Perceived Ease of Use	Perceived Usefulness
mean TG	5.14^{**}	5.19^{*}	5.00^{*}
mean CG	3.95	4.58	4.25
SD TG	1.21	1.00	1.04
SD CG	1.72	0.86	1.79
p-value	0.003	0.046	0.045
t-value	3.161	2.050	2.050
$^{*}p < 0.05$, $^{**}p < 0.01$			

Table 3. Mean and standard derivation on a 7-point Likert scale (1: low, 7: high).

Furthermore, we analyzed the students' written texts for the quality of legal writing. The quality of legal writing is determined by the extent to which students adhere to the appraisal style, which involves structuring their writing appropriately. Additionally, the ability to draw logical conclusions based on definitions and specific facts about the case solution is crucial for persuasive writing. We utilized the assessment from our lab experiment to evaluate the quality of legal writing. The assessment was conducted by an independent legal tutor and based on the rating scales that are also used for a German law exam. Following these, the assessments were assigned on a scale of 1 to 18, with 1 indicating poor quality and 18 indicating high quality. The analysis showed that the students who used the adaptive system (mean = 10.05^{*}) showed a higher quality in legal writing (p-value = 0.015), compared to the control group (mean =

7.97). All in all, we assume that the proof-of-concept evaluation is successful so that the development of *LegalWriter* can continue (Venable et al. 2016). However, to show long-term effects and that the students improve their writing even without system support, we conducted a field experiment.

Step 6: Evaluation in the Field

To provide evidence of *LegalWriter's* effectiveness in continuous use, we evaluated *LegalWriter* in a naturalistic ex-post evaluation designed as a field experiment in Step 7 (Venable et al. 2016). The field experiment aims to answer the second research question (RQ2): *To what extent does an AI-based learning system help law students improve their structural and persuasive writing skills*? Hence, we evaluated the system in a natural use-case scenario (Venable et al. 2016). To do so, we implemented the system in a legal tutorial, which was offered in addition to a law lecture at a European university. The students were asked to write three legal case solutions to a given problem, each one week apart from the other. Students had the opportunity to participate in the experiment voluntarily and were randomly allocated to two groups. The control group engaged with the cases conventionally, followed by subsequent feedback from a tutor provided in the form of a sample solution[6]. This tutor's feedback, which is grounded in the example solution, is emblematic of the current gold standard in German legal education. The treatment group employed *LegalWriter* and obtained automated intelligent feedback guided by our design principles. After randomization, we had 19 participants in the treatment and 24 in the control group. Participants of the treatment group had an average age of 19.57 (SD= 5.20), 9 were male and 14 were female. In the control group, participants' average age was 18.32 (SD=1.47), 6 were male and 12 were female. A minor variation in age exists between the control and treatment groups, but this disparity lacks significance ($p = 0.2011$). This discrepancy could be attributed to the relatively high standard deviation within the treatment group (SD = 5.20). The experiment phases took five weeks. It consisted of three main phases: *1) pre-test phase, 2) individual writing phase* and 3) *post-test phase*.

Pre-test Phase: In the pre-test phase, students were given a survey with 12 questions. In the survey, we collected demographics, students' experiences with AI-based learning systems, and students' experiences in writing legal case solutions by using individual items. To measure their proficiency in composing legal case solutions, students were tasked with resolving a legal issue within a time frame of around 25 minutes. The case solutions were evaluated by an independent tutor who supported the lecture.

Individual Writing Phase: In the individual writing phase, students were each asked to solve a legal problem in three tutorials. In each tutorial, students had 60 minutes to solve the legal problems and write a case solution. The treatment group received feedback from *LegalWriter* to improve their case solutions, and the control group received feedback from an independent tutor.

