

# Abschlussbericht “Public Interest AI”

**Autorin:** Dr. Theresa Züger

**Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01IS20058 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei der Autorin.**

**Zuwendungsempfänger:** Alexander von Humboldt Institut für Internet und Gesellschaft

**Förderkennzeichen:** 01IS20058

**Vorhabenbezeichnung:** Nachwuchsforscher\*innengruppe Public Interest AI (PIAI)

**Laufzeit des Vorhabens:** Oktober 2020 bis September 2024

## **I. Kurze Darstellung**

### **Ursprüngliche Aufgabenstellung**

Das Projekt verfolgte das Ziel, ein operationalisierbares Gemeinwohlkonzept vorzulegen, das es ermöglicht, die Gemeinwohlorientierung von KI-Anwendungen auf verschiedenen Ebenen von der konzeptuellen theoretischen Fundierung, über die technische Umsetzung, bis hin zum sozio-technischen Anwendungskontext anhand von definierten Kriterien zu beurteilen. Dabei sollten neben einem tragfähigen und fundierten Gemeinwohlverständnis auch technische Anwendungen umgesetzt werden, die das Konzept praktisch erproben und illustrieren.

### **Wissenschaftlich-technischer Stand an den angeknüpft wurde**

Der wissenschaftlich-technische Stand, an den das Projekt anknüpfte, variierte in Bezug auf verschiedene Ebenen des Projektes. Zunächst wurde allgemein an den theoretischen Diskurs zum Gemeinwohlbegriff angeknüpft, als auch an wissenschaftliche Literatur, die sich mit der Gestaltung von AI for Good und gemeinwohlorientierter KI auseinandersetzt. In Bezug auf die Anwendungsprojekte, die beide im Bereich Natural Language Processing angesiedelt sind, wurde auf jeweils relevante technische Fortschritte als auch den jeweiligen Diskurs zum Themenfeld aufgebaut (in AP 6 wurde der aktuelle Forschungsstand zur Automatisierung der Vereinfachung deutscher Sprache herangezogen, in AP 5 wurde zur Erkennung von Misinformation auf den State of the Art gesetzt), wobei in beiden Fällen LLMs zur Basis der Entwicklung von Prototypen herangezogen wurden. Bezüglich AP 4 zur partizipativen Data Governance wurde einerseits an die konzeptionellen Debatten des Feldes angeknüpft, zum anderen wurden mit etablierten Methoden Case Studies auf ihre empirische Umsetzung befragt.

### **Ablauf des Vorhabens**

Der Ablauf des Forschungsvorhabens hat sich grundlegend am geplanten Zeit- und Meilensteinplan orientiert. Alle APs wurden durchgeführt und erfolgreich abgeschlossen. Einzige Veränderung zum angestrebten Zeitplan ist die finale Einreichung von zwei der drei Promotionsprojekte, die im Jahr 2025 zur Verteidigung angemeldet werden.

### **Wesentliche Ergebnisse**

Zunächst hat das Projekt durch ein konzeptuelles Paper (im Journal AI & Society) einen wichtigen Beitrag zum Diskurs um gemeinwohlorientierte KI geleistet und darin grundlegende Kriterien für gemeinwohlorientierte KI definiert. Das entwickelte Konzept wurde auf zahlreichen Konferenzen und öffentlichen Veranstaltungen vorgestellt und diskutiert. Insgesamt konnten im Projektverlauf 23 Veröffentlichungen realisiert werden und zahlreiche Ergebnisse in zusätzlichen Wissenstransferformaten vermittelt werden. Auf Basis der empirischen Befunde der Forschung wurde ein Datensatz zu internationalen Beispielen von gemeinwohlorientierter KI erarbeitet und Ergebnisse hierzu veröffentlicht.

Lead und Post-Doc verankerten das Thema Public Interest AI erfolgreich als fortlaufenden Workshop auf einer führenden deutschen KI-Konferenz (KI der Gesellschaft für Informatik). Das erarbeitete Konzept von gemeinwohlorientierter KI wurde darüber hinaus in Form einer Webseite als Demonstrator veranschaulicht ([www.publicinterest.ai](http://www.publicinterest.ai)). Zum Abschluss des Projektes editierten und veröffentlichten Lead und Post-Doc als Gast-Editor\*innen ein Special Issue im Internet Policy Review zum Forschungsthema, unter dem Titel "AI systems in the public interest", das weitreichende Aufmerksamkeit erzielte.

