

# KI-gestützte Klauselkontrolle in allgemeinen Geschäftsbedingungen

Wie künstliche Intelligenz dabei helfen kann, den  
Verbraucherschutz beim Onlineshopping zu stärken

Daniel Braun

## *Abstract*

*Vier von fünf Menschen in Deutschland haben schon einmal online eingekauft. Dabei stimmen sie oft AGB zu, ohne diese gelesen zu haben, und akzeptieren so unwissentlich für sie nachteilige Regeln, die nicht selten gegen Gesetze verstoßen. Verbraucherschützer:innen können in solchen Fällen zwar helfen, meist aber erst, wenn Konsument:innen aktiv Hilfe suchen. Für eine flächendeckende Marktbeobachtung fehlen die Ressourcen. Im Rahmen des Projekts „AGB-Check“ wurden KI-Methoden entwickelt, die in Zukunft dabei helfen können, unwirksame Klauseln in AGB effizient zu erkennen und so den Verbraucherschutz zu stärken.*

---

Dieser Beitrag erscheint unter der Creative-Commons-Lizenz: CC BY 4.0  
Kurzform | <https://creativecommons.org/licenses/by/4.0/>  
Lizenztext | <https://creativecommons.org/licenses/by/4.0/legalcode>

# 1 Hintergrund

„Ja, ich habe die AGB gelesen.“ Das ist eine Lüge, der sich die meisten schon einmal schuldig gemacht haben. Verschiedene Studien haben gezeigt, dass lediglich zwischen 4 Prozent (Hillman 2005) und 0,1 Prozent (Bakos et al. 2014) der Verbraucher:innen beim Onlineeinkauf AGB lesen. Auch Gesetzgeber, sowohl auf nationaler als auch auf europäischer Ebene, sind sich dessen bewusst. Entsprechend sind die gesetzlichen Regelungen, was den Gestaltungsspielraum für AGB angeht, besonders restriktiv. So erklärt zum Beispiel die Richtlinie 93/13/EWG des Rates der Europäischen Union vom 5. April 1993 Vertragsklauseln für nichtig, die „entgegen dem Gebot von Treu und Glauben zum Nachteil des Verbrauchers ein erhebliches und ungerechtfertigtes Missverhältnis der vertraglichen Rechte und Pflichten der Vertragspartner verursacht“. Eine vage formulierte, in der Praxis aber weitreichend ausgelegte Regelung.

Während die Gesetzgebung tendenziell aufseiten der Verbraucher:innen steht, gibt es in der Praxis trotzdem ein Kräfteungleichgewicht zum Vorteil von Unternehmen, das durch die häufig mangelnde Durchsetzung der Rechte entsteht. In einer Befragung unter 100 Konsument:innen gaben 52 Prozent an, dass sie bereits mindestens einmal Probleme beim Onlineshopping erlebt haben, wie bezahlte, aber nicht gelieferte oder defekte Ware. Während gut zwei Drittel die Probleme am Ende lösen konnten, konnte immerhin knapp ein Drittel das Problem nicht zufriedenstellend lösen. Trotzdem haben nur 2 Prozent aller Teilnehmer:innen jemals professionelle juristische Beratung im Zusammenhang mit Onlineshopping in Anspruch genommen. Nicht zuletzt wegen häufig geringer Streitwerte schrecken Verbraucher:innen vor der Durchsetzung ihrer Rechte zurück (Braun 2021).

Um einen flächendeckenden Verbraucherschutz im Internet gewährleisten zu können, wäre es daher nötig, AGB proaktiv, also bevor Schaden entsteht, aus Sicht des Verbraucherschutzes zu prüfen. Angesichts Hunderttausender Onlineshops ist das jedoch eine Aufgabe, die manuell kaum zu bewältigen ist. Im Rahmen des vom Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz geförderten Projekts „AGB-Check“ hat der Lehrstuhl für Software Engineering betrieblicher Informationssysteme der

Technischen Universität München in Zusammenarbeit mit den Verbraucherzentralen Brandenburg und Hamburg untersucht, wie KI-Technologien eingesetzt werden können, um Verbraucherschützer:innen bei der Kontrolle von Klauseln in AGB von Onlineshops zu unterstützen. Die Ergebnisse der Forschung zeigen, dass es mit modernen Technologien möglich ist, nicht nur automatisiert das Thema einer Klausel zu erkennen, sondern auch eine zuverlässige Vorhersage zu treffen, ob eine gegebene Klausel potenziell unwirksam ist. In Zukunft können solche Technologien verwendet werden, um die Einhaltung des Verbraucherschutzes im Internet sicherzustellen.

