

Die Umsetzung der Europäischen Charta der Regional- oder Minderheitensprachen mithilfe von künstlicher Intelligenz

Europarat,
Sekretariat der Europäischen Charta
der Regional- oder Minderheitensprachen

COUNCIL OF EUROPE



CONSEIL DE L'EUROPE

Veröffentlichung des Europarats

MIN-LANG(2022)4

Autorin: Miriam Gerken

Konzept und Redaktion: Sekretariat der Europäischen Charta der Regional- oder
Minderheitensprachen

Titelbild: Shutterstock

Februar 2022

Inhalt

Einleitung	2
1. Allgemeine Gründe für den Gebrauch von künstlicher Intelligenz zur Erleichterung der Charta-Umsetzung.....	3
2. Maschinelle Übersetzung	5
2.1 Verschiedene Methoden für maschinelle Übersetzung	5
2.2 Einsatz von vorhandenen maschinellen Übersetzungssystemen	7
2.3 Entwicklung neuer maschineller Übersetzungssysteme.....	8
3. Weitere Anwendungen der natürlichen Sprachverarbeitung und ihr Gebrauch für die Charta-Umsetzung.....	8
3.1 Regional- oder Minderheitensprachen im Privatleben (Artikel 7.1.d)	8
3.2 Regional- oder Minderheitensprachen im Bildungswesen (Artikel 7.1.g, 8.1.f.ii, iii)	9
3.3 Regional- oder Minderheitensprachen in Justizbehörden (Artikel 9.1.a.i-iv, 9.1.b.i-iii, 9.1.c.i-iii, 9.1.d, 9.3)	12
3.4 Regional- oder Minderheitensprachen in Verwaltungsbehörden und öffentlichen Dienstleistungsbetrieben	12
3.4.1 Chatbots (Artikel 10.1.a.i-iv, 10.2.a, 10.2.b, 10.3.a-c)	12
3.4.2 Intelligente Suchmethoden (Artikel 10.2.g).....	13
3.4.3 Sprachsynthese für die Ansage von Straßennamen (Artikel 10.2.g)	14
3.4.4 Maschinelle Übersetzung (Artikel 10.1.a.i-v, 10.1.b, 10.1.c, 10.2.a-f, 10.3.a-c, 10.4.a)	15
3.5 Regional- oder Minderheitensprachen in den Medien	15
3.5.1 Automatische Generierung von Untertiteln (Artikel 11.1.a.i-iii, 11.1.c.i-ii).....	15
3.5.2 Automatische Extraktion von Informationen (Artikel 11.1.e.i-ii)	16
3.6 Regional- oder Minderheitensprachen in kulturellen Tätigkeiten und Einrichtungen	17
3.6.1 Datenstrukturierung (Artikel 12.1.g, 12.1.h)	17
3.6.2 Maschinelle Übersetzung (Artikel 12.1.a, 12.1.b, 12.1.c)	17
3.6.3 Automatische Generierung von Untertiteln (Artikel 12.1.b, 12.1.c)	18
3.7 Regional- oder Minderheitensprachen im wirtschaftlichen und sozialen Leben	18
3.7.1 Sentimentanalyse (Artikel 13.1.c, 13.1.d, 13.2.b).....	18
3.7.2 Maschinelle Übersetzung (Artikel 13.1.a, 13.1.d, 13.2.a, 13.2.b, 13.2.d, 13.2.e).....	19
3.8 Regional- oder Minderheitensprachen im grenzüberschreitenden Austausch (Artikel 7.1.i, 14).....	19
Fazit	20

Danksagung

Dieser Bericht wurde von Miriam Gerken in Zusammenarbeit mit dem Sekretariat der Europäischen Charta der Regional- oder Minderheitensprachen während ihres Studienaufenthalts beim Europarat im Februar 2020 recherchiert und verfasst und im März 2022 aktualisiert. Miriam Gerken hat 2018 ihr Studium der Übersetzungswissenschaft an der Universität Hildesheim und 2021 ihr Studium der Computerlinguistik an der Universität Bielefeld abgeschlossen, in beiden Fällen mit einem Schwerpunkt auf maschineller Übersetzung. Sie arbeitet zurzeit als Expertin für Sprachtechnologien bei der Deutschen Bahn. Das Charta-Sekretariat bedankt sich bei Miriam Gerken, die dank ihrer fundierten Kenntnisse im Bereich der Sprachverarbeitung mithilfe von künstlicher Intelligenz (KI) einen umfassenden und praktischen Leitfaden für politische Entscheidungsträger und Akteure, die an der Umsetzung der Charta und der Förderung des täglichen Gebrauchs von Regional- oder Minderheitensprachen im öffentlichen und privaten Leben arbeiten, erstellt hat.

Einleitung

In den meisten europäischen Staaten werden Regional- oder Minderheitensprachen seit jeher in Teilen des Staatsgebiets von einer Minderheit der Bevölkerung verwendet. Die demografische und rechtliche Lage der Sprachen ist sehr unterschiedlich. Die meisten haben jedoch gemeinsam, dass sie zu einem bestimmten Grad gefährdet sind.

Der internationale rechtliche Bezugsrahmen für die Förderung dieser Sprachen ist die Europäische Charta der Regional- oder Minderheitensprachen des Europarats. Die Charta fördert den Gebrauch von Regional- oder Minderheitensprachen in verschiedenen Bereichen des öffentlichen Lebens: Bildung, Justizbehörden, Verwaltungsbehörden und öffentliche Dienstleistungsbetriebe, Medien, kulturelle Tätigkeiten und Einrichtungen, wirtschaftliches und soziales Leben sowie grenzüberschreitender Austausch. Die Charta wurde von 25 Staaten ratifiziert;¹ weitere neun Staaten haben sie unterzeichnet, aber noch nicht ratifiziert.²

Die Erarbeitung der Charta begann 1984. Seit ihrer Annahme durch das Ministerkomitee des Europarats 1992 haben verschiedene neue Technologien die Bedingungen für die Umsetzung durch die Vertragsstaaten verändert. So hat beispielsweise das Internet zu einer Erweiterung des Medienangebots beigetragen und damit Anträge auf Rundfunkfrequenzen oder bessere Sendezeiten für Sendungen in Regional- oder Minderheitensprachen in gewissem Maße unnötig gemacht.

Aufgrund des zunehmenden Fortschritts von KI-Technologien ergeben sich für die Vertragsstaaten neue Möglichkeiten, aber auch neue Verpflichtungen hinsichtlich der Charta-Umsetzung. Künstliche Intelligenz - ein Begriff, der früher nur Science-Fiction war, ist heute im Alltag vieler Menschen angekommen: sei es durch Benachrichtigungen, das Haus früher zu verlassen, wenn viel Verkehr auf dem Weg zum Arbeitsplatz herrscht, durch automatische Empfehlungen von neuen Filmen oder die Möglichkeit, Texte in unbekannt Sprachen zu lesen. Aber was genau ist künstliche Intelligenz? Und wie kann diese Technik Lösungen für ein spezifisches Problem, die Unterstützung von Regional- und Minderheitensprachen, bieten?

KI beschreibt im Allgemeinen intelligente Maschinen, die in der Lage sind, ihr Umfeld zu analysieren und auf Grundlage dieser Analysen Entscheidungen zu treffen. Dabei imitieren KI-Systeme menschliche Fähigkeiten wie Lernen oder Problemlösung durch statistische Datenverarbeitung in neuronalen Netzen. Die Verwendung von KI zur Verarbeitung von sprachlichen Daten wird natürliche Sprachverarbeitung (NSV bzw. engl. NLP/natural language processing) genannt. NSV ist eines der wichtigsten Teilgebiete der KI und Informatik. Sie befasst sich mit der Mensch-Maschine-Interaktion durch natürliche Sprachen. Ziel der NSV ist die Entwicklung von Programmen, die in der Lage sind, natürliche Sprachen in ihrer Komplexität zu lesen, zu verarbeiten, zu analysieren und schlussendlich zu verstehen. Um dies zu erreichen, sind große Mengen an natürlichen Sprachdaten notwendig.

Der Europarat unterstützt die Vertragsstaaten bei der Umsetzung der Charta. Der vorliegende Bericht trägt einen Teil dazu bei. Sein Ziel ist es, aufzuzeigen, wie verschiedene NSV-Anwendungen und damit KI den alltäglichen Gebrauch und die Förderung von Regional- oder Minderheitensprachen erleichtern und somit die Vertragsstaaten bei der Umsetzung der von ihnen ratifizierten Bestimmungen der

¹ Armenien, Bosnien und Herzegowina, Dänemark, Deutschland, Finnland, Kroatien, Liechtenstein, Luxemburg, Montenegro, Niederlande, Norwegen, Österreich, Polen, Rumänien, Schweden, Schweiz, Serbien, Slowakei, Slowenien, Spanien, Tschechische Republik, Ukraine, Ungarn, Vereinigtes Königreich und Zypern.

² Aserbaidschan, Frankreich, Island, Italien, Malta, Moldau, Nordmazedonien, Portugal und Russische Föderation.

Charta unterstützen können. Dabei folgt der Bericht dem Aufbau der Charta und konzentriert sich auf die Artikel 7-14.

Das Rahmenübereinkommen zum Schutz nationaler Minderheiten des Europarats umfasst Rechte von Angehörigen nationaler Minderheiten in verschiedenen Bereichen, einschließlich Sprache und Kultur. Es wurde von 39 Staaten ratifiziert, von denen mehrere die Charta noch nicht ratifiziert haben. Da sich Maßnahmen zugunsten der Umsetzung der Charta auch positiv auf die Umsetzung der sprachlichen Rechte des Rahmenübereinkommens auswirken werden, wird in diesem Bericht auf die einschlägigen Bestimmungen des Rahmenübereinkommens verwiesen.

