

## DISKUSSION

ZEITSCHRIFT FÜR DIALEKTOLOGIE UND LINGUISTIK

DOI 10.25162/ZDL-2026-0001, online first [16.01.2026], (CC BY-NC-ND 4.0)

NETZWERK REGIONALE SPRACHE UND KÜNSTLICHE INTELLIGENZ

# Regionale Sprache und Künstliche Intelligenz im Zeitalter der digitalen Transformation

Positionspapier zur Zukunft der Regionalsprachenforschung

## Regional Language and Artificial Intelligence in the Era of Digital Transformation

Position Paper on the Future of Regional Language Research

**KURZFASSUNG:** Für die Regionalsprachenforschung ergibt sich durch die rapide Entwicklung der KI-Technologie der letzten Jahre ein großes Innovationspotenzial. Zugleich kommt dem Fach hinsichtlich der Repräsentation regionaler Sprache im digitalen Raum eine besondere Verantwortung zu. Die digitale Transformation stellt das Fach vor die drei übergeordneten Entwicklungsziele, (1) die Potenziale von KI zu nutzen, (2) die kulturelle Vielfalt abzubilden und (3) die digitale Transformation verantwortlich zu gestalten. Das Positionspapier definiert diese Entwicklungsziele und zeigt über eine Standortbestimmung auf, welche Voraussetzungen und Rahmenbedingungen für den Einsatz von KI-basierten Verfahren im Fach aktuell gegeben sind. Die darüber erschlossenen Herausforderungen und Desiderate führen zu zentralen Leitlinien, die den Entwicklungszielen dienen. Mit konkreten Empfehlungen werden schließlich Maßnahmen für die Weiterentwicklung der Regionalsprachenforschung vorgestellt.<sup>1</sup>

**SCHLAGWÖRTER:** Regionalsprachenforschung, Dialektologie, Künstliche Intelligenz, FAIR-Prinzip, Forschungsdateninfrastruktur, Open Science

**ABSTRACT:** For research on regional languages, the rapid development of AI technology in recent years offers significant innovative potential. At the same time, the discipline carries a particular responsibility when it comes to representing regional language in the digital sphere. The digital transformation confronts the field with three overarching development goals: (1) harnessing the potential of AI, (2) depicting cultural diversity, and (3) shaping digital transformation responsibly. This position paper defines these developmental goals and, through an assessment of the current situation, outlines the prerequisites and conditions currently in place

- 1 Das Positionspapier ist das Ergebnis eines interdisziplinären Scoping-Workshops mit Vertreterinnen und Vertretern der Regionalsprachenforschung des Deutschen, der Digitalen Linguistik und der KI-Forschung. Der Scoping-Workshop fand mit Unterstützung der VolkswagenStiftung vom 27. bis 29. August 2025 auf Schloss Herrenhausen bei Hannover statt. Beteiligt waren 35 Wissenschaftlerinnen und Wissenschaftler, die über das Thema „Künstliche Intelligenz und regionale Identität – Potenziale der Dialektologie im Zeitalter digitaler Transformation“ diskutierten. Gemeinsam wurde das Netzwerk *Regionale Sprache und Künstliche Intelligenz* gegründet, dessen Mitglieder als Autorinnen und Autoren für dieses Positionspapier verantwortlich zeichnen.

for the use of AI-based methods in the field. The resulting challenges and desiderata lead to central guidelines that support these goals. Finally, concrete recommendations present measures for advancing research on regional languages.

KEYWORDS: regional language studies, dialectology, artificial intelligence, FAIR principle, research data infrastructure, open science

## Übersicht

<b>Ausgangspunkt</b>	Für die Regionalsprachenforschung ergibt sich durch die rapide Entwicklung der KI-Technologie der letzten Jahre ein großes <b>Innovationspotenzial</b> . Zugleich kommt dem Fach hinsichtlich der Repräsentation regionaler Sprache im digitalen Raum eine besondere <b>Verantwortung</b> zu. Die digitale Transformation stellt das Fach vor drei übergeordnete Entwicklungsziele.
<b>Entwicklungsziele</b>	1 – Potenziale von KI nutzen 2 – Kulturelle Vielfalt abbilden 3 – Digitale Transformation gestalten
<b>Positionspapier</b>	Das Positionspapier definiert die <b>Entwicklungsziele</b> (Kap. 1) für das Fach und zeigt über eine <b>Standortbestimmung</b> (Kap. 2) auf, welche Voraussetzungen und Rahmenbedingungen für den Einsatz von KI-basierten Verfahren im Fach aktuell gegeben sind. Die darüber erschlossenen <b>Herausforderungen</b> und Desiderate (Kap. 3) führen zu zentralen <b>Leitlinien</b> (Kap. 4), die den Entwicklungszielen dienen. Mit den konkreten <b>Empfehlungen</b> (Kap. 5) werden schließlich Maßnahmen für die Weiterentwicklung der Regionalsprachenforschung vorgestellt.
<b>Leitlinien</b>	<ul style="list-style-type: none"> <li>– FAIRe regionalsprachliche Daten in hoher Qualität bereitstellen</li> <li>– Fachspezifische Infrastruktur ausbauen und langfristig vernetzen</li> <li>– Wissenstransfer durch Kooperationen zwischen Regionalsprachenforschung und KI-Forschung verstetigen</li> <li>– Doppelten Kompetenzerwerb bei Studierenden und Forschenden in der Regionalsprachenforschung forcieren</li> <li>– Wandel der eigenen Fachkultur hin zu Open Science gestalten</li> <li>– Dialektsprecherinnen und -sprecher partizipativ einbinden</li> <li>– Diskurs über regionale Sprache und KI aktiv mitgestalten</li> <li>– Gesellschaftliche Verantwortung durch Einbringen der eigenen regionalsprachlichen Expertise in die Diskurse übernehmen</li> <li>– Nachhaltigkeit in Ressourcennutzung anstreben</li> </ul>

<b>Adressaten</b>	Das Positionspapier richtet sich an die Fachgemeinschaft der Regionalsprachenforschung, ihre Fachgesellschaften, Infrastruktureinrichtungen sowie Fördermittelgeber.
<b>Fallbeispiele</b>	Über drei Fallbeispiele aus den Bereichen Datenbearbeitung, Datenanalyse und Datengenerierung wird der Einsatz von KI in der regionalsprachlichen Forschung konkretisiert. Die Potenziale und Risiken werden diskutiert und Perspektiven ausgewiesen.

## 1. Einleitung: Ausgangspunkt und Aufgaben

Die digitale Transformation der vergangenen Jahre sowie der Umbruch, den KI-Systeme und vor allem text- und akustikbasierte große Sprachmodelle (LLMs sowie SpeechLLMs) gegenwärtig auslösen, stellt die Variationslinguistik, insbesondere die Regionalsprachenforschung,<sup>2</sup> vor eine doppelte Aufgabe: Die Transformation eröffnet neue methodische Möglichkeiten, bringt aber zugleich erhebliche Herausforderungen mit sich. Die rapide Weiterentwicklung leistungsfähiger Verfahren der Künstlichen Intelligenz (KI)<sup>3</sup> hat die Möglichkeiten der Generierung, Aufbereitung und Analyse von textlichen und akustischen Sprachdaten in kurzer Zeit grundlegend erweitert. Für ein Fach, das traditionell auf empirische Datenerhebung, ressourcenintensive Datenaufbereitungen, fundierte qualitative wie quantitative Analysen, sorgfältige Kontextualisierung und differenzierte Interpretation setzt, bedeutet dies sowohl einen methodischen Umbruch als auch die Notwendigkeit zur kontinuierlichen, die Arbeitsprozesse begleitenden kritischen Reflexion. Es gilt, die gewachsene Expertise in der Dokumentation, Analyse und Interpretation sprachlicher Vielfalt zu bewahren, sie aktiv auszubauen, methodisch zu erweitern und in neue Kontexte zu überführen. Die Regionalsprachenforschung kann hier weichenstellend für andere Disziplinen der Sprachwissenschaft (und darüber hinaus) aktiv werden, da sie per se Fragen zu sprachlicher Variation jenseits von Normierung und Standardisierung adressieren muss.

- 2 Regionalsprachenforschung referiert im Positionspapier auf das gesamte Feld der Erforschung regionaler Varietäten (in Gegenwart und Vergangenheit) mit den Objektbereichen der lokalen Basisdialekte als traditionelle Sprachvarietäten bis hin zu den sich großräumig erstreckenden regionalen Akzenten und der arealen Variation im Standard. Die Betrachtung schließt neben den regionalen Sprechweisen des binnendeutschen Sprachraums in Mitteleuropa auch alle extraterritorialen Sprachgebiete des Deutschen ein. Wir beziehen uns damit auf das, was die Regionalsprachenforschung im engeren Sinne ausmacht: die Erforschung der räumlich gebundenen Variation des Deutschen in all ihren Ausprägungen. Beteiligt sind insbesondere die germanistische Sprachwissenschaft mit ihren Fachausrichtungen Regionalsprachenforschung, Dialektologie und Soziolinguistik, aber auch die Computerlinguistik, Korpuslinguistik, Phonetik und andere spezialisierte Teildisziplinen, wie sie etwa in der DFG-Fächergruppe 1.14 angesprochen sind.
- 3 Der Begriff „Künstliche Intelligenz“ wird im Sinne der DFG weit gefasst und schließt auch etablierte Methoden wie Maschinelles Lernen (ML) oder Text- und Datamining (TDM) ein (vgl. z. B. die Förderlinie „Datenkorpora für Künstliche Intelligenz“).