Ebenso wurden wie geplant zwei Anwendungen entwickelt und veröffentlicht: *Simba*, ein Tool zur Vereinfachung deutscher Sprache und *Claimspotting*, eine Anwendung für Faktenchecker\*innen, die dabei unterstützt werden, mögliche Desinformation auf Telegram zu filtern und zu untersuchen. Mit dem Projekt wurden erfolgreich drei Promotionsvorhaben entwickelt, von denen eines bereits in der Projektlaufzeit abgeschlossen und eingereicht wurde (AP 6). Die beiden anderen befinden sich in der finalen Phase (AP 5 und AP 4) und werden voraussichtlich innerhalb des Jahres 2025 eingereicht und abgeschlossen.

### **Zusammenarbeit mit anderen Stellen**

Im Projektverlauf wurden zahlreiche Kontakte geschaffen und der Diskurs zum Thema gemeinwohlorientierter KI verstärkt. Wichtiger Austausch fand mit verschiedensten Organisationen statt, wie dem Civic Coding Innovationsnetz, welches von BMAS, BMUV und BMFSFJ ins Leben gerufen wurde, der GIZ und insbesondere dem Projekt FAIR forward, welches sich mit KI-Anwendungen in Entwicklungskontexten auseinandersetzt. Ebenso relevante Austauschpartner\*innen waren Akteure aus Zivilgesellschaft, Journalismus und öffentlicher Hand, wie die Diakonie, die Deutsche Welle, die Bundesfachstelle für Barrierefreiheit als auch Wikimedia e.V.. Der Austausch fand in verschiedenen Formaten statt, von Workshops bis hin zu gemeinsamen Panels oder Veranstaltungen und Hintergrundgesprächen.

## II. Eingehende Darstellung

### 1. Im Rahmen des Vorhabens durchgeführte Arbeiten (vgl. zu ursprünglicher Vorhabenbeschreibung)

#### Aufzählung der wichtigsten wissenschaftlich-technischen Ergebnisse und anderer wesentlicher Ereignisse

- Präsentation eines Poster zur Forschung über gemeinwohlorientiert KI auf Neurips 2024 (einer A\* rankten Konferenz zu Machine Learning und Neuroscience)
- Veröffentlichung eines Special Issues zum Thema “AI-systems in the public interest” editiert durch Lead und Post-Doc im Internet Policy Review
- Zwei halbtägige Konferenz-Workshops auf internationalen Fachkonferenzen wurden realisiert, die als wichtiger Teil der Dissemination und Reflektion des Projektes dienen (AP 7) und anstelle einer einzelnen Abschlusskonferenz durchgeführt wurde. Der eine Workshop wurde auf der KI Konferenz der Gesellschaft für Informatik konnte jährlich wiederkehrend verstetigt werden.
- Auf dem Public Interest AI Interface wurde ein Survey veröffentlicht, das von 34 internationalen Projekten ausgefüllt wurde und öffentlich zugänglich ist
- Lead und Post-Doc sind zu oft angefragten Expert\*innen in ihrem Forschungsfeld geworden und haben sehr erfolgreiche Publikationen veröffentlicht und zahlreiche öffentliche Vorträge gehalten
- Ein Dissertationsprojekt abgeschlossen, die beiden weiteren in finaler Phase

#### → **Participatory Data Governance in the Process of Developing Public Interest Tech and AI (AP4)**

- Präsentation im Doctoral Colloquium der AoIR (Association of Internet Researchers Konferenz) in Philadelphia
- Veröffentlichung “Fassbender, J. (2021). Particles of a whole: design patterns for transparent and auditable AI-systems.” in den Proceedings der 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing
- Erfolgreiche Teilnahme an der internationalen Konferenz ‘Big Data Discourses: Communicating, Deliberating, and Imagining Datafication’ an der Universität Leipzig, mit einem Beitrag unter dem Titel: “The Who, How and Why of Participation in Databased Projects A Systematic Literature Review”
- Veröffentlichung “Fassbender, J., & Henderson, T. (2024). Participatory to what end? Mapping motivations for participatory approaches in data-driven projects.” in den Proceedings der internationalen Konferenz ‘GoodIT ’24: International Conference on Information Technology for Social Good’
- Teilnahme an der internationalen Konferenz ‘EASST/4S 2024: Making and Doing Transformations. European Association for the Study of Science and Technology and the Society for Social Studies of Science’ in Amsterdam auf dem Panel “Datafied Publics: Organising participation in data-based projects”
- insgesamt vier wissenschaftliche Veröffentlichungen in verschiedenen relevanten Journals, davon eine als alleinige Autorin und eine weitere als Erstautorin, sowie drei andere Publikationen
- Eingereichte Veröffentlichung “Fassbender J., Kühnlein, I. & Henderson, T. (forthcoming). Facing the ambiguities of participation in data-driven projects: A systematic literature