Der Sonderausschuss für künstliche Intelligenz des Europarats hat, gestützt auf dessen Normen zu Menschenrechten, Demokratie und Rechtsstaatlichkeit, mögliche Bestandteile eines Rechtsrahmens für künstliche Intelligenz geprüft. Während seiner Arbeit von 2019 bis 2021 hat er festgestellt, dass KI-Anwendungen Personen, die Minderheitensprachen sprechen, dazu befähigen können, aktiver am öffentlichen Leben, an Debatten, Beratungen oder Entscheidungsfindungen teilzunehmen, aber auch das Bewusstsein für diese Sprachen zu schärfen. Es wurde insbesondere darauf hingewiesen, dass die Verwendung von Minderheitensprachen in KI-Anwendungen ein Weg sein könnte, sprachliche Diskriminierung zu vermeiden.³

Der Bericht kann somit auch einen Beitrag zu den weitergehenden Betrachtungen des Europarats zum Thema KI leisten.

Aus Gründen der leichten Lesbarkeit wird in diesem Bericht die männliche Sprachform bei personenbezogenen Substantiven und Pronomen verwendet. Dies impliziert jedoch keine Benachteiligung des weiblichen Geschlechts, sondern soll im Sinne der sprachlichen Vereinfachung als geschlechtsneutral zu verstehen sein.

1. Allgemeine Gründe für den Gebrauch von künstlicher Intelligenz zur Erleichterung der Charta-Umsetzung

Bevor auf die spezifischen Möglichkeiten eingegangen wird, wie KI bei der Umsetzung einzelner Bestimmungen der Charta helfen kann, soll in diesem Abschnitt aufgezeigt werden, wie KI dazu beitragen kann, die allgemeinen Ziele der Charta zu erreichen.

Forscher sprechen von **digitalem Sprachenverlust**, wenn es einer Sprache nicht gelingt, sich in den neuen Kommunikationsmitteln und Technologien zu etablieren. Dies betrifft nicht nur weniger verbreitete („kleine“) Sprachen, sondern auch Sprachen, die zwar in einem Land die Mehrheitssprache sind, sich in einem anderen Land aber in einer bedrohten Minderheitenlage befinden. Um dieser Herausforderung entgegenzutreten, ist großer Einsatz zur Förderung von Regional- oder Minderheitensprachen in neuen Technologien erforderlich. Wenn sie richtig eingesetzt werden, bieten die neuen Technologie- und Kommunikationsmittel enorme Möglichkeiten zur Förderung dieser Sprachen.

³ Siehe Ad-hoc Committee on Artificial Intelligence (CAHAI): Feasibility Study, CAHAI(2020)23, S. 19; Analysis of the Multi-Stakeholder Consultation, CAHAI(2021)07, S. 22; Compilation of responses to the Multi-Stakeholder Consultation (F to M), CAHAI(2021)06, S. 80.

Die in diesem Bericht beschriebenen Anwendungen sind allesamt einfach zu handhaben, praxisnah und auf reale Situationen anwendbar. Des Weiteren tragen viele von ihnen dazu bei, Regional- oder Minderheitensprachen einem größeren Publikum nahezubringen. Wenn Regional- oder Minderheitensprachen in alltäglichen KI-Anwendungen wie z. B. Sprachassistenzsystemen verwendet werden können, werden die Möglichkeiten zur Verwendung der Sprache im Alltag erheblich erweitert. Außerdem werden neue Sprachlerner mobilisiert und motiviert, was dazu beiträgt, die Sprecherzahl zu erhöhen. Darüber hinaus erhöht die Präsenz von Regional- oder Minderheitensprachen auf verschiedenen Internetplattformen ihre Sichtbarkeit. Ein Beispiel dafür ist Katalanisch, das zu den 20 am häufigsten verwendeten Sprachen auf Wikipedia gehört.

Mithilfe von KI können Behörden mit relativ geringem Kostenaufwand schnell ein Angebot für Nutzer von Regional- oder Minderheitensprachen bereitstellen. Auf diese Weise unterstützt KI die Behörden bei einem „entschlossenen Vorgehen [...] zur Förderung von Regional- oder Minderheitensprachen, um diese zu schützen“, was eines der zentralen Ziele und Grundsätze der Charta ist (Artikel 7.1.c).

Alle in diesem Bericht erwähnten Anwendungen sind grenzüberschreitend und können in anderen Ländern mit derselben Regional- oder Minderheitensprache verwendet werden. Zudem kann eine Anwendung, die für einen bestimmten Bereich und eine bestimmte Sprache entwickelt wurde, leicht an andere Bereiche und andere Regional- oder Minderheitensprachen angepasst werden. Wie später im Bericht ausführlicher beschrieben, sind dies die Gründe, warum die Verwendung von KI zur Umsetzung der Charta den Austausch und die Vernetzung von Gruppen und Behörden aus verschiedenen Staaten, die dieselbe Regional- oder Minderheitensprache verwenden, fördert.

Die Fachgebiete, die sich mit KI und NSV beschäftigen, sind relativ neu und zukunftsorientiert. Die Einbeziehung von Regional- oder Minderheitensprachen in diese Forschungsfelder kann daher eine bedeutende Stütze für ihr Überleben und ihre Entwicklung als lebende Sprachen darstellen. Die Förderung von Studium und Forschung zu diesen Sprachen ist in Artikel 7.1.h festgelegt, der für alle Regional- oder Minderheitensprachen in den Vertragsstaaten gilt. Der Bericht wird an mehreren Stellen verschiedene Ansatzpunkte für zukünftige Forschung aufzeigen. Die wichtigste Grundlage für die Entwicklung hilfreicher NSV-Anwendungen ist jedoch die Erhebung relevanter natürlingsprachlicher Daten der Regional- oder Minderheitensprachen in **Textkorpora** (große Datensätze aus verschiedenen Texten in einer oder mehreren Sprache(n), abhängig von der jeweiligen NSV-Anwendung). Ohne ausreichende Daten ist die Entwicklung von funktionierenden NSV-Anwendungen nicht möglich. Daher muss die Datensammlung natürlicher Sprachdaten von Regional- oder Minderheitensprachen der erste Schritt sein, um Studium und Forschung zu fördern.

Es gibt bereits mehrere Forschungsprojekte und -gruppen in diesem Bereich. Ein Beispiel ist die Arbeitsgruppe SALTMIL, die das Ziel verfolgt, Forschung und Entwicklung im Bereich der Sprachtechnologie für Minderheitensprachen, insbesondere den europäischen, zu fördern. Eine weitere Forschungsgruppe ist Ixa an der Universität des Baskenlands (Spanien), die daran arbeitet, Algorithmen, Werkzeuge und sprachliche Ressourcen zu entwickeln und zur Verfügung zu stellen, die es Computern ermöglichen, menschliche Sprachen, insbesondere das Baskische, zu verarbeiten und zu verstehen. Die Gruppe hat bereits mehrere NSV-Werkzeuge für das Baskische veröffentlicht wie z. B. ein Wordnet (lexikalische Datenbank für semantische Beziehungen zwischen Wörtern), ein Rechtschreibkorrekturprogramm und einen Parser (Programm zur Satzbauanalyse) sowie Werkzeuge zur Informationsabfrage und -gewinnung, zur maschinellen Übersetzung und zum Sprachenlernen. Als drittes Beispiel steht die Universität Helsinki/Helsingfors, die Studienprojekte zu Sprachtechnologien

wie Baubanken (Erstellen eines Textkorpus mit Informationen über die syntaktischen und semantischen Strukturen der Sätze) für verschiedene finnische Minderheitensprachen durchführt.

Ein positives und erstrebenswertes Beispiel für die staatliche Unterstützung von Sprachtechnologien für Regional- oder Minderheitensprachen ist der walisische Sprachtechnologieplan von 2018. Er ist Bestandteil der allgemeinen Strategie der walisischen Regierung zur Förderung des alltäglichen Gebrauchs des Walisischen und zur Erhöhung der Sprecherzahl. Der Plan soll technologische Entwicklungen fördern und die Richtung für die Arbeit an walisischen Sprachtechnologien vorgeben. Um ihr Ziel zu erreichen, finanziert die walisische Regierung verschiedene Projekte in diesem Bereich wie z. B. die Programmierung von walisischen Sprachlernanwendungen für Kinder, die Erweiterung von OpenStreetMap (eine kostenlose kollaborative Open-Source-Weltkarte) um eine walisische Version, die Einrichtung einer walisischen Sprachdatenbank oder die Entwicklung von Maccsen, einem digitalen Assistenten in Walisisch. Die Regierung fördert auch die Verwendung bereits bestehender Parallelkorpora und anderer Hilfsmittel wie Stoppwortlisten (Listen häufig verwendeter Wörter einer Sprache wie Präpositionen oder Artikel) oder Programmen zur Syntaxanalyse für Forschungsprojekte. Darüber hinaus veranstaltet die walisische Regierung jährliche Konferenzen für Forscher an der Universität Bangor und schafft für diesen Forschungsbereich reservierte Stellen an den Universitäten. Dank dieses Projekts wird das öffentliche Bewusstsein für gegenwärtige und zukünftige walisische Sprachtechnologien und die walisische Sprache im Allgemeinen deutlich erhöht. Angesichts dieser positiven Auswirkungen kann es als Modellprojekt für andere Länder dienen.

2. Maschinelle Übersetzung

Maschinelle Übersetzungstechnologien können für die Umsetzung der Charta von besonders großem Nutzen sein. Viele der Verpflichtungen der Charta beziehen sich auf die Übersetzung verschiedener Arten von Dokumenten in Regional- oder Minderheitensprachen. Alle diese Maßnahmen könnten durch den Einsatz von maschineller Übersetzung mit einem geringeren zeitlichen und finanziellen Aufwand umgesetzt werden. Angesichts der Bedeutung der Übersetzung für die Umsetzung der Charta konzentrieren sich die folgenden Kapitel ausschließlich auf die maschinelle Übersetzung, ihre verschiedenen Methoden, welche Anwendungen bereits bestehen und welche Regional- oder Minderheitensprachen sie unterstützen sowie darauf, wie neue Übersetzungsanwendungen für spezifische Bereiche entwickelt werden können.

