

Byron Meißner

3791829

9. Fachsemester

Seminar: „Der Grundsatz der Verhältnismäßigkeit: Grundlagen,  
Verbreitung, Problemlagen  
und Chancen“

SoSe 2024

Prof. Dr. Mattias Wendel

Willkür durch Abwägung? Der Grundsatz der Verhältnismäßigkeit als  
Instrument rechtlicher Beliebigkeit?

Danksagung .....	1
Abstract .....	1
Einleitung .....	1
Begriffsbestimmungen .....	6
Verhältnismäßigkeitsprinzip .....	6
Willkür .....	8
Subjektivität .....	9
Konventionalismus .....	9
These .....	10
Methode .....	14
Der Datensatz .....	15
Web-Scraping .....	15
Text-Mining .....	16
Legal Argument Mining .....	17
Transformer-Model .....	18
Datasets .....	19
Training .....	20
Inference .....	21
Ergebnisse & Auswertung .....	21
Ausblick .....	23
Glossar .....	25
Anhang .....	30
Literaturverzeichnis .....	33

# Danksagung

Mein besonderer Dank gilt Shashank Shetty dafür, dass er mir die zeitweise Nutzung der Rechenressourcen des Max-Planck-Instituts für Kognitions- und Neurowissenschaften ermöglicht hat und meinem Vater, dafür, dass er meine Arbeit geduldig Korrektur gelesen hat.

## Abstract

Die Jurisprudenz strebt stets danach, bessere und gerechtere Mittel für die erfolgreiche Abstimmung konfligierender Interessen bereitzustellen. Insoweit die Anwendung eines Rechtsprinzips für eine bessere Abstimmung dieser widerstreitenden Interessen sorgt, ist es gegenüber alternativen Rechtsprinzipien vorzugswürdig. Ein Maßstab für die Qualität eines Rechtsprinzips ist das Ausmaß der rechtlichen Willkür, das durch die Anwendung des Prinzips entsteht. In dieser Arbeit wird der Frage nachgegangen, inwieweit ein konkretes Rechtsprinzip - das Prinzip der Verhältnismäßigkeit - ein Instrument rechtlicher Willkür darstellt. Es wird ein methodischer Ansatz entwickelt, um eine fünfstellige Anzahl von Gerichtsurteilen des EGMR mittels modernster NLP<sup>1</sup> Technologien zu analysieren. Die für die Untersuchung erschaffenen Algorithmen, sowie eine Anleitung für die eigenständige Implementierung der für die Untersuchung nötigen Prozesse sind auf einem GitHub-Repository einsehbar.<sup>2</sup>

## Einleitung

Die klassische juristische Methode betrachtet für die Beantwortung einer Frage nach der Qualität eines rechtlichen Prinzips die in einem bestimmten Fachbereich gängigen Formalisierungen von legalen Prozessen und versucht durch die Kontrastierung verschiedener Argumente zu ermitteln, welche Gesetzgebung, Gesetzesauslegung und Gesetzesvollstreckung am ehesten den in einem Gesellschaftsvertrag angelegten Grundprinzipien entsprechen. Sie verlässt sich dabei auf die in Lehre und Rechtsprechung kultivierten Prinzipien nach denen, Akteuren, Rechtsgütern und Ereignissen entlang gebräuchlicher

---

<sup>1</sup> Siehe Glossar, Stichwort: Natural Language Processing

<sup>2</sup> <https://github.com/ByronMeissner/ArbitrarinessOfProportionality>.

Dimensionen abstrakte Werte und Qualitäten zugewiesen werden (Dogmatik).<sup>3</sup>

In der Betrachtung einschlägiger Einzelfälle erörtern Juristen, wie sich eine abstrakt und generell formulierte Gesetzgebung oder Lesart auf die konkrete, spezifische Realität überträgt. Zur Beantwortung der Frage, inwieweit das Verhältnismäßigkeitsprinzip ein Instrument rechtlicher Willkür darstellt, bestimmt ein klassischer Rechtswissenschaftler, auf welche Akteure, Ereignisse und Rechtsgüter das Prinzip seine Wirkung entfaltet, in welchem Bereich der Urteilsfindung das Prinzip zur Anwendung kommt und entlang welcher dogmatischen Dimensionen die Betroffenen bzw. das Betroffene einzuordnen sind.

Im öffentlichen Recht, dem wohl wichtigsten Anwendungsbereich des Verhältnismäßigkeitsprinzips, muss das gesellschaftliche Interesse an der Umsetzung einer Maßnahme dem individuellen Interesse am Ausbleiben dieser Maßnahme gegenübergestellt werden. Ein möglicher dogmatischer Rahmen für eine Abwägung in diesem Kontext liefert die Gewichtsformel Alexys.<sup>4</sup> Die Formel enthält drei unterschiedliche Parameter: Die Wertigkeit der betroffenen Interessen, die Wirkintensität der Maßnahme und die epistemische Ungewissheit des Wirkungsverlaufs. Je höherwertiger das individuelle Rechtsgut im Verhältnis zum öffentlichen Rechtsgut ist, je intensiver eine Maßnahme das individuelle Rechtsgut beschränkt und je ungewisser eine Förderung des öffentlichen Rechtsguts bei Umsetzung der Maßnahme ist, desto eher müsste die Jurisprudenz die Durchsetzung einer zu prüfenden Maßnahme unterbinden. Im nächsten Schritt der Analyse würde der Wissenschaftler Spezifika tatsächlicher und imaginärer Sachverhalte Werte entlang dieser drei Dimensionen zuweisen und anschließend teleologische, systematische, historische und semantische Auslegungsmethoden verwenden, um zu diskutieren, ob die Anwendung des Verhältnismäßigkeitsprinzips innerhalb dieser Dimensionen für eine optimale Abstimmung der konfligierenden Interessen sorgt. Der Wortlaut der Verfassung, historische Rechtsprechung oder ein Vergleich mit anderen Jurisdiktionen, die sich bei vergleichbaren Entscheidungen auf ein anderes Rechtsprinzip verlassen, können dem Juristen Stoff für eine solche Diskussion liefern. Diese Herangehensweise ist altbewährt.

---

<sup>3</sup> Kühl/Reichold/Ronellenfitsch, Einführung in die Rechtswissenschaft (2019), S.24.

<sup>4</sup> Alexy, Die Gewichtsformel (2003), S. 140.

Sie hat jedoch auch einige offensichtliche Schwächen. Zum einen gibt es keine einheitliche Formalisierung rechtlicher Dogmatik. Bereits auf regionaler und nationaler Ebene gibt es erhebliche Uneinigkeiten darüber, wie ein Prüfungsschema zu strukturieren sei und welche Dimensionen man zu unterscheiden habe. So könnte ein Jurist gegen die beispielhaft angeführte Formalisierung einwenden, dass die Unterscheidung zwischen der epistemischen Ungewissheit, der Wirkintensität der Maßnahme und der abstrakten Wertigkeit der betroffenen Interessen eine völlig willkürliche und zudem wenig sinnvolle sei. Er könnte argumentieren, dass eine solche Formalisierung die Beurteilung von Sachverhalten unnötig verkomplizieren würde und man diese drei Dimensionen besser in einem „rechtlichen Erwartungswert“ zusammenfassen sollte, der die theoretisch zu erwartende relative Häufigkeit eines bestimmten Wirkungseintritts ins Verhältnis zu dem Einfluss auf das betroffene Rechtsinteresse setzt. Es ist unmöglich eine solche Behauptung zu falsifizieren.

Aber sogar dort, wo es den Juristen gelingt, sich auf eine Formalisierung zu einigen, entsteht das nächste Problem, wenn den abstrakten Dimensionen konkrete Werte zugeordnet werden sollen. Zwei unterschiedliche Rechtsgüter, wie bspw. Vermögen und körperliche Integrität, sind stets inkommensurabel. Sogar, wenn man eine Kommensurabilität der Rechtsgüter *per se* fingiert, wie es Juristen zu tun genötigt sind, lassen sich keine mathematischen Gleichungen aufstellen, weil die betroffenen Menschen inkommensurabel bleiben. Nur weil man in einem Fall zu dem Ergebnis gekommen ist, dass eine Körperverletzung mit 5000€ Schmerzensgeld zu kompensieren sei, folgt daraus nicht, dass eine solche Form der Körperverletzung 5000€ pro Person „kostet“.

Die Sorge, dass ein Rechtssystem Willkür erzeugt, lässt sich in wesentlichen Teilen auf die Erkenntnis zurückführen, dass die Fähigkeit des Menschen, objektive Entscheidungen zu treffen, begrenzt ist. Gemeint ist nicht allein die Informationsaufnahme- und Verarbeitungskapazität des einzelnen Wissenschaftlers. Denn derlei Begrenzungen kann man teilweise dadurch überwinden, dass man mehr Wissenschaftler, besser ausbildet. Gemeint ist auch die Unvollkommenheit des menschlichen Beurteilens generell.

Einem aufmerksamen Beobachter entgeht nicht wie sehr sich die Einschätzung eines Ereignisses in Bezug auf seine epistemische Ungewissheit auf der

individuellen und der kollektiven Ebene in Abhängigkeit bestimmter Umstände verändert. Wenn eine Gesellschaft durch eine ungewisse Bedrohung in Aufruhr versetzt ist, neigen die Menschen dazu, Gefahren als größer einzuschätzen als sie es in Zeiten des Friedens tun würden (Risikobias). Und je einfacher es für Menschen ist eine bestimmte Information zu erinnern (z.B., weil sie diese besonders häufig präsentiert bekommen), desto höher schätzen sie auch die Relevanz dieser Information ein (Availability Bias)<sup>5</sup>. Graphik 1 zeigt die deutsche Wahrnehmung terroristischer Bedrohung in Abhängigkeit von einschlägigen Terrorakten mit besonders großer Medienresonanz.

**Graphik 1:**



<https://de.statista.com/infografik/5034/angst-vor-terrorismus-hat-sich-seit-9-11-verdoppelt/>,  
abgerufen: 21.05.2024.

Zuletzt hat die Coronapandemie die Risikoeinschätzung der Bevölkerung deutlich verändert.<sup>6</sup> Die verwaltungsrechtlichen Konflikte, die im Rahmen der Corona-Pandemie entstanden sind, haben zu anhaltenden Diskussionen unter Juristen darüber geführt, ob die traditionelle Anwendung des Verhältnismäßigkeitsprinzips die staatliche Gewalt zu sehr in ihrer

<sup>5</sup> Kahnemann/Tversky, Judgment under Uncertainty: Heuristics and Biases (1974).

<sup>6</sup> <https://de.statista.com/statistik/daten/studie/1287784/umfrage/auswirkungen-der-corona-pandemie-auf-globale-risiken/>, abgerufen: 21.05.2024.