Die digitale Transformation eröffnet zugleich neue Forschungs- und Anwendungsfelder, in denen regionalsprachliche Erkenntnisse substanzielle Beiträge leisten können: zum Beispiel bei der Entwicklung sprach-, variations- und varietätensensitiver KI-Systeme und damit auch der Sichtbarmachung kultureller Identität in digitalen Räumen oder der Stärkung gesellschaftlicher Teilhabe. Besonders deutlich wird hierbei die enge Verknüpfung von Sprache und Identität. Dialekte und regionale Varietäten sind Träger kultureller Symbolik und Ausdruck regionaler Zugehörigkeit. Wenn solche Daten mithilfe von KI-Systemen nicht hinreichend exakt erkannt, analysiert oder generiert werden, stellt sich mit Nachdruck die Frage, wie Authentizität, Vielfalt und die komplexen sozialen Dimensionen von Sprache gewahrt werden können. Es ergibt sich u. a. folgendes Problem: Während weit verbreitete Standardsprachen wie z. B. Englisch, Chinesisch und Spanisch, aber auch Deutsch über viele digitale Ressourcen verfügen und damit fest im Repertoire von KI-Systemen verankert sind, bleiben ressourcenärmere Sprachen bislang unterrepräsentiert. Hierzu gehören nicht nur kleinere Nationalsprachen wie z. B. Estnisch, Slowenisch oder Ungarisch, sondern insbesondere auch regionale Varietäten wie Dialekte oder regionale Akzente. Grund hierfür sind insbesondere fehlende Trainingsdaten – sowohl im Hinblick auf Menge als auch auf Qualität. Dies hat zur Folge, dass ein zentraler Teil sprachlicher und kultureller Vielfalt im digitalen Raum nicht sichtbar ist.

Vor diesem Hintergrund ergeben sich für die Regionalsprachenforschung **drei übergeordnete Ziele**, die auf die technische Innovation von KI-Verfahren, ihre gesellschaftliche Relevanz und die erforderliche kritische Reflexion abzielen.

**Ziel 1 – Potenziale von KI nutzen:** Die Regionalsprachenforschung macht KI-basierte Technologien für die Wissenschaft verstärkt nutzbar, um digitale Arbeitsprozesse effizienter zu gestalten und neue Erkenntnismöglichkeiten zu gewinnen.

**Ziel 2 – Kulturelle Vielfalt abbilden:** Die Regionalsprachenforschung übernimmt Verantwortung für eine angemessene und authentische Repräsentation von Regionalität in KI-generierter Sprache, um die Vielfalt des Deutschen angemessen abzubilden.

**Ziel 3 – Digitale Transformation gestalten:** Die Regionalsprachenforschung bringt ihre Expertise aktiv in die wissenschaftspolitischen und gesellschaftlichen Diskurse zum Einsatz von KI in Forschung und Alltag ein, um die digitale Transformation in Wissenschaft und Gesellschaft nachhaltig und im Sinne der gesellschaftlichen Werte zu gestalten.

Ziel des Positionspapiers ist es, eine **Standortbestimmung** des Fachs im Hinblick auf die genannten Ziele vorzunehmen: Wie kann die Regionalsprachenforschung das Potenzial KI-basierter Technologien produktiv in ihre Forschungsprozesse integrieren, um neue methodische und inhaltliche Erkenntnisse zu gewinnen? Auf welche Weise kann die Regionalsprachenforschung ihre fachliche Expertise in die wissenschafts- und gesellschaftspolitischen Diskurse zur digitalen Transformation einbringen, um diese

kritisch und konstruktiv mitzugestalten? Wie verändert die Auseinandersetzung mit Künstlicher Intelligenz das Selbstverständnis, die Methoden und die gesellschaftliche Rolle der Regionalsprachenforschung insgesamt?

Aus der Standortbestimmung lassen sich die **Herausforderungen, Desiderate und Entwicklungsziele** erschließen, denen sich die Regionalsprachenforschung aktuell gegenüberstellt. Auf diese identifizierten Desiderate reagiert das Positionspapier schließlich mit einer Reihe von **Leitlinien** und **Handlungsempfehlungen**. Diese richten sich an unterschiedliche Akteure: alle Forschenden und Lehrenden in der Regionalsprachenforschung in einem weiten Sinne und den an ihr beteiligten und mit ihr kooperierenden Fächern wie der Computerlinguistik und der Informatik (s. Fußnote 2), die Fachgesellschaften, die Forschungs- und Fördereinrichtungen sowie die Wissenschaftspolitik.

## 2. Standortbestimmung

### 2.1 Forschungsdaten

Die Regionalsprachenforschung verfügt über einen enorm umfangreichen, heterogen angelegten Datenbestand aus über 200 Jahren Fachtradition. Der Bestand umfasst u. a. Sprachaufnahmen (z. B. aus Sprachatlas- und Sprachdokumentationsprojekten), Texte (z. B. Textsammlungen, Social-Media-Korpora), strukturierte Textdaten (z. B. Wörterbücher, Grammatiken, Datenbanken von Erhebungen und Online-Umfragen) und Bilder (z. B. Sprachkarten, Scans von Fragebögen). Das besondere Potenzial dieser Daten für die KI-unterstützte Forschung liegt in ihrer räumlichen, zeitlichen und in weiteren sozial-kulturellen Dimensionen umfassenden Reichweite: Die Daten ermöglichen Einblicke in historische Entwicklungen ebenso wie in die aktuelle Dynamik von Sprachwandel. Weiterhin dokumentieren sie die regionalen Sprachformen und -strukturen, die Repräsentationen für die regionale Identität der Sprechenden darstellen. Aufgrund ihrer Heterogenität und Vielschichtigkeit sind sie wertvolle Ressourcen für die Modellierung sprachlicher Variation. Entscheidend ist dabei weniger die bloße Masse der Daten, sondern deren Qualität; gerade im digitalen Kontext wird aber beides benötigt: Standardisierte Datenformate, einheitliche Metadaten und hochwertige Annotationen sind unverzichtbar, um Vergleichbarkeit und Nachnutzbarkeit bei höchst diversen Sprachformen zu gewährleisten. Hier liegt ein aktuelles Problem: Trotz der großen Potenziale liegen die regionalen Forschungsdaten weder in ausreichender Quantität noch in hinreichender Qualität vor, um für KI-basierte Anwendungen dienlich zu sein.

Die Defizite im Hinblick auf die gesamte Menge an verfügbaren regionalsprachlichen Forschungsdaten sind klar benennbar: fehlende Digitalisierung analoger Daten und Materialien, komplexe rechtliche Rahmenbedingungen (insbesondere im Kontext älterer Datenbestände), unzureichende Annotationen der Daten, oft mangelnde Normierung und Standardisierung der bisherigen digitalen Praxis und geringe Beachtung der FAIR-Prinzipien (Findable, Accessible, Interoperable, Reusable). Was fehlt, sind

große Mengen an FAIRen Daten, die rechtssicher zugänglich und nutzbar sind und in hoher Qualität aufbereitet wurden. Dazu gehören die Digitalisierung und Aufbereitung historischer Bestände ebenso wie die systematische Sammlung gesprochener Alltagssprache, aber auch eine umfassende Erschließung der Metadaten und offene Lizenzierung. Ohne eine deutliche Steigerung des Umfangs FAIRer und hochwertiger Datenbestände bleibt die Regionalsprachenforschung in der KI-Anwendung randständig. Ihre Daten müssen unter freien Lizenzen zugänglich und KI-fähig werden, indem sie transparent und interoperabel erschlossen, aufbereitet und distribuiert werden. Zugleich muss der Schutz von Urheber- und Persönlichkeitsrechten gewahrt bleiben, digitale Selbstbestimmungsrechte und die Datenhoheit gegenüber kommerziellen Akteuren müssen gewährleistet sein. Nur wenn diese allgemeinen Standards der Digital Humanities für regionalsprachliche Forschungsdaten erreicht werden, können sie für KI-Anwendungen nutzbar gemacht werden und zu einer höheren Repräsentation von Regionalität führen.

## 2.2 Werkzeuge

KI-basierte Werkzeuge bzw. Modelle versprechen, zentrale Arbeitsprozesse auch der Regionalsprachenforschung – Datendigitalisierung, Datenaufbereitung, Datenanalyse und Datengenerierung – in erheblichen Umfang zu automatisieren und damit einerseits Forschung produktiver bzw. kostengünstiger zu gestalten und andererseits neue Erkenntnismöglichkeiten zu schaffen. Die Sprachwissenschaft verfügt bereits heute über eine breite Palette an KI-basierten Werkzeugen. Systeme wie Whisper oder Kaldi haben sich für die wichtige Aufgabe der automatischen Spracherkennung (ASR) bewährt, Annotationstools wie WebMAUS oder Prodigy unterstützen bei der Segmentierung, während Modelle wie BERT oder SBERT komplexe sprachliche Analysen ermöglichen. Für klassifikatorische Aufgaben (der Standardsprache) werden zudem vermehrt LLMs eingesetzt. Dazu gehören sowohl Open-Weight-Modelle, die an universitären Rechenzentren datenschutzkonform betrieben werden, als auch Modelle kommerzieller Anbieter, die beide über Schnittstellen angesprochen werden können. Echte Open-Source-Modelle, die volle Datentransparenz gewährleisten und auch ihre Trainingsdaten und ihren Trainingscode veröffentlichen, sind äußerst rar. Zur Verfügung stehen Spezialanwendungen sowie kollaborative Plattformen. Diese Werkzeuge bilden eine tragfähige Grundlage, wurden jedoch primär auf Basis von standardsprachlichen Daten entwickelt. Gleiches gilt für Systeme zur Texterkennung und Transkription. Um die Bedarfe der Regionalsprachenforschung zu erfüllen, müssen solche Werkzeuge gezielt auf regionale Varietäten zugeschnitten werden: Sprachvariation muss systematisch durch große Mengen von Trainingsdaten abgebildet werden, um damit Sprachmodelle zu trainieren, die auch die regionale Vielfalt berücksichtigen. Weiterhin müssen neue Verfahren und Herangehensweisen entwickelt werden, die auch bei beschränkten Datenmengen gezielt eine variationssensible Abbildung des Sprachgebrauchs erreichen.

Die größten Defizite betreffen bislang Bereiche, die für die Regionalsprachenforschung besonders wichtig sind: semantische und pragmatische Analysen, prosodische Indizierungen und Lautanalysen sowie präzise Annotationen einschließlich Transkriptionen. Auch bei der Automatisierung von Routineaufgaben, der Interoperabilität und bei der Verfügbarkeit standardisierter Metadaten bestehen erhebliche Lücken. Damit verbunden ist eine eingeschränkte Analysequalität, die insbesondere Dialekte betrifft, die in bestehenden Trainingsdaten nur unzureichend repräsentiert sind.