review“ akzeptiert für das Journal ‘Data & Policy’ (Cambridge University Press)

#### → Claim Spotting und Public Interest AI (AP 5)

- Präsentation bei Public Interest AI Workshop auf der KI 2023
- Poster Präsentation “Propositional Claim Detection” auf der IC2S2 in Kopenhagen.
- Präsentation in Colloquium/Workshop an der TU Dortmund und Modus – Zentrum für angewandte Deradikalisierungsforschung
- Bereitstellung eines Machine Learning Datensatzes für “Propositional Claim Detection” auf [GESIS](#)<sup>1</sup>
- Teilnahme an der ‘European Conference on Machine Learning and Data Mining (ECML PKDD) 2024’ im Workshop “Advances in Interpretable Machine Learning and Artificial Intelligence” mit dem Beitrag “Asghari, H., Nenzo, S. (forthcoming). Mechanistic Interpretability of Socio-Political Frames in Language Models: an Exploration” (Veröffentlichung in den Proceedings folgt)
- Teilnahme an der internationalen Konferenz ‘MISDOOM 2024: 6th Multidisciplinary International Symposium on Disinformation in Open Online Media’ mit dem Beitrag “Random Evidence. Study on the Types of Evidence used to debunk Misinformation and the Implications for Automated Fact-checking” in Münster
- Veröffentlichung 2024 im ‘Communication Methods and Measures’ Journal mit dem Beitrag “Bootstrapping public entities. Domain-specific NER for public speakers.”
- Veröffentlichung 2024 im ‘Weizenbaum Journal of the Digital Society’ mit dem Beitrag “Potentials and Limitations of Active Learning for the Reduction of Energy Consumption during Model Training”
- insgesamt fünf wissenschaftliche Veröffentlichungen, davon vier als alleiniger Autor in verschiedenen relevanten Journals, sowie zehn andere Publikationen
- Veröffentlichung des *Claimspotting* Tools auf der Public Interest AI Website
- Die weitere Nutzung und im besten Fall einer weiteren Förderung und auch die Weiterentwicklung des *Claimspotting* Tool ist in Zusammenarbeit mit der Deutschen Welle (dw) geplant

#### → Natural Language Processing and Public Interest AI (AP6)

- ein annotierter Datensatz und die zugehörige Analyse wurden auf der ACL (Association for Computational Linguistics) Konferenz in Toronto präsentiert und anschließend veröffentlicht
- Erkenntnisse, die bei der Entwicklung des NLP Prototypen gewonnen wurden, wurden bei dem “AI Systems for the Public Interest” Workshop auf der AoIR (Association of Internet Researchers Konferenz) in Philadelphia präsentiert
- Veröffentlichung im A-Level Journal COLING mit dem Beitrag “Hewett, F. & Stede, M. (2022). Extractive Summarisation for German-language Data: A Text-level Approach with Discourse Features.”

---

<sup>1</sup> Nenzo, S.: Proportional Claim Detection (NLP Datensatz). (2023). GESIS.  
[https://search.gesis.org/research\\_data/SDN-10.7802-2538?doi=10.7802/2538](https://search.gesis.org/research_data/SDN-10.7802-2538?doi=10.7802/2538)

- Teilnahme an der internationalen Konferenz ‘SIGDIAL 2024: 25th Meeting of the Special Interest Group on Discourse and Dialogue’ in Kyoto, Japan mit dem Beitrag “Hewett, F., Asghari, H., Stede, M. (2024). Elaborative Simplification for German-Language Texts”. Der Beitrag wurde in den Proceedings der Konferenz veröffentlicht
  - insgesamt sechs wissenschaftliche Veröffentlichungen in relevanten Journals, davon eine als alleinige Autorin und vier weitere als Erstautorin, sowie vier andere Publikationen
  - offizielle Veröffentlichung der Browser-Erweiterung “*Simba*” für intelligente Textvereinfachung im September 2024
- Weitreichender Impact der Forschung, sowohl in der Forschungscommunity als auch in den politischen, zivilgesellschaftlichen und privaten Sektor hinein.