Post-test Phase: The post-test phase started with the final exam of the law lecture. All students who participated in the experiment agreed that their exam results may be evaluated during the experiment. After the exam, students received a post-survey in which we asked qualitative questions: *"What did you like most about interacting with LegalWriter?", "What would you improve about LegalWriter?"* and *"Do you have any additional ideas? What would you like to add to the system?"*. In total, we asked six questions in the post-test. To evaluate the data from the field experiment, we analyzed the *quality of legal writing*. The assessment was conducted by an independent tutor on a 1-18 scale (1: poor, 18: high), which is typical for German legal education.

Results

For data analysis, we performed a double-tailed t-test with equal variances to assess whether differences between both groups are statistically significant[7]. To assess the normal distribution of the data, we employed the Shapiro-Wilk test in addition to conducting a graphical analysis. To verify the homogeneity of variances, we utilized the Levene's test. To mitigate the possible influence of confounding variables given our relatively limited sample size and to determine the effectiveness of randomization, we compared the scores of the case solutions from the pretest between the two groups. We received p-values larger than 0.05 between the treatment and the control group (see Table 4). Nevertheless, we would like to clarify that a

[6] Typically, the tutor independently writes the sample solution, which is subsequently collaboratively elaborated with the students.

[7] The data collection and analysis were conducted according to the ethical guidelines of our university.

marginal significance is shown ($p<0.1$) (see Table 4). This implies that the observed differences between the groups concerning the measured variable exist but are statistically only slightly above the random level. This marginal significance indicates that the effects may be due to natural variation or other influences. It is important to interpret such results cautiously and to possibly consider a larger sample or adjusted research designs in future studies to gain clearer insight into the observed differences. However, we assume that the marginal significance indicates that there is no explainable difference in the quality of the legal writing between the control and the treatment group in the pre-test. The analysis of the texts from the exams shows that the treatment group (mean = 11.08^{**}) performed significantly better in the *quality of legal writing* than the control group (mean = 8.84). Since we are aware that the group in the field experiment is relatively small, we have additionally specified effect sizes. According to Cohen's d, the effect on the quality of the appraisal style shows a strong effect size (greater than 0.8) (Cohen 2016). A higher Cohen's d value signifies a more considerable disparity between the groups, thus reinforcing our observation that the treatment group achieves better outcomes in the quality of legal writing than the control group.

Group (n = 42)	Quality of Legal Writing (Pre-test)	Quality of Legal Writing (Post-test)
mean TG	6.83	11.08^{**}
mean CG	5.42	8.84
SD TG	2.43	2.19
SD CG	2.99	2.59
p-value	0.095	0.004
t-value	1.712	3.079
Cohen’s d	-	0.95
$^{**}p < 0.01$		

Table 4. Mean and standard derivation based on the standardized 1-18 German law scale (1: poor, 18: high).

Step 7: Revising Design Knowledge

Considering the qualitative feedback received from the two experiments, we have revised the design features of the system. One key aspect that emerged from the feedback was the importance of accuracy in classifying the individual components of the appraisal style and the quality of recommendations. Consequently, we prioritized the development of the ML-models to address this concern. To enhance the capabilities of the system, we expanded our own corpus by including an additional 200 case solutions. This expansion enables the training of the system in various areas of law (**DP1**). Moreover, we have taken steps to improve the discriminatory power of the components by addressing sentence separation issues prevalent in German legal texts (Glaser et al. 2021). These improvements aim to enhance the accuracy and precision of the system's feedback. By incorporating these revisions, we aim to ensure that the AI-based learning system provides users with more accurate and reliable feedback. After carefully considering the improvements made to the system, we have concluded that they contribute to its enhancement without necessitating significant changes in design knowledge. As a result, we have made the decision not to adapt or extend the design principles. While the improvements have brought valuable enhancements, they have not introduced substantial alterations that would require corresponding modifications to the existing design principles. This decision allows us to focus on refining the system's performance and functionality based on the feedback and insights gained from the implemented improvements.