Zentrale Herausforderungen liegen folglich in der Anpassung vorhandener Werkzeuge an regionale Sprachdaten, im Aufbau ausgewogener und repräsentativer Trainingskorpora sowie in der Entwicklung verbindlicher Standards für Formate, Annotationen und Metadaten. Ohne diese Maßnahmen bleibt die Vergleichbarkeit der Ergebnisse begrenzt, und die Potenziale der Werkzeuge können nicht ausgeschöpft werden. Zudem müssen geeignete Evaluationsmethoden und Benchmarks nicht nur hinsichtlich allgemeiner Leistungen in regionalsprachlichen Daten entwickelt werden, sondern auch in Bezug auf spezifische Leistungen wie z. B. Evaluierungen zur Passgenauigkeit regionalsprachlicher Textgenerierung oder Kleinräumigkeit von regionalen Sprachmodellen. Hinzu kommt die Notwendigkeit, Tools, Codes, Trainingsdaten und Informationen zu deren Zusammensetzung nachhaltig, datenschutzkonform und kollaborativ verfügbar zu machen, um eine langfristige Nutzung und Weiterentwicklung der Werkzeuge zu sichern und methodische Innovation zu fördern.

Die Priorität liegt nicht nur auf der Neuentwicklung, sondern auch auf der spezifischen Anpassung bestehender Open-Weight-Modelle und Open-Source-Werkzeuge. Entscheidend dafür ist die Verfügbarmachung spezifischer Trainingsdaten. Erst durch den offenen Zugang zu den zentralen Sprachdaten des Fachs ist eine gezielte Entwicklung regionaler Sprachmodelle möglich. Grundvoraussetzung ist hierfür eine enge und vertrauensvolle Kooperation der Forschenden – nicht nur innerhalb der Regionalsprachenforschung, sondern auch mit der KI-Forschung und Computerlinguistik.

### 2.3 Infrastruktur

Die auch für die Regionalsprachenforschung nutzbare digitale Infrastruktur ist in den vergangenen Jahren zu einem tragfähigen Fundament herangewachsen. Archive, Datenzentren und digitale Repositorien bilden heute die zentrale Basis für die Sammlung, Sicherung, Archivierung und Bereitstellung sprachlicher Ressourcen (z. B. AGD, LinguRep, ARCHE, REDE). Mit der Einbindung in die Nationale Forschungsdateninfrastruktur (NFDI) über das Konsortium Text+ ist die für regionalsprachliche Belange relevante Forschungsdateninfrastruktur in übergreifende Strukturen der Digital Humanities in Deutschland integriert. Ähnliches gilt für z. B. Österreich und die Schweiz über die nationalen CLARIN-Konsortien oder für Südtirol über Eurac Research.

Auch wenn die Einrichtungen bereits etabliert sind, ist auch weiterhin ein kontinuierlicher Ausbau erforderlich, um den wachsenden Anforderungen der Forschung gerecht zu werden. Die Infrastruktur ist konsequent interoperabel zu gestalten, durch

### *Fallbeispiel 1 – Pipeline zur Aufbereitung von Tonaufnahmen*

**Hintergrund:** Teil der Grundlagenforschung ist die phonetische Analyse der Lautsysteme der Regionalsprachen. Erst diese ermöglicht Erkenntnisse über Vokalräume, Lautdistinktionen und Lautvariation und -wandel. Dafür ist ein groß angelegtes Korpus gesprochener regionaler Varietäten erforderlich. Die manuelle Aufbereitung hunderter Stunden Audiomaterial ist jedoch kaum leistbar.

**Einsatz von KI:** KI-basierte Pipelines ermöglichen erstmals eine effiziente Verarbeitung: Sprecherdiarisierung, standardnahe und phonetische Transkription, Segmentierung mit Zeitstempeln sowie weitere Annotationen können automatisiert und in hoher Qualität durchgeführt werden und dienen als Grundlage für weiterführende Analysen. Modelle wie Whisper haben Standards gesetzt, stoßen bei Dialekten jedoch rasch an ihre Grenzen. Forschungsnahe Alternativen wie CrisperWhisper oder Implementierungen in Huggingface-Umgebungen bieten die Chance, Varietäten besser abzubilden, lokale Rechenzentren zu nutzen und damit Datenschutz sowie digitale Souveränität zu sichern. Durch die linguistische Bearbeitung von Trainingsdaten und die analytische Evaluation von Output bringen sich die Forschenden aus der Regionalsprachenforschung aktiv im Entwicklungsprozess von KI-Tools ein und bereichern die Kooperation mit der KI-Forschung durch die linguistische Bewertung, Methodik und Interpretation der Ergebnisse.

**Potenziale:** Der Nutzen reicht weit über die technische Entlastung hinaus: Routinetätigkeiten werden reduziert, Forschende können sich stärker auf die inhaltliche Analyse konzentrieren, Hilfskräfte können gezielter eingesetzt werden. Über KI-Systeme können dialektale Feinheiten erkannt werden, die in manueller Arbeit oft übersehen würden, und die damit robuste Grundlagen für Spracherkennungsmodelle jenseits der Standardsprache liefern. Damit könnten ein Korpusvolumen und eine Datenqualität erreicht werden, die ohne KI nicht denkbar wären.

**Rahmenbedingungen & Risiken:** Risiken ergeben sich aus der fehlenden Kontrolle der Analysealgorithmen. Neben ungenauen oder falschen Segmentierungen besteht die Gefahr darin, dass KI-erzeugte fehlerhafte Klassifikationen unkontrolliert übernommen oder dialektale Merkmale künstlich erzeugt bzw. „halluziniert“ werden. Umso wichtiger ist eine transparente Evaluation, die Kombination automatischer Verfahren mit menschlicher Kontrolle und die Sicherstellung datenschutzkonformer Infrastrukturen. Ein solcher Ansatz schafft die Grundlage für eine zentrale, sichere Plattform zur Erstellung großer Sprachkorpora. Er unterstützt die Entwicklung innovativer Sprachtechnologien und trägt zugleich zum Erhalt kultureller Vielfalt bei.



standardisierte Metadaten und Schnittstellen anschlussfähig zu halten und für den Austausch mit anderen (interdisziplinären) Infrastrukturen zu öffnen. Notwendig sind die konsequente Standardisierung von Formaten und Metadaten, die Implementierung von Schnittstellen sowie die Einbindung in übergreifende digitale Forschungsinfrastrukturen. Für die Entwicklung von regionalen Sprachmodellen ist zudem der Zugang zu Hochleistungsservern erforderlich. Auch eine verstärkte Einbindung der fachwissenschaftlichen Infrastrukturen, Ressourcen und Standards in europäische Forschungsinfrastrukturen wie CLARIN-ERIC oder die European Open Science Cloud (EOSC) würde die internationale Sichtbarkeit und Kooperation sowie die Förderfähigkeit im Rahmen von Horizon Europe erhöhen.

Allerdings müssen die Infrastrukturen und Repositorien den Forschenden in ihren konkreten Nutzungsmöglichkeiten bekannt sein und aktiv von ihnen genutzt werden. Nur wenn sie ihre Forschungsdaten etwa in offenen Formaten publizieren, sind diese für die Weiterentwicklung von KI-Tools verfügbar. Hierfür müssen Angebote für rechtliche und technische Beratungen eingerichtet werden, Standards für Forschungsdaten müssen transparent und verständlich kommuniziert werden. Neben den sprachlichen Primärdaten auch Programmcodes und Skripte zu veröffentlichen, muss eine Selbstverständlichkeit werden. Voraussetzung dafür ist eine Veränderung der Fachkultur hin zu mehr Sharing und Transparenz.

Weiterhin darf der Ausbau von Infrastrukturen nicht zu einer zusätzlichen ökologischen Belastung werden. Die energieaufwändige Bereitstellung und Verarbeitung von Daten in Repositorien und Hochleistungsrechenzentren erhöht den Ressourcenbedarf der Wissenschaft erheblich und braucht klimaschonende und nachhaltige Lösungen.

Damit lässt sich zusammenfassen: Digitale Infrastrukturen sind für die Regionalsprachenforschung vielfältig vorhanden und funktionsfähig, ihr Potenzial kann aber nur dann vollständig ausgeschöpft werden, wenn sie durch Standardisierung, Interoperabilität und Kooperation gezielt ausgebaut, erweitert und geöffnet werden. Neue Initiativen wie das vom BMFT im Rahmen der Hightech-Agenda in Aussicht gestellte Akademienzentrum zur Digitalen Lexikographie sind vielversprechend.

## 2.4 Kompetenzen

Als empirische Wissenschaft hat die Regionalsprachenforschung in den letzten zwei Jahrzehnten eine deutliche datenwissenschaftliche Ausrichtung angenommen. Diese Entwicklung wurde durch digitale Analyseverfahren benachbarter Disziplinen, etwa der Computerlinguistik, Phonetik, Soziolinguistik oder Geographie, beschleunigt. Statistische Analysen und andere quantitative Verfahren gehören mittlerweile zum etablierten Methodenrepertoire der regional orientierten Forschung. Im Gegensatz dazu lässt sich feststellen, dass Forschende in diesem Bereich bislang eher nicht bzw. nur eingeschränkt über methodische KI-Kompetenzen verfügen. Auch grundlegende Kenntnisse im Natural Language Processing (NLP) sind fachwissenschaftlich nicht übergreifend vorauszusetzen. Hier fehlt eine größere Offenheit des Fachs für entsprechende Metho-

den, aber auch eine Integration in die germanistisch-linguistischen Studiengänge sowie eine fachwissenschaftliche Qualifizierung. In der Computerlinguistik, wo entsprechende methodische Kompetenzen vermittelt werden, spielen wiederum regionalsprachliche Inhalte bisher eine nur untergeordnete Rolle. Eine linguistische Grundausbildung findet in der Informatik nicht statt, wodurch insgesamt eine ungleiche und dispartete Kompetenzverteilung in der Kooperation der Disziplinen entsteht.

Eine nachhaltige Integration KI-basierter Verfahren erfordert die Zusammenführung unterschiedlicher Expertisen – nicht allein durch interdisziplinäre Kooperation, sondern ebenso durch integrative Ausbildungsangebote in der akademischen Lehre. Die spezifischen Förderformate zur Stärkung der Datenkompetenzen des wissenschaftlichen Nachwuchses und die Einrichtung von Datenkompetenzzentren in den deutschsprachigen Ländern stellen bereits wertvolle, wenn auch nur vereinzelte und befristet geförderte Unterstützungsangebote dar.