## 2 Technologie

Seit vielen Jahren sind bei der automatischen Verarbeitung von natürlicher Sprache, dem sogenannten Natural Language Processing (NLP), Verfahren des maschinellen Lernens der Standard. Im Gegensatz zu den klassischen, regelbasierten Ansätzen, bei denen Expert:innenwissen mithilfe einer Programmiersprache oder anderer Formalismen für die Maschine aufbereitet wird, werten Verfahren des maschinellen Lernens Daten statistisch aus, um Erkenntnisse zu gewinnen und selbstständig Regeln abzuleiten.

Grundsätzlich können zwei Arten des maschinellen Lernens unterschieden werden: überwachtes (oder englisch *supervised*) und unüberwachtes (oder englisch *unsupervised*) Lernen. Beim überwachten Lernen versucht der Algorithmus einen Zusammenhang zu finden zwischen den zugrundeliegenden Daten und zugehörigen Annotationen, die dem Algorithmus zur Verfügung gestellt werden. Das können zum Beispiel Bilder sein als Daten und die Information, welches Tier sich auf einem Bild befindet, als Annotation. Der Algorithmus versucht dann einen statistischen Zusammenhang zu finden, zwischen den Bildern und den Annotationen. Stehen genug Daten zum Lernen bereit, dann kann das angelegerte Modell später für neue Bilder, die nicht Teil des Lernprozesses waren, eine Einschätzung abgeben, welches Tier sich auf einem gegebenen Bild befindet.

Beim unüberwachten Lernen stehen dem Algorithmus nur Daten und keine Annotationen zur Verfügung. Auch hier wird er versuchen, Gemeinsamkeiten statistisch auszuwerten, um zum Beispiel Bilder zu finden, die sich ähnlich sind. Der Vorteil des unüberwachten Lernens im Vergleich zum überwachten Lernen, ist, dass keine Annotationen benötigt werden. Das Erzeugen von Annotationen geschieht in der Regel manuell durch Menschen, um sicherzustellen, dass der Algorithmus eine zuverlässige Grundlage für das Lernen hat. Insbesondere zum Beispiel bei juristischen Fragestellungen ist die Annotation von Daten daher zeit- und kostenintensiv. Der Nachteil des unüberwachten Lernens ist jedoch, dass nicht immer kontrollierbar oder nachvollziehbar ist, anhand welcher Merkmale der Algorithmus Ähnlichkeit feststellt. Es kann zum Beispiel passieren, dass Tierbilder anhand der Hintergrundfarbe sortiert werden statt des abgebildeten Tieres oder sogar anhand eines Merkmals, das für das menschliche Auge überhaupt nicht erkennbar ist.

Zur automatischen Verarbeitung natürlicher Sprache wird heute häufig ein hybrider Ansatz gewählt, bei dem zunächst Grundlagen der Sprache unüberwacht erlernt und aufgabenspezifische Informationen dann überwacht werden. Das ist notwendig, da die Standarddarstellung von Texten, als Kette einzelner Buchstaben beziehungsweise Zeichen, nicht geeignet ist, um Worte in ihrer Bedeutung zu vergleichen. Die beiden Worte „Hose“ und „Rose“ unterscheiden sich zum Beispiel zwar nur in einem Buchstaben, haben semantisch aber wenig Ähnlichkeit. Die Wörter „Automobil“ und „Fahrzeug“ haben dagegen zwar kaum eine optische Ähnlichkeit, sind dafür aber semantisch eng miteinander verwandt.

Um die Bedeutung eines Wortes zu erfassen, nutzen sogenannte Transformer Modelle und andere Ansätze, zu denen auch BERT und GPT gehören, die sogenannte Verteilungshypothese von Harris (1954). Diese besagt, dass Wörter, die im gleichen Kontext auftauchen, eine ähnliche Bedeutung haben. Alle Wörter, die zum Beispiel die Lücke im Satz „Das \_\_\_\_\_ hat 40 PS und kann bis zu drei Personen befördern.“ füllen können, haben eine ähnliche Bedeutung. So können Algorithmen anhand riesiger, nicht annotierter Textkorpora mathematische Repräsentationen für die Bedeutung von Worten basierend auf deren Kontext entwickeln. Diese mathematische Darstellung kann dann genutzt werden, um überwachtes Lernen mit einer vergleichsweise kleinen Anzahl an annotierten Daten zu ermöglichen.