Step 8: Documenting of Design Knowledge

We summarize our theoretical contributions from the conducted design process by documenting our design knowledge in accordance with the seven core components of a design theory (Gregor and Jones 2007), as shown in Table 5. Through this approach, we communicate our insights to the scientific knowledge base and capture the outcomes of our project. Our goal is to formulate a "design and action" theory grounded in solid principles that can guide the process of designing legal learning systems (Gregor and Jones 2007).

1) Purpose and scope	*LegalWriter* aims to enable students to learn how to write persuasive and structured case solutions in the appraisal style independently of a teacher through learning support based on recent advances in NLP and ML.

2) Constructs	Text editor (DF1); Useful paragraphs (DF2); Colored marking of components that belong to the appraisal style (DF3, DF4); Checklist (DF5); Explanations of the individual elements of their appraisal style (DF6); functionalities and explanations on how to use them effectively (DF8).
3) Principles of form and function	**DP1:** *... embed realistic case studies into the system so that students can solve legal problems from different areas of law individually and thus digitally apply an established learning technique in law.* **DP2:** *... provide individualized feedback on errors in adherence to the structure of legal writing so that students can intuitively identify errors in adherence to the structure of legal writing.* **DP3:** *... integrate support that gives students recommendations on how to adhere to the structure of the appraisal style, how to formulate the components of the appraisal style and how to build a stringent legal argumentation so that students can improve their legal texts in a self-directed manner.*
4) Artifact mutability	Core design features, such as the individual feedback algorithm, might be adapted to different pedagogical scenarios e. g., the content and language of the texts.
5) Testable propositions	(1) Using *LegalWriter* increases the quality of legal writing. (2) Using *LegalWriter* increases the skill of writing a structured and persuasive case solution.
6) Justificatory knowledge	Learning from Errors (Metcalfe 2017)
7) Principles of implementation	The feedback algorithm for structural and persuasive legal writing needs to be linked to course content and language.

Table 5. Documentation of our design knowledge (Gregor and Jones 2007).

Discussion and Conclusion

In this project, we followed the DSR approach (Hevner 2007) to design, develop, and evaluate the AI-based learning system *LegalWriter*. *LegalWriter* helps students learn structured and persuasive legal writing by identifying individual errors. Our learning system differs from existing systems in persuasive writing by providing individualized feedback on written texts and natural errors. Traditional writing support systems usually only support general argumentation approaches (Osborne et al. 2016), which is insufficient when writing legal case solutions (see Section Legal Writing). In addition, *LegalWriter* differs from systems like Grammarly, as Grammarly tends to specialize in improving spelling and grammar (Bailey and Lee 2020). However, our system has a more specific focus and helps students to follow the formal structure of legal writing (appraisal style). Moreover, our system adheres to pertinent learning theories in its design. This is the reason why our system strives for enduring learning success, rather than merely assisting with writing during system usage, as seen in platforms like Grammarly. To align the design of our learning system in accordance with established learning theories and user requirements, we extracted requirements from 64 scholarly papers and conducted ten semi-structured interviews with law students. This process enabled us to formulate a collection of design principles for AI-based learning systems. We evaluated these design principles as an instantiated artifact (*LegalWriter*) based on two experiments with 104 users. In a short-term laboratory experiment with 62 participants, we evaluated how students perceive the usefulness of our design principles for writing tasks in a legal context and demonstrated the short-term support effects of the system. In a field experiment with 42 students, we demonstrated the learning effects' effectiveness in structured and persuasive legal writing. Our contribution extends beyond the development of an AI-based learning system as a software artifact. We have also derived valuable design knowledge, which we have documented in our final step of our design research (Table 5). This design knowledge is not only applicable to our specific case but can also be utilized in other contexts where structured writing and persuasive writing are essential. For instance, the concept of *LegalWriter* can be easily adapted to courses covering different subjects or languages that require structured and persuasive writing. This would involve modifying the backend algorithm to cater to the specific scenario. Existing corpora and ML-models in the literature can be incorporated into *LegalWriter* to handle, for example, English legal cases, as demonstrated by (Mochales and Ieven 2009). Another example would be writing research papers, essays or term papers where clear structuring of writing is of great importance (Resch and Yankova 2019). In these cases, the design principles of form and function, as well as the overall system design, do not require significant adaptation. The transferability of our design knowledge allows us to contribute not only at Level 1 of DSR by providing an implementation of a situated artifact but also at Level 2 by contributing to an emerging design theory (Gregor and Hevner 2013). This indicates the broader significance of our research in advancing the understanding and application of AI-based learning systems. Our results demonstrate that state-of-the-art