Gestaltungsmacht begrenzt.<sup>7</sup> Auch diejenigen, die das Prinzip nicht gleich gänzlich abschaffen wollen, fordern oft, zumindest bei solchen Fällen, die im Rahmen der Verhältnismäßigkeitsabwägung nicht eindeutig entschieden werden können, nicht wie in der Vergangenheit der Freiheit, sondern stattdessen dem öffentlichen Interesse Vorzug zu gewähren.<sup>8</sup> Nichtsdestotrotz führte die deutsche Rechtsprechung führte der Krise wiederholt Verhältnismäßigkeitsabwägungen an, um Verwaltungsmaßnahmen im Interesse der Volksgesundheit zu delegitimieren. So geschah es bspw. in einem Urteil des Mannheimer Verwaltungsgerichts, in dem eine Testpflicht für die Mitarbeiter eines Schlachthauses für unverhältnismäßig bestimmt wurde<sup>9</sup> oder in einem Urteil des Bayrischen Verwaltungsgerichtshofs in München, in dem ein Verbot öffentlicher Grillveranstaltungen ausgeschlossen wurde.<sup>10</sup> S Unabhängig davon, wo man sich als Jurist in diesem Diskurs verorten mag, sollte man anerkennen, dass wir als Menschen und deshalb auch als Richter und Gesetzgeber nicht frei sind von der Wirkung, welche die genannten kognitiven Phänomene auf uns haben. Die Jurisprudenz hat daher die Verantwortung den Werkzeugkasten ihrer Wahrheitsfindung, um solche methodischen Ansätze zu ergänzen, die eine unbefangene Betrachtung der rechtlichen Materie ermöglichen. Neue Technologien der Textverarbeitung, getragen von der enormen Rechenkapazität moderner Computer erlauben es innerhalb von Minuten Text-Corpora zu verarbeiten, die ein einzelner Mensch sogar dann nicht prüfen könnte, wenn er diesem Unterfangen sein ganzes Leben widmen würde. Die Betrachtung einer sehr großen Zahl von legalen Dokumenten über einen ausgedehnten Zeitraum, kann und sollte für den Rechtswissenschaftler der Gegenwart ein Hilfsmittel sein, um ihm seine eigene Befangenheit vor Augen zu führen. Die Analyse neuer Muster und Strukturen auf der „Makroebene“ der Rechtswissenschaft kann neue Erkenntnisse über die Qualität rechtlicher Prinzipien zu Tage fördern. Und sie kann diese auch in Bezug setzen zu solchen Ereignissen, die unsere kollektive Urteilsfähigkeit trüben.

Es ist jedoch wichtig über die Euphorie für die neuen Möglichkeiten umgekehrt deren Begrenzungen nicht aus den Augen zu verlieren. Solche

---

<sup>7</sup> BVerfGE vom 19. November 2021 - 1 BvR 781/21 -, Rn. 1-306.

<sup>8</sup> Pavel/Horák, Proportionality during Times of Crisis (2024).

<sup>9</sup> VGH Baden-Württemberg 30.06.2020, 1 S 2087/20.

<sup>10</sup> VGH Bavaria 01.09.2020, 20 NE 20.1754.

Begrenzungen existieren nicht nur, weil die wissenschaftlichen Prozesse der Selektion, Verarbeitung, Darstellung und Interpretation von großen Datenmengen ebenso sehr durch Befangenheit geprägt sind, wie die analytischen Prozesse der klassischen Rechtswissenschaft.

Sie existieren auch nicht nur, weil die juristische Analyse nicht mathematisch formalisiert werden kann. Sie existieren insbesondere auch deshalb, weil die klassische juristische Methode sich in anderer Weise, als die datenanalytische Betrachtungsweise es vermag, das unübertroffene Erkenntnispotential menschlicher Intuition zu Nutze macht. Es mag dem Selbstverständnis des aufrichtig wahrheitssuchenden Juristen widersprechen, doch nicht selten, kennt er sein Ergebnis, bevor er seine Nachforschungen anstellt. Die Begründung, auf die er seine Erkenntnis stellt, ist oft keine wertungsfreie Betrachtung möglicher Ursachenzusammenhänge, sondern vielmehr ein überzeugter Überzeugungsversuch, der aus seiner Erkenntnis erwachsen ist und nicht umgekehrt seine Erkenntnis erst erzeugt hat. Anders als es das Ideal des hyperrationalen, losgelösten Wissenschaftlers vermuten lassen würde, ist das Vertrauen in diese Intuition nicht nur ein Mangel der juristischen Methode, es ist zugleich Quelle ihres Reichtums. Denn es ist gerade auch die menschliche Intuition eine der seltenen Qualitäten, die uns heute noch von der Maschine unterscheiden. Die tiefgehende, qualitative Analyse des rechtlich geschulten Juristen kann daher nicht einfach durch die Interpretation statistischer Zusammenhänge in großen Datenmengen ersetzt werden. Ein interdisziplinärer Ansatz verspricht jedoch die juristische Methode wertschöpfend zu ergänzen.

## **Begriffsbestimmungen**

Für eine sinnvolle Beantwortung der Fragestellung ist zunächst zu bestimmen, was „Willkür“ und „Verhältnismäßigkeitsprinzip“ jeweils bedeuten.

## **Verhältnismäßigkeitsprinzip**

Eine wichtige Grundlage für das Verständnis des modernen Menschenrechts liegt in der Unterscheidung zwischen dem Schutzbereich eines konstitutionellen Rechts auf der einen und der Rechtmäßigkeit seiner Einschränkung auf der anderen Seite. Die meisten konstitutionellen Rechte sind relative und keine absoluten Rechte. Das heißt, dass ihre Einschränkung



gerechtfertigt werden kann. Es ist die Verhältnismäßigkeit, die über das Bestehen und das Ausmaß solcher Rechtfertigungen entscheidet.<sup>11</sup> Das Verhältnismäßigkeitsprinzip als ein juristischer Formalismus ist ein Prinzip rechtlicher Entscheidungsfindung, das in den 50er Jahren des letzten Jahrhunderts aus dem deutschen Verwaltungsrecht hervorgegangen ist und heute in den allermeisten Demokratien einen festen Bestandteil der Rechtssystematik darstellt.<sup>12</sup> Die USA sind die wohl prominenteste Ausnahmen zu dieser Regel. Wenn man die juristische Entscheidungsfindung als einen Prozess versteht, der den Algorithmen der rechtlichen Dogmatik folgt, so ist das Verhältnismäßigkeitsprinzip eine Operation, die aus vier Schritten besteht und determiniert, inwieweit einer öffentlich-rechtlichen Maßnahme den individuellen Interessen eines Rechtssubjekts Vorrang zu gewähren ist. Im ersten Schritt wird festgestellt, ob die zu prüfende Maßnahme ein berechtigtes Interesse verfolgt (*Legitimer Zweck*). Im zweiten Schritt wird geprüft, ob die Maßnahme dieses Interesse auch tatsächlich fördern kann (*Geeignetheit*). Im dritten Schritt, wird untersucht, ob außer der zu prüfenden Maßnahme noch andere, gleich effektive Mittel existieren, um das durch die Maßnahme verfolgte Interesse zu fördern, ohne das Interesse des betroffenen Rechtssubjekts in gleicher Weise zu belasten (*Erforderlichkeit*). Wenn die Anforderungen der ersten drei Schritte erfüllt sind, wird im vierten und letzten Schritt eine Abwägung zwischen den konfligierenden Interessen vorgenommen, in der sie nach ihrer abstrakten Wertigkeit und dem Grad ihrer (möglichen) Betroffenheit gegeneinander aufgewogen werden (*Verhältnismäßigkeit strictu sensu*). Das Verhältnismäßigkeitsprinzip soll ein flexibles Rechtssystem ermöglichen, in dem die Staatsgewalt möglichst frei von den starren Zwängen abstrakter Regeln, für konkrete Einzelfälle vernünftige Entscheidungen treffen kann.<sup>13</sup> Doch es ist auch gerade diese Flexibilität der Richter- und Verwaltungsentscheidung, die das Verhältnismäßigkeitsprinzip unter seinen Kritikern als ein Instrument rechtlicher Willkür gelten lässt. Im Kern der Kritik steht die Behauptung, dass abstrakte Rechtsgüter inkommensurabel seien.<sup>14</sup> Es heißt außerdem die Entscheidung über das öffentliche Interesse den

---

<sup>11</sup> Barak, Proportionality: Constitutional Rights and Their Limitations (2012), S. 131.

<sup>12</sup> Barak, Proportionality: Constitutional Rights and Their Limitations (2012), S. 10.

<sup>13</sup> Hufen, Staatsrecht II (2020), S.113.

<sup>14</sup> Schlink, Abwägung im Verfassungsrecht (1976), S.151.

subjektiven Wertungen einzelner Richter und Verwaltungsbeamter zu überlassen, würde die Gestaltungsmacht der Legislative zu sehr beschränken<sup>15</sup> und dem Rechtssubjekt zudem Gewissheit über die rechtlichen Folgen seiner Handlungen nehmen.<sup>16</sup>

## Willkür

Der Begriff ist seit dem 13. Jahrhundert belegt<sup>17</sup> und ist mutmaßlich von dem mittelhochdeutschen „Willekür“ abgeleitet. Im alltäglichen Sprachgebrauch assoziiert man mit dem Wort i.d.R. eine beliebige, ggf. sogar despotische Willensausübung. Jemand handelt willkürlich, wenn er seinen Willen unabhängig oder entgegen der in einem bestimmten Kontext geltenden Regeln manifestiert. In der deutschen Jurisdiktion soll das abgeschlossene System von Gesetzgebung und Auslegungsregeln eine beliebige subjektive Einflussnahme auf das rechtliche Ergebnis unmöglich machen – es gilt ein allgemeines Willkürverbot für alle in den Rechtsfindungsprozess involvierten Parteien und Institutionen. Im Zivilrecht kommt dies implizit zum Beispiel in § 626 BGB zum Ausdruck, der das Recht einer Privatpartei zur fristlosen Kündigung nur bei Vorliegen eines einschlägigen Grundes vorsieht. Wesentlich wichtiger noch ist das Willkürverbot in der Beziehung zwischen privaten Personen und der hoheitlichen Gewalt. So verpflichtet § 39 VwVfG bspw. die Exekutivgewalt der Bekanntgabe eines Verwaltungsaktes eine schriftliche Begründung beizufügen. In einem Urteil von 2018 über eine mutmaßliche Verletzung des Art. 3 I GG durch richterliche Willkür bei der Entscheidung über einen Schadensersatzbetrag beschreibt das Bundesverfassungsgericht rechtliche Willkür wie folgt:

*„Eine gerichtliche Entscheidung ist dann willkürlich, wenn sie unter keinem denkbaren Aspekt rechtlich vertretbar ist und sich daher der Schluss aufdrängt, dass sie auf sachfremden Erwägungen beruht. [...] Willkür liegt [...] vor, wenn eine offensichtlich einschlägige Norm nicht berücksichtigt oder der Inhalt einer Norm in krasser Weise missdeutet wird.“<sup>18</sup>*

---

<sup>15</sup> Dąbrowska-Kłosińska, The Protection of Human Rights in Pandemics (2021).