Im konkreten Fall haben wir ein BERT Transformer Modell verwendet, das auf über 10 Gigabyte an deutschen Texten trainiert wurde, darunter auch ein großer Korpus von Rechtstexten. So kann das Modell unüberwacht zum Beispiel lernen, dass die beiden Worte „Kostenpauschale“ und „Gebühr“ eine ähnliche Bedeutung haben. Lernt es dann später überwacht noch, dass eine Gebühr in Höhe von 10 Euro für das Versenden einer Mahnung unwirksam ist, so kann das Modell daraus ableiten, dass auch eine Kostenpauschale von 10 Euro für das Versenden einer Mahnung unwirksam ist, ohne jemals eine entsprechende Klausel mit Annotation gesehen zu haben.

### 3 Datenbasis

Für die KI-gestützte Klauselkontrolle bestehen die annotierten Daten aus den Klauseltexten aus AGB von Onlineshops, die jeweils mit einem Thema und einer juristischen Einschätzung annotiert wurden, ob die jeweilige Klausel potenziell unwirksam ist. Insgesamt haben die Expert:innen der Verbraucherzentralen Brandenburg und Hamburg auf diese Weise mehr als 10.200 Klauseln annotiert. Der Großteil der Klauseln (knapp 10.000) stammt dabei aus der Annotation von 285 kompletten AGB-Regelwerken. Die restlichen Klauseln wurden bewusst einzeln ausgewählt, zu Themen, die in der Beratung eine besondere Relevanz haben, wie zum Beispiel Mahnkostenpauschalen und Haftungsausschlüsse. Die Auswahl der Anbieter und Prüfung der AGB erfolgt nach Maßgabe, ob es zu den Anbietern bereits Verbraucherbeschwerden gab. Daneben wurden auch weitere Anbieter, für die keine Verbraucherbeschwerden vorlagen, geprüft. Die Einzelklauseln werden in Tabellenform übertragen, auf ihre Rechtmäßigkeit untersucht und bestimmten Themenbereichen zugeordnet.

Die Überprüfung der Rechtmäßigkeit folgt einem binären Schema, indem eine Klausel entweder als potenziell unwirksam (1) oder wirksam (0) markiert werden kann. Da eine komplett trennscharfe Einordnung in der Praxis nicht immer möglich ist und eine letztendliche Entscheidung nur von Gerichten getroffen werden kann, galt als Maßgabe beim Annotieren im Zweifelsfall eher als poten-

ziell ungültig zu annotieren, um sicherzustellen, dass bei einer späteren Prüfung durch Menschen keine potenziell ungültigen Klauseln übersehen werden.

Zur thematischen Annotation stand eine zweistufige Taxonomie zur Verfügung, die über 27 Klassen und 37 Unterklassen verfügt. Diese Klassen beschreiben die am häufigsten auftretenden Klauseln in den AGB von Onlineshops und wurden basierend auf Formularbüchern und Muster-AGB entwickelt. Die Klassen umfassen zum Beispiel die Themen Rücktritt, Bezahlung, Garantie und Lieferung. Unterklassen für die Klasse „Bezahlung“ sind zum Beispiel Zahlungsmethoden, Zahlungsgebühren und Verzugszinsen.

Tabelle 1 zeigt am Beispiel zweier Klauseln, wie die fertig annotierten Daten vorliegen. Neben dem Klauseltext und den Annotationen verfügt jede Klausel über eine eindeutige ID und eine weitere ID, die sie eindeutig einem AGB-Vertragswerk zuordnet.

ID	AGB	Text	Thema	Unterthema	Unwirksam
135	12	Wenn Sie Zahlungen trotz Fälligkeit nicht leisten, behält sich der Veranstalter zudem vor, für die zweite Mahnung eine Mahnkostenpauschale von € 20,- zu erheben.	Bezahlung	Mahnkosten	1
765	25	Bis zur vollständigen Bezahlung bleibt die Ware Eigentum des jeweiligen Verkäufers.	Bezahlung	Eigentumsvorbehalt	0

Tabelle 1: Beispiel für annotierte AGB-Klauseln (Quelle: Eigene Darstellung).