→ **Wissenstransfer Highlights:**

- Zahlreiche Vorträge auf nicht-akademischen Konferenzen zu gemeinwohlorientierter KI
- Publicinterest.ai bietet einen Stakeholder Index, der relevant Akteure im Feld gemeinwohlorientierter KI strukturiert listet und zugänglich macht
- re:publica 2023 /2024: Theresa Züger war Speakerin auf einem Panel zu Gemeinwohlorientierter Digitalpolitik und zu der Frage, ob und wie KI im Klimawandel helfen kann
- Judith Faßbender hat ein Panel zu ‘Data Care’ auf der re:publica 2023 gehostet
- Theresa Züger war Co-Host eines Workshops auf dem Digitalgipfel 2023 in Jena
- Beitritt in die Community of Practice der Digital Public Goods Alliance<sup>2</sup> und Präsentation unseres Vorgehens in einer Sitzung
- Entwicklung einer Poster Journey als Open Educational Ressource zum Thema Bias und Bias Mitigation mit der Mathematikerin Irma Mastenbroek

## 2. Wichtigste Positionen des zahlenmäßigen Nachweises

Der größte Anteil der Zuwendung wurde für wissenschaftliches Personal und studentische Hilfskräfte aufgewendet. Darüber hinaus wurden Mittel für technische Anschaffungen, sowie für Reise- und Sachmittel für z.B. die Durchführung von Workshops, Teilnahme an Konferenzen und die Publikation der Ergebnisse verwendet. Eine genaue Verteilung der projektbezogenen Mittel findet sich im zahlenmäßigen Nachweis.

Im Projektzeitraum wurden zwei Umwidmungen (Bescheide vom 17.03.2022 und 08.12.2022) umgesetzt, die die Anschaffung eines Grafikprozessors als Erweiterung des vorhandenen Servers und die Aufstockung einer Projektassistenz für das Team für 12 Monate ermöglichten.

Geringfügige Verzögerungen im Laufe des Projekts (z.B. Verschiebung von Veranstaltungen durch die Corona-Pandemie, den Weggang eines Doktoranden und entsprechende Stellennachbesetzung) konnten ausgeglichen werden.

---

<sup>2</sup> Digital Public Goods Alliance: Unlocking the potential of open-source technologies for a more equitable world. (o.D.). <https://digitalpublicgoods.net/>

In Bezug auf die geplanten Partnerschaften in der Praxis ergaben sich - immer in Absprache mit dem Projektträger - einige Änderungen, da sich andere Partnerschaften für unsere Praxisprojekte als vorteilhafter entwickelt haben. Zum Beispiel hat sich für das AP 5 nach personellem Wechsel ein neuer thematischer Schwerpunkt entwickelt, weshalb die ursprünglich geplante Kooperation mit Wheelmap im Bereich Computer Vision nicht mehr zielführend war. Mit der veränderten Konstellation der Praxispartner\*innen hat sich auch der Zeitpunkt der Kollaboration gegenüber unserer ursprünglichen Planung verändert. Es zeichnete sich ab, dass wir mit den neu gewonnenen Partnerorganisationen erst ab Juni 2024 bis Ende 2024 zusammenarbeiten würden. Unter anderem aufgrund dieser Tatsache wurde ein Antrag auf kostenneutrale Verlängerung (11.03.2024) gestellt, der allerdings negativ beschieden wurde (17.07.2024). In der Folge wurde beantragt die gesperrten Mittel für die Weitergabe an nicht-wissenschaftliche Partner i.H.v. 59.669,92€ zu kürzen (24.09.2024).

Die Arbeiten wurden dennoch bis zum Projektende am 30.09.2024 fertiggestellt. Es war zielführender, weniger eng mit einem Partner (wie ursprünglich geplant CorrelAID) und dafür breiter vernetzt mit verschiedenen Partnern (Deutsche Welle, ARD Faktenfinder oder der Bundesfachstelle für Barrierefreiheit) zusammenzuarbeiten. Aufgrund der nicht erfolgten Verlängerung konnte die geplante Disseminationsphase allerdings nicht in dem Umfang und der Intensität umgesetzt werden. Die Anwendungen *Simba* und *Claimspotting* hätten zudem intensiver mit Partner\*innen evaluiert und weiterentwickelt werden können.