<sup>16</sup> Barak, Proportionality: Constitutional Rights and Their Limitations (2012), S. 490.

<sup>17</sup> Kluge, Friedrich, Etymologisches Wörterbuch der deutschen Sprache (2001), S. 990.

<sup>18</sup> BVerfGE vom 14. Februar 2018 - 1 BvR 2120/16 -, Rn. 12.

Die Willkür ist auch in anderen Domänen als ein Fachbegriff anerkannt. In der Medizin bezeichnet Willkür die bewusste Betätigung einer Körperfunktion - also die Kontrollierung des somatischen Nervensystems.<sup>19</sup> Der morgendliche Griff zur Kaffeetasse ist demnach medizinisch gesehen zunächst ein Willkürakt. In der Naturwissenschaft bezeichnet die Willkür eine unbegründete Festlegung.<sup>20</sup> So ist z.B. die aus der Mathematik bekannte Entscheidung die Kreiskonstante als  $\pi$  zu bezeichnen ein Willkürakt. Der Sprachwissenschaftler Ferdinand de Saussure postulierte in seinem Grundkurs der Linguistik *Das Prinzip von der Willkür des Zeichens*<sup>21</sup>, demzufolge die Beziehung zwischen einem Konzept (dem Signifikat) und einem Ausdruck dafür (dem Signifikanten) eine ist, die nicht im Wesen des Konzepts bereits angelegt ist, sondern vielmehr durch den willkürlichen Gebrauch von Sprache über die Zeit von den Sprechenden erschaffen wird.

Wenn man die Bedeutungsfacetten des Willkürbegriffs über verschiedene Domänen hinweg betrachtet, scheint es abstrakt zwei wesentliche Merkmale zu geben, die für Willkür über diese unterschiedlichen Bedeutungskontexte hinweg bestehen: Die initiale Subjektivität der (willkürlichen) Entscheidung, sowie die Möglichkeit eine Willkür durch Konvention zum Gesetz zu machen.

## **Subjektivität**

Die Willkür setzt stets eine Entscheidung voraus, die keine Folge systemischen Zwangs ist, sondern den Willen des handelnden Subjekts oder einer Gruppe von Subjekten realisiert. Ebenso wenig wie der morgendliche Griff zur Kaffeetasse ein Resultat physiologischen Zwangs ist, besteht eine mathematische Notwendigkeit die Kreiskonstante  $\pi$  zu nennen oder eine natürliche Notwendigkeit, das neugeborene Kind eines Hundes als Welpen zu bezeichnen.

## **Konventionalismus**

Willkür in jedem der benannten Kontexte kann sich über die Zeit zu festen Gesetzen verhärten. So ist die zunächst willkürlich getroffene Entscheidung die Kreiskonstante in der Mathematik mit dem Buchstaben  $\pi$  zu betiteln über die Zeit so sehr zum Brauch geworden, dass das Zeichen den meisten Leuten wohl

---

<sup>19</sup> <https://de.wikipedia.org/wiki/Willk%C3%BCr>, abgerufen am 31.05.24.

<sup>20</sup> <https://de.wiktionary.org/wiki/Willk%C3%BCr>, abgerufen am 31.05.24.

<sup>21</sup> Saussure, *Les Sources Manuscrites Du Cours De Linguistique Générale* (1979), S.31.

eher bekannt ist als die Konstante selbst. Der morgendliche Griff zur Kaffeetasse kann über die Zeit so sehr zur Gewohnheit werden, dass das handelnde Subjekt ihn als eine physiologische Notwendigkeit erfährt. Und auch die Jurisprudenz kennt diese Macht des Konventionalismus. So können sich über Zeit Wertungen, die auf willkürlichen Entscheidungen beruhen und keiner haltbaren rechtlichen Logik folgen (wie z.B. das Inzestverbot, der Tatbestand des Mordes oder die historisch lange Zeit wirksame Einschränkung von Frauenrechten etc.) als Sittengesetze manifestieren.

## These

Die Frage, inwieweit das Verhältnismäßigkeitsprinzip das Kriterium des Konventionalismus erfüllt, ist empirisch leicht zu beantworten. Das ist der Fall. Verwaltung, Rechtsprechung und Lehre beziehen sich regelmäßig explizit oder implizit auf vergangene Verhältnismäßigkeitsabwägungen der obersten Gerichte. Berühmte Fälle wie das Lüth-Urteil (1958) zur Meinungsfreiheit, das Volkszählungsurteil (1983) zur informationellen Selbstbestimmung, die Brokdorf-Entscheidung (1985), der Kruzifixbeschluss (1995) zur Religionsfreiheit und das Sterbehilfe-Urteil (2020) zur Disponibilität des Lebens haben nicht nur das deutsche Recht, sondern darüber hinaus auch die europäische und internationale Rechtsprechung nachhaltig beeinflusst. Bestimmte rechtliche Wertungen, die auf Grundlage der gleichen Gesetzestexte vor 75 Jahren selbstverständlich gewesen wären, sind heute unvorstellbar geworden. Oft sind es Verhältnismäßigkeitsabwägungen des Bundesverfassungsgerichts gewesen, die auf rechtlicher Ebene die Weichen für solcherlei Veränderungen gestellt haben. Neuartige Wertungen sind durch das Verhältnismäßigkeitsprinzip zur Konvention geworden.

Deutlich schwieriger zu beantworten ist die Frage nach der Subjektivität der Verhältnismäßigkeitsabwägungen. Das Verhältnismäßigkeitsprinzip wird oft als ein Ausdruck praktischer Vernunft verstanden. Diese Ansicht mag intuitiv durch die Beobachtung bestärkt werden, dass die „Rationalität“ das „Verhältnis“ in ihrem Wortstamm bereits enthält (von lat. *ratio*). Als eine Rationalitätskontrolle dient das Verhältnismäßigkeitsprinzip den Anhängern dieser Ansicht nicht in erster Linie dem Zweck gesetzgeberische Wertungen zu korrigieren. Es ist für sie stattdessen vor allem ein Korrektiv für die

Verfehlungen und Unbedachtheiten, die ein natürliches Beiprodukt des komplexen Gesetzgebungsprozesses darstellen.<sup>22</sup> Oft werden derlei Begründungen von abenteuerlichen Versuchen begleitet, rechtliche Problematiken mathematisch zu formalisieren.<sup>23</sup> Zur Verteidigung dieser Betrachtungsweise wird auch angeführt, dass Rationalität sich nicht durch eine absolute Determinierbarkeit auszeichne. Eine rationale Entscheidung ist demnach nicht die einzig richtige, sondern eine von ggf. mehreren rational begründbaren Entscheidungen.<sup>24</sup>

Auch die Menge der rationalen Entscheidungen zu erweitern, sodass sie nicht nur eine einzig richtige Entscheidung enthält, ändert jedoch nichts daran, dass nicht objektiv festgestellt werden kann, ob eine spezifische Entscheidung den Minimalanforderungen für Rationalität gerecht wird oder nicht. Es ist gerade dies die Tragik der Rationalitätswaage, dass sie ohne eine initiale Kalibrierung nicht funktionieren kann. Die Relevanz dieser Erkenntnis wird besonders deutlich, wenn man sich vor Augen führt, dass autokratische Länder wie China und Russland ihre Verfassungsrechtsprechung ebenso sehr unter Verwendung von Verhältnismäßigkeitsabwägungen durchführen, wie demokratische Länder wie Deutschland oder Frankreich es tun.<sup>25</sup> Das Problem wird verlagert. Die Inkommensurabilität der Rechtsgüter wird ersetzt durch die Inkommensurabilität der Differenz zwischen einem Richterurteil und den abstrakten Weisungen der objektiven Rechtsgrundlagen, auf die das Urteil gestellt wird. Genauso wie für die Abwägung zwischen verschiedenen Rechtsgütern lässt sich für die Einordnung verschiedener Rechtsurteile entlang der (fiktiven) „Rationalitätsskala“ das berühmte Diktum des US Supreme Court Richters Scalias anführen, nach dem eine solche Einschätzung dem Urteil gleichkommt *„that a particular line is longer, than a particular rock is heavy.“*<sup>26</sup> Eine andere Ansicht, die das Verhältnismäßigkeitsprinzip für ein rationales Kontrollinstrument hält und ebenfalls die Gefahr einer übermäßigen Subjektivität bestreitet, tut dies paradoxerweise nicht, indem sie auf die

---

<sup>22</sup> Petersen, Verhältnismäßigkeit als eine Rationalitätskontrolle (2015), S. 161.

<sup>23</sup> Petersen, Verhältnismäßigkeit als eine Rationalitätskontrolle (2015), S. 65.

<sup>24</sup> Alexy, Theorie der Grundrechte (2011), S. 144.

<sup>25</sup> Barak, Proportionality: Constitutional Rights and Their Limitations (2012), S. 200.

<sup>26</sup> BENDIX AUTOLITE CORP. v. MIDWESCO ENTERPRISES, INC., et al., 486 U.S. 888 Supreme Court.

Rationalität der Richterentscheidung abstellt, sondern führt dies stattdessen auf die Sorge der Richter zurück, die öffentliche Legitimität zu verlieren, die für ihre Autorität entscheidend ist.<sup>27</sup> Ungeachtet der Tatsache, dass es zumindest fragwürdig erscheint, inwieweit die Angst vor dem öffentlichen Achtungsverlust eine gute Grundlage für rationale Entscheidungen bilden kann, bleibt in jedem Fall ein Problem bestehen: Es gibt in der klassischen Rechtswissenschaft keinen Maßstab für den Grad, zu dem eine subjektive Entscheidung von ihren objektiven Rechtsgrundlagen oder der kollektiven Wertung abweicht. Wenn man einen Maßstab entwickeln könnte, der zumindest als eine Annäherung dafür dient, wäre viel gewonnen.

In den wirkmächtigsten Gerichten, die Verhältnismäßigkeitsabwägungen anstellen, sitzen i.d.R. mehrere Richter, die gemeinsam über das Urteil entscheiden. Vor dem US Supreme Court sind es 9, vor dem Bundesverfassungsgericht 8 und vor dem Europäischen Gerichtshof für Menschenrechte sind es 3, 7 oder 17, je nach der Relevanz eines zu fällenden Urteils. Damit ein Richter eine Stelle an einem dieser Gerichte erhält muss er eine langwierige juristische Ausbildung durchlaufen und seine herausragende rechtliche Kenntnis dabei vielfach unter Beweis stellen. Sachlichkeit, Unbefangenheit und eine uneingeschränkte Loyalität gegenüber den objektiven Rechtsgrundlagen seiner jeweiligen Jurisdiktion sind unverzichtbare Kernkompetenzen eines obersten Richters. Wenn einer oder mehrere der urteilenden Richter sich dem Leiturteil der Richtermehrheit partout nicht anschließen wollen, gibt es gelegentlich abweichende Meinungen. Eine größere Uneinigkeit zwischen den urteilenden Richtern könnte ein Indiz für einen größeren Einfluss des subjektiven Richterwillens auf die Entscheidung sein. Uneinigkeit zwischen den urteilenden Richtern lässt sich quantifizieren. Sie kommt zum einen im Anteil der Dissens-Entscheidungen an der Gesamtmenge der Urteile generell zum Ausdruck. Sie lässt sich zum anderen auch in der statistischen Varianz der Stimmverhältnisse unter den Dissens-Entscheidungen erkennen. Die Varianz ist ein statistisches Maß für die Streuung der Datenpunkte in einem Datensatzes um den Mittelwert.

