

## FAIRmedia

Fair and Trusted Datasets for Media Computing

<b>Programm / Ausschreibung</b>	Digitale Technologien, Digitale Technologien, Digitale Schlüsseltechnologien: Ausschreibung 2022	<b>Status</b>	abgeschlossen
<b>Projektstart</b>	01.11.2023	<b>Projektende</b>	31.12.2025
<b>Zeitraum</b>	2023 - 2025	<b>Projektlaufzeit</b>	26 Monate
<b>Keywords</b>	AI; datasets; trustworthiness; media; journalism		

### Projektbeschreibung

Die Medien, und damit die Demokratie, stehen aufgrund von sich rasch verändernden Konsumgewohnheiten und der Zunahme von Fehlinformationen, vor riesigen Herausforderungen, die durch Krisen und intransparent (oder sogar autokratisch) kontrollierte soziale Mediendienste angeheizt werden. Künstliche Intelligenz (KI) kann als Werkzeug dienen, um Medien zu helfen, auf diese Herausforderungen zu reagieren und ihre Rolle in der Demokratie wahrzunehmen. Die meisten verfügbaren und ausreichend großen Datensätze für das Training von KI-Systemen werden jedoch von Organisationen in den USA und China bereitgestellt. Dies schafft nicht nur Abhängigkeiten von Technologieunternehmen außerhalb der EU, mit dem Risiko, dass kritische Datensätze jederzeit zurückgezogen werden können, sondern kann auch zu KI-Tools führen, die nicht mit der DSGVO und der kommenden KI-Regulierung konform sind (oder bestenfalls in einer Grauzone liegen). Die Quellen der Datensätze sind nicht immer gut dokumentiert, und den verfügbaren Datensätzen fehlen oft regional relevante Daten, so dass sie für das Verständnis von Inhalten und die Beantwortung von Sachfragen auf nationaler oder regionaler Ebene irrelevant sind.

FAIRmedia hat daher folgende Ziele: (i) die Schaffung eines Rahmens für die effiziente Sammlung von vertrauenswürdigen Datensätzen aus aktuellen und archivierten Inhalten österreichischer Medienorganisationen, (ii) die Erstellung von No-/Low-Code-Tools für das Training von KI-Modellen, die von Journalist:innen, Redaktionsleiter:innen, Publikumsmanager:innen und Archivar:innen verwendet werden können, und (iii) die Entwicklung von post-hoc erklärbaren KI-Tools, die zum Verständnis der trainierten Modelle und Biasanalyse verwendet werden können. Für die Datensatzerstellung werden relevante aktuelle Inhalte aus den täglichen Produktionsprozessen und aus den professionell dokumentierten Archiven österreichischer Medienunternehmen verwendet und in Bezug auf ausgewogene Berichterstattung und sachliche Korrektheit bewertet. Die Datensätze werden, wo immer möglich, aus Inhalten gewonnen, die potenzielle Probleme mit dem Urheberrecht und dem Datenschutz reduzieren. Um Medienschaffende und Forscher:innen in die Lage zu versetzen, KI-Modelle auf diesen Datensätzen selbst zu testen oder Ad-hoc-Modelle für aktuelle Produktions- oder Forschungsaufgaben zu erstellen, wird eine Reihe von Werkzeugen für das Training von KI-Modellen geschaffen, die keine Programmierkenntnisse erfordern. Im Gegensatz zu anderen Lösungen wird es möglich sein, das Training intern durchzuführen und die Daten unter Kontrolle des Medienunternehmens zu halten. Diese Werkzeuge werden auch aktuell Explainable AI (XAI)-Methoden enthalten, die eine post-hoc-Erklärung der erstellten Modelle ermöglichen. Außerdem richten sich diese Tools an Medienschaffende, die damit in

die Lage versetzt werden, die Modelle und möglichen Bias selbst zu analysieren. Die erstellten Tools und Datensätze werden in Anwendungsfällen für die Inhaltsklassifizierung und Question Answering validiert.

Das Projekt hat das Potenzial, die europäische Datensouveränität zu erhöhen, indem es Datensätze schafft, die vollständig unter der Kontrolle österreichischer Medienorganisationen stehen. Es ist beabsichtigt, diese Datensätze für andere Organisationen im Mediensektor verfügbar zu machen. Die Werkzeuge werden die Fähigkeit der Medienorganisationen zur Selbsterstellung fördern und sie in die Lage versetzen, KI-Technologien effektiv zu nutzen. Das Projekt kann als Sprungbrett dienen, um österreichische Medienorganisationen dabei zu unterstützen, Teil des europäischen Media Data Spaces zu werden, der im Rahmen des Programms Digitales Europa geschaffen wird.

## **Abstract**

Media and democracy face unprecedented challenges due to rapidly changing consumption habits, and the rise of misinformation fueled by crisis and intransparently (or even autocratically) controlled social media services. Artificial intelligence (AI) can serve as a tool for the media sector to react to these challenges, and enable it to fulfil its role in democracy. However, most available large-scale datasets for training AI systems are provided by organisations in the US and China. This does not only create dependencies on non-EU tech companies, with the risk that mission-critical datasets can be withdrawn at any time, but also may result in AI tools that are not compliant with GDPR and the emerging AI regulation (or in a grey zone at best), as the sources of the datasets are not always well documented. In addition, available datasets may lack regionally relevant data, and thus render them irrelevant for content understanding and factual question answering tasks on a national or regional level.