2.1 Verschiedene Methoden für maschinelle Übersetzung

Als maschinelle Übersetzung werden Programme bezeichnet, die einen Text automatisch von einer Sprache in eine andere übersetzen. Es handelt sich dabei um einen der am häufigsten erforschten Bereiche der **Computerlinguistik**. Maschinelle Übersetzung ist eine komplexe Herausforderung, da der Algorithmus in der Lage sein muss, zuerst den Sinn aus der Eingabe- bzw. Ausgangssprache zu entschlüsseln und dann diesen Inhalt in der Zielsprache neu zu kodieren. Um den Sinngehalt einer Phrase zu entschlüsseln, wäre jedoch ein Programm mit einem semantischen Verständnis der jeweiligen Sprache erforderlich, was technisch (noch) nicht möglich ist - lediglich die Verarbeitung natürlicher Sprache ist möglich, nicht aber deren Verständnis. Dieses fehlende semantische Verständnis der Software führt zu verschiedenen Übersetzungsschwierigkeiten: der Umgang mit Mehrdeutigkeiten (lexikalisch wie syntaktisch), die Übersetzung von Eigennamen (z. B. von Institutionen) und die Übertragung komplexer Satzstrukturen, ganz zu schweigen von der Beachtung

kultureller Unterschiede zwischen Ausgangs- und Zielsprache. All diese Probleme lassen sich auf die Tatsache zurückführen, dass Übersetzung nicht einfach nur die Übertragung von Wörtern in eine andere Sprache bedeutet, sondern vielmehr die Vermittlung der Satzbedeutung unter Berücksichtigung des jeweiligen sprachlichen und kulturellen Hintergrunds.

Es gibt jedoch bereits begrenzte Bereiche, in denen mit maschineller Übersetzung gute Ergebnisse erzielt werden können. Dies gilt vor allem für Bereiche, in denen eine standardisierte, formelhafte Sprache mit ähnlich strukturierten Sätzen und eindeutigen Formulierungen verwendet wird. Dies ist z. B. bei Wetterberichten oder Antragsformularen der Fall.

Es gibt drei wesentliche methodische Ansätze für maschinelle Übersetzung.

Der erste Ansatz ist die **regelbasierte maschinelle Übersetzung**. Für diese Methode benötigt man eine linguistische Beschreibung der jeweiligen Ausgangs- und Zielsprache mit einem vordefinierten Lexikon und expliziten syntaktischen Regeln. Aufgrund der Komplexität dieser Beschreibungen und der damit einhergehenden Unzuverlässigkeit der Übersetzungsergebnisse ist die regelbasierte maschinelle Übersetzung weitgehend nicht mehr zeitgemäß.

Der zweite Ansatz ist die **statistische maschinelle Übersetzung**. Hier werden Parallelkorpora (übersetzte Daten sowohl für die Ausgangs- als auch Zielsprache) verwendet, um daraus Wahrscheinlichkeiten abzuleiten, z. B., dass auf das Wort „Film“ häufig das Wort „war“ folgt. Bei einer neuen Eingabe verwendet der Übersetzungsalgorithmus diese Wahrscheinlichkeiten, um die statistisch gesehen beste Übersetzung zu finden, das heißt die Übersetzung mit der höchsten Wahrscheinlichkeit. Dieser Ansatz ist einfacher als der erste, da er keine manuell erstellten Regeln, sondern nur Parallelkorpora benötigt. Die statistische maschinelle Übersetzung kann jedoch nicht für alle Sprachkombinationen überzeugende Ergebnisse liefern, vor allem nicht für solche mit großen syntaktischen Unterschieden.

Der dritte Ansatz wird als **neuronale maschinelle Übersetzung** bezeichnet und verwendet KI und neuronale Netze, um die Verknüpfung von Sätzen zu „lernen“, ähnlich wie biologische Neuronen beim Lernen Verknüpfungen zwischen Informationen herstellen. Im Übersetzungsprozess kodiert das Netzwerk zunächst den Ausgangssatz in sogenannte **Wortvektoren**. Durch diese Vektoren kann das System semantische Beziehungen zwischen Wörtern in einen mathematischen Raum übertragen. Eine Bedeutungsdistanz zwischen zwei Wörtern wird dann durch die Differenz zwischen den Werten ihrer Vektorrepräsentationen erfasst. Wenn man z. B. die Werte des Wortvektors für „Mann“ vom Wortvektor für „König“ subtrahiert und die Vektorwerte für „Frau“ addiert, erhält man einen Vektor, der mit dem Wortvektor für „Königin“ identisch ist. Die Verwendung von Wortvektoren bedeutet einen erheblichen Fortschritt für maschinelle Übersetzung und ist einer der Gründe, warum neuronale maschinelle Übersetzung heute so gute Ergebnisse erzielt. Nach diesem Kodierungsprozess dekodiert das Netzwerk die Wortvektoren dann in die Zielsprache. Dies ist möglich durch die bereits erwähnten Neuronen, die die Verbindungen zwischen Ausgangs- und Zielsprache mithilfe von Parallelkorpora „gelernt“ haben. Die Methode, bei der die aus verbundenen Neuronen bestehenden neuronalen Netzwerke aus Daten lernen, wird als „training“ bezeichnet. Die neuronale maschinelle Übersetzung ist in der Lage, bessere Ergebnisse als je zuvor zu erzielen und gleichzeitig weniger Speicherplatz zu beanspruchen als statistische maschinelle Übersetzung. Deshalb wird die neuronale maschinelle Übersetzung heute von den meisten maschinellen Übersetzungsanwendungen verwendet.

2.2 Einsatz von vorhandenen maschinellen Übersetzungssystemen

Viele der bekannten maschinellen Übersetzungsprogramme unterstützen bereits verschiedene Regional- oder Minderheitensprachen, die in den Vertragsstaaten der Charta durch diese geschützt werden. In der folgenden Tabelle sind die wichtigsten Übersetzungssysteme und die von ihnen angebotenen Regional- oder Minderheitensprachen in alphabetischer Reihenfolge aufgeführt.

Maschinelles Übersetzungssystem	Unterstützte Regional- oder Minderheitensprachen	Anzahl der Sprachen
DeepL	Bulgarisch, Dänisch, Deutsch, Finnisch, Französisch, Griechisch, Italienisch, Litauisch, Polnisch, Rumänisch, Russisch, Schwedisch, Slowakisch, Slowenisch, Tschechisch, Ungarisch	16
Google Translate	Albanisch, Armenisch, Baskisch, Bosnisch, Bulgarisch, Dänisch, Deutsch, Finnisch, Französisch, Friesisch, Galicisch, Griechisch, Irisch, Italienisch, Jiddisch, Katalanisch, Kroatisch, Kurdisch, Litauisch, Mazedonisch, Polnisch, Rumänisch, Russisch, Schottisch-Gälisch, Schwedisch, Serbisch, Slowakisch, Slowenisch, Tschechisch, Tatarisch, Türkisch, Ukrainisch, Ungarisch, Walisisch, Weißrussisch	35
Microsoft Translator	Bosnisch, Bulgarisch, Dänisch, Deutsch, Finnisch, Französisch, Griechisch, Irisch, Italienisch, Katalanisch, Kroatisch, Kurdisch, Litauisch, Mazedonisch, Polnisch, Rumänisch, Russisch, Schwedisch, Serbisch, Slowakisch, Slowenisch, Tatarisch, Tschechisch, Türkisch, Ukrainisch, Ungarisch, Walisisch	27
PROMT	Deutsch, Finnisch, Französisch, Griechisch, Italienisch, Russisch, Tatarisch, Türkisch, Ukrainisch	9
Watson Language Translator (IBM)	Baskisch, Bosnisch, Bulgarisch, Dänisch, Deutsch, Finnisch, Französisch, Griechisch, Irisch, Italienisch, Katalanisch, Kroatisch, Litauisch, Polnisch, Rumänisch, Russisch, Schwedisch, Serbisch, Slowakisch, Slowenisch, Tschechisch, Türkisch, Ukrainisch, Ungarisch, Walisisch	25
Yandex	Albanisch, Armenisch, Baskisch, Bosnisch, Bulgarisch, Dänisch, Deutsch, Finnisch, Französisch, Galicisch, Griechisch, Irisch, Italienisch, Jiddisch, Katalanisch, Kroatisch, Litauisch, Mazedonisch, Polnisch, Rumänisch, Russisch, Schottisch-Gälisch, Schwedisch, Serbisch, Slowakisch, Slowenisch, Tatarisch, Tschechisch, Türkisch, Ukrainisch, Ungarisch, Walisisch, Weißrussisch	33

Diese Liste ist keineswegs vollständig. Es gibt einige andere Unternehmen, die maschinelle Übersetzungsdienste anbieten, und da der Markt sich schnell verändert, besteht die Möglichkeit, dass neue Unternehmen die Systemlandschaft rasch verändern. Beispielsweise wurde DeepL erst 2017 eingeführt und hat sich in dieser kurzen Zeit bereits einen guten Ruf erworben.

Alle oben aufgeführten Dienstleistungen sind frei zugänglich. Um die Dienste von Microsoft Translator und Watson Language Translator nutzen zu können, ist es erforderlich, ein Konto zu erstellen. Die Nutzung ist weiterhin kostenlos, allerdings im Fall von Watson Language Translator nur bis zum Erreichen einer bestimmten Anzahl von zu übersetzenden Wörtern.

Die Genauigkeit der verschiedenen Anwendungen variiert je nach der gewählten Ausgangs- und Zielsprache. Für einige Sprachen, insbesondere solche mit einer geringen verfügbaren Datenmenge, sind diese Übersetzungsdienste möglicherweise nur dazu geeignet, ein grundlegendes Verständnis des Inhalts eines Textes oder Dokuments zu erhalten, um dann zu entscheiden, ob eine weiterführende Übersetzung durch einen professionellen Übersetzer erforderlich ist. Dieser Prozess einer Grobübersetzung wird auch als „Gisting“ bezeichnet. Wie aus den obigen Angaben ersichtlich ist, haben alle Anwendungen andere Schwerpunkte sowie unterschiedliche Vor- und Nachteile. Aus diesem Grund ist es nicht möglich, eine Lösung für alle Verwendungszwecke zu empfehlen.