Neben dem Erwerb neuer digitaler Kompetenzen ist aber auch der Erhalt „klassischer“ fachlicher, insbesondere dialektologischer Expertise in der Regionalsprachenforschung unverzichtbar, um eine präzise sprachwissenschaftliche Analyse und kritische Bewertung KI-basierter Ergebnisse zu gewährleisten. Die zunehmende Automatisierung von Forschungsprozessen durch KI-Methoden birgt das Risiko, dass linguistische Analyseschritte nur noch in Ansätzen beherrscht werden und es insgesamt zu einem sprachwissenschaftlichen Kompetenzverlust bei Forschenden und Studierenden kommt. Regionalsprachliche und dabei v. a. dialektologische Inhalte müssen weiterhin Kernbestand einer germanistisch-linguistischen Ausbildung bleiben.

Fachübergreifend ist eine Auseinandersetzung mit wissenschaftstheoretischen Grundlagen erforderlich, um in den gesellschaftlichen und wissenschaftspolitischen Diskursen zu den Chancen und Herausforderungen von KI-Verfahren differenziert, kritisch und sprachlich exakt Stellung beziehen zu können.

## 2.5 Kooperationen

Punktuell hat die produktive Zusammenarbeit von KI-Forschung bzw. Computerlinguistik und Regionalsprachenforschung bereits stattgefunden. Bisher beruht eine solche Kooperation jedoch vor allem auf dem Engagement und der Vernetzung einzelner Forschender. Institutionell geförderte Zusammenarbeit – etwa durch die Einrichtung von Special Interest Groups oder durch Kooperationen zwischen sprachwissenschaftlichen und informatischen Fachgesellschaften – ist bislang nicht breit etabliert.

Wie produktiv die Vernetzung sein kann, zeigt die seit mehr als zehn Jahren bestehende Workshopreihe VarDial, in der NLP-basierte Analyseverfahren für Dialekte diskutiert werden und die regelmäßig im Rahmen internationaler Tagungen zur Computerlinguistik (bei ACL, der Association for Computational Linguistics) stattfindet. VarDial ist jedoch eine klare Ausnahme; der Aufbau ähnlicher Kooperationen wie auch die Integration solcher methodischer Sektionen in die einschlägigen Fachgesellschaften und Konferenzen steht bis auf wenige Ausnahmen (e.g. ICLaVE 13) noch aus. Dabei

erweist sich die interdisziplinäre Kooperation nicht nur für die Regionalsprachenforschung als gewinnbringend, auch die Informatik profitiert von der Zusammenarbeit, da die sprachliche Variation und Dynamik für theoretische wie methodische Fragestellungen gleichermaßen zentrale Herausforderungen in der KI-Forschung darstellen.

Kooperation ist jedoch nicht nur innerhalb der Wissenschaft erforderlich, sondern auch mit der Öffentlichkeit im Allgemeinen und den Dialektsprecherinnen und -sprechern im Besonderen. Mit Citizen Science steht ein etablierter Forschungsansatz zur Verfügung, der die Bürgerbeteiligung bei der Generierung, Qualitätssicherung und Verbreitung von Wissen fördert und dadurch auch die Beteiligung und Mitbestimmung der Bevölkerung an gesellschaftlich relevanten Fragestellungen ermöglicht. In der Regionalsprachenforschung wird die Bevölkerung traditionell in die Erhebung von Sprachdaten eingebunden und vielerorts ist ein Austausch von Forschenden mit Dialektvereinen oder interessierten Einzelpersonen etabliert. Vorhaben, die eine strukturierte Einbindung der Öffentlichkeit in die Forschungsprozesse und -diskurse anstreben und neben Datenerhebung auch die Evaluierung von Sprachdaten anzielen, stehen jedoch noch am Anfang. Die strukturierte Einbindung der Öffentlichkeit kann nur über Internetplattformen stattfinden, die speziell auf die bei der interessierten Gruppe vorauszusetzenden Kompetenzen zugeschnitten sind. Traditionelle dialektologisch-linguistische Arbeitsabläufe, Instrumente und Ressourcen sind der Öffentlichkeit weder zugänglich noch vermittelbar.

Eine verstärkte Kollaboration ist auch mit Akteuren der Wissenschaftspolitik notwendig. In der öffentlichen Wahrnehmung besteht oft das Missverständnis, dass KI-Verfahren die Arbeit der Forschenden ersetzen werden. Stattdessen muss die Entwicklung und Anwendung von KI-Verfahren auch in der Regionalsprachenforschung durch fachliche Expertise abgesichert und reflektiert werden, um eine angemessene Repräsentation und Authentizität der regionalen Sprache und die Akzeptanz durch die Sprecherinnen und Sprecher zu gewährleisten. Daraus folgt, dass sich die Regionalsprachenforschung verstärkt und aktiv in die politischen und gesellschaftlichen Diskurse über KI einbringen muss.

## 2.6 Förderung und Institutionalisierung

Die einschlägigen Förderinstitutionen haben auf die digitale Transformation der Wissenschaften mit spezifischen Förderlinien, Schwerpunktbildungen und aktiven Netzwerken für Entwicklung, Training und Forschung in den Digital Humanities reagiert. Hier scheint jedoch Deutschland weiter vorangeschritten zu sein als z. B. Österreich oder die Schweiz, wo entsprechende Förderprogramme für die Geisteswissenschaften bislang nicht eingerichtet sind.

An den verfügbaren Förderungen partizipiert die Regionalsprachenforschung bereits aktiv, z. B. im NFDI-Konsortium Text+, an verschiedenen DFG-Förderlinien oder BMFT-Förderschwerpunkten. Allerdings stehen die Förderbedingungen dem Aufbau einer nachhaltigen und langfristigen Infrastruktur für Forschungsdatenmanagement, Kooperation und Kompetenzentwicklung entgegen: Die Infrastrukturförderungen

sind auf wenige Jahre befristet. So wurden bisher weder die NFDI-Infrastruktur noch die Datenkompetenzzentren verstetigt, die Förderlinie zum Aufbau von „Datenkorpora für Künstliche Intelligenz“ sieht lediglich eine Förderzeit von zwei Jahren vor. Diese kurzen Förderzeiträume mit fehlender Verstetigungsperspektive sind nur bedingt geeignet, um Expertise und Datenmanagementsysteme langfristig aufzubauen und entsprechend qualifiziertes Personal in der Wissenschaft zu halten. Für Dialekte und andere Low-Resource-Sprachen stellt sich dieses Problem noch drastischer als für Standardsprachen. Weiterhin schränken die Förderbedingungen den Zugang zu den Förderungen ein: So adressiert die DFG-Förderinitiative „Künstliche Intelligenz“ nur Emmy-Noether-Nachwuchsgruppen oder Forschungsgruppen. Die LIS-Förderlinien der DFG sehen in der Regel einen beachtlichen Eigenanteil der Antragstellenden vor und setzen die Beteiligung von Bibliotheken oder Archiven voraus – Anforderungen, die von vielen Forschenden und Institutionen nicht erfüllt werden können.

Es zeigt sich, dass es aktuell vor allem an offenen und niedrigschwelligen Förderlinien fehlt, die es den Forschenden ermöglichen, KI-Kompetenzen aufzubauen, Kooperationen anzustreben, Forschungsdaten KI-geeignet aufzubereiten und entsprechend den fachlichen und technischen Standards zu publizieren. Dieser Bedarf ist insbesondere vor dem Hintergrund einer sinkenden Grundfinanzierung der Universitäten umso drängender.

Weiterhin sind spezifische Förderlinien für Citizen-Science-Vorhaben erforderlich, um eine gesellschaftliche Partizipation zu ermöglichen. Auf diese Weise könnten z. B. Dialekt sprecherinnen und -sprecher verstärkt in die Erhebung, Aufbereitung und Evaluierung sprachlicher Ressourcen sowie in den Diskurs über die Repräsentation von Regionalität in KI-erzeugter Sprache eingebunden werden. Gerade bei Befragungen zu regionaler Sprache und Identität sind Bürgerinnen und Bürger in hohem Maße bereit, sich einzubringen. Eine Einbindung ist jedoch von Seiten der Wissenschaft personalintensiv und erfordert eigenständige Förderinstrumente, die über die bisherigen Förderformate (z. B. Citizen Science Award des OeAD, universitäre Seed Grants, beantragbare Module für Wissenschaftskommunikation) hinausgehen.

Zur fehlenden ganzheitlichen Institutionalisierung trägt auch die Kulturpolitik der Länder bei. Hier zeigt sich eine uneinheitliche Förderlandschaft. Während einzelne Staaten und Gliedstaaten (z. B. Bundesländer) Dialektpflege institutionell absichern und mitunter als Staatsziel verankern, fehlt andernorts eine kontinuierliche Förderung regionaler Identität gänzlich. Uneinheitlichkeit wird dabei über politische Rahmenbedingungen forciert: Während das Niederdeutsche eine im Sinne der Europäischen Charta der Regional- oder Minderheitensprachen anerkannte Regionalsprache ist und damit einen besonderen Schutz genießt, trifft dies auf die hochdeutschen Varietäten nicht zu.