NLP and ML techniques are well-suited for designing sophisticated systems capable of providing support for students based on individual errors (Metcalfe 2017). In general, our research can stimulate further work in AI-based and technology-mediated learning (Gupta and Bostrom 2009).

With respect to our work, it is important to acknowledge certain limitations. While it is plausible to assume that the transferability of *LegalWriter* to other areas of law is feasible without significant modifications, we are unable to establish this conclusively with our current research design. This limitation arises from the fact that the system has been trained on only two specific areas of law thus far. However, as mentioned in step 7, we have introduced a new area of law that will be evaluated in future studies. In relation to the field experiment, it's important to note that the experiment involves a modest sample size of 42 participants, comprising 19 in the treatment group and 24 in the control group. Despite the relatively small sample, we would like to highlight the noteworthy p-value ($p = 0.004$) and Cohen's d ($d = 0.95$), both indicating substantial effect sizes concerning the quality of legal writing (see Table 4). However, our aim for the future is to expand our field experiments by employing a more extensive sample size to replicate and validate our findings. Despite the positive outcomes of our research, it is important to acknowledge and address the ethical concerns that may arise in relation to our AI-based learning system. One primary concern is the potential impact of automating feedback and assessments on the learning process and individual student development. We want to emphasize that our intention is to enhance the learning environment and support students in their academic growth. We understand that AI-generated feedback can be perceived as crucial, and students may strongly respond to it. Thus, it is crucial to view AI feedback as a supplement to, rather than a substitute for, human feedback. Additionally, it is important to note that the current state of research does not rule out the possibility of model misses. This means that the AI feedback may not capture all relevant aspects accurately, and there is a need for ongoing improvement and refinement. Furthermore, we acknowledge the potential bias that may exist within the AI feedback system. It is possible that the feedback is limited to specific stylistic preferences or patterns, which may not encompass the diverse cultural, linguistic, or individual backgrounds of all students. This could result in certain students being disadvantaged if they do not conform to the predefined standard.

For future research endeavors, we intend to delve further into understanding how *LegalWriter* can effectively support students in improving their case solutions. We aim to evaluate the individual design principles and explore different combinations of these principles to determine if certain combinations yield more favorable results compared to utilizing all the design principles simultaneously (Abbasi et al. 2010). Furthermore, we would like to work with large language models in the future since they could be domain independent to counteract the corpus limitations. The large language models could replace our feedback algorithm; however, our design theory would still be valid as it is independent of the underlying corpus. All in all, our research provides design knowledge to improve further AI-based learning systems based on techniques from NLP and ML. With further technological advances, we expect our work to stimulate researchers to design more AI-based learning systems for other learning scenarios in the IS field. Furthermore, we contribute to an established learning theory (Metcalfe 2017), which has shown its effectiveness in digital learning scenarios.

Acknowledgement

The results presented in this article were developed in the research project Komp-HI funded by the German Federal Ministry of Education and Research (BMBF, grant 16DHBKI073).

References

Abbasi, A., Zhang, Z., Zimbra, D., Chen, H., and Nunamaker Jr, J. F. 2010. "Detecting Fake Websites: The Contribution of Statistical Learning Theory," *Mis Quarterly* (43:3), JSTOR, pp. 435–461.