FAIRmedia thus has the objectives to (i) create a framework for the efficient collection of trusted datasets from recent and archived content of Austrian media organisations, (ii) create no- and low-code tools for training AI models usable by journalists, head of (corporate) newsrooms, audience managers and archivists and (iii) develop post-hoc explainable AI tools that can be used to understand the trained models and analysis biases. For dataset creation, highly relevant current content from the daily production processes and from the professionally annotated archives of Austrian media companies will be used, and assessed for balanced coverage and factual correctness. Datasets will be mined from archival content where possible, that reduces the potential copyright and privacy issues to be addressed. In order to enable media professionals and researchers to test AI models on these datasets themselves, or to create ad-hoc models addressing current production or research needs, a set of tools for training AI models for common tasks without the requirement of programming skills will be created. In contrast to comparable solutions, it will be possible to run the training in-house, and keep the data in the media organisation's domain. These tools will also include state-of-the art explainable AI methods, enabling post-hoc explainability of the created models. Also these tools are targeted at media professionals, empowering them to assess the models and potential bias themselves. The created tools and datasets will be validated in use cases for content classification and question answering.

The project has the potential to increase European data sovereignty by creating datasets fully under control of Austrian media organisations. It is intended to share these datasets with relevant stakeholders in the sector. The tools will foster the self-creation capability of media organisations, enabling them to embrace AI technologies. The project can serve as a springboard to support Austrian media organisations to become part of the European Media Data Space being created under the Digital Europe programme.

## **Endberichtkurzfassung**

Die Medien, und damit die Demokratie, stehen aufgrund von sich rasch verändernden Konsumgewohnheiten und der

Zunahme von Fehlinformationen, vor riesigen Herausforderungen, die durch Krisen und intransparent (oder sogar autokratisch) kontrollierte soziale Mediendienste angeheizt werden. Künstliche Intelligenz (KI) kann als Werkzeug dienen, um Medien zu helfen, auf diese Herausforderungen zu reagieren und ihre Rolle in der Demokratie wahrzunehmen. Die meisten verfügbaren und ausreichend großen Datensätze für das Training von KI-Systemen werden jedoch von Organisationen in den USA und China bereitgestellt. Dies schafft nicht nur Abhängigkeiten von Technologieunternehmen außerhalb der EU, mit dem Risiko, dass kritische Datensätze jederzeit zurückgezogen werden können, sondern kann auch zu KI-Tools führen, die nicht mit der DSGVO und der KI-Regulierung konform sind. Die Quellen der Datensätze sind nicht immer gut dokumentiert, und den verfügbaren Datensätzen fehlen oft regional relevante Daten, so dass sie für das Verständnis von Inhalten und die Beantwortung von Sachfragen auf nationaler oder regionaler Ebene irrelevant sind.

FAIRmedia hat daher Lösungen erarbeitet: (i) für die Schaffung eines Rahmens für die effiziente Sammlung von vertrauenswürdigen Datensätzen aus aktuellen und archivierten Inhalten österreichischer Medienorganisationen, (ii) für die Erstellung von No-/Low-Code-Tools für das Training von KI-Modellen, die von Journalist:innen, Redaktionsleiter:innen, Publikumsmanager:innen und Archivar:innen verwendet werden können, und (iii) die Entwicklung von post-hoc erklärbaren KI-Tools, die zum Verständnis der trainierten Modelle und Biasanalyse verwendet werden können. Für die Datensatzerstellung werden relevante aktuelle Inhalte aus den täglichen Produktionsprozessen und aus den professionell dokumentierten Archiven österreichischer Medienunternehmen verwendet und in Bezug auf ausgewogene Berichterstattung und sachliche Korrektheit bewertet. Die Datensätze werden, wo immer möglich, aus Inhalten gewonnen, die potenzielle Probleme mit dem Urheberrecht und dem Datenschutz reduzieren. Um Medienschaffende und Forscher:innen in die Lage zu versetzen, KI-Modelle auf diesen Datensätzen selbst zu testen oder Ad-hoc-Modelle für aktuelle Produktions- oder Forschungsaufgaben zu erstellen, wurde eine Reihe von Werkzeugen für das Training von KI-Modellen geschaffen, die keine Programmierkenntnisse erfordern. Im Gegensatz zu anderen Lösungen wurde darauf bedacht gelegt, das Training intern durchführen zu können und die Daten unter Kontrolle des Medienunternehmens zu halten. Außerdem richten sich diese Tools an Medienschaffende, die damit in die Lage versetzt werden, die Modelle und möglichen Bias selbst zu analysieren. Die erstellten Tools und Datensätze wurden in Anwendungsfällen für die Inhaltsklassifizierung und Question Answering validiert. Im Laufe des Projektes wurden zwei Datensätze verfügbar gemacht.

## **Projektkoordinator**

- JOANNEUM RESEARCH Forschungsgesellschaft mbH

## **Projektpartner**

- Hochschule für Angewandte Wissenschaften St. Pölten Forschungs GmbH
- RedLink GmbH
- Technisches Museum Wien mit Österreichischer Mediathek
- Universität Wien
- APA-Comm GmbH
- Österreichischer Rundfunk