### 3. Notwendigkeit und Angemessenheit der geleisteten Projektarbeit

Mit der Durchführung unserer Forschungsarbeit haben wir erfolgreich einen zentralen Debattenbeitrag zur Entwicklung von KI-Systemen geleistet, indem wir spezifisch zeigen konnten, wie KI-Systeme gemeinwohlorientiert entwickelt werden können. Unser Vorgehen war dabei besonders erforderlich, da es einerseits auf theoretischer Ebene eine fundierte Herleitung leistet, als auch empirische Erkenntnisse hervorgebracht hat, z.B. zur Praxis gemeinwohlorientierter KI. Weiter haben wir mit unseren Prototypen einen zentralen Beitrag geleistet, um zu verstehen, wie die Umsetzung von gemeinwohlorientierter KI gestaltet werden kann.

### 4. Nutzen des Projektes/ Verwertbarkeit der Ergebnisse/ Verwertungsplan zu wissenschaftlichen und technischen Erfolgsaussichten

Wissenschaftskommunikation		
Veranstaltungen	zeitliche Planung	Umsetzung
Transdisziplinäres Fachsymposium (Teil von AP 3): Hierzu wurden neben den Projekt- und Praxispartnern fachlich relevante Expert*innen eingeladen. Ziel: Vorstellung erster Arbeitsergebnisse, Feedback aus Anwendungsdomänen und von technischen Expert*innen.	Q4/2021	Das Fachsymposium fand am 13.06.2022 statt.

### Vorträge auf relevanten Veranstaltungen:

**Asghari, H.:** *Amplifying Privacy: Scaling Up Transparency Research Through Delegated Access Requests*. ConPro '21; online, San Francisco; 27. Mai 2021.

**Asghari, H.:** *Approximating Accessibility of Regions from Incomplete Volunteered Data*. CHI EA '22. Association of Computing Machinery. Online, New Orleans, USA; 03. Mai 2022.

**Asghari, H.:** *Poster Pitch: Are AI standardisation processes inclusive enough to deliver tech that works for everyone?* European Think Tank Conference 2022. Deutsche Gesellschaft für Auswärtige Politik (DGAP), Berlin; 20. September 2022.

**Asghari, H.:** *The Opportunities of AI in Addressing Human Rights Risks*. UN GCD Annual Conference 2023, Session: Menschenrechte und Arbeitsnormen. UN Global Compact Netzwerk Deutschland (UCD). Humboldt Carré, Berlin, Deutschland: 07.11.2023.

**Asghari, H.:** *On the Prevalence of Leichte Sprache on the German Web*. 15th ACM Web Science Conference 2023. Association of Computing Machinery. 15th ACM Web Science Conference 2023, Evanston, USA: 30.04.2023.

**Asghari, H., & Nenno, S.:** *Eine Bestandsaufnahme des CO2-Fußabdrucks von AI: Was wissen wir und wie können wir ihn verringern?* KI Camp 2023. Gesellschaft für Informatik. Factory Berlin Görlitzer Park, Berlin, Deutschland: 26.04.2023.

**Asghari, H., & Nenno, S.:** *Mechanistic Interpretability of Socio-Political Frames in Language Models: an Exploration*. Advances in Interpretable Machine Learning and Artificial Intelligence (AIMLAI). INRIA. Radisson Blu Vilnius Hotel, Vilnius, Lithuania: 09.09.2024.

**Faßbender, J.:** *Particles of a whole: design patterns for transparent and auditable AI-systems*. UbiComp 21: The 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing; online, New York; 25. September 2021.

**Faßbender, J.:** *PhD Workshop on Artificial Intelligence and Machine Learning research and Democracy*. Universität Tübingen, Tübingen; 02. April 2022.

**Faßbender, J.:** *Design Patterns – Prozessformate für eine verantwortungsvolle Umsetzung von KI-Systemen*. KI-Lab. Denkfabrik digitale Arbeitsgesellschaft des Bundesministeriums für Arbeit und Soziales, Berlin; 30. September 2022.