### 3. Herausforderungen und Desiderate

Die Standortbestimmung hat folgende Kernprobleme aufgezeigt, die den übergeordneten Zielen der Regionalsprachenforschung entgegenstehen:

#### Verfügbarkeit von Forschungsdaten und -infrastrukturen

- Unzureichende Forschungsdaten sowohl im Hinblick auf Quantität und Ausgewogenheit als auch im Hinblick auf Qualität
- Mangelnde Einhaltung der FAIR-Standards bei bestehenden Sprachdaten
- Unterrepräsentation regionaler Varietäten in bestehenden Sprachmodellen
- Unzureichende Anpassung vorhandener Tools und Infrastrukturen an regionale Sprachformen, fehlende spezifische Benchmarks für regionale Sprache
- Rechtliche Unsicherheiten (Datenschutz, Lizenzierung) bei der Datenpublikation
- Fehlende langfristige Förderstrukturen
- Insbesondere in Österreich und der Schweiz: fehlende Förderlinien für Infrastrukturprojekte

#### Kooperation und fehlende Kompetenzentwicklung

- Kaum systematische Kooperation zwischen Regionalsprachenforschung und KI-Forschung
- Disparate Kompetenzverteilung bei Forschenden aus Regionalsprachenforschung, Computerlinguistik und KI-Forschung
- Fehlende integrative Ausbildungsformate
- Geschlossene und intransparente Fachkultur: fehlende Etablierung von Sharing, Transparenz, Kollaboration und Risikobereitschaft
- Fehlende interdisziplinäre Fachgesellschaften, Netzwerke und Austauschformate

#### Beteiligung an Diskursen

- Geringe Beteiligung von Forschenden an den politischen, wissenschaftspolitischen und gesellschaftlichen Diskursen über KI
- Fehlende systematische Einbindung der Zivilgesellschaft in die Forschungsprozesse
- Geringe Wissenschaftskommunikation zur KI-basierten Forschung in der Regionalsprachenforschung
- Fehlende Förderlinien für partizipative Forschung in Form von Citizen Science, in der Bürgerinnen und Bürger an den wissenschaftlichen Prozessen und Diskursen beteiligt werden

Aus den genannten Herausforderungen und Desideraten heraus ergeben sich **Leitlinien** für die Regionalsprachenforschung. Sie bilden die Grundlage für die Weiterentwicklung des Faches und fokussieren die notwendigen Rahmenbedingungen, um die übergeordneten Ziele – Potenziale von KI nutzen, kulturelle Vielfalt erhalten und digitale Transformation gestalten – zu erreichen.

### *Fallbeispiel 2 – Grammatische Muster in Texten erkennen*

**Hintergrund:** Regionale Sprachvariation ist das Ergebnis historischer Sprachwandelprozesse im Zusammenspiel mit inner- und außersprachlichen Faktoren. Komplexe grammatische Muster in großen heterogenen Textkorpora zu identifizieren, zu klassifizieren und in ihrem historischen Wandel nachzuzeichnen, gibt uns Aufschluss über sprachkognitive und sozio-kommunikative Prozesse, die Teil der Kulturgeschichte der Regionen sind.

**Einsatz von KI:** Bisher waren sprachraum- und sprachstufenübergreifende korpuslinguistische Analysen aufgrund der enormen Datenmengen und der Vielfalt sprachlicher Strukturen praktisch nicht realisierbar. Erst durch KI-gestützte Verfahren wird es möglich, regionale Strukturen in Morphologie und Syntax systematisch zu erfassen und deren Entwicklung über Jahrhunderte hinweg zu quantifizieren. Dies setzt jedoch vielfach technisches Spezialwissen voraus. Abhilfe können KI-basierte Analysesysteme und Analysemodelle leisten, die den gesamten Forschungsprozess, von der Formulierung komplexer Suchanfragen über die Fehlerprüfung und Nachannotation bis hin zur Mustererkennung und Klassifikation, unterstützen. Anforderungen an solche Systeme sind die Einbindung bestehender Werkzeuge sowie die Auswahl geeigneter Methoden, um so linguistische Forschung effizienter und breiter zugänglich zu machen, während die Forschenden die Ergebnisse kontrollieren und validieren.

**Potenziale:** Der Nutzen besteht darin, dass es durch die Leistungsfähigkeit von KI-basierten Verfahren möglich ist, bislang unentdeckte Muster und Strukturen in den Daten effizient zu erschließen. Das eröffnet neue Erkenntnisse über Variation und Wandel und macht hochspezialisierte Analysen auch ohne technisches Spezialwissen nutzbar. Damit entsteht nicht nur ein Innovationsschub für die Sprachwissenschaft, sondern auch ein Beitrag zur Sichtbarkeit regionaler Identität und ein niedrigschwelliger Zugriff auf das kulturelle Erbe Sprache. Für die KI-Forschung stellen sich aufgrund der historischen Sprachdaten besondere Herausforderungen, und es ergeben sich interessante Fallbeispiele für die Entwicklung von Analyse- und Klassifikationsverfahren.

**Rahmenbedingungen & Risiken:** Erforderlich für die Entwicklung entsprechender KI-Systeme sind qualitativ hochwertige und tief annotierte Trainingsdaten, in denen die unterschiedlichen Sprachlandschaften angemessen repräsentiert sind. Die Entwicklung der Tools erfordert nicht nur offene Formate und Interoperabilität sowie die entsprechende technische Infrastruktur, sondern auch eine enge Zusammenarbeit von Forschenden aus der KI-Forschung und Regionalsprachenforschung. Risiken bestehen in der Validität und Reliabilität der Ergebnisse, die entscheidend für die Akzeptanz im Fach sein werden.

#### 4. Leitlinien

##### **FAIRe Daten in hoher Qualität.**

Die Grundlage jeder nachhaltigen Kooperation von Regionalsprachenforschung und KI-Forschung ist der Zugang zu umfangreichen und hochwertigen regionalen Daten. Erforderlich ist eine konsequente Umsetzung der FAIR-Prinzipien: Die Daten müssen auffindbar, für die wissenschaftliche Nutzung zugänglich, interoperabel und nachnutzbar zur Verfügung gestellt werden. **Leitlinie ist es daher, Daten nicht nur aktiv verfügbar zu machen, sondern sie so nach internationalen DH-Standards aufzubereiten, dass sie nachhaltig und für KI-Anwendungen unmittelbar als Trainings- und Evaluationsdaten (Benchmarks) nutzbar sind.**

##### **Datenhoheit.**

Bei der Bereitstellung von regionalen Sprachdaten müssen Urheber- und Persönlichkeitsrechte gewahrt bleiben, insbesondere gegenüber kommerziellen Akteuren, die ihre Trainingsdaten bislang nicht ausreichend ausweisen. Wissenschaftspolitische und juristische Initiativen, die eine rechtssichere Veröffentlichung und Nutzung von Daten fördern, sind zu unterstützen und zu berücksichtigen. **Leitlinie ist die Gewährleistung von möglichst offenen Lizenzierungen bei gleichzeitig gewährten Persönlichkeits- und Urheberrechten.**

##### **Lizenzierung und Nutzungsklarheit.**

Open-Data-Strategien sind zentral für wissenschaftliche Innovation, zugleich erfordert der Einsatz von KI-Systemen eine verantwortungsvolle Steuerung der Datennutzung. Zwischen der gebotenen Offenheit gemäß FAIR-Prinzipien und dem Schutz vor unkontrollierter kommerzieller Verwertung entsteht ein Spannungsfeld, das nur durch transparente, eindeutig formulierte und technisch wie juristisch belastbare Lizenzierungspraktiken aufgelöst werden kann. **Leitlinie ist daher eine Lizenzierung, die wissenschaftliche Nachnutzbarkeit maximal erleichtert, zugleich aber klare und maschinenlesbare Nutzungsvorgaben formuliert, die den Schutz sensibler Sprachdaten gegenüber proprietären KI-Anwendungen gewährleisten.**

##### **Regionale Sprachmodelle.**

Dialekte und andere regionale Varietäten müssen in Sprachmodellen aus Sicht der Sprecherinnen und Sprecher authentisch repräsentiert sein. Notwendig sind spezifische Sprachmodelle, die die Eigenheiten regionaler Sprachformen erfassen und sichtbar machen. Hierzu müssen bestehende Tools gezielt angepasst und neue Modelle entwickelt werden. **Leitlinie ist die Verbindung linguistischer Expertise mit technologischer Innovation.**

##### **Infrastruktur.**

Infrastrukturen müssen „KI-ready“ sein. Die vorhandene Infrastruktur ist tragfähig, aber nur eingeschränkt kompatibel mit KI-Systemen. Notwendig ist eine systematische Öffnung für KI-Anwendungen durch Bereitstellung von Daten über API-Schnittstellen

sowie Schnittstellen zu übergreifenden Forschungsinfrastrukturen und leistungsstarken Rechenzentren. Ebenso entscheidend sind Kompetenzzentren, die den rechtlichen Rahmen, Best Practices und technische Standards vermitteln. **Leitlinie ist der Ausbau und die Verstetigung der bestehenden Infrastruktur, die zugleich verlässlich, nachhaltig und innovativ ist.**

### **Kooperation.**

Regionalsprachenforschung und KI-Forschung müssen gleichberechtigt kooperieren. Die Herausforderungen lassen sich nur im Verbund meistern. Dies umfasst gemeinsame Workshops, interdisziplinäre Forschungsprojekte, wechselseitige Kompetenzvermittlung und die Etablierung dauerhafter institutioneller Schnittstellen. **Die Leitlinie ist daher: Wissenstransfer verstetigen, individuelle Expertise zusammenführen und durch institutionalisierte Kooperation neue Forschungsfelder erschließen.**

### **Doppelter Kompetenzerwerb.**

Sprachwissenschaftliche Kompetenz und KI-Kompetenz müssen zusammengeführt werden. Regionalsprachliche Expertise bildet das Fundament, doch ohne KI-Kompetenz droht sie im digitalen Raum wirkungslos zu werden. Umgekehrt führt reine KI-Kompetenz ohne linguistisches Wissen zu oberflächlichen und verzerrten Ergebnissen. **Leitlinie ist daher ein doppelter Kompetenzerwerb: der Erhalt und die Stärkung der regionalsprachlichen, insbesondere der dialektologischen Fachausbildung sowie der systematische Aufbau von KI-Know-How für Forschende und Studierende in der Regionalsprachenforschung.**

### **Fachkultur.**

Open Science muss das übergeordnete Leitprinzip sein. Eine integrative Fachkultur setzt auf Transparenz, offene Bereitstellung von Daten, Tools und Ergebnissen sowie auf gemeinsames Lernen, auch aus nicht erfolgreichen Projekten. Daher braucht es eine verstärkte Anerkennung von entsprechenden Leistungen im akademischen Lebenslauf. Honoriert werden müssen der Mut zum Experimentieren, das Zulassen von Risiken, das Testen neuer Wege, die Schaffung von Freiräumen für Innovation. **Leitlinie ist die Abkehr von einer reinen Ergebnisorientierung hin zu einer prozessorientierten Wissenschaftskultur, die durch Transparenz und kollektives Lernen getragen ist.**

### **Zivilgesellschaft.**

Die Zivilgesellschaft<sup>4</sup> muss als Partner eingebunden werden. Regionale Sprache betrifft nicht nur die Wissenschaft, sondern vor allem die kulturelle Praxis. Vereine, Schulen,