2.3 Entwicklung neuer maschineller Übersetzungssysteme

Obwohl die bestehenden maschinellen Übersetzungsdienste bereits eine Vielzahl von Regional- oder Minderheitensprachen anbieten, gibt es auch einige Regional- oder Minderheitensprachen, für die es keinen maschinellen Übersetzungsdienst gibt. Dennoch ist eine maschinelle Übersetzung auch für diese Sprachen möglich.

Wie im ersten Kapitel erwähnt, funktionieren maschinelle Übersetzungsanwendungen am besten, wenn sie für ein spezielles Thema mit begrenztem Vokabular und formelhaften Sätzen verwendet werden. Da dies bei vielen in der Charta erwähnten Übersetzungen (Rechtsdokumente, Verwaltungsformulare usw.) der Fall ist, ist es möglich, einen maschinellen Übersetzungsdienst für diese Bereiche zu entwickeln, der mit großer Wahrscheinlichkeit zuverlässige Ergebnisse liefern wird.

Die wichtigste Voraussetzung für die Entwicklung eines solchen Systems sind Paralleldaten derselben Texte in Ausgangs- und Zielsprache, die möglichst von professionellen Übersetzern übersetzt wurden. Als nächstes ist es möglich, entweder ein bereits funktionierendes Programm mit diesen Daten zu trainieren (es gibt mehrere Angebote, bei denen man eigene Trainingsdaten in ein bereits bestehendes Programm einfügen kann) oder mithilfe einer der vielen online verfügbaren Anleitungen ein eigenes Übersetzungsprogramm zu entwickeln. Dies könnte auch eine gute Gelegenheit sein, um Forschung und Studium zu Regional- oder Minderheitensprachen zu fördern, da mit verfügbaren Daten die Entwicklung eines solchen Systems eine geeignete Aufgabe z. B. für Studentengruppen aus verschiedenen Programmier- oder Computerstudiengängen ist.

3. Weitere Anwendungen der natürlichen Sprachverarbeitung und ihr Gebrauch für die Charta-Umsetzung

3.1 Regional- oder Minderheitensprachen im Privatleben (Artikel 7.1.d)

Heute findet ein großer Teil des gesellschaftlichen Lebens, insbesondere der jüngeren Generationen, online in **sozialen Netzwerken** statt, die viele KI-Technologien in ihre Dienste integrieren. Vor allem während der Covid-19-Pandemie waren die sozialen Medien ein wichtiges Kommunikationsinstrument, um soziale Kontakte aufrechtzuerhalten. Um den Gebrauch von Regional- oder Minderheitensprachen im Privatleben zu fördern, ist es naheliegend, sich auch mit

diesen sozialen Netzwerken zu befassen. Tatsächlich sind viele dieser Netzwerke bereits in Regional- oder Minderheitensprachen verfügbar. Die unterstützten Sprachen der drei wichtigsten sozialen Netzwerke sind in der folgenden Tabelle aufgeführt.

Soziales Netzwerk	Unterstützte Regional- oder Minderheitensprachen	Anzahl der Sprachen
Facebook	Albanisch, Armenisch, Baskisch, Bosnisch, Bulgarisch, Dänisch, Deutsch, Finnisch, Französisch, Friesisch, Galicisch, Griechisch, Irisch, Italienisch, Katalanisch, Kurdisch, Kroatisch, Litauisch, Mazedonisch, Polnisch, Rumänisch, Russisch, Schwedisch, Serbisch, Slowakisch, Slowenisch, Tatarisch, Tschechisch, Türkisch, Ukrainisch, Ungarisch, Walisisch, Weißrussisch	33
Instagram	Bulgarisch, Dänisch, Deutsch, Finnisch, Französisch, Griechisch, Italienisch, Kroatisch, Polnisch, Rumänisch, Russisch, Serbisch, Slowakisch, Schwedisch, Tschechisch, Türkisch, Ukrainisch	17
Twitter	Dänisch, Deutsch, Finnisch, Französisch, Griechisch, Italienisch, Polnisch, Rumänisch, Russisch, Schwedisch, Tschechisch, Türkisch, Ukrainisch, Ungarisch	14

Es gibt hauptsächlich vier Maßnahmen, die Behörden und Privatpersonen ergreifen können, um soziale Medien in Regional- oder Minderheitensprachen für deren Schutz und Förderung zu gebrauchen. Die erste ist die persönliche Verwendung von sozialen Netzwerken in Regional- oder Minderheitensprachen und die Ermutigung junger Menschen, z. B. im Unterricht von Regional- oder Minderheitensprachen, das Gleiche zu tun. So wird die Sprache zu einem aktiven Teil ihres alltäglichen sozialen Lebens und die Aktualität, Reichweite und Sichtbarkeit der Sprache wird erhöht. Zweitens können Sprecher von Regional- oder Minderheitensprachen, für die bereits eine Übersetzung auf sozialen Netzwerken zur Verfügung steht, dazu beitragen, diese Übersetzungen zu verbessern, z. B. in Online-Übersetzungsforen. Besonders auf Facebook gibt es eine große **Übersetzungsgemeinschaft**, die dazu beiträgt, weitere Regional- oder Minderheitensprachen in die Plattform aufzunehmen. Benutzer können eigene Übersetzungen einreichen oder bereits vorhandene verbessern. Drittens können Sprecher von Regional- oder Minderheitensprachen, die noch nicht über eine Übersetzung verfügen, diese anfordern und andere dazu ermutigen, das Gleiche zu tun, um den sozialen Netzwerken zu zeigen, dass ein Bedarf an Übersetzungen für diese Sprache besteht. Dies ist eine wichtige Möglichkeit für Fürsprecher von Regional- oder Minderheitensprachen, den Gebrauch dieser Sprachen in sozialen Medien zu fördern. Schließlich bieten soziale Medien eine neue Kommunikationsform für Gruppen von Sprechern oder Lernern von Regional- oder Minderheitensprachen, auch aus verschiedenen Ländern. Sie können Treffen oder Initiativen durchführen, in der Regional- oder Minderheitensprache kommunizieren und sich miteinander vernetzen. So kann der grenzüberschreitende Austausch erleichtert und gefördert werden.

3.2 Regional- oder Minderheitensprachen im Bildungswesen (Artikel 7.1.g, 8.1.f.ii, iii)

Das Angebot von Sprachunterricht ist eine der wichtigsten Maßnahmen für das Verständnis, den Gebrauch in verschiedenen sozialen Situationen und damit den Fortbestand einer Sprache. Mit verschiedenen **Online-Lernplattformen** kann KI das **Lernen von Regional- oder Minderheitensprachen** vereinfachen. Insbesondere während der Covid-19-Pandemie wurde deutlich,

dass solche Lernplattformen ein wichtiges Arbeitsmittel sind, in das diese Sprachen aufgenommen werden sollten. Sie können auch zur Umsetzung von Artikel 14.2 des Rahmenübereinkommens beitragen.

Online-Lernplattformen sind entweder Webseiten oder mobile Anwendungen, die sich an Menschen richten, die eine Sprache lernen oder ihre Sprachkenntnisse verbessern wollen. Die Plattformen nutzen verschiedene Medien (Texte, Tonaufnahmen etc.) und oft auch Methoden der sogenannten Gamifizierung (Verwendung von spielerischen Elementen wie Belohnungen oder Wettbewerbe), um den Lernprozess für die Nutzer unterhaltsamer und effizienter zu gestalten. Zudem nutzen die meisten Plattformen die Daten der Lerner, um ihre Produkte zu verbessern. Viele der bestehenden Plattformen unterstützen bereits verschiedene Regional- oder Minderheitensprachen, die in den Vertragsstaaten der Charta durch diese geschützt werden, wie der untenstehenden Tabelle zu entnehmen ist.

Anwendung	Unterstützte Regional- oder Minderheitensprachen	Anzahl der Sprachen
Beelingu	Deutsch, Französisch, Italienisch, Russisch, Schwedisch, Türkisch	6
Busuu	Deutsch, Französisch, Italienisch, Polnisch, Russisch, Türkisch	6
Clozemaster	Albanisch, Armenisch, Baskisch, Bulgarisch, Dänisch, Deutsch, Finnisch, Französisch, Galicisch, Griechisch, Irisch, Italienisch, Jiddisch, Katalanisch, Kornisch, Kroatisch, Litauisch, Mazedonisch, Polnisch, Rumänisch, Russisch, Schottisch-Gälisch, Schwedisch, Serbisch, Slowakisch, Slowenisch, Tschechisch, Türkisch, Ukrainisch, Ungarisch, Walisisch, Weißrussisch	32
Drops	Bosnisch, Dänisch, Deutsch, Finnisch, Französisch, Griechisch, Italienisch, Kroatisch, Polnisch, Russisch, Schwedisch, Serbisch, Türkisch, Ungarisch	14
Duolingo	Dänisch, Deutsch, Finnisch, Französisch, Griechisch, Irisch, Italienisch, Jiddisch (Testphase), Polnisch, Rumänisch, Russisch, Schottisch-Gälisch, Schwedisch, Tschechisch, Türkisch, Ukrainisch, Ungarisch (Testphase), Walisisch	18
Memrise	Dänisch, Deutsch, Französisch, Italienisch, Polnisch, Russisch, Schwedisch, Slowenisch, Türkisch	9
Mondly	Bulgarisch, Dänisch, Deutsch, Finnisch, Französisch, Griechisch, Italienisch, Katalanisch, Kroatisch, Litauisch, Polnisch, Rumänisch, Russisch, Schwedisch, Slowakisch, Tschechisch, Türkisch, Ukrainisch, Ungarisch	19

Die aufgeführten Lernplattformen sind alle als mobile Anwendungen zugänglich. Mit Ausnahme von Beelingu und Drops bieten alle Plattformen ihre Dienste auch als Webseite an. Für die Nutzung der Anwendungen wird ein Konto benötigt, um den Lernprozess zu speichern und anschließend an der letzten Speicherposition fortzusetzen. Alle oben genannten Plattformen sind kostenlos. Es gibt auch

weitere gebührenpflichtige Sprachlernangebote wie Babbel oder Rosetta Stone, die nicht in der obigen Liste aufgeführt sind, da eine kostenpflichtige Mitgliedschaft einige Sprachlerner ausschließen könnte.