**Hewett, F.:** *Automatically evaluating the conceptual complexity of German texts*. KONVENS 2021: German Society for Computational Linguistics and Language Technology, the Special Interest Group on Computational Linguistics of the German Linguistic Society and the Austrian Society for Artificial Intelligence, Heinrich-Heine-Universität Düsseldorf; Düsseldorf; 09. September 2021.

**Hewett, F.:** *APA-RST: A Text Simplification Corpus with RST Annotations*. ACL 23: The 61st Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics. The Westin Harbour Castle, Toronto, Kanada: 13.07.2023.

**Hewett, F.:** *Komplexitätsmerkmale auf der Textebene: Eine systematische Korpusanalyse*. Tagung: Einfach oder komplex? Befunde zur Passung geschriebener Sprache für verschiedene Zielgruppen. Institut für Deutsch als Fremdsprachenphilologie. Universität Heidelberg, Heidelberg, Germany: 22.03.2024

**Hewett, F.:** *Elaborative Simplification for German-Language Texts*. SIGDIAL 2024: 25th Meeting of the Special Interest Group on Discourse and Dialogue. Kyoto University. Kyoto University, Kyoto, Japan: 18.09.2024

**Nenno, S.:** *Propositional Claim Detection*. International Conference for Computational Social Science IC2S2'23 Copenhagen. International Society for Computational Social Science (ISCSS). Mærsk Tower of the University of Copenhagen, Kopenhagen, Dänemark: 20.07.2023.

**Nenno, S., & Puschmann, C.:** *All the (fake) news that's fit to share? News values in perceived misinformation across 24 countries*. ECREA 2024 Communication & social (dis)order: 10th European Communication Conference (Session: From sharing to discussing: Online media experiences). European Communication Research and Education Association (ECREA). University of Ljubljana, Ljubljana, Slovenia: 25.09.2024

**Stolberg-Larsen, J.:** *AI with Volunteering Communities*. International Joint Conferences on Artificial Intelligence; Montreal; 21. August 2021.

**Züger, T., Asghari, H., Hewett, F., Fassbender, J., Stolberg, J.:** *AI in the public interest: What does this mean and how can we build it?* KI Camp der Gesellschaft für Informatik; 24. Februar 2021.

**Züger, T.:** *KI für das Gemeinwohl. Was ist „Künstliche Intelligenz“ und was kann sie für die Gesellschaft leisten?* sowie: *Dialog: Der Mensch im Spiegelbild der KI*. KI Tagung der Diakonie Deutschland, des V3D und des vediso; 1. Dezember 2021.

**Züger, T.:** *KI und Gesellschaft? Können wir künstlicher Intelligenz moralische Grenzen setzen?* sowie: *Mehr Chancen, Mehr Gefahren?* Europäische Akademie Berlin; 25. Februar 2021.

**Züger, T.:** *Vortrag beim All-hands-meeting der BMBF-geförderten KI-Forschungsprojekte*. DLR; Online; 10. Januar 2022.

**Züger, T.:** *Künstliche Intelligenz & Gemeinwohl*. Zusammen: Über Medien des Miteinanders; Fachhochschule Potsdam; 17. Februar 2022.

**Züger, T.:** *Die hybride Gesellschaft*. POCHEN Biennale 2022: Die (neue) Vermessung der Welt. Spinnerei e.V., Online, Chemnitz; 06. Oktober 2022.

**Züger, T., Fassbender, J., Kuper, F.:** *Civic Coding – KI für das Gemeinwohl nutzen*.

Bundesministerium für Arbeit und Soziales. Impact HUB Berlin, Berlin; 18. Oktober 2022.

**Züger, T.:** *Was bedeutet Transparenz für KI-Systeme?* ZVKI Fachforum. iRights.Lab, Online, Berlin; 10. November 2022.

**Züger, T.:** *Public Interest AI.* UNESCO Fachausschuss Kommunikation und Information. Deutsche UNESCO-Kommission, Brüssel, Belgien; 15. November 2022.

**Züger, T.:** *AI und Social Sustainability.* AI for Sustainability / Sustainable AI. Technische Universität München (TUM), Hochschule München. Online, München, Deutschland: 17.11.2023.

**Züger, T.:** *KI im Sinne des Gemeinwohls.* #ZukunftWohlfahrt: Mit Daten & KI die Wohlfahrt von morgen gestalten. Deutsches Rotes Kreuz. Online, Dresden, Deutschland: 15.11.2023.