4 „Zivilgesellschaft“ wird im Positionspapier entsprechend der Definition des Bundesministeriums für wirtschaftliche Zusammenarbeit und Entwicklung als präzisierende Erweiterung von „Öffentlichkeit“ verstanden: „Die Zivilgesellschaft umfasst die Gesamtheit des Engagements der Bürgerinnen und Bürger eines Landes – zum Beispiel in Vereinen, Verbänden und vielfältigen Formen von Initiativen und sozialen Bewegungen. Dazu gehören alle Aktivitäten, die nicht profitorientiert und nicht abhängig von parteipolitischen Interessen sind.“ (<https://www.bmz.de/de/service/lexikon/zivilgesellschaft-14976>, 12/2023).



regionale Initiativen und interessierte Bürgerinnen und Bürger sind Träger und Multiplikatoren dialektaler Vielfalt. Ihre Einbindung sichert Kontrolle, Legitimität und gesellschaftliche Wirkung von KI-Technologien. **Leitlinie ist, die Zivilgesellschaft nicht nur als Empfänger, sondern durch Beteiligung, Mitgestaltung und Dialog als Partner in Forschungsprozessen zu begreifen.**

### **Diskurs.**

Fachwissenschaft muss im KI-Diskurs sichtbar sein. Nur wenn die Regionalsprachenforschung die kulturelle Dimension von Sprache in die Debatte einbringt, kann verhindert werden, dass regionale Identität in technikgetriebenen Entwicklungen unsichtbar bleibt. **Leitlinie ist die aktive Mitgestaltung öffentlicher und politischer Debatten über KI, um sprachliche Vielfalt als kulturellen Wert zu verankern und als Resource in die digitale Transformation einzubringen.**

### **Verantwortung.**

Forschende müssen ihre sprachwissenschaftliche Expertise in den öffentlichen Diskurs über die Chancen und Risiken der digitalen Transformation einbringen. Nur eine kritische und differenzierte Diskussion über die Leistungen und Gefahren von KI-Verfahren schärft eine realistische Bewertung und ist Grundlage für soziale Angemessenheit und rechtliche Sicherheit. **Leitlinie ist die Bereitschaft, die eigenen Erfahrungen und Expertisen in den Gremien der akademischen Selbstverwaltung, der Fachinfrastruktur und Wissenschaftspolitik sowie im gesellschaftlichen Diskurs engagiert einzubringen.**

### **Nachhaltigkeit.**

Die Verantwortung der Regionalsprachenforschung umfasst auch den Bereich der gesellschaftlichen und ökologischen Nachhaltigkeit. Aufwand, Kosten und natürliche Ressourcen müssen in einem sinnvollen Verhältnis zum erwartbaren Nutzen der zu entwickelnden Technologien stehen. Entsprechende Entscheidungen sollten auf Grundlage informierter und ausgewogener Aushandlungsprozesse erfolgen. **Leitlinie ist das Bemühen um von Nachhaltigkeit und maßvoller Ressourcennutzung geprägte Forschungsprozesse.**

### *Fallbeispiel 3 – Dialekte verstehbar machen*

**Hintergrund:** Mit zunehmender überregionaler Kommunikation verlieren lokale Basisdialekte an Bedeutung. Weil Dialekt heute oft nicht mehr als Familiensprache weitergegeben wird, geht die Dialektkompetenz in der Bevölkerung zurück. Dadurch werden relevante Teile der öffentlichen Kommunikation, aber auch historisch gewachsene Orientierungspunkte – etwa Mundartliteratur, Briefe früherer Generationen, Haus-, Flur- und Straßennamen – für viele unzugänglich. Das erschwert den Zugang zu lokalem Wissen, mindert die Verständlichkeit des öffentlichen Raums und schwächt regionale bzw. kulturelle Bindung und gesellschaftliche Teilhabe. Die genannten Quellen sind jedoch ein wichtiger Teil des gesellschaftlichen Gedächtnisses und damit Grundlage für eine identitätsstiftende demokratische Öffentlichkeit.

**Einsatz von KI:** Mit der Entwicklung eines Übersetzungstools für Dialekte können regionalsprachliche Texte ohne eigene Dialektkompetenz erschlossen und zugleich standardsprachliche Texte in den gewählten Dialekt übertragen werden. Grundlage ist die Entwicklung eines Sprachmodells, das die regionalen Eigenheiten der Dialekte in geographischer Ausdehnung erfasst und sichtbar macht. Für das Training kann auf bestehende Erhebungsdaten, Wörterbuch- und Grammatikdaten und Paralleltexte zurückgegriffen werden, weitere Daten gilt es systematisch zu erschließen.

**Potenziale:** Dialekte übersetzbar und damit auch für Außenstehende verstehbar zu machen, eröffnet zahlreiche Möglichkeiten, die weit über Anwendungen in der Wissenschaft hinausreichen. Die Einsatzbereiche eines solchen Tools sind vielseitig und reichen von der wissenschaftlichen Analyse dialektaler Ausdrucksformen und ihrer sozio-kommunikativen Funktion über die private und behördliche Nutzung bis hin zum Einsatz im Bildungsbereich. Nicht nur lassen sich historische Dokumente (z. B. Urkunden) übersetzen und der Sprachgebrauch früherer Generationen erschließen, sondern Dialekte werden lern- und reproduzierbar und erhalten – z. B. auch gerade in Sprachinseln – in ganz neuer Form Sichtbarkeit und eine identitätssichernde Relevanz für die Sprachgemeinschaft.

**Rahmenbedingungen & Risiken:** Die Güte eines dialektalen Sprachmodells ist abhängig von der Qualität, Menge und Differenziertheit seiner Trainingsdaten, die in offener und transparenter Weise für das (Weiter-)Trainieren eines Basismodells zur Verfügung stehen müssen. Für die Aufbereitung und Erschließung möglicher Trainingsdaten sind neben leistungsfähigen Modellen zur automatischen Texterkennung (OCR) auch zuverlässige Verfahren für die Handschriftenerkennung (HTR) erforderlich. Der Geltungsbereich der Trainingsdaten muss klar bestimmt werden, um kontextsensible Textausgaben zu erreichen. Archaisierende Formen historischer Dialektstände würden die Akzeptanz bei den Nutzenden schmälern. Weiterhin müssen Richtlinien für den Umgang mit rechtlich oder ethisch problematischen Inhalten entwickelt werden. Die Qualitätssicherung der Modelle muss zwar durch Wissen-

schaftlerinnen und Wissenschaftler gewährleistet werden, profitiert aber in hohem Maße von der Expertise der Dialektsprecherinnen und -sprecher, die – zum Beispiel im Rahmen von Citizen-Science-Projekten – erheblich zur Evaluierung der Trainingsdaten beitragen können.

## 5. Empfehlungen

Aus diesen Leitlinien erfolgen Empfehlungen für die Akteure in wissenschaftlichen Handlungsfeldern in ihren bisweilen unterschiedlichen Rollen:

### 5.1 Empfehlungen an **Fachgesellschaften** in der Regionalsprachenforschung

1. Kooperationen mit Fachgesellschaften der Informatik, Computerlinguistik und Digital Humanities initiieren und disziplinenübergreifende Special Interest Groups einrichten, um den interdisziplinären Austausch zu institutionalisieren
2. Beauftragte für Open Science benennen oder entsprechende Sektionen einrichten, um einen Fachkulturwandel zu fördern
3. Workshops zu Open Data und Shared-Task-Aktionen im Rahmen von Konferenzen veranstalten, um Open Science in der Community stark zu machen
4. Tool-Pools und Methoden-Archive anlegen, um Kompetenzaufbau und Transfer zu erhöhen
5. Die Anliegen des Fachs auch im Hinblick auf Datenhoheit und Schutz von Persönlichkeits- und Urheberrechten in wissenschaftspolitischen Gremien unter Berücksichtigung fachspezifischer Eigenheiten vertreten, um rechtliche Rahmenbedingungen für eine FAIRe Datennutzung zu gewährleisten
6. Kommunikationsformate zu KI und Regionalsprachenforschung initiieren, um die Diskurse um KI und ihre wissenschafts- wie gesellschaftspolitische Relevanz aktiv mitzugestalten

### 5.2 Empfehlungen an **Infrastruktureinrichtungen** und -verbände

1. Zentrale Dienste einrichten, die die Fachgemeinschaft über KI-bezogene Infrastrukturen, Ressourcen, Standards und Services informieren und im Hinblick auf technische und rechtliche Fragen beraten, um den Informationsfluss im Fach zu erhöhen
2. FAIRe Forschungsdaten in Kooperation mit der Fachgemeinschaft kuratieren und langfristig bereitstellen, damit diese als Trainings- und Evaluationsdaten verfügbar sind
3. Forschende bei der Kuratierung und Aufbereitung von Forschungsdaten unterstützen, um die Publikation von FAIRen Daten zu erleichtern

4. Forschungsdatenrepositorien ausbauen, weiterentwickeln und vernetzen, um die notwendige Infrastruktur verfügbar zu halten
5. Datenschutzrechtliche Aspekte klären und transparente Lizenzierungsmodelle entwickeln, um klare Grundlagen für Datenpublikationen zu schaffen und Datenhoheit zu gewährleisten
6. Standardisierte Metadatenschemata für raumbezogene Forschungsdaten ausarbeiten, um darüber bestehende Ressourcen in interoperablen Systemen zu verknüpfen
7. Regionalsprachliche Datenbestände in nationalen und europäischen Katalogen indizieren, um sie grenzüberschreitend auffindbar zu machen
8. Kooperationen mit KI-Zentren eingehen, um für die Entwicklung und Anwendung nicht-proprietärer Sprachmodelle strukturiert nutzbare Rechenkapazitäten bereitzustellen
9. Klimafreundliche Energieversorgung von Hochschulen und KI-Zentren einfordern, um ressourcenschonende Forschung zu stärken
10. Fortbildungsangebote (Schulungen, Schulungsmaterial) bereitstellen, um Kompetenzaufbau zu ermöglichen und Nachwuchsförderung zu stärken