Zu den oben genannten Lernplattformen gibt es einige zusätzliche Informationen zu beachten.

- **Beelingu** ist eine besondere Sprachlernanwendung, da keine normalen Kurse angeboten werden, sondern stattdessen Texte und Audiodateien, die das parallele Lesen und Hören der Sprache des Lerners und der zu lernenden Fremdsprache ermöglichen.
- **Clozemaster** bietet eine breite Palette von Sprachen an. Das Programm kann vor allem fortgeschrittenen Lernern empfohlen werden, da ebenfalls kein klassischer Online-Sprachkurs angeboten wird, sondern neue Vokabeln durch das Einfügen dieser Wörter in verschiedene Kontexte erlernt werden.
- **Drops** bietet eine Sprachlernanwendung für Kinder namens Droplets an. Sie vermittelt die Sprachen mit Zeichenspielen und ohne Werbung.
- **Duolingo** bietet Kurse von vielen Regional- oder Minderheitensprachen mit verschiedenen innovativen Sprachlernmethoden wie z. B. Geschichtenerzählen oder Podcasts für Sprachlerner an. Darüber hinaus gibt es eine spezielle Duolingo-Version für den Einsatz im Sprachunterricht an Schulen. Lehrer können gleichsam Klassenzimmer mit den Konten ihrer Schüler einrichten, deren Lernprozess überwachen und ihnen bestimmte Aufgaben zuweisen.
- **Memrise** bietet zusätzlich zu den oben genannten Sprachen weitere Sprachkurse an, die von Nutzern erstellt wurden. Nutzererstellte Kurse gibt es bereits für mehrere Regional- oder Minderheitensprachen wie Albanisch, Armenisch, Baskisch, Finnisch, Griechisch, Irisch, Katalanisch, Litauisch, Nordsamisch, Rumänisch, Schottisch-Gälisch, Ungarisch oder Walisisch. Die Aufnahme weiterer Sprachen ist einfach, da jedes registrierte Mitglied eigene Sprachkurse erstellen kann, die dann von jedem anderen Nutzer der Plattform verwendet werden können.
- **Mondly** bietet auch eine Sprachlernanwendung für Kinder namens Mondly Kids an. Sie vermittelt die Sprachen mithilfe von einfachen, kindgerechten und spielerischen Lektionen.

Neben diesen eher klassischen Sprachlernanwendungen gibt es auch verschiedene Programme (wie z. B. HelloTalk), die Sprachlerner mit Muttersprachlern der jeweiligen Sprache verbinden, so dass sie miteinander sprechen können. Diese **Online-Sprach-Tandemanwendungen** können auch für Lerner von Regional- oder Minderheitensprachen von Interesse sein.

Ein großer Nachteil der genannten Lernplattformen ist, dass nicht alle Kombinationen von Lernsprache und zu lernender Sprache bestehen. Die meisten Plattformen bieten nur einige wenige Lernsprachen an, meist Englisch. Daher können Online-Sprachlernplattformen bisher nur die allgemeine Sprachausbildung in den Vertragsstaaten unterstützen, nicht aber ersetzen. Die verschiedenen Lernanwendungen sind jedoch besonders für ältere Sprachlerner aus anderen Ländern interessant, die bereits über Grundkenntnisse z. B. in Englisch verfügen und eine Regional- oder Minderheitensprache lernen wollen.

Zusammenfassend lässt sich sagen, dass es bereits viele Sprachlernangebote gibt, die Regional- oder Minderheitensprachen anbieten und unterstützen. Die Nutzung dieser Dienste hat viele Vorteile wie z. B. die Möglichkeit, dass Menschen, die nicht in der Region oder nicht einmal im jeweiligen Land leben, eine Regional- oder Minderheitensprache lernen können oder dass die Sichtbarkeit von Regional- oder Minderheitensprachen verbessert wird. Die Vertragsstaaten können diese Möglichkeiten nutzen, indem sie entweder für bereits bestehende Online-Sprachkurse auf diesen Anwendungen werben oder neue Sprachkurse für die noch nicht von einer Sprachlernanwendung abgedeckten Regional- oder Minderheitensprachen erstellen. Dadurch würden die Sprachlernmittel

einer großen Anzahl neuer potenzieller Interessenten zur Verfügung gestellt und somit die Sichtbarkeit der jeweiligen Sprachen erhöht.

3.3 Regional- oder Minderheitensprachen in Justizbehörden (Artikel 9.1.a.i-iv, 9.1.b.i-iii, 9.1.c.i-iii, 9.1.d, 9.3)

Artikel 9 befasst sich mit dem Gebrauch von Regional- oder Minderheitensprachen in einem der Kernbereiche des Staates, den Justizbehörden. In diesem Zusammenhang können NSV-Anwendungen, insbesondere die maschinelle Übersetzung, bei der Aufbereitung von Dokumenten in Regional- oder Minderheitensprachen von Nutzen sein. Maschinelle Übersetzung ist besonders im Hinblick auf Artikel 9.1.d von Interesse, da sie die hohen Kosten, die oft mit einer professionellen Übersetzung verbunden sind, deutlich verringern kann.

Wie maschinelle Übersetzung funktioniert und welche Arten von Anwendungen es gibt, wurde bereits in Kapitel 2 erläutert. Wie dort beschrieben, sind maschinelle Übersetzungsergebnisse im Bereich der amtlichen Gerichtsdokumente aufgrund ihrer formelhaften Struktur und Sprache vermutlich vielversprechend. Eine Investition in diesem Bereich könnte daher zu lohnenden Ergebnissen führen.

Maschinelle Übersetzung kann jedoch auch dann nützlich sein, wenn die bereits vorhandenen maschinellen Übersetzungsprogramme für eine Grobübersetzung („Gisting“) verwendet werden. Dies ist sogar für Regional- oder Minderheitensprachen mit einem geringen Grad an maschineller Übersetzungsgenauigkeit möglich.

Maschinelle Übersetzung in Justizbehörden kann auch zur Umsetzung von Artikel 10.3 des Rahmenübereinkommens beitragen.

3.4 Regional- oder Minderheitensprachen in Verwaltungsbehörden und öffentlichen Dienstleistungsbetrieben

NSV-Anwendungen können die Verwendung von Regional- oder Minderheitensprachen durch und im Verkehr mit Verwaltungsbehörden und öffentlichen Dienstleistungsbetrieben erleichtern. Relevante Anwendungen sind Chatbots, verschiedene intelligente Suchmethoden, die Verwendung von Sprachsynthese zur Generierung von Ansagen von Straßennamen und erneut maschinelle Übersetzung. Die in diesem Abschnitt erwähnten Anwendungen können auch zur Umsetzung der Artikel 10.2 (Gebrauch von Minderheitensprachen im Verkehr mit Behörden) und 11.3 (topografische Namen) des Rahmenübereinkommens beitragen.

3.4.1 Chatbots (Artikel 10.1.a.i-iv, 10.2.a, 10.2.b, 10.3.a-c)

Chatbots sind Softwareprogramme, die durch die Nachahmung von Gesprächssituationen unmittelbar auf Kundenwünsche und -anfragen eingehen können. Sie sind in der Lage, Standardfragen zu beantworten oder den Kunden an weitere Angestellte zu verweisen. Diese **Gesprächsagenten** ahmen die geschriebene oder gesprochene menschliche Sprache nach, um eine Unterhaltung und Interaktion mit einer realen Person zu simulieren. Dies wird erreicht, indem die Chatbots zunächst die Texteingaben des Kunden verarbeiten und anschließend auf der Grundlage eines Algorithmus antworten, der das Gesagte des Kunden interpretiert und davon ausgehend eine angemessene Antwort ermittelt. Es gibt zum einen Chatbots, die Schlüsselwörter und -sätze erkennen können, um daraufhin eine vorbereitete oder vorprogrammierte Antwort auszugeben. Diese Methode ist

besonders nützlich für die Abfrage eindeutiger Informationen, die sich in klare Kategorien einordnen lassen. Es gibt jedoch auch Chatbot-Programme, die durch KI dazu in der Lage sind, neue Antworten auf Grundlage von vorherigen Kundeninteraktionen zu erlernen. Verschiedene Möglichkeiten bestehen, Kunden eine Chatbot-Software anzubieten: in beliebigen Nachrichten-Anwendungen, per SMS, in eigenständigen Anwendungen oder auf einer Webseite. Zudem bieten viele Firmen ihre Dienste für die Programmierung von Chatbots an. Wenn man bereits über vorgefertigte Dialoge verfügt, ermöglichen Webseiten, einfache Chatbots durch Ziehen und Ablegen zu erstellen.

Chatbots sind eine beliebte und weit verbreitete Methode, um einfache Anfragen mithilfe einer Konversationsoberfläche zu automatisieren, was die Automatisierung menschlicher erscheinen lässt. Sie werden heute in vielen Bereichen der **Kundeninteraktion** eingesetzt und können eine große Unterstützung für Behörden sein, da Nutzer über sie Termine vereinbaren, Formulare herunterladen, Antworten auf Standardfragen erhalten, Terminerinnerungen setzen oder wichtige Informationen z. B. über Öffnungszeiten erhalten können. Viele Behörden haben bereits Chatbots eingeführt, um ihr Dienstleistungsangebot zu erweitern und zu vereinfachen: die Städte Berlin, Bonn und Würzburg (Deutschland) sowie verschiedene Ämter in Finnland (Einwanderungs- und Finanzbehörde, Patent- und Meldeamt) oder das Londoner Verkehrsbüro in Großbritannien. Darüber hinaus hat die Europäische Kommission einen Bericht veröffentlicht, um Behörden zu helfen, die Chatbot-Programme in ihre Dienste integrieren wollen.⁴

Da Chatbots meist vollständig schriftlich kommunizieren, ist eine **kombinierte Nutzung mit maschineller Übersetzung** denkbar. Es wäre dementsprechend einfach, bereits vorhandene Chatbots auch an Regional- oder Minderheitensprachen anzupassen. Durch den Einsatz von Chatbot-Technologien können die Behörden sicherstellen, dass Sprecher von Regional- oder Minderheitensprachen problemlos bei den jeweiligen Ämtern schriftliche Anträge stellen und eine schriftliche Antwort in ihrer Sprache erhalten können.