---

<sup>27</sup> Peterson, Proportionality and Political Activism (2018).

Je größer die Varianz in einem Datensatz ist, desto weiter sind die einzelnen Datenpunkte um den Mittelwert verteilt. Die Varianz der Stimmverhältnisse in einer gegebenen Entscheidungsmenge gibt Auskunft darüber, wie häufig sehr gut ausgebildete Juristen über den gleichen Sachverhalt unterschiedlich urteilen. Wenn die Varianz der Stimmverhältnisse größer ist, kann man annehmen, dass die einzelnen Richterentscheidungen, weniger gut objektiv legitimierbar gewesen sind. Der Einfluss des subjektiven Richterwillens ist dann größer gewesen und das Maß an Willkür entsprechend höher. Diese Überlegungen lassen sich wie folgt formalisieren:

Sei  $D$  eine Teilmenge der Urteilsmenge  $U$ , die alle diejenigen Urteile  $u$  umfasst, bei denen es zu einem materiellen Dissens zwischen den beteiligten Richtern gekommen ist:

$$D = \{u \in U \mid \text{Dissens}(u)\}$$

und seien  $M$  und  $m$  respektive jeweils die Größe des Mehrheitsvotums und des Minderheitenvotums, dann lässt sich der Anteil der Dissens-Urteile  $\alpha$  ausdrücken als:

$$\alpha = \frac{D}{U}$$

die Varianz der Stimmverhältnisse  $\sigma^2$  in einer bestimmten Urteilsmenge lässt sich dann beschreiben als:

$$\sigma^2 = \frac{1}{D} \sum_{i=1}^D \left( \frac{m_i}{M_i} - \frac{\sum_{i=1}^D \frac{m_i}{M_i}}{D} \right)$$

Wenn die Annahme zutrifft, dass das Verhältnismäßigkeitsprinzip die Ursache größerer Willkürlichkeit in der Rechtsprechung ist, dann ist zu erwarten, dass bei einer häufigeren Verwendung des Verhältnismäßigkeitsprinzips für die Entscheidungsfindung in  $U$ ,  $\alpha(U)$  und  $\sigma^2(U)$  größer werden.

Um eine solche Beobachtung anzustellen, gilt es für eine große Anzahl an Gerichtsurteilen zu bestimmen:

1. Beinhaltet ein Urteil eine Dissens-Entscheidung?
2. Was ist das Verhältnis vom Mehrheitsvotum zum Minderheitenvotum?
3. Wie häufig wurde für die Entscheidungsbegründung eines Urteils das Verhältnismäßigkeitsprinzip verwendet?

## Methode

Die Untersuchung ist in vier wesentlichen Schritten erfolgt. Zunächst habe ich ein Web-Scraping-Tool<sup>28</sup> verwendet, um 63202 Urteile herunterzuladen und aufzubereiten, die auf der offiziellen Datenbank des EGMR<sup>29</sup> zur allgemeinen Sichtung freigestellt sind. Als nächstes habe ich einen Text-Mining<sup>30</sup> Algorithmus konzipiert, um die aufbereiteten .txt Dateien nach bestimmten Kriterien zu filtern. So konnte ich die Fälle zum Beispiel danach klassifizieren welches Land jeweils beteiligt war, ob das Gericht im Sinne des Klägers entschieden hat und inwieweit unter den Richtern Einigkeit über die getroffene Entscheidung bestand. Nach dem zweiten Schritt bestand der Datensatz noch aus 15741 Gerichtsurteilen. Im dritten Schritt habe ich einen Datensatz von Habernal et. al<sup>31</sup> verwendet, um das LLaMA3 Transformer-Model auf die Klassifizierung juristischer Argumente zu trainieren, sodass ich die 15471 Gerichtsurteile danach kategorisieren konnte, ob sie nach einer Verhältnismäßigkeitsabwägung *strictu sensu* gefällt wurden oder nicht. Im vierten und letzten Schritt habe ich die Ergebnisse aus dem zweiten und dritten Schritt in verschiedenen Weisen miteinander in Beziehung gesetzt, um dadurch Aufschluss über den unterschiedlichen Willkürgehalt der Entscheidungen in Abhängigkeit davon zu bekommen, ob sie unter Verwendung einer Verhältnismäßigkeitsabwägung *strictu sensu* getroffen worden sind. Die Algorithmen für die Untersuchung wurden in Python 3.12 geschrieben und in einem Miniconda-Environment ausgeführt. Das Transformer Modell wurde mit der PyTorch Bibliothek instruiert und auf einer leistungsfähigen NVIDIA A100-SXM4-40GB GPU der Max Planck Computing and Data Facility trainiert. Die relevanten Algorithmen, Datensätze, Workflows und Transformer-

---

<sup>28</sup> Siehe Glossar, Stichwort: Web-Scraping.

<sup>29</sup> <https://hudoc.echr.coe.int/>, abgerufen: 14.05.2024.

<sup>30</sup> Siehe Glossar, Stichwort: Text-Mining.

<sup>31</sup> Habernal et. al, Mining Legal Arguments in Court Decisions (2023).



Models<sup>32</sup> liegen in einem Github Repository zu Einsicht bereit das ich für diese Untersuchung erstellt habe.<sup>33</sup>

## Der Datensatz

Die massive Schrumpfung des Datensatzes im Vergleich zur ursprünglich heruntergeladenen Menge an Gerichtsurteilen (von 63202 auf 15741) ist auf verschiedene Ursachen, in erster Linie aber auf den Umstand zurückzuführen, dass ich den Datensatz im Rahmen des Textaufbereitungsschritts auf diejenigen Schriftsätze reduziert habe, die vollständig in englischer Sprache vorliegen, um sprachliche Einheitlichkeit zu gewähren. Die Schriftsätze sind jeweils einem konkreten Fall gewidmet. Der Zuständigkeit des EGMR entsprechend geht es in allen Fällen um eine Streitigkeit zwischen einem Mitgliedsstaat des Europarats und einem Rechtssubjekt, dass in diesem Staat bereits sämtliche Revisionsinstanzen (erfolglos) durchlaufen hat. Alle Fälle sind zwischen 1960 und dem Zeitpunkt des Downloads (12.05.2024) vor dem EGMR ausgehandelt worden. Jeder Schriftsatz beginnt mit einer Auflistung der beteiligten Richter, der Nennung des Antragstellers, sowie weiterer prozessbeteiligter Parteien. Als nächstes werden der prozessuale Rahmen und der Sachverhalt des Falles geschildert. Im letzten Teil legt das Gericht die Argumente der beteiligten Parteien dar um anschließend sein Urteil zu verkünden. Der Datensatz ist nicht materiell-rechtlich auf die Behandlung bestimmter Rechtsprobleme beschränkt worden. Urteile zur Meinungs- und Ausdrucksfreiheit finden sich darin genauso wie Urteile zum Diskriminierungsverbot etc.

## Web-Scraping

Ich habe das Web-Scraping-Tool von Quemy und Wrembel verwendet<sup>34</sup>. Das Tool besteht aus vielen einzelnen Algorithmen, die auf die Webseite des EGMR zugreifen, die Urteilsdateien herunterladen, sie für eine Datenanalyse aufbereiten und anschließend nach bestimmten Kriterien vorsortieren.

Die Funktion *get\_documents\_function*<sup>35</sup> ist das Herzstück des

---

<sup>32</sup> Siehe Glossar, Stichwort: Transformer

<sup>33</sup> <https://github.com/ByronMeissner/Arbitrariness-of-Proportionality>

<sup>34</sup> Quemy, A. and Wrembel, R., On Integrating and Classifying Legal Text Documents (2020).

<sup>35</sup> Siehe Anhang 1.

*get\_documents.py* Algorithmus, mit dem die Urteilsdateien heruntergeladen werden können. Zunächst werden darin einige Kategorisierungsparameter festgelegt und ein Speicherpfad für die heruntergeladenen Dokumente bestimmt. Anschließend werden .docx und .pdf Dokumente als Zielobjekte und die URL des EGMR als Zieladresse definiert. Schließlich wird eine logische Schleife durchlaufen, bis alle Dokumente entweder heruntergeladen oder als beschädigt aussortiert worden sind.

Die Algorithmen *normalize\_documents.py* und *process\_documents.py* schaffen eine vollständige Infrastruktur für die Aufbereitung der heruntergeladenen Textdateien. Der Inhalt der Texte wird zunächst tokenisiert<sup>36</sup>, normalisiert<sup>37</sup> und lemmatisiert<sup>38</sup>. Basierend auf den normalisierten Tokens werden dann N-Gramme<sup>39</sup> erstellt und deren Häufigkeiten berechnet. Die Dokumente werden anschließend in Bag-of-Words Repräsentationen<sup>40</sup> umgewandelt, die die Häufigkeit jedes Wortes im Dokument darstellen. Schließlich wird ein TF-IDF Format<sup>41</sup> erstellt, das die relative Wichtigkeit von Wörtern im Kontext des gesamten Dokumentenkörpus bewertet. Die Texte sind damit ideal für die Weiterverarbeitung durch ein Transformer Model aufbereitet.

## Text-Mining

Im nächsten Schritt habe ich den Text-Mining Algorithmus *Phrasengold.py* entwickelt, der die heruntergeladenen Gerichtsurteile als Input verwendet und die Urteilstexte mit RegEx-Pattern<sup>42</sup> nach relevanten Meta-Daten durchsucht. Der Algorithmus erstellt dann eine Tabelle, die für jedes Dokument die folgenden Meta-Daten erfasst:

Name der Tabellenspalte	Enthaltene Information	Datentyp <sup>43</sup>
<i>country</i>	Name des beteiligten Mitgliedstaats des Europarats	string

<sup>36</sup> Siehe Glossar, Stichwort: Tokenization.

<sup>37</sup> Siehe Glossar, Stichwort: Normalization.

<sup>38</sup> Siehe Glossar, Stichwort: Lemmatisation.

<sup>39</sup> Siehe Glossar, Stichwort: N-Gramme.

<sup>40</sup> Vgl. Glossar, Stichwort: Bag-of-Words.

<sup>41</sup> Siehe Glossar, Stichwort: TF-IDF.

<sup>42</sup> Siehe Glossar, Stichwort: RegEx.

<sup>43</sup> Siehe Glossar, Stichwort: Datentyp.