Eine statische Analyse der Klauseln aus Verträgen, die komplett annotiert wurden, ergab, dass circa 6 Prozent aller Klauseln von den Expert:innen als potenziell unwirksam eingeschätzt wurden. Da im Beratungsalltag fast nie ganze Verträge geprüft werden können, sondern in der Regel nur punktuell geprüft wird, war dies auch für die Expert:innen der Verbraucherzentrale eine neue Erkenntnis.

## 4 Thematische Klassifizierung

Basierend auf den annotierten Daten wurden verschiedene Ansätze zur automatischen thematischen Klassifikation von Klauseln untersucht: angefangen von sehr einfachen Methoden wie einer klassischen Stichwortsuche über einfache neuronale Netze bis hin zu den zuvor erwähnten Transformer-Modellen. Während die einfache Stichwortsuche in den Experimenten eine Genauigkeit (englisch *accuracy*) von 78 Prozent erreichte, konnte ein einfaches neuronales Netz eine Genauigkeit von 81 Prozent erreichen und ein Transformer-Modell sogar eine Genauigkeit von 91 Prozent. Zwar hat das komplexeste Verfahren damit, wie zu erwarten war, am besten abgeschnitten, allerdings war dafür auch ein erheblicher Ressourceneinsatz sowohl im Bezug auf benötigte Trainingsdaten als auch im Bezug auf Rechenzeit nötig. Im Vergleich dazu hat ein relativ einfaches Verfahren des „klassischen“ maschinellen Lernens (logistische Regression) bei einem Bruchteil der benötigten Rechenzeit sogar eine Genauigkeit von 87 Prozent erreichen können.

Eine Hypothese, warum auch einfachere Verfahren die Aufgabe vergleichsweise gut lösen, basiert auf der häufigen Verwendung von Nominalkomposita. Nominalkomposita spielen in der deutschen Sprache allgemein eine wichtige Rolle, in der juristischen Sprache im Besonderen. Wörter wie „Widerrufsfrist“ oder „Mahnkostenpauschale“ sind eindeutige Identifizierungsmerkmale, die es selbst für primitive Verfahren wie die Stichwortsuche einfach machen, eine korrekte thematische Klassifikation durchzuführen. Die getrennte Verwendung der Nomen, zum Beispiel „Frist für den Widerruf“, bietet dagegen mehr Potenzial für Mehrdeutigkeit, da die einzelnen Bestandteile, hier insbesondere „Frist“, auch in anderen Kontexten auftauchen können.

## 5 Juristische Klassifizierung

Die juristische Klassifizierung einer Klausel ist aus menschlicher Sicht deutlich komplexer als die thematische Klassifizierung und benötigt neben Sprachverstehen auch juristische Expertise. Eine naheliegende Herangehensweise ist der Versuch, diese juristische Expertise in maschinenlesbaren Regeln festzuhalten. Zum Beispiel in der Form: Wenn es sich um einen Fernabsatzvertrag handelt und die Widerrufsfrist kürzer als 14 Tage ist, dann ist die Klausel potenziell unwirksam. Allerdings gilt das nur, wenn der:die Käufer:in Verbraucher:in ist. Außerdem gilt dies ebenfalls nicht, wenn es sich um Hygieneprodukte handelt, dessen Versiegelung geöffnet wurde, oder wenn es sich um Spezialanfertigungen handelt. Schnell wird klar, wie komplex die Abbildung der notwendigen rechtlichen Regelungen wäre. Die gesamte Breite der in AGB enthaltenen Regelungen damit abzudecken, erscheint praktisch unmöglich.

Beim überwachten maschinellen Lernen dagegen, werden dem Algorithmus tausende Klauseln übergeben, zusammen mit der rechtlichen Einschätzung von Annotator:innen. Basierend auf diesen Trainingsdaten entwickelt der Algorithmus dann selbst Heuristiken, nach denen er die Klassifikation in wirksam und unwirksam vornimmt. Konkret haben wir auch hierfür das BERT Transformer Modell verwendet. Auch hier diente als Grundlage der große, nicht annotierte, Textkorpus. Im zweiten Schritt wurde das Modell dieses Mal aber nicht mit den thematischen Annotationen trainiert, sondern mit den rechtlichen Annotationen.