Afrin, T., Kashefi, O., Olshefski, C., Litman, D., Hwa, R., and Godley, A. 2021. "Effective Interfaces for Student-Driven Revision Sessions for Argumentative Writing," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–13.

Agarwal, R., and Prasad, J. 1998. "A Conceptual and Operational Definition of Personal Innovativeness in the Domain of Information Technology," *Information Systems Research* (9:2), Informs, pp. 204–215.

Akram, W., and Kumar, R. 2017. "A Study on Positive and Negative Effects of Social Media on Society," *International Journal of Computer Sciences and Engineering* (5:10), pp. 351–354.

Alarie, B., Niblett, A., and Yoon, A. H. 2016. "Using Machine Learning to Predict Outcomes in Tax Law," *Can. Bus. LJ* (58), HeinOnline, pp. 231–253.

Aletras, N., Tsarapatsanis, D., Preoţiuc-Pietro, D., and Lampos, V. 2016. "Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective," *PeerJ Computer Science* (2), PeerJ Inc., pp. 1–19.

Aleven, V. 2003. "Using Background Knowledge in Case-Based Legal Reasoning: A Computational Model and an Intelligent Learning Environment," *Artificial Intelligence* (150:1–2), Elsevier, pp. 183–237.

Arkorful, V., and Abaidoo, N. 2015. "The Role of E-Learning, Advantages and Disadvantages of Its Adoption in Higher Education," *International Journal of Instructional Technology and Distance Learning* (12:1), United States of America, pp. 29–42.

Ashford, S. J., Blatt, R., and Walle, D. V. 2003. "Reflections on the Looking Glass: A Review of Research on Feedback-Seeking Behavior in Organizations," *Journal of Management* (29:6), Sage Publications Sage CA: Thousand Oaks, CA, pp. 773–799.

Ashley, K. D., and Brüninghaus, S. 2009. "Automatically Classifying Case Texts and Predicting Outcomes," *Artificial Intelligence and Law* (17), Springer, pp. 125–165.

Bailey, D., and Lee, A. R. 2020. "An Exploratory Study of Grammarly in the Language Learning Context: An Analysis of Test-Based, Textbook-Based and Facebook Corpora.," *TESOL International Journal* (15:2), ERIC, pp. 4–27.

Bala, H., and Venkatesh, V. 2007. "Assimilation of Interorganizational Business Process Standards," *Information Systems Research* (18:3), pp. 340–362.

Beurskens, M. 2016. "Neue Spielräume Durch Digitalisierung? E-Learning in Der Deutschen Rechtslehre," *ZDRW* (3:1), Nomos Verlagsgesellschaft mbH & Co. KG, pp. 1–17.

Bouki, V., Economou, D., and Kathrani, P. 2014. "'Gamification' and Legal Education: A Game Based Application for Teaching University Law Students," in *2014 International Conference on Interactive Mobile Communication Technologies and Learning (IMCL2014)*, IEEE, pp. 213–216.

Cagiltay, K. 2006. "Scaffolding Strategies in Electronic Performance Support Systems: Types and Challenges," *Innovations in Education and Teaching International* (43:1), pp. 93–103.

Cannon, A. M. 1955. *The Case Method at the Harvard Business School*, (30:1), The Accounting Review, pp. 178–180.

Carr, C. S. 2003. "Using Computer Supported Argument Visualization to Teach Legal Argumentation," *Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making*, Springer, pp. 75–96.

Carter, M. J., and Harper, H. 2013. "Student Writing: Strategies to Reverse Ongoing Decline," *Academic Questions* (26:3), pp. 285–295.

Cohen, J. 2016. "A Power Primer," in *Methodological Issues and Strategies in Clinical Research*, A. E. Kazdin (ed.), pp. 279–284. (https://doi.org/10.1037/14805-018).

Cooper, H. M. 1988. "Organizing Knowledge Syntheses: A Taxonomy of Literature Reviews," *Knowledge in Society* (1:1), pp. 104–126.