<i>applicant</i>	Name des beteiligten Antragstellers	string
<i>decisiondate</i>	Datum des Urteilsspruchs	string
<i>article</i>	Nummer der im Rahmen des Urteils (materiell-rechtlich) einschlägigen Artikel der EMRK	integer
<i>violation</i>	Hat das Gericht auf eine Verletzung des einschlägigen Artikels entschieden?	boolean
<i>seperateopinion</i>	War die Entscheidung unter den anwesenden Richtern einheitlich?	boolean
<i>voteratio</i>	Der Quotient aus Mehrheits- und Minderheitsmeinung der anwesenden Richter	float

Die Funktion *analyze\_voting*<sup>44</sup> enthält z.B. ein RegEx-Pattern, das alle geläufig vom Gerichtshof verwendeten Phrasen erkennt, die eine Dissens-Entscheidung zum Ausdruck bringen. Aus den durch das RegEx-Pattern gefilterten Phrasen extrahiert sie anschließend die Stimmzahlen, wandelt diese in integer um, bildet ihr mathematisches Verhältnis und gibt dann einen float aus, der später der Meta-Daten-Tabelle angehängt werden kann.

## Legal Argument Mining

Bereits 2007 verwendeten Moens et al. verschiedene statistische Methoden zur automatischen Erkennung von Argumenten in juristischen Texten. Mithilfe von N-Grammen und Statistiken zu Satzlänge und Wortfrequenz gelang es ihnen Klassifizierungen von Argumenttypen in juristischen Texten vorzunehmen.<sup>45</sup> Zhong et. al verwendeten 2019 ein CNN, um die Vorhersagekraft einzelner

<sup>44</sup> Siehe Anhang 2.

<sup>45</sup> Moens et. al, Automatic detection of arguments in legal texts (2007).

Sätze für das Ergebnis eines Gerichtsurteils zu bestimmen. Sie versteckten dafür während des Trainingsprozesses wiederholt Teilausschnitte des Urteils vor der KI, und ließen diese so Stück für Stück, die wichtigsten Sätze herausfiltern.<sup>46</sup> Seit 2019 haben Chalkidis et. al in einer Reihe von Projekten den Einfluss verschiedener Trainingsmethoden auf die Vorhersagegenauigkeit des BERT Transformer-Models untersucht. 2020 veröffentlichten sie das von ihnen trainierte Model Legal-BERT. 2020 verwendeten Poudyal et. al ebenfalls das BERT Transformer-Model für die Erkennung von Argumentsätzen, Satzrelationen sowie Prämissen und Schlussfolgerungen in 42 Urteilen des EGMR. Die Forscher konnten dabei Vorhersagegenauigkeiten von respektive 76.5%, 51,1% und 85,9%/62.8% verzeichnen.<sup>47</sup> 2023 trainierten Habernal et. al das den BERT Transformer unter Verwendung einer verbesserten Form des Labelings auf die Erkennung von legalen Argumenttypen und Akteuren.<sup>48</sup> Die Vorhersagegenauigkeit wurde in allen angeführten Studien ausschließlich auf kleine Datensätze getestet, die in ihrer Formatierung exakt den Trainingsdatensätzen für das jeweils verwendet Transformer-Model entsprachen.

### **Transformer-Model**

Ich habe das Model LLaMA-8Billion-Instruct für die Untersuchung ausgewählt, da es eines der fortschrittlichsten Sprachmodelle auf dem Markt ist und durch seine hohe Präzision und Effizienz in der Verarbeitung natürlicher Sprache überzeugt. LLaMA3 bietet eine verbesserte Textgenerierung und ein tiefes Verständnis komplexer Kontexte. Zudem zeichnet sich das Model durch seine Skalierbarkeit und Anpassungsfähigkeit an unterschiedliche Anwendungsbereiche aus. Diese Eigenschaften machen es besonders geeignet für die Analyse großer Datenmengen. Außerdem gibt es eine große Entwickler-Community rund um die LLaMA Sprachmodelle, sodass es oft einfacher ist als bei anderen Modellen für die Lösung von Problemen Beratung einzuholen. Ich habe LLaMA3 in Kombination mit LoRA (Low Rank Adaptation) verwendet, einer Methode zur effizienten Abstimmung großer Sprachmodelle, die 2021 von Hu et. al vorgeschlagen wurde.<sup>49</sup> Anstatt die Parameter des

---

<sup>46</sup> Zhong et. al, How Does NLP Benefit Legal System (2019).

<sup>47</sup> Poudyal et. al, ECHR: Legal Corpus for Argument Mining.

<sup>48</sup> Habernal et. al, Mining Legal Arguments in Court Decisions (2023).

<sup>49</sup> Hu et. al, LoRA: Low Rank of Adaption of Large Language Models (2021).

gesamten Modells zu aktualisieren, zerlegt LoRA die Parameter in Unterklassen und reduziert so die Anzahl der zu trainierenden Parameter. Dies geschieht durch die Zerlegung der Gewichtungsmatrix des Modells in zwei kleinere Matrizen. Dieser Ansatz hat mehrere Vorteile: Er reduziert den Speicherbedarf und die Rechenleistung und beschleunigt das Training und die Feinabstimmung des Modells erheblich. Außerdem können die Modelle schnell und flexibel an spezifische Aufgaben angepasst werden, ohne dass die gesamte Modellstruktur neu gelernt werden muss. Die Kombination von LLaMA3 und LoRA ermöglichte es mir sehr effizient gute Trainingsergebnisse zu erzielen.

### **Datasets**

Die Trainings- und Testdatensätze für die habe ich von Habernal et. al.<sup>50</sup> übernommen. Die hatten 2021 ein älteres Transformer Model darauf trainiert Argumenttypen in Gerichtsurteilen des EGMR zu erkennen. Die Forscher haben in ihrer Studie in Kollaboration mit Rechtswissenschaftlichen Experten ein Klassifizierungssystem mit 16 juristischen Argumenttypen entwickelt. Darunter findet sich auch die Klasse: „Verhältnismäßigkeitsprinzip – Angemessenheit“. Die Wissenschaftler haben den Text von insgesamt 347 Gerichtsurteilen des EGMR mit den Argumentationstypen ihres Klassifikationssystems annotiert. Sie beschäftigten dafür 6 Jurastudenten über einen Zeitraum von 12 Monaten. Die Studenten sollten unter der Supervision zwei rechtswissenschaftlichen Professoren 373 Urteile annotieren. Die Forscher haben die Datensätze in einem Github-Repository<sup>51</sup> frei zur Verfügung gestellt. Aus dem Datensatz der Wissenschaftler habe ich außerdem durch manuelles Lesen ein Kontrollset mit 195 Gerichtsurteilen erstellt. Davon waren 101 Fälle, in denen das Urteil unter Einbezug einer Verhältnismäßigkeitsabwägung gefällt wurde. Die übrigen 94 Fälle enthielten eine solche Abwägung nicht. Mit diesem Kontrollset konnte ich die Vorhersagegenauigkeit des LLaMA3-Modells überprüfen. Bemerkenswert ist, dass ich das Kontrollset anders als in den erwähnten Studien ausschließlich aus unformatierten .txt Dateien zusammengesetzt, habe weil ich eine Vorhersagegenauigkeit des Modells nicht nur auf den „künstlichen“ Datensätze, sondern auch auf die tatsächlichen Gerichtsurteile bewerten können wollte.

---

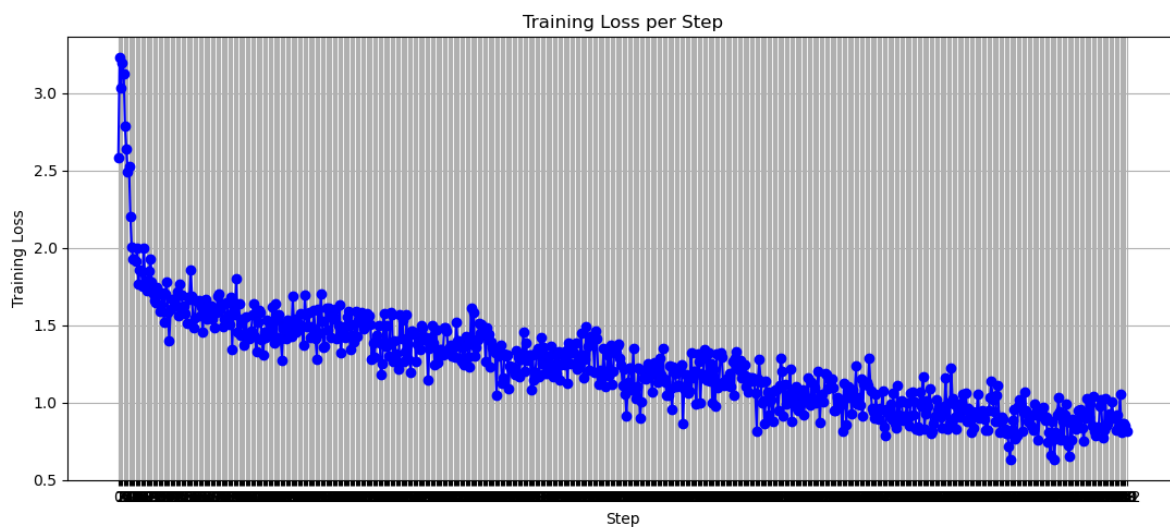
<sup>50</sup> Habernal et. al, Mining Legal Arguments in Court Decisions (2023).

<sup>51</sup> <https://github.com/trusthlt/mining-legal-arguments>, abgerufen: 12.05.2024.

## Training

Ich nutzte die Datensätze der Forscher, um das LLaMA3 Model darauf zu trainieren für ein beliebiges Gerichtsurteil zu erkennen, ob darin eine Verhältnismäßigkeitsabwägung *strictu sensu* vorgenommen wurde oder nicht. Das heißt das Model lernte zu erkennen, ob ein präsentierter Textinhalt ein solcher ist, in dem geschulte Juristen mindestens eine Textpassage als ein Argument der Klasse „Verhältnismäßigkeitsprinzip – Angemessenheit“ annotieren würden. Das *Llamadompteur.ipynb* Jupyter-Notebook enthält alle Arbeitsschritte, um das Training des Modells einzurichten. Ein großes Problem bei der Untersuchung war, dass sogar die enormen Rechenkapazitäten der verfügbaren NVIDIA A100-SXM4-40GB GPU nicht für eine direkte Verarbeitung des Trainingsdatensatzes ausgereicht haben. Die Urteile des EGMR sind teilweise so umfangreich, dass sie nach der Tokenization eine Länge von 200.000 Tokens übersteigen. Die Trainingsdaten mussten daher in umständlicher Weise aufgebrochen nach der Verarbeitung auf ihr Ursprungsdokument zurückbezogen werden. Der (negative) Einfluss auf die Präzision der Vorhersagen ist nur schwer einzuschätzen. Die Parameter, die ich für das finale Training gewählt habe, sind in *Trainer.py* definiert.<sup>52</sup> Die Trainings-Loss<sup>53</sup> Kurve für den finalen Trainingsprozess ist in Graphik 2 zu sehen.

Graphik 2



<sup>52</sup> Siehe Anhang 3.

<sup>53</sup> Siehe Glossar, Stichwort: Trainings-Loss.