3.4.2 Intelligente Suchmethoden (Artikel 10.2.g)

Das Internet wird heute für viele Aktivitäten genutzt, die mit der **Suche nach Ortsnamen** verbunden sind wie Reisebuchungen, Suchen nach Wegbeschreibungen oder Wettervorhersagen. Auf einigen Webseiten ist es möglich, diese Suchen unter Verwendung von Ortsnamen in Regional- oder Minderheitensprachen durchzuführen. Dies lässt sich auf verschiedene sogenannte intelligente Suchtechniken („smart search“) zurückführen. Wenn man beispielsweise einen Ortsnamen in einer Regional- oder Minderheitensprache in das Suchfeld von booking.com eingibt, führt die Webseite automatisch eine Quersuche in anderen Suchmaschinen wie Google durch, um weitere Ergebnisse zu finden. Auf diese Weise können verschiedene Versionen von Ortsnamen durch Suchergebnisse von anderen Webseiten miteinander verbunden werden.

Intelligente Suchmethoden verbinden dementsprechend Ergebnisse verschiedener Suchmaschinen und verknüpfen Suchergebnisse mit realen Entitäten wie Objekten, Personen oder Orten. Mithilfe von KI haben sich diese Suchmethoden erheblich weiterentwickelt und können auch mit Abweichungen, z. B. in der Rechtschreibung wie Tippfehler, umgehen. Intelligente Suchmethoden umfassen unter anderem:

⁴ European Commission, Directorate-General for Informatics & ISA2 Programme. (2019). Architecture for public service chatbots. Verfügbar unter: https://joinup.ec.europa.eu/sites/default/files/news/2019-09/ISA2_Architecture%20for%20public%20service%20chatbots.pdf.

- Extraktion von Schlüsselwörtern („keyword extraction“): Bestimmung von wichtigen Wörtern und Ausdrücken in einem Text durch Analyse der Worthäufigkeiten und deren Vorkommen mit anderen Wörtern;
- Extraktion von Entitäten („entity extraction“): Klassifikation von klar benennbaren Entitäten in einem Text in vordefinierte Kategorien wie z. B. Person, Ortsname, Zeitausdruck usw.;
- Verknüpfung von Entitäten („entity linking“): Verbindung eines Wortes mit dem Wissensdatensatz der realen Entität;
- „Web crawling“: automatische und systematische Suche auf anderen Webseiten, in der Regel zum Zweck der Web-Indizierung, um den Inhalt von Webseiten zu extrahieren, damit die Informationen bei Bedarf abgerufen werden können und um den Inhalt oder die Verweise auf den Inhalt durch andere Webseiten zu aktualisieren;
- Semantische Ähnlichkeit/Wortvektoren: Übersetzung von semantischen Beziehungen zwischen Wörtern in den mathematischen Raum, wie in Kapitel 2 über maschinelle Übersetzung beschrieben.

Diese Methoden beweisen, dass Webseiten nicht im luftleeren Raum arbeiten, sondern vielmehr voneinander abhängig sind, um relevante Ergebnisse für ihre Nutzer zu finden. Dies erklärt auch, warum Ortsnamen in Regional- oder Minderheitensprachen in einigen Suchmaschinen akzeptiert werden. Für Personen außerhalb des jeweiligen Unternehmens ist es recht schwierig, diesen Suchprozess zu beeinflussen. Die weitere Verwendung intelligenter Suchmethoden in denjenigen Suchmaschinen, die diese Methoden noch nicht einsetzen, sollte gefördert werden.

3.4.3 Sprachsynthese für die Ansage von Straßennamen (Artikel 10.2.g)

Sprachsynthese beschreibt die **künstliche Erzeugung menschlicher Sprache** durch die Verkettung von verschiedenen Sprachaufnahmen aus einer Datenbank. Je spezifischer die Domäne, desto länger können diese aufgenommenen Phrasen sein, d. h. sie können entweder nur aus Kombinationen von Vokalen und Konsonanten oder aus vollständigen Wörtern wie etwa für Zeitansagen bestehen. Durch maschinelles Lernen hat die Sprachsynthese große Fortschritte gemacht und liefert gute Ergebnisse, insbesondere für eingeschränkte Bereiche wie die Ansage von Straßennamen.

Sprachsynthese kann auf verschiedene Weise zur Erleichterung der Charta-Umsetzung eingesetzt werden. Eine Möglichkeit ist die Verwendung von **zweisprachigen Ansagen im öffentlichen Nahverkehr**. Die Ansagen, die zur Ankündigung der nächsten Haltestelle in öffentlichen Verkehrsmitteln verwendet werden, werden heute bereits häufig mithilfe von Sprachsynthese generiert und könnten leicht um die entsprechenden Straßennamen in Regional- oder Minderheitensprachen erweitert werden. Eine weitere Möglichkeit wäre die Einbindung von Regional- oder Minderheitensprachen in barrierefreie Fußgängersignale. Diese Signale vermitteln Blinden oder Personen mit eingeschränktem Sehvermögen Informationen über „Geh“- und „Halt“-Intervalle an Kreuzungen in nicht-visuellen Formaten, z. B. in Form von Hörsignalen. Einige barrierefreie Fußgängersignale vermitteln zudem Informationen, z. B. über den Straßennamen oder die Richtung des Fußgängerübergangs. Wenn diese Angaben durch gesprochene Signale vermittelt werden, können sie auch in der jeweiligen Regional- oder Minderheitensprache präsentiert werden.

Da für Sprachsynthese eine recht große Datengrundlage von vorab aufgezeichneten Äußerungen benötigt wird, ist die Nutzung dieser Methode zur Erleichterung der Charta-Umsetzung eher für diejenigen Regional- oder Minderheitensprachen von Interesse, die bereits über eine derartige Datengrundlage verfügen, etwa weil sie die Mehrheitssprache eines anderen Landes sind. Auf diese Weise wird ebenfalls der grenzüberschreitende Forschungsaustausch gefördert, da bereits bestehende Datenbanken aus Ländern, in denen die betreffende Sprache die Mehrheitssprache ist, in

Länder weitergeleitet werden könnten, wo sie eine Regional- oder Minderheitensprache ist. Wenn bereits Datenbanken für eine Sprache bestehen, sind Ansagen von Straßennamen in weiteren Ländern leicht umsetzbar und können Straßenschilder in Regional- oder Minderheitensprachen unterstützen und in manchen Fällen sogar ersetzen.

3.4.4 Maschinelle Übersetzung (Artikel 10.1.a.i-v, 10.1.b, 10.1.c, 10.2.a-f, 10.3.a-c, 10.4.a)

Wie in den vorangegangenen Kapiteln beschrieben, können maschinelle Übersetzungsanwendungen für Grobübersetzungen („Gisting“) oder für die vollautomatische Übersetzung von eingereichten Dokumenten, Verwaltungstexten, Formularen oder anderen amtlichen Schriftstücken verwendet werden. Aufgrund des eingeschränkten Bereichs und der formelhaften Struktur und Sprache dieser ist es auch möglich, neue maschinelle Übersetzungsanwendungen für diejenigen Regional- oder Minderheitensprachen zu entwickeln, die noch nicht von großen maschinellen Übersetzungsdiensten angeboten werden.

3.5 Regional- oder Minderheitensprachen in den Medien

Die Präsenz von Regional- oder Minderheitensprachen in verschiedenen Massenmedien ist für ihren Schutz und ihre Förderung von entscheidender Bedeutung. Um Regional- oder Minderheitensprachen einen Platz in den Medien zu sichern, kann KI durch die automatische Generierung von Untertiteln bei Fernsehsendungen und die automatische Extraktion von Informationen bei Zeitungsartikeln helfen. Diese Anwendungen können auch zur Umsetzung des Artikels 9.4 des Rahmenübereinkommens beitragen.

3.5.1 Automatische Generierung von Untertiteln (Artikel 11.1.a.i-iii, 11.1.c.i-ii)

Die automatische Generierung von Untertiteln ist ein Teilbereich der **automatischen Spracherkennung** („speech-to-text“). Es beschreibt den Prozess der automatischen Transkription von Audiodateien, z. B. von einem Video, in Untertitel. Bei Untertiteln geht es dabei nur um die gesprochene Sprache einer Tondatei, während bei sogenannten „closed captions“ auch andere Audiosignale wie Musik oder Hintergrundgeräusche einbezogen werden. Automatische Spracherkennung hat eine lange Forschungstradition und konnte in letzter Zeit von technologischen Fortschritten profitieren.

Zur automatischen Generierung von Untertiteln werden ein akustisches als auch ein schriftliches Sprachenmodell benötigt. Diese werden heute meist gleichzeitig am selben Korpus trainiert, um eine höhere Genauigkeit zu erreichen. Bei Spracherkennung handelt es sich um ein komplexes Problem, da diese in der Lage sein muss, mit Variationen bezüglich Akzent, Aussprache, Artikulation, Tonhöhe, Lautstärke oder Geschwindigkeit sowie mit Hintergrundgeräuschen, Echos usw. umzugehen. Spracherkennung funktioniert daher besser bei Sprachen mit umfangreichen Trainingsdaten von verschiedenen Sprechern und Kommunikationssituationen, um natürliche Kommunikation, die auch sich überlappende oder improvisierte Äußerungen umfasst, verarbeiten zu können.