**Züger, T.:** *KI im Sinne des Gemeinwohls.* Fachtagung Münster. Diakonie Deutschland. Online, Münster, Deutschland: 21.09.2023.

**Züger, T.:** *AI and Public Interest.* I-AI #2. Künstlerische Lehre und künstliche Intelligenz @ UdK Berlin. Universität der Künste Berlin. Online, Berlin, Deutschland: 10.07.2023.

**Züger, T.:** *How AI can serve the Public Interest.* Women in Data Science Conference Chemnitz. Women in Data Science Chemnitz. Universitätsbücherei Chemnitz, Chemnitz, Deutschland: 04.05.2023

**Züger, T.:** *Public Interest AI.* Feministische Digitalpolitik (Session: Public Interest AI). Superrr Lab. Superrr Lab, Berlin, Deutschland: 20.02.2023

**Züger, T.:** *Wie verändert generative Künstliche Intelligenz den Journalismus?* Medienpolitische Tagung 2024: KI, Big Tech & Co. – Was wird aus dem Journalismus?. ver.di, DGB. ver.di-Haus, Berlin, Germany: 16.10.2024

**Züger, T.:** *AI and the Public Interest.* AI Grid Summit 2024. AI Grid, EIT ICT Labs Germany GmbH. Forum Digitale Technologien, Berlin, Germany: 12.11.2024

**Züger, T.:** *Public Interest AI.* FARI Brussels Conference 2024: AI, a Public Good?. FARI - AI for the Common Good Institute Brussels. Studio Flagey, Brussels, Belgium: 18.11.2024

**Züger, T.:** *Public Interest AI in Action.* Women in Machine Learning (WiML) Workshop @ NeurIPS 2024. WiML. Vancouver Convention Center, Vancouver, Canada: 10.12.2024

**Lehre:**

**Asghari, H., & Nenno, S.:** *AI for Sustainability & Sustainable AI.* CIFE Master programmes in Global Economic Governance & Public Affairs (GEGPA) and EU Trade & Climate Diplomacy (EUDIPLD). Centre International de Formation Européenne (CIFE). Centre International de Formation Européenne (CIFE), Berlin, Deutschland: 17.03.2023.

**Asghari, H.:** *AI for the Public Interest*. Bachelor's Course: "Künstliche Intelligenz als Herausforderung der Demokratie". Universität Erfurt. Universität Erfurt, Erfurt, Deutschland: 28.06.2023.

**Asghari, H.:** *Problems of Technology for Society*. Ethical, Social, and Legal Foundations of Data Science Bachelors Course. Universität Göttingen. Universität Göttingen, Göttingen, Deutschland: 15.02.2023.

### Organisation von Veranstaltungen

**Asghari, H., Faßbender, J., Hewett, F., & Züger, T.:** *AI-systems for the public interest*. Halbtages-Workshop bei der AolR 2023. Sonesta Hotel, Philadelphia, USA. 18.10.2023.

Züger, T.: AI-systems for the public interest (Vortrag)

Hewett, F.: Public Interest AI in practice (Vortrag)

Sessionmoderation: J. Faßbender

**Asghari, H., Faßbender, J., Hewett, F., Nenno, S., & Züger, T.:** *AI-systems for the Public Interest*. Halbtagesworkshop bei der KI 2023 – 46th German Conference on Artificial Intelligence. Gesellschaft für Informatik (GI). Hochschule für Technik und Wirtschaft Berlin, Berlin, Deutschland. 26.09.2023.

Nenno, S.: Fake Values (Vortrag)

Züger, T.: Public Interest AI in practice (Vortrag)

Sessionmoderation: T. Züger und H. Asghari

**Asghari, H., Faßbender, J., Hewett, F., Pothmann, D., Spitz, S., & Züger, T.:** *Hackathon. Entwicklung von KI-Erklärungsmodellen*. Humboldt Institut für Internet und Gesellschaft, Berlin, Deutschland. Co-organisiert von: neuland & gestalten. 03.02.2023.

**Asghari, H., Nenno, S., Hewett, F., Faßbender, J., Züger, T.:** *AI and the Public Interest*. Humboldt Institute for Internet and Society, Berlin, Germany; 13.06.2022.