## **Inference**

Die Skripte für die Anwendung des Models sind ebenfalls in *Llamadompteur.ipynb* enthalten. Obwohl die Trainings-Loss Kurve für das finale Training zunächst vielversprechend aussah, waren die tatsächlichen Resultate des Trainings ernüchternd. Nach dem Training erreichte das Model auf das Kontrollset eine Vorhersagegenauigkeit von 61.5% bei den Fällen in denen keine Verhältnismäßigkeitsabwägung strictu sensu durchgeführt wurde und 38.1% bei den Fällen, die eine solche Abwägung enthielten. Eine Anwendung auf den großen Datensatz der 15741 Gerichtsurteile erschien in Anbetracht der geringen Vorhersagegenauigkeit des Models obsolet. Auch eine statistische Extrapolation vom Kontrolldatenset und eine anschließende Bestimmung der Konfidenzintervalle für die Stimmverhältnisvarianzwerte hätte keine aussagekräftigen Ergebnisse geliefert.

## **Ergebnisse & Auswertung**

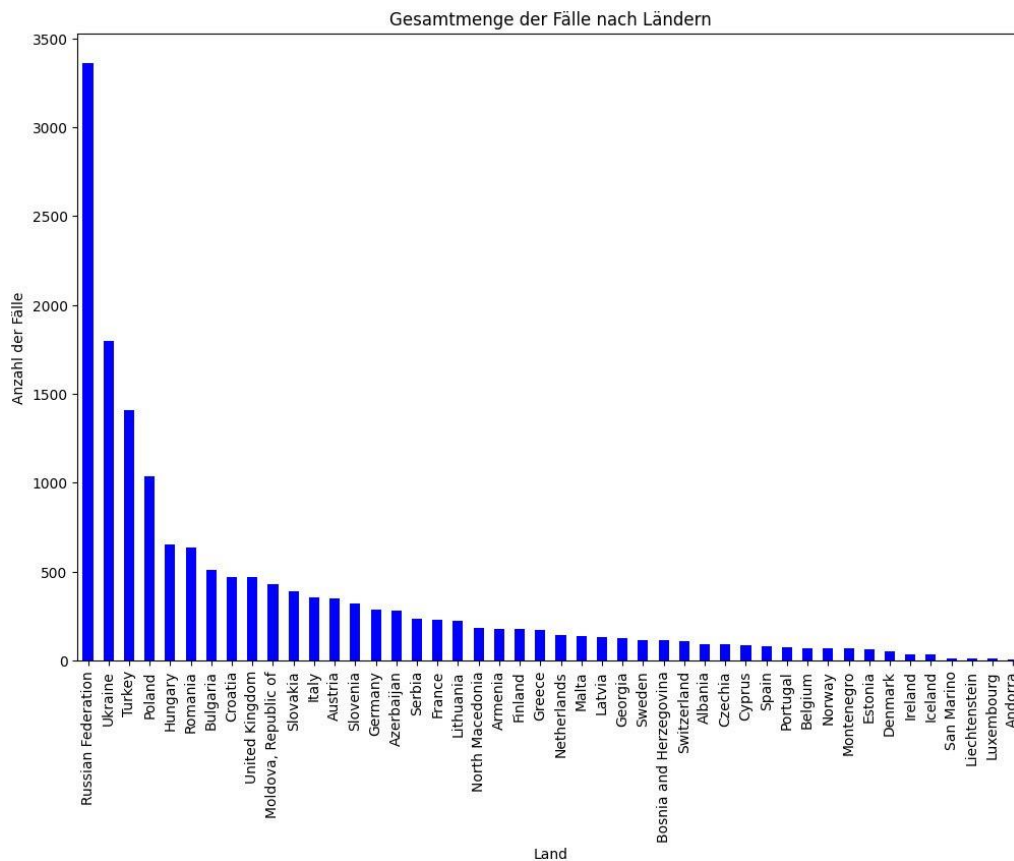
Ergebnisse aus anderen Studien ließen sich nicht reproduzieren. Mehr Zeit, eine bessere GPU und eine größere Expertise im Bereich der technischen Implementierung wären nötig gewesen, um ein aussagekräftigeres Forschungsergebnis zu erzielen. Eine Anwendung auf den großen Datensatz musste ausbleiben, weil sie zu keinerlei brauchbaren wissenschaftlichen Ergebnissen geführt hätte. Sogar, wenn die Vorhersagegenauigkeit des Models hinreichend groß gewesen wäre, um eine statistische Extrapolation zu ermöglichen, hätte man das Ergebnis kritisch reflektieren müssen.

Die Varianz der Stimmverhältnisse der Entscheidungsmengen vermag zwar Einsicht darüber zu liefern, wie sehr die Anwendung des Verhältnismäßigkeitsprinzips mit einer größeren Uneinigkeit unter den Richtern korreliert. Der Beweis dafür ob und inwieweit dies auf einen höheren Grad der Willkürlichkeit zurückgeführt werden kann, wäre jedoch auch bei einer höheren Vorhersagegenauigkeit keineswegs erbracht worden. Sogar, wenn eine Korrelation zwischen der Anwendung des Verhältnismäßigkeitsprinzips und der Varianz der Stimmverhältnisse beobachtet worden wäre, hätte man mögliche Konfundierungsfaktoren identifizieren und diese, soweit möglich isolieren müssen. Für eine größere Varianz der Stimmverhältnisse könnte neben einem höheren Maß an

Willkürlichkeit bspw. auch die Komplexität einer Entscheidung ursächlich sein. Eine größere Uneinigkeit könnte auch ein Indiz dafür sein, dass die Entscheidungen eine größere Menge an rechtlichen Konflikten und Teilproblemen enthalten. Auch der historische Wandel der Jurisprudenz wäre in seinem Einfluss nicht zu unterschätzen. Im Jahr 1960, zum Zeitpunkt, an dem der Datensatz beginnt, war die Entwicklung des Verhältnismäßigkeitsprinzips nur zwei Jahre vergangen und seine Anwendung wurde daher von den Jurisdiktionen vieler Mitgliedsstaaten des Europarats noch gar nicht geteilt. Außerdem hat sich der Europarat seitdem laufend verändert. Während er 1960 noch aus nur 15 Mitgliedern zusammengesetzt war, umfasst er gegenwärtig bereits 46 Staaten. Jedes Mitglied des Rats ist entsprechend den Regelungen aus Art. 20ff. EMRK durch einen eigenen Richter im EGMR vertreten. Dazugekommen sind Länder wie Armenien, Bosnien & Herzegowina, Spanien, Rumänien und die Ukraine. Auch Russland war zwischen 1996 und 2022 ein Mitglied des Europarats. Der Einfluss solcher Veränderungen ist nicht nur beachtlich, weil jeder neue Richter den speziellen Wertekatalog seines eigenen Landes in die Rechtsprechung des EGMR einbringt. Sofern man eine völlige Unbefangenheit der hohen Richter ausschließt, könnten geopolitische Konflikte, in denen einige Mitgliedsstaaten deutlich andere Positionen beziehen, als andere sich durchaus auf die Uneinigkeit auswirken, die zwischen den entscheidenden Richtern besteht. Ein größeres Maß an Dissens könnte in Anbetracht dieser Überlegungen auch die natürliche Folge eines pluralistischeren Wertesystems sein, das zwischen seinen weniger pluralistischen Einzelteilen ständig aufs Neue ausgehandelt wird. Ein Indiz dafür, dass ein solcher Faktor in seiner konfundierenden Wirkung nicht zu unterschätzen ist, liefert Graphik 3 in der zu sehen ist in wie vielen Fällen die Mitgliedsstaaten jeweils verwickelt gewesen sind.



**Graphik 3**



Der Umstand, dass die Verbreitung des Verhältnismäßigkeitsprinzips zeitlich mit dem Zuwachs der Mitgliederzahl einhergeht, verkompliziert eine akkurate Analyse der Ursachen zusätzlich. Es reicht nicht den Faktor Zeit zu isolieren, um den Einfluss des Verhältnismäßigkeitsprinzips auf die Willkürlichkeit einer Entscheidungsmenge zu bestimmen, weil eine denkbare zeitliche Veränderung der Varianz der Stimmverhältnisse ihre Ursache sowohl in dem wachsenden Einfluss des Verhältnismäßigkeitsprinzips als auch in der strukturellen Veränderung des Europarats als auch in der Wirkung weiterer geopolitischer Ereignisse haben könnte.

## Ausblick

Im Fokus zukünftiger Untersuchungen muss die genauere Betrachtung einzelner Einflussfaktoren liegen. Eine statistische Isolation möglicher Störfaktoren, würde eine bessere Bestimmung derjenigen Wirkung ermöglichen, welche die Anwendung des Verhältnismäßigkeitsprinzips tatsächlich hat. Durch eine gesonderte Betrachtung einzelner Zeitabschnitte, Kammerzusammensetzungen und Rechtsproblematiken könnten relevante

Störfaktoren identifiziert werden. Die Varianz der Stimmverhältnisse in Abhängigkeit von einschlägigen geopolitischen Ereignissen zu betrachten könnte einen weiteren Störfaktor liefern. Es ist auch denkbar einen Indikator für die Komplexität einzelner Rechtsurteile zu entwickeln, den man z.B. aus der Zahl der darin behandelten rechtlichen Probleme, der Länge des Urteilsspruchs und den Ergebnissen einer möglichen Sentimentwertanalyse zusammensetzen könnte. Wäre ein solcher Indikator für die Komplexität eines Urteils einmal konzipiert, könnte man auch dessen Einfluss auf die Varianz der Stimmverhältnisse separat betrachten.

In der Zukunft könnte außerdem ein Vergleich mit den Text-Corpora anderer Jurisdiktionen Aufschluss über den Einfluss des Verhältnismäßigkeitsprinzip auf die Willkürlichkeit der Rechtsprechung geben. Insbesondere die Rechtsprechung des US-Amerikanischen Supreme Courts, sowie die des Chinesischen *Zui-Gao-Ren-Min-Fa-Yuan* erscheinen aufschlussreiche Untersuchungsobjekte zu sein. Erstere, weil sie zumindest nach dem Selbstverständnis des Supreme Courts auf die Anwendung des Verhältnismäßigkeitsprinzips verzichtet und sich stattdessen des Rational-Basis-Tests bedient.<sup>54</sup> Letztere, weil sie das Verhältnismäßigkeitsprinzip zwar anwendet, dies aber im Kontext eines fundamental anderen Werte- und Staatsorganisationssystems tut. Zumindest für die USA ist eine noch größere Menge an Gerichtsurteilen im Rahmen des Harvard Case Law Projects, bereits öffentlich zugänglich gemacht worden, sodass reichlich Material für aufschlussreiche Datenanalysen verfügbar ist.<sup>55</sup>

---

<sup>54</sup> Barak, Proportionality: Constitutional Rights and Their Limitations (2012), S. 284.

<sup>55</sup> <https://case.law/>, abgerufen am 03.06.2024.