Im Endergebnis war das Modell in der Lage potenziell unwirksame Klauseln mit einer Genauigkeit von 90 Prozent zu erkennen. Zumindest zum Teil lässt sich dieser, für alle Beteiligten überraschend, hohe Wert damit erklären, dass es manche Arten von Klauseln gibt, die, zumindest in den vorliegenden Daten, nahezu immer unwirksam waren. Darunter fallen zum Beispiel Klauseln, die automatische Preiserhöhungen vorsehen.



Ein inhärenter Nachteil der genutzten Methode ist es, dass das Modell keine Angaben darüber macht, warum es zum Schluss kommt, dass eine Klausel ungültig ist, sondern nur eine binäre Einschätzung abgibt.

## 6 Prototyp

Die Ergebnisse zeigen, dass sich sowohl die thematische Klassifikation als auch die juristische Einordnung von AGB-Klauseln im Bereich des Onlineshoppings zuverlässig mit KI durchführen lässt. Damit die Technologie Verbraucherschützer:innen im Arbeitsalltag aber aktiv unterstützen und insbesondere dazu beitragen kann, eine breitere Marktbeobachtung zu gewährleisten, reicht es nicht, einzelne Klauseln analysieren zu können, die wohlformatiert in Tabellen vorliegen. Die KI muss in ein Tool integriert werden, das die Arbeitsprozesse der Expert:innen unterstützt und vom automatischen Auffinden von AGB bis zur juristischen Auswertung alle Schritte unterstützen kann. Hierzu haben wir einen Softwareprototyp entwickelt, der sämtliche Schritte in einer sogenannten Pipeline-Architektur, wie in Abbildung 1 gezeigt, miteinander verbindet und über eine einfach zu bedienende Web-Schnittstelle zur Verfügung stellt.

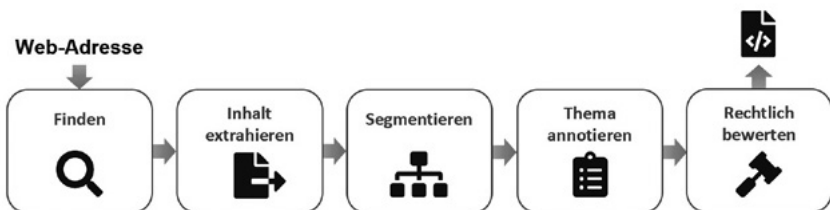


Abbildung 1: Pipeline-Architektur (Quelle: Eigene Darstellung).

Als Eingabe erwartete der Prototyp die Adresse eines Webshops, also zum Beispiel `shop.example.com`, oder eine Liste solcher Adressen. Im ersten Schritt wird dann automatisch die Unterseite ausfindig gemacht, die die AGB enthält. Dafür werden zunächst alle von der Startseite ausgehenden Verlinkungen

untersucht. Aus Gründen der Suchmaschinenoptimierung verwenden viele Webseiten sogenannte „sprechende“ Links, die innerhalb der Adresse die Abkürzung „AGB“ oder das Wort „Geschäft[s]bedingungen“ enthalten. Sollte kein solcher Link vorhanden sein, wird der Inhalt der verlinkten Seiten nacheinander auf AGB geprüft.

Sobald die AGB-Seite identifiziert wurde, wird zunächst der Inhalt der Seite extrahiert, also der eigentliche Vertragstext von anderen Elementen der Webseite getrennt, die die KI stören könnten. Dazu gehören zum Beispiel Navigations-elemente oder die Fußzeile. Hierzu kommt die Standardbibliothek „Trafilatura“ zum Einsatz. Als Ergebnis dieses zweiten Schrittes liegt der Vertrag in reiner Textform vor.

Im nächsten Schritt wird der Text nun in eine maschinenlesbare Struktur gebracht. Zwar haben die AGB in ihrer Ursprungsform auf der Webseite meist bereits eine für Menschen erkennbare Struktur, eingeteilt in Klauseln, mit Überschriften auf verschiedenen Ebenen, oft liegt diese Struktur aber nur optisch und damit nicht maschinenlesbar vor. Daher wird in diesem Schritt, durch Analyse der Aufzählungsstruktur von Überschriften und weiteren Merkmalen wie Schriftgröße und Absätzen, eine maschinenlesbare Strukturierung des Textes vorgenommen. Dieser vermeintlich einfache Schritt ist entscheidend für die Qualität der späteren thematischen und rechtlichen Klassifizierung. Werden die Klauseln hier nicht korrekt getrennt, wird es für den Algorithmus später schwer, Themen korrekt zu erkennen und eine juristische Einschätzung abzugeben. Die meisten Fehler, die der Prototyp zurzeit noch produziert, haben ihren Ursprung daher in diesem Schritt.