Es gibt verschiedene Anwendungen, die sich mit der automatischen Untertitelung befassen. Eine kostenlose, weit verbreitete und einfach zu bedienende Anwendung ist der **Algorithmus zur automatischen Spracherkennung von YouTube (Google)**. Er unterstützt derzeit neun europäische Sprachen, und zwar Deutsch, Englisch, Französisch, Italienisch, Niederländisch, Portugiesisch, Russisch, Spanisch und Türkisch. Um ihn zu verwenden, muss ein Video in den YouTube-Editor hochgeladen und dann die automatische Untertitelung aktiviert werden. Auf der Hilfeseite von YouTube findet sich eine Anleitung, wie dies je nach Art des Videos durchzuführen ist. Nach der

automatischen Generierung ist es möglich, die Untertitel zu überprüfen und als Textdatei herunterzuladen. Da der YouTube-Algorithmus zur Generierung von Untertiteln auf maschinellem Lernen basiert, wird er ständig verbessert und für neue Sprachen aktualisiert. Eine weitere hilfreiche YouTube-Funktion ist die **automatische zeitliche Anpassung der Untertitel**. Dazu muss eine Textdatei mit allen in einem Video vorkommenden Texten bereitgestellt werden. Der Algorithmus findet dann die genauen Zeitpunkte im Video, an denen der entsprechende Text gesprochen wird. Es findet sich hierfür auch eine ausführliche Anleitung auf der Hilfeseite von YouTube. Dies funktioniert für die gleichen Sprachen wie die automatische Spracherkennung und kann insofern nützlich sein, als dass die Erstellung einer Abschrift eines Videos weitaus weniger zeitaufwändig ist als die zeitliche Terminierung der Untertitel.

Die Verwendung der automatischen Generierung von Untertiteln kann dazu beitragen, Fernsehsendungen in Regional- oder Minderheitensprachen auch für Hörgeschädigte zugänglich zu machen. Darüber hinaus fördert sie die Sichtbarkeit von Regional- oder Minderheitensprachen. Zuletzt ermöglicht die automatische Generierung von Untertiteln auch die maschinelle Übersetzung von Fernsehsendungen. Denkbar sind z. B. übersetzte Untertitel in Regional- oder Minderheitensprachen für wichtige Fernsehsendungen oder Untertitel in der Mehrheitsprache für Sendungen, die in Regional- oder Minderheitensprachen produziert wurden. Es gibt bereits mehrere Fernsehsender, die daran arbeiten, KI für die automatische Untertitelung zu nutzen wie der deutsch-französische Fernsehsender Arte oder der katalanische Fernsehsender Betevé. Daher ist dies erneut auch eine Möglichkeit für den grenzüberschreitenden wissenschaftlichen Austausch.

3.5.2 Automatische Extraktion von Informationen (Artikel 11.1.e.i-ii)

Automatische Informationsextraktion beschreibt die Aufgabe, aus unstrukturierten maschinenlesbaren Dokumenten, meist natürlichsprachlichen Texten, automatisiert Informationen zu extrahieren. Dies ist besonders nützlich für Dokumentengruppen, die auf einer bestimmten Vorlage basieren, wobei jedes Dokument innerhalb der Gruppe Entitäten oder Ereignisse in ähnlicher Form, aber mit unterschiedlichen Einzelheiten beschreibt. Um die gewünschten Informationen zu extrahieren, wird das Dokument zunächst durch verschiedene Verarbeitungsprozesse aufbereitet wie z. B. die automatische Wortarterkennung („part-of-speech tagging“) oder das Ermitteln der Verbstämme. Danach werden verschiedene Teilaufgaben (wie die Erkennung von benennbaren Entitäten oder die Analyse von Koreferenzen) durchgeführt, um die gewünschten Informationen zu extrahieren.

Die **automatische Erstellung von Zusammenfassungen** ist eine der Hauptanwendungen der automatischen Informationsextraktion. Sie beschreibt den Prozess der computergestützten Reduzierung eines Datensatzes, um eine Teilmenge der besagten Daten, also eine Zusammenfassung, zu erstellen, die die wichtigsten Informationen des ursprünglichen Inhalts enthält. Dies kann entweder extraktions- oder abstraktionsbasiert erfolgen. Das Ziel des ersten Ansatzes ist es, die Schlüsselphrasen im Dokument zu finden und die relevantesten Sätze zu extrahieren. Es werden dementsprechend Sätze bewertet und nach ihrer Relevanz sortiert. Im zweiten Ansatz erzeugt ein (maschinelles Lern-)Algorithmus zunächst eine interne semantische Repräsentation des ursprünglichen Inhalts und erstellt dann eine kurze, präzise und sprachlich korrekte Zusammenfassung. Dies ist rein rechnerisch weitaus anspruchsvoller als der erste Ansatz, der bereits länger erforscht wird und gute Ergebnisse liefert.

Für die automatische extraktionsbasierte Zusammenfassung stehen mehrere Online-Ressourcen zur Verfügung. Mithilfe dieser Ressourcen könnten kürzere Fassungen von Nachrichtenartikeln in

Regional- oder Minderheitensprachen auf weiteren Seiten veröffentlicht werden oder Nachrichtenartikel in der Mehrheitssprache automatisch gekürzt werden, so dass sie schneller und einfacher übersetzt werden können. Beides kann dazu beitragen, die regelmäßige **Veröffentlichung von Zeitungsartikeln** in Regional- oder Minderheitensprachen zu vereinfachen.

3.6 Regional- oder Minderheitensprachen in kulturellen Tätigkeiten und Einrichtungen

Regional- oder Minderheitensprachen stehen für vielfältige kulturelle Identitäten und reiche Geschichte. Deren Erhaltung und Förderung ist ein wesentlicher Bestandteil der Unterstützung einer Regional- oder Minderheitensprache. Darüber hinaus können moderne kulturelle Initiativen das Image einer Sprache, insbesondere bei jüngeren Generationen, verbessern und eine wichtige Rolle bei ihrer Weiterentwicklung als eine lebende Sprache spielen. Die Bedeutung von KI in diesem Bereich ist vielleicht nicht so klar wie in anderen Bereichen, aber NSV-Anwendungen wie Datenstrukturierung, maschinelle Übersetzung und automatische Untertitelung können auch hier von Nutzen sein. Die Anwendungen können auch zur Umsetzung des Artikels 5.1 des Rahmenübereinkommens beitragen.

3.6.1 Datenstrukturierung (Artikel 12.1.g, 12.1.h)

Datenstrukturierung beschreibt den Prozess der Strukturierung oder des Bündelns großer Datenmengen mithilfe von KI. Die Aufgabe, **Dokumente zu klassifizieren**, ist eine der Hauptanwendungen der NSV.

Es gibt verschiedene Algorithmen und Techniken⁵, die zur Erkennung von Mustern in Datenbanken verwendet werden. Für jedes neue Dokument, das der Datenbank hinzugefügt wird, entscheidet der Algorithmus auf Grundlage zuvor klassifizierter Dokumente, in welche Kategorie das neue Dokument gehört. Das Training für diese Art von Aufgaben kann entweder durch **beaufsichtigtes oder unbeaufsichtigtes Lernen** erfolgen. Im ersten Fall wird ein neuronales Netz mit einem manuell gekennzeichneten Datensatz trainiert, im zweiten Fall erfasst das Netzwerk die inhaltlichen Unterschiede zwischen den Dokumenten, indem es sie in einen mathematischen Raum abbildet und so den „Dokumentabstand“ berechnet. Diese Methode wurde in Kapitel 2 über maschinelle Übersetzung genauer erläutert und wird in diesem Gebiet als Clusteranalyse bezeichnet.

Methoden zur Datenstrukturierung können insbesondere bei großen unstrukturierten Daten, aber auch bei der Organisation bereits strukturierter Daten hilfreich sein.

Zur Erleichterung der Umsetzung von Artikel 12 können sie für die **Klassifikation von Publikationen** in Regional- oder Minderheitensprachen und für die **Strukturierung terminologischer Datenbanken** von Regional- oder Minderheitensprachen verwendet werden. Der Einsatz intelligenter Datenstrukturierungsmethoden für die Entwicklung und Pflege solcher Datenbanken kann verantwortlichen Stellen dabei helfen, den Zeit- und Kostenaufwand erheblich zu verringern. Je nach Größe, Sprache und Inhalt der Daten sowie dem betreffenden Land gibt es verschiedene Dienstleister und Anleitungen für Datenstrukturierung. Darüber hinaus können diese Datenstrukturierungsprojekte weitere Möglichkeiten für Studium und Forschung für Studenten aus verschiedenen Studiengängen im Bereich der Datenanalyse sein.

3.6.2 Maschinelle Übersetzung (Artikel 12.1.a, 12.1.b, 12.1.c)

Die Verwendung maschineller Übersetzung für literarische Werke ist äußerst schwierig. Selbst für professionelle Übersetzer ist es nicht einfach, Literatur zu übersetzen. Bei einigen stark

⁵ „naive Bayes“, „tf-idf (term frequency-inverse document frequency)“, „support vector machines“ usw.

sprachabhängigen Werken wie z. B. Gedichten kann die Übersetzung sogar unmöglich sein. Natürlich ist es deshalb umso schwieriger, literarische Werke maschinell zu übersetzen. Maschinelle Übersetzung kann hier dennoch von Nutzen sein, z. B. für die **Übersetzung von Zusammenfassungen oder Untertiteln**. Durch übersetzte Zusammenfassungen könnte ein größeres Interesse an Werken in Regional- oder Minderheitensprachen erzeugt werden, weil dadurch, dass die grundlegende Handlung bereits übersetzt ist, die Sprecher anderer Sprachen stärker motiviert werden könnten, Werke in Regional- oder Minderheitensprachen zu lesen. Natürlich funktioniert diese Methode auch umgekehrt, um den Zugang zu literarischen Werken in anderen Sprachen für Sprecher von Regional- oder Minderheitensprachen zu fördern. Darüber hinaus könnten literarische Werke, von denen die übersetzten Zusammenfassungen häufig angefordert werden, dann manuell übersetzt werden, da ein allgemeines Interesse an dem betreffenden Werk festgestellt werden konnte.

3.6.3 Automatische Generierung von Untertiteln (Artikel 12.1.b, 12.1.c)

In Artikel 12 werden „Tätigkeiten auf dem Gebiet der Untertitelung“ erwähnt. Wie bereits beschrieben, können diese mithilfe von Spracherkennung automatisch durchgeführt werden. Dies reduziert den zeitlichen und finanziellen Aufwand, da die Untertitel nur noch manuell überprüft und nicht mehr komplett erstellt werden müssen. Bereits funktionierende Beispiele für die automatische Untertitelung und die verschiedenen Möglichkeiten in diesem Bereich wurden im vorherigen Kapitel zu Artikel 11 ausführlicher beschrieben.