Driver, R., Newton, P., and Osborne, J. 2000. "Establishing the Norms of Scientific Argumentation in Classrooms," *Science Education* (84:3), Wiley Online Library, pp. 287–312.

Enqvist-Jensen, C., Nerland, M., and Rasmussen, I. 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), Elsevier, pp. 38–49.

Ericsson, K. A., Krampe, R. T., and Tesch-Römer, C. 1993. "The Role of Deliberate Practice in the Acquisition of Expert Performance.," *Psychological Review* (100:3), American Psychological Association, pp. 363–406.

Flower, L., and Hayes, J. R. 1981. "A Cognitive Process Theory of Writing," *College Composition and Communication* (32:4), National Council of Teachers of English, pp. 365–387.

Glaser, I., Moser, S., and Matthes, F. 2021. "Sentence Boundary Detection in German Legal Documents.," in *ICAART (2)*, pp. 812–821.

Gordon, T. F., Prakken, H., and Walton, D. 2007. "The Carneades Model of Argument and Burden of Proof," *Artificial Intelligence* (171:10–15), Elsevier, pp. 875–896.

Gregor, S., Chandra Kruse, L., and Seidel, S. 2020. *Research Perspectives: The Anatomy of a Design Principle*, (21:6), Journal of the Association for Information Systems, pp. 1622–1652.

Gregor, S., and Hevner, A. R. 2013. "Positioning and Presenting Design Science Research for Maximum Impact," *MIS Quarterly* (37:2), JSTOR, pp. 337–355.

Gregor, S., and Jones, D. 2007. "The Anatomy of a Design Theory," *Journal of the Association for*