### 5.3 Empfehlungen an **Fördermittelgeber**

1. Förderlinien mit mehrjähriger Förderdauer zur FAIRen Aufbereitung von Sprachkorpora und Forschungsdaten niedrigschwellig gestalten (z. B. durch den Verzicht auf die Einbringung strukturelle Eigenmittel), um den Zugang zu Fördermitteln zu öffnen und zu vereinfachen
2. Förderlinien mit mittelfristiger Förderdauer für impulsgebende Infrastrukturprojekte einrichten, die ausdrücklich auch den initialen Aufbau neuer Sprachkorpora und Datenbestände einschließen, um eine tragfähige Grundlage für zukünftige Dateninfrastrukturen zu schaffen
3. Förderlinien für regionale KIs einrichten, um die Entwicklung varietätensensitiver Sprachmodelle zu beschleunigen
4. Fachspezifische Datenzentren verstetigen, um Langzeitarchivierung und Verfügbarmachung von Daten in Repositorien abzusichern
5. KI-Zentren mit Hochleistungsservern ausbauen und mit entsprechendem Personal für Service und Schulungen ausstatten, um Forschenden den Zugang zu KI-Technologien zu ermöglichen
6. Citizen-Science-Förderlinien einrichten, um die systematische und nachhaltige Einbindung von Bürgerinnen und Bürgern in die wissenschaftlichen Forschungsprozesse zu gewährleisten

#### 5.4 Empfehlungen an **Herausgeberinnen und Herausgeber von Fachpublikationen**

1. Öffnung klassischer sprachwissenschaftlicher Publikationsorgane auch für stärker digital-methodische und transdisziplinäre Beiträge an der Schnittstelle von Regionalsprachenforschung und Informationstechnologie bzw. Computerlinguistik
2. Bei Publikationen stets Supplementary Materials (Angabe von Trainingsdaten, KI-Modellversionen, Prompts, Programmcode, Repositories etc.) einfordern, damit KI-Verfahren unabhängig geprüft, repliziert und weiterentwickelt werden können
3. Special Issues zu KI-Anwendungen in der Regionalsprachenforschung und zu Replikationsstudien ausschreiben, um Austausch und Transparenz zu erhöhen
4. Publikationsmöglichkeiten für Studien mit negativen Ergebnissen schaffen, um methodischen Austausch und Innovationskraft zu stärken

#### 5.5 Empfehlungen für **Gutachten und Entscheidungsgremien**

1. Bei der Bewertung von wissenschaftlichen Leistungen in der Regionalsprachenforschung neue Maßstäbe in Hinblick auf Transparenz und Reproduzierbarkeit entwickeln, um die Fachkultur im Hinblick auf Open Science und Interdisziplinarität zu verändern
2. Die Daten- und Methodenentwicklung im akademischen Lebenslauf anerkennen, um Leistungen im Bereich von Open Data angemessen zu würdigen
3. Förderungen für Vorhaben, die FAIRe Daten bereitstellen, priorisieren, um zügig Datenquantität und -qualität im Fach zu verbessern
4. Kooperationen fördern, um Transdisziplinarität zu stärken

#### 5.6 Empfehlungen an **Forschende**

1. Projekte zur Digitalisierung von Erhebungsmaterialien und Quellen (z. B. Fragebücher, Wörterbuchbelege, Grammatiken) entwickeln, um bestehende analoge Forschungsdaten digital verfügbar zu machen
2. Eigene Forschungsdaten KI-geeignet aufbereiten und in Repositorien zur Verfügung stellen, um zu angemessenen Repräsentationen regionaler Sprache durch vielfältige Daten beizutragen
3. Kerndatenbestände der Regionalsprachenforschung identifizieren und diese als Referenzkorpora „KI-ready“ aufbereiten, um zentrale Trainingsdaten verfügbar zu machen
4. Als Gemeinschaftsleistung regionale Sprachmodelle entwickeln (vgl. 5.3) (z. B. durch Adaption vortrainierter Modelle), um KI-basierte Werkzeuge an die spezifischen Bedarfe anzupassen und Regionalität angemessen zu repräsentieren:
  - a. kurzfristig: erste Prototypen regionaler Sprachmodelle für jeweils eine Region kooperativ entwickeln und veröffentlichen, um methodische Fortschritte schnell verfügbar zu machen

- b. mittelfristig: den LLM-Prototyp für ausgewählte Regionen ausbauen, um so kontinuierlich ein größeres Sprachmodell aufzubauen
- c. langfristig: die Entwicklung eines integrativen variations- und wandelsensitiven multimodalen Sprachmodells der deutschen Regionalsprachen aufnehmen und zielführende Maßnahmen zur Umsetzung entwickeln
- 5. Aktiv interdisziplinäre Netzwerke mit Koordinationsstellen aufbauen, um die Kooperation zwischen Forschenden der verschiedenen Fächer nachhaltig zu etablieren
- 6. Internationale Kooperationen anstreben, um Forschende ohne Zugriff auf Infrastrukturförderungen in bestehende Initiativen und Infrastrukturen einzubinden, an internationalen Entwicklungen zu partizipieren und darüber Datenbestände verfügbar zu machen
- 7. Die Auswirkungen von KI-generierter Sprache auf die Entwicklung und Bewertung von regionalen Markern und Stereotypen erforschen, um die Dynamik in Sprachgebrauch und -bewertung wissenschaftlich zu erfassen
- 8. Neue KI-basierte Methoden aktiv nutzen und kritisch reflektieren, um Herausforderungen und Potenziale offenzulegen
- 9. Engagiert die eigenen Erfahrungen und Expertisen in die wissenschaftlichen Organisationsformen und öffentlichen Diskurse einbringen, um eine differenzierte und reflektierte Bewertung des Einsatzes von KI sicherzustellen
- 10. Aktiv Kooperationen mit der Zivilgesellschaft initiieren, um die Partizipation der Öffentlichkeit an der KI-basierten regional orientierten Forschung zu erhöhen
- 11. Aktionen und KI-basierte Exponate gemeinsam mit Kulturschaffenden entwickeln und bei öffentlichen Ausstellungen, Festen und Messen präsentieren, um regional-sprachlich relevante KI-Anwendungen in der Gesellschaft sichtbar zu machen
- 12. KI-Anwendungen bedarfsgerecht und ressourcenschonend einsetzen, indem bei Routineaufgaben bevorzugt etablierte, energieeffiziente NLP-Verfahren genutzt und nur dort große KI-Modelle eingesetzt werden, wo sie einen klaren qualitativen Mehrwert bieten

### 5.7 Empfehlungen an **Lehrende**

- 1. Grundprinzipien von Open Science in der universitären Lehre über Prüfungsordnungen verankern, um den transparenten und offenen Umgang mit Daten in studentischen Arbeiten zum Standard zu machen
- 2. KI-Verfahren und -Anwendungen in linguistische Lehrformate integrieren, um Reflexionskompetenzen anzubahnen und Kritikfähigkeit auszubilden
- 3. Replikationsstudien mit Studierenden durchführen, um Datenkompetenzen aufzubauen und eine offene Fachkultur zu etablieren
- 4. In Kooperation mit KI-Forschenden integrative Ausbildungsformate und Lehrmaterialien entwickeln, um einen doppelten Kompetenzerwerb zu gewährleisten
- 5. Unterrichtsmaterialien zu regionaler Sprache und KI für die Schule bereitstellen, um den Wissenstransfer in die Schule sicherzustellen

*Künstliche Intelligenz in der Regionalsprachenforschung: Vertrauen und Vorbehalte*

Die Diskussion um den Einsatz von KI-Verfahren in der Regionalsprachenforschung ist von Ambivalenz geprägt. Einerseits eröffnen KI-Anwendungen neue Chancen für Effizienz, Standardisierung und Erkenntnisgewinne. Andererseits bestehen erhebliche Bedenken hinsichtlich gesellschaftlicher Fairness, regionaler Identität, wissenschaftlicher und methodischer Transparenz und Nachhaltigkeit. Regionale Sprachdaten sind besonders sensibel: Authentizität, kulturelle Identität und wissenschaftliche Integrität müssen gesichert sein. Vertrauen in und informierte Reflexion von KI-Methoden ist daher keine technische Nebensächlichkeit, sondern eine zentrale Voraussetzung für ihre Akzeptanz und ihren langfristigen Nutzen.

Tab. 1: Künstliche Intelligenz in der Regionalsprachenforschung

<p><b>Reproduzierbarkeit</b></p> <p>Ergebnisse müssen stabil, überprüfbar und langfristig archivierbar sein. Wenn Modelle nicht reproduzierbare Ausgaben erzeugen oder sich durch Updates unvorhersehbar verändern, wird wissenschaftliches Arbeiten erschwert.</p>	<p><b>Bias vs. Ausgewogenheit</b></p> <p>Trainingsdaten müssen regional, sozial und im Hinblick auf den situativen und zeitlich gebundenen Sprachgebrauch ausgewogen sein. Unzureichende Repräsentation oder stereotype Muster gefährden die Legitimität und marginalisieren regionale Varietäten.</p>
<p><b>Authentizität</b></p> <p>Ergebnisse von KI-Systemen müssen durch die Fachwissenschaft überprüft werden. Generative Inhalte dürfen nicht unkritisch als Originaldaten behandelt werden. Die Einbindung von Sprecherinnen und Sprechern regionaler Varietäten aus der Zivilgesellschaft ist notwendig, um Akzeptanz und Legitimität sicherzustellen. Ohne diese Überprüfung droht ein Verlust an Glaubwürdigkeit.</p>	<p><b>Datensouveränität</b></p> <p>Dialektdaten sind kulturelles Erbe. Ihre Nutzung muss rechtssicher, transparent und unter Kontrolle der Wissenschaft erfolgen, nicht durch kommerzielle Akteure. Die Datenhoheit muss gewahrt bleiben. Die Abschöpfung von Daten durch kommerzielle Akteure sowie die Dominanz proprietärer Systeme werfen Fragen nach Datenschutz, Machtasymmetrien und Abhängigkeiten auf.</p>
<p><b>Explainable AI</b></p> <p>Black-Box-Modelle sind vielfach unzureichend für den Erkenntnisgewinn. Verfahren und Trainingsdaten müssen so weit wie möglich transparent, erklärbar und reproduzierbar sein. Wenn weder Forschende noch Anwenderinnen und Anwender nachvollziehen und überprüfen können, wie ein durch KI-Anwendungen erzeugtes Ergebnis zustande kommt, sinkt die Bereitschaft, sich auf diese Ergebnisse zu verlassen.</p>	<p><b>Nachhaltigkeit</b></p> <p>Maßnahmen zur Ressourcenschonung, wie zum Beispiel die Verfügbarmachung hochwertiger Trainingsdaten zur energieeffizienten Erreichung hoher Benchmarks oder die Entwicklung kleinerer, aber spezifischerer Modelle, müssen unterstützt werden. Angepasste Methoden und Modelle können helfen, ressourcenschonend zu forschen. Ökologische Aspekte müssen in die Abwägung einfließen.</p>
<p>Vertrauen entsteht durch offene Standards, transparente Evaluation und kritische Distanz zu proprietären und kommerziellen Systemen. KI darf Forschung nicht domi-</p>	

nieren, sondern muss in klaren normativen Rahmenbedingungen verantwortungsvoll genutzt werden. Die Regionalsprachenforschung muss die Vorgaben des EU AI Acts zur Datenqualität, Dokumentation und Transparenz aktiv berücksichtigen, um die Konformität und Förderfähigkeit in europäischen Projekten sicherzustellen. Die Regionalsprachenforschung braucht aber auch maßgeschneiderte Lösungen, die kulturelle Eigenheiten respektieren und die Interessen von Wissenschaft, Bildung und Zivilgesellschaft gleichermaßen berücksichtigen.