# Glossar

## **Natural Language Processing (NLP)<sup>56</sup>**

Natural Language Processing ist ein Bereich der künstlichen Intelligenz, der sich mit der Interaktion zwischen Computern und Menschen durch natürliche Sprache befasst. Es geht um die Entwicklung von Algorithmen und Modellen, die es Computern ermöglichen, menschliche Sprache zu verstehen, zu interpretieren, zu generieren und auf nützliche Weise zu verarbeiten.

## **Tokenization**

Tokenization ist der Prozess, einen Text in kleinere Einheiten wie Wörter, Phrasen oder Zeichenfolgen herunterzubrechen (Tokens). Zum Beispiel könnte der Satz „Der Mann geht Samstag nach Hause.“ in „Der“, „Mann“, „geht“, „Samstag“, „nach“, „Hause.“ zerlegt werden. Tokenization ist wichtig für die Verarbeitung natürlicher Sprache (NLP), da sie den Text für weitere Analysen vorbereitet.

## **Normalization**

Normalization ist der Prozess, einen Text in ein einheitliches Standardformat umzuwandeln. Je nach der konkreten Implementierung umfasst die Normalization verschiedene Schritte wie z.B. Stemming, Lemmitization, die Umwandlung in Kleinbuchstaben, sowie die Entfernung von Satzzeichen und Füllwörtern.

## **Lemmitization/Stemming**

Durch Lemmitization werden die Wörter eines Textes auf ihre gemeinsamen Wortstämme reduziert, die sich zumindest augenscheinlich voneinander unterscheiden (aus „Bäume“ wird „Baum“). Stemming ist eine einfachere Form der Lemmitization, die durch das Entfernen von Suffixen funktioniert (aus „gehen“ wird „geh“).

## **N-Gramme**

Ein n-Gramm ist eine Folge von n Wörtern: ein 2-Gramm (Bigramm) ist eine Zwei-Wort-Folge wie z.B. „Papageien essen“ oder „selten dort“ etc. Ein 3-

---

<sup>56</sup> Sofern nicht anders gekennzeichnet, sind die Glossareinträge sinngemäß Jurafski, Daniel, Martin, James H., Speech and Language Processing (2017) entnommen.

Gramm (Trigramm) ist eine Drei-Wort-Folge wie z.B. „keiner weiß davon“ usw.

### **Bag-of-Words**

Die Bag-of-Words (BoW)-Darstellung ist eine Repräsentation von Texten, bei der die Wortreihenfolge, Grammatik und Syntax ignoriert werden. Stattdessen wird der Text als eine Sammlung von Wörtern und ihrer jeweiligen Häufigkeiten dargestellt. Dies lässt sich mathematisch wie folgt formalisieren:

Sei  $V$  die Menge der unterschiedlichen Wörter im Korpus:

$$V = \{w_1, w_2, \dots, w_N\}$$

und  $N$  die Kardinalität der Menge  $V$  (also die absolute Anzahl unterschiedlicher Wörter):

$$N = |V|$$

Dann kann ein Dokument  $j$  durch einen Vektor dargestellt werden,

$$x = (x_1, x_2, \dots, x_N)$$

wobei  $x_i$  die Häufigkeit des Wortes  $w_i$  im Dokument  $d$  angibt. Ein solcher Vektor ist die BoW-Darstellung eines Textdokuments.

### **Termfrequenz**

Die Termfrequenz  $TF_{i,j}$  eines Wortes  $w_i$  in einem Dokument  $d_j$  ist das Verhältnis der Häufigkeit des Wortes  $w_i$  zur Gesamtanzahl der Wörter im Dokument  $d_j$ :

$$TF_{i,j} = \frac{f_{i,j}}{\sum_k f_{k,j}}$$

wobei  $f_{i,j}$  die Häufigkeit des Wortes  $i$  im Dokument  $j$  und  $\sum_k f_{k,j}$  die Gesamtanzahl der Wörter im Dokument  $j$  ausdrücken.

### **Inverse Dokumentfrequenz**

Die inverse Dokumentenfrequenz  $IDF_i$  misst die Relevanz eines Wortes  $w_i$  im

gesamten Corpus. Sie ist umgekehrt proportional zur Anzahl der Dokumente, in denen das Wort  $w_i$  vorkommt:

$$IDF_i = \log \left( \frac{M}{|\{d_j : w_i \in d_j\}|} \right)$$

wobei  $M$  die Gesamtanzahl der Dokumente im Korpus und  $|\{d_j : w_i \in d_j\}|$  die Anzahl der Dokumente bezeichnet, die das Wort  $w_i$  enthalten.

### **TF-IDF**

Das TF-IDF-Maß eines Wortes  $w_i$  in einem Dokument  $d_j$  ist das Produkt aus seiner Termfrequenz und seiner inversen Dokumentfrequenz:

$$TF-IDF_{i,j} = TF_{i,j} \times IDF_i$$

Das TF-IDF-Maß gibt die Relevanz eines Wortes im Kontext eines konkreten Dokuments an, indem es die Häufigkeit seiner Verwendung in diesem Dokument mit der Häufigkeit seiner Verwendung im gesamten Korpus in Beziehung setzt. Dies berücksichtigt, dass einige Wörter generell häufiger verwendet werden als andere.

### **Datentyp**

In der Computerwissenschaft werden im Wesentlichen drei Datentypen unterschieden. Ein *string* ist eine Zeichenfolge (z.B. „Katze“ oder „slkejd“). Ein *integer* ist eine Zahl (z.B. „1“ oder „13“). Ein *float* ist eine Gleitkommazahl (z.B. „0,3457“ oder „0,4“). Ein *boolean* ist ein Datentyp, der nur zwei mögliche Werte annehmen kann (z.B. „wahr“ / „falsch“ oder „an“ / „aus“).

### **RegEx**

RegEx, kurz für "Regular Expressions", ist eine formale Sprache, die verwendet wird, um Muster für das Suchen und Bearbeiten von Text zu definieren. RegEx-Pattern heißen die Muster, die verwendet werden, um Textzeichenfolgen zu durchsuchen und zu manipulieren.

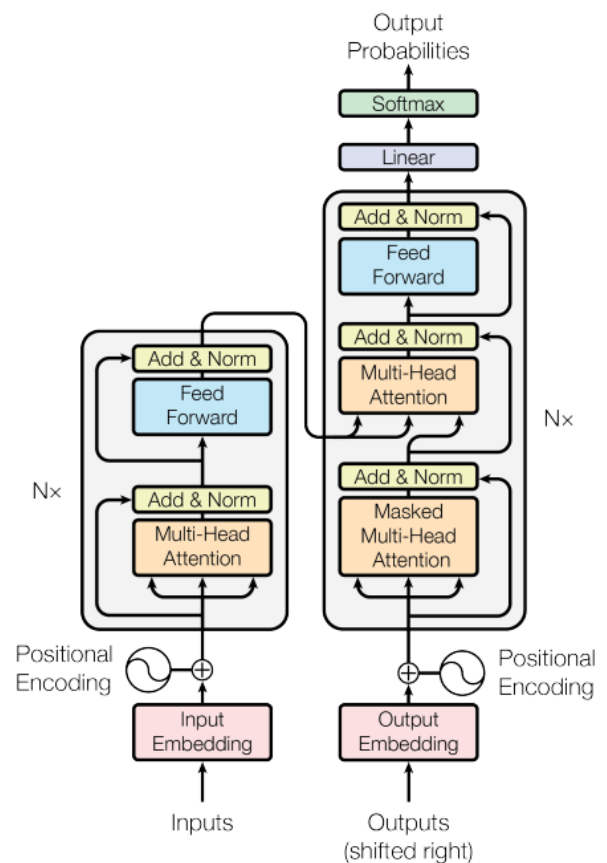
### **CNN**

Convolutional Neural Networks (CNNs) sind eine spezielle Art von neuronalen

Netzwerken, die besonders gut für die Verarbeitung von Bilddaten geeignet sind. Sie basieren auf dem Prinzip der Faltung (Convolution), das dazu dient, wichtige Merkmale in den Daten zu extrahieren. Die Faltungsschicht ist das Herzstück eines CNN. Ein Faltungsfilter oder Kern (Kernel) wird über das Eingabebild geschoben, und an jeder Position wird eine Faltung (Convolution) durchgeführt.

### Transformer Model<sup>57</sup>

Ein Transformer ist ein neuronales Netzwerk, das aus sogenannten Attention- und Feed-Forward-Schichten besteht. Erstere nutzen die Position von Wörtern in hochdimensionalen Vektorräumen (Embeddings), um die Bedeutung der Wörter in Abhängigkeit von anderen Wörtern im Satz sowie ihrer Position im Kontext zu bestimmen. Der Attention-Mechanismus berechnet dabei für jedes Wort Gewichtungen, die angeben, wie stark es auf andere Wörter im Satz "achtet". Durch den Einsatz von Multi-Head-Attention können verschiedene Aspekte der Beziehungen zwischen Wörtern parallel betrachtet werden. Die Feed-Forward-Schichten verarbeiten dann die durch die Attention-Schichten ermittelten Informationen weiter. Transformer-Modelle bestehen aus einer Encoder-Decoder-Architektur, wobei der Encoder die Eingabesequenz verarbeitet und eine kontextuelle Repräsentation erzeugt, während der Decoder diese Repräsentation nutzt, um die Ausgabe zu generieren. Bei Aufgaben wie maschineller Übersetzung wird die Eingabesequenz vom Encoder verarbeitet und der Decoder erzeugt schrittweise die übersetzte Sequenz. Diese Schichten werden mehrmals wiederholt, um die endgültige Transformer-Architektur zu bilden, die die Eingabe sequenziell verarbeitet und die relevanten Wortbeziehungen und Bedeutungen lernt.



<sup>57</sup> Vgl. Vaswani et. al, Attention is all you need (2017).

## **Trainings-Loss**

Der Begriff „Lernverlust“ bezieht sich auf eine Metrik, die zur Bewertung der Leistung eines Modells während des Trainings eines maschinellen Lernmodells verwendet wird. Der Lernverlust quantifiziert, wie gut ein Modell die Trainingsdaten vorhersagt, indem die Differenz (Fehler) zwischen den vorhergesagten und den tatsächlichen Werten des Modells gemessen wird. Der Lernverlust wird mithilfe von Verlustfunktionen berechnet, die je nach Art des zu lösenden Problems variieren. Typische Verlustfunktionen sind zum Beispiel der Mean Squared Error (MSE) für Regressionsprobleme und der Cross-Entropy für Klassifizierungsprobleme. Die Werte der Verlustfunktion werden über alle Trainingsinstanzen aggregiert, um einen einzigen Verlustwert zu erzeugen, der den aktuellen Zustand des Modells darstellt. Ein kleiner Trainingsverlust bedeutet, dass das Modell die Trainingsdaten gut approximiert, während ein großer Trainingsverlust bedeutet, dass das Modell Schwierigkeiten hat, die Trainingsdaten zu lernen. Sehr kleine Trainingsverluste können jedoch auch auf Overfitting hinweisen, d. h. auf die Möglichkeit, dass das Modell die Trainingsdaten zu gut gelernt hat und nicht gut auf neue, unbekannte Daten generalisieren kann.