Information Systems (8:5), pp. 312–335. (https://doi.org/DOI: 10.17705/1jais.00129).
Gupta, S., and Bostrom, R. 2009. "Technology-Mediated Learning: A Comprehensive Theoretical Model," *Journal of the Association for Information Systems* (10:9), pp. 686–714.
Hachey, B., and Grover, C. 2005. "Automatic Legal Text Summarisation: Experiments with Summary Structuring," in *Proceedings of the 10th International Conference on Artificial Intelligence and Law*, pp. 75–84.
Hevner, A. R. 2007. "A Three Cycle View of Design Science Research," *Scandinavian Journal of Information Systems* (19:2), pp. 87–92.
Houy, C., Niesen, T., Fettke, P., and Loos, P. 2013. "Towards Automated Identification and Analysis of Argumentation Structures in the Decision Corpus of the German Federal Constitutional Court," in *2013 7th IEEE International Conference on Digital Ecosystems and Technologies (DEST)*, IEEE, pp. 72–77.
Huang, C., Chen, M., Yang, P., and Chang, J. S. 2013. "A Computer-Assisted Translation and Writing System," *ACM Transactions on Asian Language Information Processing (TALIP)* (12:4), ACM New York, NY, USA, pp. 1–20.
Jonassen, D. H., and Kim, B. 2010. "Arguing to Learn and Learning to Argue: Design Justifications and Guidelines," *Educational Technology Research and Development* (58), Springer, pp. 439–457.
Kabudi, T. 2021. "Identifying Design Principles for an AI-Enabled Adaptive Learning Systems.," in *Proceedings of the 2022 Annual Pacific Asia Conference on Information Systems*, pp. 1–16.
Kendeou, P., and Van Den Broek, P. 2007. "The Effects of Prior Knowledge and Text Structure on Comprehension Processes during Reading of Scientific Texts," *Memory & Cognition* (35:7), pp. 1567–1577.
Kornell, N., Hays, M. J., and Bjork, R. A. 2009. "Unsuccessful Retrieval Attempts Enhance Subsequent Learning.," *Journal of Experimental Psychology: Learning, Memory, and Cognition* (35:4), American Psychological Association, pp. 989–998.
Krippendorf, K. 1980. "Validity in Content Analysis," in *Computerstrategien Für Die Kommunikationsanalyse*, E. Mochmann (ed.), Frankfurt: Campus, pp. 69–112.
Landis, J. R., and Koch, G. G. 1977. "The Measurement of Observer Agreement for Categorical Data," *Biometrics* (33:1), JSTOR, pp. 159–174.
Lorenzet, S. J., Salas, E., and Tannenbaum, S. I. 2005. "Benefiting from Mistakes: The Impact of Guided Errors on Learning, Performance, and Self-Efficacy," *Human Resource Development Quarterly* (16:3), Wiley Online Library, pp. 301–322.
Man, J. I. N. 2022. "The Appraisal-Based Case Teaching Method in China's Legal Education," *Canadian Social Science* (18:2), pp. 1–4.
Mayring, P. 2015. *Qualitative Inhaltsanalyse*, (12th ed.), Weinheim: Beltz.
Metcalfe, J. 2017. "Learning from Errors," *Annual Review of Psychology* (68), pp. 465–489.
Metcalfe, J., and Xu, J. 2018. "Learning from One's Own Errors and Those of Others," *Psychonomic Bulletin & Review* (25), Springer, pp. 402–408.
Meth, H., Mueller, B., and Maedche, A. 2015. "Designing a Requirement Mining System," *Journal of the Association for Information Systems* (16:9), pp. 799–837. (https://doi.org/10.17705/1jais.00408).
Mochales, R., and Ieven, A. 2009. "Creating an Argumentation Corpus: Do Theories Apply to Real Arguments? A Case Study on the Legal Argumentation of the ECHR," in *Proceedings of the 12th International Conference on Artificial Intelligence and Law*, pp. 21–30.
Moens, M.-F., Boiy, E., Palau, R. M., and Reed, C. 2007. "Automatic Detection of Arguments in Legal Texts," in *Proceedings of the 11th International Conference on Artificial Intelligence and Law*, pp. 225–230.
Ohlsson, S. 1996. "Learning from Performance Errors.," *Psychological Review* (103:2), American Psychological Association, pp. 241–262.
Osborne, J. F., Henderson, J. B., MacPherson, A., Szu, E., Wild, A., and Yao, S.-Y. 2016. "The Development and Validation of a Learning Progression for Argumentation in Science," *Journal of Research in Science Teaching* (53:6), Wiley Online Library, pp. 821–846.
Palau, R. M., and Moens, M.-F. 2009. "Argumentation Mining: The Detection, Classification and Structure of Arguments in Text," in *Proceedings of the 12th International Conference on Artificial Intelligence and Law*, pp. 98–107.
Persing, I., and Ng, V. 2016. "End-to-End Argumentation Mining in Student Essays," in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 1384–1394.
Pinkwart, N., Ashley, K. D., Aleven, V., and Lynch, C. F. 2008. "Graph Grammars: An ITS Technology for Diagram Representations.," in *FLAIRS Conference*, pp. 433–438.

Pinkwart, N., Ashley, K., Lynch, C., and Aleven, V. 2009. "Evaluating an Intelligent Tutoring System for Making Legal Arguments with Hypotheticals," *International Journal of Artificial Intelligence in Education* (19:4), pp. 401–424.

Potts, R., and Shanks, D. R. 2014. "The Benefit of Generating Errors during Learning.," *Journal of Experimental Psychology: General* (143:2), American Psychological Association, pp. 644–667.

Poudyal, P., Gonçalves, T., and Quaresma, P. 2019. "Using Clustering Techniques to Identify Arguments in Legal Documents.," *In Proceedings of the Third Workshop on Automated Semantic Analysis of Information in Legal Texts Co-Located with the 17th International Conference on Artificial Intelligence and Law*, pp. 1–8.

Reed, C. 2006. "Preliminary Results from an Argument Corpus," *Linguistics in the Twenty-First Century*, Cambridge Scholars Press, pp. 185–196.