## 6. Fazit

Die Regionalsprachenforschung steht im Zeitalter der digitalen Transformation vor einer doppelten Aufgabe: Sie muss ihre gewachsene Expertise in der Dokumentation, Analyse und Interpretation sprachlicher Vielfalt bewahren und zugleich die Potenziale der Künstlichen Intelligenz für ihre Weiterentwicklung nutzen. Das Positionspapier zeigt, dass zentrale Herausforderungen – von der Aufbereitung hochwertiger Forschungsdaten über die Anpassung bestehender Werkzeuge bis hin zur Schaffung nachhaltiger Infrastrukturen, Kompetenzen und Förderstrukturen – noch nicht ausreichend adressiert sind. Zugleich wird deutlich, dass die Regionalsprachenforschung einen unverzichtbaren Beitrag zur verantwortungsvollen Gestaltung der digitalen Gesellschaft leisten kann.

Die herausgearbeiteten Leitlinien und Handlungsempfehlungen sind drei übergeordneten Entwicklungszielen zugeordnet:



Abb. 1: Entwicklungsziele der Regionalsprachenforschung

Als Aufgabe für die Zukunft muss die Regionalsprachenforschung ihre Rolle als Brückenfach zwischen sprachwissenschaftlicher Expertise, technologischer Innovation und gesellschaftlicher Verantwortung aktiv ausfüllen. Offene, qualitativ hochwertige



Datenbestände, an regionale Sprachrealitäten angepasste Modelle und transparente Verfahren bilden die Grundlage. Ebenso entscheidend sind Kooperationen, sowohl interdisziplinär mit der KI-Forschung als auch mit der Zivilgesellschaft, die als Partner an der Generierung, Bewertung und Nutzung sprachlicher Daten teilhaben muss.

Die Fallbeispiele in diesem Positionspapier deuten an, welches Innovationspotenzial sich durch KI-Einsatz eröffnet, sei es bei der automatisierten Aufbereitung gesprochener Sprache, bei der Analyse komplexer grammatischer Strukturen oder bei der Übersetzbarkeit von Dialekttexten. Gleichzeitig machen sie die Risiken sichtbar: Fragen der Authentizität, Verzerrungen in Trainingsdaten, mangelnde Transparenz oder Abhängigkeiten von proprietären Systemen. Ein zukunftsweisender Umgang mit KI-Anwendungen in der Regionalsprachenforschung erfordert daher Vertrauen, das nur durch fachübergreifende Standards, nachvollziehbare Verfahren, kritische Evaluation und nachhaltige Infrastrukturen entstehen kann.

Der Ausblick ist vor diesem Hintergrund ebenso herausfordernd wie vielversprechend. Gelingt es, die sprachwissenschaftliche Expertise mit technologischer Innovationskraft und gesellschaftlicher Verantwortung zu verbinden, kann die Regionalsprachenforschung zu einem Modellfach für den reflektierten Umgang mit Künstlicher Intelligenz werden. Sie trägt nicht nur zum wissenschaftlichen Erkenntnisgewinn bei, sondern auch dazu, kulturelle Vielfalt im digitalen Raum sichtbar zu machen und damit einen Beitrag zu digitaler Teilhabe, kultureller Identität und gesellschaftlichem Zusammenhalt zu leisten.

## Netzwerk Regionale Sprache und Künstliche Intelligenz

Hans **Boas**, UT Austin College of Liberal Arts  
 Magnus Breder **Birkenes**, Norwegische Nationalbibliothek  
 Daniel **Braun**, Universität Marburg  
 Anne **Breitbarth**, Universiteit Gent  
 Caroline **Döhmer**, Universität Luxemburg  
 Christoph **Draxler**, LMU München, Institut für Phonetik und Sprachverarbeitung  
 Lisa **Dücker**, Universität Marburg, Forschungszentrum Deutscher Sprachatlas  
 Sophie **Ellsäßer**, Universität Osnabrück  
 Michael **Elementaler**, Christian-Albrechts-Universität Kiel, Germanistisches Seminar  
 Robert **Engsterhold**, Universität Marburg, Forschungszentrum Deutscher Sprachatlas  
 Stephanie **Evert**, Friedrich-Alexander-Universität Erlangen-Nürnberg  
 Hanna **Fischer**, Universität Marburg, Forschungszentrum Deutscher Sprachatlas  
 Lisa **Gerlach**, Universität Marburg, Forschungszentrum Deutscher Sprachatlas  
 Alexander **Geyken**, Berlin-Brandenburgische Akademie der Wissenschaften  
 Anja **Hasse**, Universität Zürich  
 Sarah **Ihden**, Universität Rostock  
 Felicitas **Kleber**, Universität des Saarlandes  
 Mareike **Krause**, Universität Marburg, Forschungszentrum Deutscher Sprachatlas  
 Alfred **Lameli**, Universität Marburg, Forschungszentrum Deutscher Sprachatlas  
 Adrian **Leemann**, Universität Bern  
 Alexandra N. **Lenz**, Universität Wien / Österreichische Akademie der Wissenschaften  
 Nathalie **Mederake**, Universität Marburg, Forschungszentrum Deutscher Sprachatlas  
 Georg **Oberdorfer**, Universität Marburg, Forschungszentrum Deutscher Sprachatlas  
 Barbara **Plank**, LMU München  
 Christoph **Purschke**, Universität Luxemburg, Department of Humanities  
 Manuel **Raaf**, Bayerische Akademie der Wissenschaften  
 Stefan **Rabanus**, Universität Verona  
 Yves **Scherrer**, Universität Oslo  
 Thomas **Schmidt**, [linguisticbits.de](http://linguisticbits.de) / Universität Duisburg-Essen  
 Beat **Siebenhaar**, Universität Leipzig  
 Ingo **Siegert**, Otto-von-Guericke-Universität Magdeburg  
 Nico **Urbach**, Universität Marburg, Forschungszentrum Deutscher Sprachatlas  
 Ella **Wissenbach**, Universität Marburg, Forschungszentrum Deutscher Sprachatlas  
 Christian **Zimmer**, TU Dortmund

Kontakt zum Netzwerk  
 Forschungszentrum Deutscher Sprachatlas (DSA)  
 Philipps-Universität Marburg  
 Pilgrimstein 16, 35037 Marburg  
[dsa@uni-marburg.de](mailto:dsa@uni-marburg.de)  
[www.dsa.info](http://www.dsa.info)

Korrespondierende Autor:innen:

HANNA FISCHER (hanna.fischer@uni-marburg.de)

ALFRED LAMELI (lameli@uni-marburg.de) und

## Abkürzungen

ACL	– Association for Computational Linguistics, <a href="https://www.aclweb.org/">https://www.aclweb.org/</a>
AGD	– Archiv für Gesprochenes Deutsch, Leibniz-Institut für Deutsche Sprache, <a href="https://agd.ids-mannheim.de/">https://agd.ids-mannheim.de/</a>
AI	– Artificial Intelligence
API	– Application Programming Interface
ARCHE	– A Resource Centre for the HumanitiEs, <a href="https://arche.acdh.oeaw.ac.at/">https://arche.acdh.oeaw.ac.at/</a>
ASR	– Automatic Speech Recognition
BMFTR	– Bundesministerium für Forschung, Technologie und Raumfahrt, <a href="https://www.bmftr.bund.de">https://www.bmftr.bund.de</a>
CLARIN	– Common Language Resources and Technology Infrastructure
DFG	– Deutsche Forschungsgemeinschaft, <a href="https://www.dfg.de/">https://www.dfg.de/</a>
DH	– Digital Humanities
DSA	– Forschungszentrum Deutscher Sprachatlas, <a href="https://dsa.info">https://dsa.info</a>
EOSC	– European Open Science Cloud, <a href="https://open-science-cloud.ec.europa.eu/">https://open-science-cloud.ec.europa.eu/</a>
EU	– Europäische Union, <a href="https://european-union.europa.eu/">https://european-union.europa.eu/</a>
FAIR	– Findable, Accessible, Interoperable, Reusable
HTR	– Handwriting Text Recognition
ICLaVE	– International Conference on Language Variation in Europe, <a href="https://iclave13.ch/">https://iclave13.ch/</a>
KI	– Künstliche Intelligenz
LinguRep	– Linguistisches Repositorium, DSA, <a href="https://lingurep.dsa.info/home">https://lingurep.dsa.info/home</a>
LIS	– Literaturversorgungs- und Informationssystem
LLM	– Large Language Model
ML	– Maschinelles Lernen
NFDI	– Nationale Forschungsdateninfrastruktur, <a href="https://www.nfdi.de/">https://www.nfdi.de/</a>
NLP	– Natural Language Processing = maschinelle Sprachverarbeitung
OCR	– Optical Character Recognition
OeAD	– Agentur für Bildung und Internationalisierung Österreich, <a href="https://oead.at/de/">https://oead.at/de/</a>
REDE	– Regionalsprache.de, <a href="https://www.regionalsprache.de/">https://www.regionalsprache.de/</a>
TDM	– Text- und Datamining
VarDial	– Workshop on NLP for Similar Languages, Varieties and Dialects, <a href="https://aclanthology.org/venues/vardial/">https://aclanthology.org/venues/vardial/</a>