3.7 Regional- oder Minderheitensprachen im wirtschaftlichen und sozialen Leben

Regional- oder Minderheitensprachen müssen in allen Bereichen des wirtschaftlichen und sozialen Lebens verwendet werden, damit sie vollständig funktionale Sprachen bleiben. Um dieses Ziel zu erreichen, kann KI durch die Einbindung von Regional- oder Minderheitensprachen in Stimmungs- und Meinungsanalysen und erneut durch maschinelle Übersetzung helfen.

3.7.1 Sentimentanalyse (Artikel 13.1.c, 13.1.d, 13.2.b)

Sentimentanalyse beschreibt die Verwendung von NSV, um ein **Dokument mit subjektiven Informationen** (wie eine Rezension oder eine Antwort auf eine Umfrage) **zu klassifizieren**, z. B. in die Kategorien positiv/zufrieden, neutral, negativ/verärgert. Sie ist nützlich, um einen schnellen Einblick in große Textdatenmengen zu erhalten.

Die Sentimentanalyse ist ein Klassifikationsverfahren, das ähnlich wie die im obigen Kapitel zur Datenstrukturierung beschriebene Dokumentenklassifikation funktioniert. Das neuronale Netz wird zunächst beaufsichtigt oder unbeaufsichtigt trainiert, um die Zuordnung einer Eingabe (z. B. eine Filmrezension) zu einer Ausgabekategorie (z. B. „positiv“) zu erlernen. Dies kann entweder vollständig durch maschinelles Lernen erfolgen oder mit manuell festgelegten Regeln wie „toll“ = „positiv“. Im Testverfahren erhält das Netzwerk dann neue Texte und wandelt sie in Wortvektoren um. Dieser Prozess wird hier Merkmalsextraktion genannt und im obigen Kapitel 2 über maschinelle Übersetzung näher beschrieben. Anschließend werden diese Vektoren einem Klassifikationsalgorithmus übergeben, der dann eine Ausgabekategorie für das eingegebene Dokument vorhersagt und damit z. B. den Tweet „Der Film war wirklich toll!“ als „positiv“ markiert.

Heute wird Sentimentanalyse meist für die **Auswertung von Kundenbewertungen** verwendet. Programme für die Stimmungs- und Meinungsanalyse in Regional- oder Minderheitensprachen zu entwickeln würde bedeuten, die Sprecher nicht mehr aus Dialogen über Kundenzufriedenheit und Möglichkeiten zur Dienstleistungsverbesserung eines Unternehmens auszuschließen.

Es gibt mehrere Unternehmen, die Dienstleistungen im Bereich der Sentimentanalyse anbieten, je nach Land, Sprache, Datengröße und Art der gewünschten Analyse. Dies könnte ebenfalls eine Gelegenheit sein, das Studium und die Forschung über Regional- oder Minderheitensprachen zu fördern. Sentimentanalyse ist ein höchst relevantes und aktuelles Forschungsthema mit vielen Online-Anleitungen, das in allen NSV- und KI-Anfängerkursen behandelt wird. Mit den verfügbaren Daten ist die Entwicklung eines solchen Systems eine mögliche Aufgabe z. B. für Studentengruppen aus verschiedenen Programmier- oder Informatikstudiengängen.

3.7.2 Maschinelle Übersetzung (Artikel 13.1.a, 13.1.d, 13.2.a, 13.2.b, 13.2.d, 13.2.e)

In Artikel 13 werden verschiedene Arten von Dokumenten erwähnt, die eine Übersetzung in Regional- oder Minderheitensprachen erfordern: Sicherheitshinweise und Informationen über Verbraucherrechte sowie Verträge, technische Schriftstücke, Zahlungsanweisungen und andere Finanzdokumente. Durch maschinelle Übersetzung können diese Dokumente schneller und kostengünstiger in der gewünschten Sprache zur Verfügung gestellt werden. Da die Dokumente alle zu spezifischen, abgeschlossenen Bereichen gehören und in formelhafter Sprache verfasst sind, ist eine maschinelle Übersetzung möglich, entweder zur Grobübersetzung („Gisting“) oder zur vollständig automatischen Übersetzung. Es ist auch hier wieder möglich, neue maschinelle Übersetzungsanwendungen für die noch nicht von maschinellen Übersetzungssystemen abgedeckten Regional- oder Minderheitensprachen zu entwickeln. Wie bereits erwähnt, ist dies auch eine Möglichkeit für Studium und Forschung. Es ist aber hervorzuheben, dass gerade für den Bereich der Sicherheitshinweise das **Korrekturlesen** noch wichtiger ist als in anderen Bereichen. Kleine Übersetzungsfehler, wie sie bei der maschinellen Übersetzung weiterhin vorkommen, könnten fatale Folgen haben.

3.8 Regional- oder Minderheitensprachen im grenzüberschreitenden Austausch (Artikel 7.1.i, 14)

Grenzüberschreitende Zusammenarbeit vereinfacht die Förderung von Regional- oder Minderheitensprachen, die auch Amts- oder Minderheitensprachen in anderen Staaten sind, erheblich, da die vorhandene Infrastruktur in Bereichen wie Bildung (z. B. Lehrmittel, Lehrerausbildung) und Medien unmittelbar oder angepasst übernommen werden kann.

Wie im vorliegenden Bericht zu lesen war, fördert die Nutzung von KI zur Erleichterung der Umsetzung der Charta in hohem Maße **grenzüberschreitenden Austausch in Studium und Forschung** zu Anwendungen und Datensätzen. Diese Inhalte sind von Natur aus international, und Regional- oder Minderheitensprachen können von den NSV-Anwendungen anderer Staaten profitieren, vor allem von denen, wo diese Sprache die Mehrheitssprache ist. Mit den entsprechenden Sprachdaten können die Anwendungen zudem leicht an weitere Regional- oder Minderheitensprachen angepasst werden, wenn der Quellcode für andere Forscher zur Verfügung gestellt wird. Wenn z. B. eine Behörde einen Chatbot für ihre Dienstleistungen entwickelt, kann dieses Programm problemlos in anderen Ländern mit der gleichen Regional- oder Minderheitensprache verwendet oder leicht an eine andere Regional- oder Minderheitensprache angepasst werden. Auf diese Weise erleichtert KI nicht nur die Umsetzung der Charta, sondern begünstigt und fördert auch insgesamt den grenzüberschreitenden Austausch. Somit kann KI auch zur Umsetzung von Artikel 18 des Rahmenübereinkommens beitragen.

Fazit

Aufgrund des breiten Spektrums an Anwendungen und der schnellen Veränderungen im Bereich der KI-Technologien konnte dieser Bericht nur einen ersten allgemeinen Überblick und einige beispielhafte Impulse dazu geben, wie verschiedene NSV-Anwendungen den Vertragsstaaten bei der Umsetzung der Europäischen Charta der Regional- oder Minderheitensprachen und bis zu einem gewissen Grad des Rahmenübereinkommens zum Schutz nationaler Minderheiten helfen können. Der vorliegende Bericht erhebt keinen Anspruch auf Vollständigkeit, da es viele zusätzliche Ressourcen, Anwendungen und Möglichkeiten gibt, KI für die Umsetzung der Charta einzusetzen. Darüber hinaus können die in diesem Bericht gegebenen Anregungen allesamt Ausgangspunkte für neue Diskussionen und Studienprojekte sein.

Der Bericht konnte aufzeigen, dass künstliche Intelligenz und natürliche Sprachverarbeitung zahlreiche neue Möglichkeiten für den Schutz, die Erforschung und die Förderung von Regional- oder Minderheitensprachen bieten. Aufgrund der derzeitigen Dominanz des Englischen im Bereich von KI sind Regional- oder Minderheitensprachen jedoch auch der Gefahr des „digitalen Sprachenverlusts“ ausgesetzt. Um dies zu verhindern, ist ein entschlossenes Vorgehen vonnöten, insbesondere im Bereich der **Datenerhebung**. Mit genügend vorhandenen Daten wie z. B. Parallelkorpora der Regional- oder Minderheitensprache und der Amts- oder Mehrheitssprache eröffnen sich den Regional- oder Minderheitensprachen zahlreiche neue Möglichkeiten, sich als moderne, relevante und lebende Sprachen zu profilieren. Für fast jede in diesem Bericht vorgeschlagene Maßnahme ist die Sammlung relevanter natürlichsprachlicher Daten der erste Schritt und sollte daher so früh wie möglich in Angriff genommen werden, um den Stellenwert der jeweiligen Sprache in modernen Sprachverarbeitungstechnologien zu sichern.

Die Forschung auf dem Gebiet der KI und NSV ist besonders schnelllebig und alle in diesem Bericht erwähnten Anwendungen werden ständig verbessert. Wahrscheinlich werden viele der vorgeschlagenen halbautomatischen Methoden in den kommenden Jahren vollständig automatisiert werden und die Methoden, die gegenwärtig als rudimentär beschrieben werden, werden in Zukunft sehr gute Resultate erzielen. KI kann die Vertragsstaaten bereits jetzt bei der Charta-Umsetzung unterstützen und diese Unterstützung wird höchstwahrscheinlich in den kommenden Jahren nur noch größer werden. Es lohnt sich daher, jetzt in NSV-Anwendungen für Regional- oder Minderheitensprachen zu investieren, um schon heute und erst recht morgen von deren Vorteilen zu profitieren.

www.coe.int

Der Europarat ist die führende Menschenrechtsorganisation des Kontinents. Er umfasst 46 Mitgliedsstaaten, von denen 28 Mitglieder der Europäischen Union sind. Alle Mitgliedsstaaten des Europarats haben die Europäische Menschenrechtskonvention unterzeichnet, einen Vertrag zum Schutz der Menschenrechte, Demokratie und Rechtsstaatlichkeit. Der Europäische Gerichtshof für Menschenrechte überwacht die Umsetzung der Konvention in den Mitgliedsstaaten.