Reed, C., Walton, D., and Macagno, F. 2007. "Argument Diagramming in Logic, Law and Artificial Intelligence," *The Knowledge Engineering Review* (22:1), Cambridge University Press, pp. 87–109.

Resch, O., and Yankova, A. 2019. "Open Knowledge Interface: A Digital Assistant to Support Students in Writing Academic Assignments," in *Proceedings of the 1st ACM SIGSOFT International Workshop on Education through Advanced Software Engineering and Artificial Intelligence*, pp. 13–16.

Rubin, J., and Chisnell, D. 2008. *Handbook of Usability Testing: How to Plan, Design and Conduct Effective Tests*, (2nd ed.), New Jersey: John Wiley & Sons.

Schlegel, L., Schöbel, S., and Söllner, M. 2023. *Nudging Digital Learning–An Experimental Analysis of Social Nudges to Manage Self-Regulated Learning and Online Learning Success*.

Schmitt, A., Wambsganss, T., Söllner, M., and Janson, A. 2021. "Towards a Trust Reliance Paradox? Exploring the Gap Between Perceived Trust in and Reliance on Algorithmic Advice," in *International Conference on Information Systems (ICIS)*.

Stab, C., and Gurevych, I. 2014. "Annotating Argument Components and Relations in Persuasive Essays," in *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pp. 1501–1510.

Stuckenberg, C.-F. 2020. "Der Juristische Gutachtenstil Als Cartesische Methode," *ZDRW Zeitschrift Für Didaktik Der Rechtswissenschaft* (6:4), Nomos Verlagsgesellschaft mbH & Co. KG, pp. 323–341.

Toulmin, S. E. 2003. *The Uses of Argument*, (2nd ed.), Cambridge: Cambridge University Press.

Urchs, S., Mitrović, J., and Granitzer, M. 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, pp. 451–454.

Venable, J., Pries-Heje, J., and Baskerville, R. 2016. "FEDS: A Framework for Evaluation in Design Science Research," *European Journal of Information Systems* (25), Springer, pp. 77–89.

Verheij, B. 2003. "Artificial Argument Assistants for Defeasible Argumentation," *Artificial Intelligence* (150:1–2), Elsevier, pp. 291–324.

Virtucio, M. B. L., Aborot, J. A., Abonita, J. K. C., Avinante, R. S., Copino, R. J. B., Neverida, M. P., Osiana, V. O., Peramo, E. C., Syjuco, J. G., and Tan, G. B. A. 2018. "Predicting Decisions of the Philippine Supreme Court Using Natural Language Processing and Machine Learning," in *2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)* (Vol. 2), IEEE, pp. 130–135.

Vom Brocke, J., Simons, A., Riemer, K., Niehaves, B., Plattfaut, R., and Cleven, A. 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, Article 9), pp. 205–224.

Wambsganss, T., Niklaus, C., Cetto, M., Söllner, M., Handschuh, S., and Leimeister, J. M. 2020. "AL: An Adaptive Learning Support System for Argumentation Skills," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–14.

Wambsganss, T., Söllner, M., and Leimeister, J. M. 2020. "Design and Evaluation of an Adaptive Dialog-Based Tutoring System for Argumentation Skills," in *International Conference on Information Systems (ICIS).-Hyderabad, India*, pp. 1–15.

Weber, F., Neshaei, S. P., Wambsgamß, T., and Söllner, M. 2023. "Modeling Structured Persuasive Writing of Case Solutions in German Law Courses to Support Students in Legal Education," in *Findings of ACL 2023*, Toronto, Canada, pp. 1–15.

Wong, S. S. H., and Lim, S. W. H. 2019. "Prevention–Permission–Promotion: A Review of Approaches to Errors in Learning," *Educational Psychologist* (54:1), Taylor & Francis, pp. 1–19.

Xu, B., and Yan, H. 2008. "Case Study Method of Teaching under Web-Based Learning Environment," in *International Conference on Computer Science and Software Engineering*, pp. 808–811.