Am Ende der dieser Verarbeitung liegen die Vertragstexte maschinenlesbar und strukturiert vor und können dann einfach an die beiden zuvor trainierten KI-Modelle übergeben werden. Ausgabe der Pipeline ist eine maschinenlesbare, annotierte Version der AGB, die dann wiederum in der grafischen Benutzerschnittstelle angezeigt werden kann, wie in Abbildung 2 gezeigt.

AGB Check

Analysis | Menu | Expand | Open

**Table of Contents**

Terms and Conditions

1. Scope of Application
2. Definitions
3. Contract
4. Prices, price changes
5. Delivery Times
6. Delivery and transfer of risk
7. Warranty
8. Liability
9. Payment
10. Right of return
11. Delivery of used parts by the customer
12. Place of performance and jurisdiction

**Topics**

Applicability

Conclusion of Contract

Definition

Delivery

Time

Liability

Prices

VAT

Warranty

**Stats**

Void Clauses	0
No. of Clauses	40
No. of Words	1995

**Terms and Conditions**

**1. Scope of Application**

1.1 The deliveries, services and offers of LIMORA Ölölötmier GmbH are carried out exclusively on these Terms and Conditions. This shall also apply to all future business relations, even if they are not expressly agreed again.

**2. Definitions**

2.1. Consumer - In terms of these Terms and Conditions shall mean any natural person who enters into the contract for reasons that can be attributed neither to commercial activity nor relating to their independent vocational activity (§ 11 BGB)

2.2. The purpose of these Terms relates to any customer who concludes the contract as part of a commercial or professional activity (§ 14 para 1 BGB)

2.3. Means of distance communication are means of communication that can be used to initiate or enter into a contract between a consumer and a professional without the simultaneous physical presence of the parties, in particular letters, catalogs, phone calls, faxes, e-mails as well as radio, television and media services (§ 312 b paragraph 2 BGB)

**3. Contract**

**4. Prices, price changes**

**5. Delivery Times**

**6. Delivery and transfer of risk**

**7. Warranty**

**8. Liability**

**9. Payment**

**10. Right of return**

**11. Delivery of used parts by the customer**

**12. Place of performance and jurisdiction**

**Clause Inspector**

Annotation	Value	Source
Topics	Applicability	user
Valid	No	user
Add +		

Abbildung 2: Webbasierte grafische Benutzerschnittstelle  
(Quelle: Eigene Darstellung).

## 7 Übertragbarkeit

Die vorliegende Arbeit beschäftigt sich mit der Analyse von allgemeinen Geschäftsbedingungen in Onlineshops. Zwar lassen sich die verwendeten Technologien auch auf andere Arten von AGB und Verträge im Allgemeinen anwenden, aber nicht ohne entsprechende Anpassungen. So lässt sich zum Beispiel die Taxonomie zur Themenklassifikation nicht vollständig übertragen. In Versuchen mit AGB von Banken ergab sich zum Beispiel, dass sich gerade einmal 24 Prozent der Klauseln von Bank-AGB mit der bestehenden Taxonomie annotieren lassen. Dabei handelt es sich im Wesentlichen um technische Klauseln, die den Vertrag selbst betreffen, wie zum Beispiel Regelungen zu Vertragssprachen oder salvatorische Klauseln.

Regelungen, die das eigentliche Produkt betreffen, wie zum Beispiel Zinsklauseln, lassen sich dagegen mit der bestehenden Taxonomie nicht abbilden. Für die Klauseln, die von der Taxonomie abgedeckt sind, ergaben Tests jedoch, dass die Qualität der Klassifikation ähnlich gut ist wie bei den AGB aus Online-shops. Dies legt die Vermutung nahe, dass bei Vorhandensein entsprechender annotierter Daten auch andere Arten von Verträgen mit denselben technischen Mitteln automatisch analysiert werden könnten.