# Anhang

## Anhang 1

```
def get_documents_step(doc_id, progress, task):
    if doc_id[1]:
        filename = "%s.docx" % (doc_id[0].strip())
    else:
        filename = "%s.pdf" % (doc_id[0].strip())

    filename = os.path.join(folder, filename)
    if not os.path.isfile(filename) or force:
        if doc_id[1]:
            url = BASE_URL +
"docx/?library=ECHR&filename=please_give_me_the_document.docx&id=" +
doc_id[0].strip()
        else:
            url = BASE_URL +
"docx/pdf?library=ECHR&filename=please_give_me_the_document.pdf&id=" +
doc_id[0].strip()

        for j in range(MAX_RETRY):
            try:
                r = requests.get(url, stream=True, timeout=5)
                if not r.ok:
                    raise Exception()
                with open(filename, 'wb') as f:
                    for block in r.iter_content(1024):
                        f.write(block)
                error = "\n| Request complete, see [cyan]%s" % (filename)
                break
            except Exception as e:
                try:
                    # Delete the incorrect file if it exists
                    os.remove(filename)
                except OSError:
                    pass
                log.debug(e)
                error = '\n| ({} / {}) Failed to fetch document {}'.format(
                    j + 1, MAX_RETRY, doc_id[0])
                error += '\n| URL: %s' % (url)
                error += '\n| Permalink: %s' % (PERMA_URL + doc_id[0].strip())
                progress.update(task, advance=0, error=error, doc=doc_id[0])
            else:
                error = "\n| Skip as document exists already"

        progress.update(task, advance=1, error=error, doc=doc_id[0])

if id_list:
    with Progress(
        TAB + "> Downloading... [IN PROGRESS]\n",
        BarColumn(30),
        TimeRemainingColumn(),
        "| Fetching document of case [blue]{task.fields[doc]}"
[white]({task.completed}/{task.total})"
        "{task.fields[error]}",
        transient=True,
        console=console
    ) as progress:
        task = progress.add_task("Downloading...", total=len(id_list),
error="", doc=id_list[0][0])
        f = lambda x: get_documents_step(x, progress, task)
        with ThreadPoolExecutor(16) as executor:
            executor.map(f, id_list)
```



## Anhang 2

```
content_sections = data.get("content", {})
majority_patterns = [
    r'(Holds|Rejects|Finds|Declares|Decides) by (\w+) votes? to (\w+) that
    (there has been a violation of Article|there has been no violation of
    Article)',
    r'by (\w+) votes? to (\w+),? that (there has been a violation of
    Article|there has been no violation of Article)',
    r'Holds by (\w+) votes? to (\w+) that (there has been a violation of
    Article|there has been no violation of Article)',
    r'(Holds|Rejects|Finds|Declares|Decides) by (\w+) votes? to (\w+)
    (that|there has been|there has been no)',
    r'by (\w+) votes? to (\w+),? (that|there has been|there has been no)',
    r'Holds by (\w+) votes? to (\w+) (that|there has been|there has been no)'
]

content_text = json.dumps(content_sections)
dissenting = False
vote_ratio = 0
matched_sentences = []

for pattern in majority_patterns:
    matches = re.findall(pattern, content_text, re.IGNORECASE)
    for match in matches:
        if len(match) >= 3:
            matched_sentences.append(match)

if matched_sentences:
    dissenting = True
    # Using the first match for demonstration; this can be adjusted as needed
    first_match = matched_sentences[0]
    major_votes_str = first_match[1].lower()
    minor_votes_str = first_match[2].lower()

    num_dict = {
        'one': 1, 'two': 2, 'three': 3, 'four': 4, 'five': 5, 'six': 6,
        'seven': 7, 'eight': 8, 'nine': 9,
        'ten': 10, 'eleven': 11, 'twelve': 12, 'thirteen': 13, 'fourteen': 14,
        'fifteen': 15, 'sixteen': 16,
        'seventeen': 17
    }

    major_votes = num_dict.get(major_votes_str)
    minor_votes = num_dict.get(minor_votes_str)

    if major_votes is not None and minor_votes is not None:
        vote_ratio = minor_votes / (major_votes + minor_votes)

info['dissenting'] = dissenting
info['vote_ratio'] = vote_ratio

return info
```

### Anhang 3

```
from trl import SFTTrainer
from transformers import TrainingArguments
from unsloth import is_bfloat16_supported

trainer = SFTTrainer(
    model=model,
    tokenizer=tokenizer,
    train_dataset=finetuning_dataset,
    dataset_text_field="text",
    max_seq_length=max_seq_length, # Reduce for memory allocation
    dataset_num_proc=2, # 2 or 1
    packing=False, # Can make training 5x faster for short sequences
    args=TrainingArguments(
        per_device_train_batch_size=2, # 2
        gradient_accumulation_steps=4,
        warmup_steps=5,
        num_train_epochs=8, # 6 or 8
        # max_steps=60,
        learning_rate=2e-4,
        fp16=not torch.cuda.is_bf16_supported(),
        bf16=torch.cuda.is_bf16_supported(),
        logging_steps=1,
        optim="adamw_8bit",
        weight_decay=0.01,
        lr_scheduler_type="linear",
        seed=3407,
        output_dir="outputs",
        # gradient_checkpointing=True, # Enable if supported by the model
    ),
)
```

# Literaturverzeichnis

- Alexy, Robert. „Die Gewichtsformel“. In *Gedächtnisschrift für Jürgen Sonnenschein*, herausgegeben von Joachim Jickeli, Peter Kreutz, und Dieter Reuter, 771–92. Berlin: De Gruyter, 2003.
- . *Theorie der Grundrechte*. 11. Suhrkamp, 2011.
- Barak, Aharon. *Proportionality: Constitutional Rights and Their Limitations*. Cambridge Studies in Constitutional Law. Cambridge: Cambridge university press, 2012.
- Dąbrowska-Kłosińska, Patrycja. „The Protection of Human Rights in Pandemics—Reflections on the Past, Present, and Future“. *German Law Journal* 22, Nr. 6 (September 2021): 1028–38. <https://doi.org/10.1017/glj.2021.59>.
- Eichner, Daniel, Christian Fritsch, Gina-Luisa Kothe, Sina Kühner, Yannik Remond, und Jonas Warnke. „Risikowahrnehmung durch und nach Corona: Haben sich Verständnis und Wahrnehmung von Risiken verändert?“. In *Risiko im Wandel*, herausgegeben von Rolf Arnold, Marcel Berg, Oskar Goecke, Maria Heep-Altiner, und Horst Müller-Peters, 35–61. Wiesbaden: Springer Fachmedien Wiesbaden, 2023. [https://doi.org/10.1007/978-3-658-37071-8\\_3](https://doi.org/10.1007/978-3-658-37071-8_3).
- Habernal, Ivan, Daniel Faber, Nicola Recchia, Sebastian Bretthauer, Iryna Gurevych, Indra Spiecker genannt Döhmann, und Christoph Burchard. „Mining Legal Arguments in Court Decisions“. *Artificial Intelligence and Law*, 23. Juni 2023. <https://doi.org/10.1007/s10506-023-09361-y>.
- Hu, Edward J., Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, und Weizhu Chen. „LoRA: Low-Rank Adaptation of Large Language Models“. arXiv, 16. Oktober 2021. <https://doi.org/10.48550/arXiv.2106.09685>.
- Hufen, Friedhelm. *Staatsrecht II - Grundrechte*. 8. München: C.H.Beck, 2020.
- Jurafsky, Daniel, und James H. Martin. *Speech and Language Processing - An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*, 2017.
- Kahnemann, Daniel, und Amos Tversky. „Judgment under Uncertainty: Heuristics and Biases“. Zugegriffen 31. Mai 2024. <https://doi.org/10.1126/science.185.4157.1124>.
- Kluge, Friedrich. „Etymologisches Wörterbuch der deutschen Sprache“, 24:990. Berlin/NewYork: Walter de Gruyter, 2001.
- Kühl, Kristian, Hermann Reichold, und Michael Ronellenfitsch. *Einführung in die Rechtswissenschaft*. 3. München: C.H.Beck, 2019.

- Moens, Marie-Francine, Erik Boiy, Raquel Mochales, und Chris Reed. „Automatic detection of arguments in legal texts“, 225–30, 2007.  
<https://doi.org/10.1145/1276318.1276362>.
- Ondřejek, Pavel, und Filip Horák. „Proportionality during Times of Crisis: Precautionary Application of Proportionality Analysis in the Judicial Review of Emergency Measures“. *European Constitutional Law Review* 20, Nr. 1 (März 2024): 27–51. <https://doi.org/10.1017/S1574019624000051>.
- Petersen, Niels. *Verhältnismäßigkeit als Rationalitätskontrolle*. 1. Aufl. Bd. 238. Jus Publicum. Tübingen: Mohr Siebeck, 2015.
- Peterson, Niels. „Proportionality and Judicial Activism: Fundamental Rights Adjudication in Canada, Germany and South Africa“. *International Journal of Constitutional Law* 16, Nr. 3 (9. November 2018): 1035–41.  
<https://doi.org/10.1093/icon/moy081>.
- Poudyal, Prakash, Jaromir Savelka, Aagje Ieven, Marie Francine Moens, Teresa Goncalves, und Paulo Quaresma. „ECHR: Legal Corpus for Argument Mining“. In *Proceedings of the 7th Workshop on Argument Mining*, herausgegeben von Elena Cabrio und Serena Villata, 67–75. Online: Association for Computational Linguistics, 2020.  
<https://aclanthology.org/2020.argmining-1.8>.
- Quemy, Alexandre. „European Court of Human Right Open Data Project“. arXiv, 4. Februar 2019. <http://arxiv.org/abs/1810.03115>.
- Quemy, Alexandre, und Robert Wrembel. „On Integrating and Classifying Legal Text Documents“. In *Database and Expert Systems Applications*, herausgegeben von Sven Hartmann, Josef Küng, und Gabriele Kotsis, 385–99. Bratislava, 2020.
- Saussure, Ferdinand de. *Cours de linguistique générale*. Paris: Hachette, 1979.
- Schlink, Bernhard. *Abwägung im Verfassungsrecht*. Bd. 299. Schriften zum Öffentlichen Recht. Berlin: Duncker & Humblot, 1976.
- Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, und Illia Polosukhin. „Attention Is All You Need“. arXiv, 1. August 2023. <https://doi.org/10.48550/arXiv.1706.03762>.
- Zhong, Haoxi, Chaojun Xiao, Cunchao Tu, Tianyang Zhang, Zhiyuan Liu, und Maosong Sun. „How Does NLP Benefit Legal System: A Summary of Legal Artificial Intelligence“. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5218–30. Online: Association for Computational Linguistics, 2020. <https://doi.org/10.18653/v1/2020.acl-main.466>.