## 8 Zusammenfassung und Ausblick

Unsere Untersuchungen haben gezeigt, dass sich die Klauselkontrolle in AGB von Onlineshops mit KI unterstützen lässt, da moderne NLP-Methoden in der Lage sind, mit einer Genauigkeit von bis zu 90 Prozent sowohl das Thema einer Klausel zu identifizieren als auch eine rechtliche Einordnung in „potenziell unwirksam“ und „wirksam“ vorzunehmen. Nicht nur wegen der verbleibenden Fehlerquote von 10 Prozent sind solche KI-Methoden aber kein Ersatz für die Arbeit von Expert:innen. Um sinnvolle Handlungen aus einer als potenziell unwirksam erkannten Klausel abzuleiten, ist es notwendig zu verstehen, warum eine Klausel unwirksam ist. Hierzu sind die verwendeten Systeme aktuell noch nicht in der Lage.

Als Unterstützungsinstrument können die entwickelten Technologien aber schon heute die Arbeit von Verbraucherschützer:innen unterstützen: Insbesondere in der Beratung kann die automatische thematische Klassifikation dabei helfen, relevante Teile in langen Vertragsdokumenten schneller zu finden. Die automatische juristische Analyse kann insbesondere dabei helfen, eine breitere Analyse des Marktes zu ermöglichen und somit den Verbraucherschutz in der Breite zu stärken. Die KI-Analyse ermöglicht es, schnell eine große Anzahl von Vertragsdokumenten und Klauseln einer Vorprüfung zu unterziehen. Expert:innen können sich dann auf diejenigen Klauseln konzentrieren, die vom Algorithmus bereits als potenziell ungültig vorselektiert wurden.

Der zunehmende Einsatz von KI in Unternehmen und Anwaltskanzleien wird mittelfristig dafür sorgen, dass auch der Verbraucherschutz auf entsprechende Technologien angewiesen sein wird, um seiner Aufgabe angemessen nachkommen zu können. Dabei ist es jedoch wichtig im Blick zu behalten, dass Technologie keine nachhaltige Lösung für gesellschaftliche oder gesetzgeberische Probleme ist. Im Rahmen der AGB-Analyse stellt sich zum Beispiel die Frage, inwiefern weitere Informationspflichten für Unternehmen sinnvoll und zielführend sind, wenn die überwiegende Mehrheit der Verbraucher:innen schon heute keine AGB liest und deren Länge einer der Hauptgründe dafür ist (Braun 2021). Stattdessen erschiene es zielführender, Maßnahmen zu ergreifen, die die Länge von AGB begrenzen, um die Wahrscheinlichkeit zu erhöhen, dass die verbleibenden Regelungen auch tatsächlich von Verbraucher:innen gelesen werden. Denn bisher sind ausufernden Vertragstexten gesetzlich kaum Grenzen gesetzt (Oberlandesgericht Köln 2020).

## Literatur

- Bakos, Yannis, Florencia Marotta-Wurgler und David R. Trossen. 2014. Does anyone read the fine print? Consumer attention to standard-form contracts. *The Journal of Legal Studies* 43, Nr. 1: 1–35.
- Braun, Daniel. 2021. Automated semantic analysis, legal assessment, and summarization of standard form contracts. PhD Thesis – research external, graduation external. München: Technische Universität München.
- Harris, Zellig S. 1954. Distributional structure. *WORD* 10, Nr. 23: 146–162.
- Hillman, Robert A. 2005. On-line consumer standard-form contracting practices: A survey and discussion of legal implications. *Cornell Law Faculty Publications* 29.
- Rat der europäischen Union. 1993. Richtlinie 93/13/EWG des Rates vom 5. April 1993 über mißbräuchliche Klauseln in Verbraucherverträgen. <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CONSLEG:1993L0013:20111212:DE:PDF> (Zugriff: 16. März 2023).
- Oberlandesgericht Köln. 2020. Urteil vom 19.02.2020 – 6 u 184/19 [ECLI:DE:OLG K:2020:0219.6U184.19.00]. [http://www.justiz.nrw.de/nrwe/olgs/koeln/j2020/6\\_U\\_184\\_19\\_Urteil\\_20200219.html](http://www.justiz.nrw.de/nrwe/olgs/koeln/j2020/6_U_184_19_Urteil_20200219.html) (Zugriff: 16. März 2023).

## Über den Autor

**Dr. Daniel Braun;** Assistant Professor an der Universität Twente, Enschede, Niederlande, in der Sektion Industrial Engineering and Business Information Systems (IEBIS).