**Asghari, H.:** *The theory of artificial immutability: protecting algorithmic groups under anti-discrimination law*. Moderation eines Lunch-Talks mit Sandra Wachter. Humboldt Institut für Internet und Gesellschaft, Berlin, Deutschland. 27.06.2023.

**Bollmann, H-S., Faßbender, J., Köhne, S., & Spitz, S.:** *Girl's Day – Die Wissenschaft in Mädchenhänden*. Humboldt Institut für Internet und Gesellschaft, Berlin, Deutschland. Co-organisiert von: Hacker School. 27.04.2023.

**Efferenn, F., Pothmann, D., Nanno, S., Spitz, S., & Züger, T.:** *Interaktiver Workshop: KI menschlich erklärt*. Humboldt Institute for Internet and Society, Berlin, Deutschland, Moderation durch T.

Züger. 27.09.2023.

**Faßbender, J., Vukajlovic, N. & Grauvogel, C.:** *Digitaler Salon: Cookies make you lose control* Humboldt Institute for Internet and Society, Berlin, Deutschland. 25.08.2021.

**Faßbender, J., Hewett, F., & Züger, T.:** *Shifting AI Controversies: Sind die KI-Kontroversen, die wir haben, die KI-Kontroversen, die wir brauchen?* Workshop im Rahmen von Shifting AI Controversies. Humboldt Institut für Internet und Gesellschaft, Berlin, Deutschland. Co-organisiert von: Zentrum für Medien-, Kommunikations- und Informationsforschung (ZeMKI, Universität Bremen). 05.10.2023.

**Faßbender, J.:** *Data Care - Who is cleaning, sorting and maintaining data?* Mit: Roos Hopman, Franziska Peter, Nora Sagel, Freia Kuper. re:publica 2023. re:publica 2023, Berlin, Deutschland. 07.06.2023.

**Nenno, S., Züger, T.:** *Sustainable AI* with attending Vip: Friederike Rohde. Humboldt Institute for Internet and Society, Berlin, Germany; 29. April 2022.

**Nenno, S., Züger, T.:** *Sustainable AI II* with attending Vip: Lynn Kaack. Humboldt Institute for Internet and Society, Berlin, Germany; 14. Juni 2022.

**Pothmann, D., Spitz, S. & Nenno, S.:** *Digitaler Salon: Friede, Freude, Meinungsvielfalt?* Humboldt Institut für Internet und Gesellschaft, Berlin, Deutschland. 26.07.2023.

**Pothmann, D., Spitz, S. & Züger, T.:** *Digitaler Salon: Frozen Eye.* Humboldt Institut für Internet und Gesellschaft, Berlin, Deutschland. 27.09.2023.

**Pothmann, D., Spitz, S. & Kühnlein, I.:** *Digitaler Salon: Quanten statt Quaken.* Humboldt Institut für Internet und Gesellschaft, Berlin, Deutschland. 27.03.2024.

**Pothmann, D., Spitz, S. & Hewett, F.:** *Digitaler Salon: Wie bitte?* Humboldt Institut für Internet und Gesellschaft, Berlin, Deutschland. 26.06.2024.

**Züger, T., Asghari, H. & Pothmann, D.:** *2nd Workshop on 'Public Interest AI'.* Workshop auf der KI 2024 – 47th German Conference on AI. Julius-Maximilians Universität Würzburg, Würzburg, Deutschland. 23.09.2024.

Sessionmoderation und Vortrag: T. Züger

#### Teilnahme als ExpertIn

**Asghari, H.:** *12. Indo-German Frontiers of Engineering Symposium (INDOGFOE).* Alexander von Humboldt Foundation. Alexander von Humboldt Foundation, Bremen; 29.09.2022.

**Faßbender, J.:** *Panel: KI & Kulturerbe.* digiS-Jubiläumsjahreskonferenz 2023. Forschungs- und Kompetenzzentrum Digitalisierung Berlin. Zuse-Institut Berlin, Berlin, Deutschland: 24.05.2023

**Henkes, L., Bächle, T. C., Züger, T.:** *Infrastructures of Autonomy*. Humboldt Institut für Internet und Gesellschaft, Berlin, Germany. Co-Organised by: Centre for Media Studies Potsdam; 23. November bis 25.11.2022.

**Hewett, F.:** *Potenziale und Hürden inklusiver Chatbots am Beispiel des Chatbots Ina*. Bertelsmann Stiftung & betterplace lab. Online, Berlin; 18.08.2022.

**Hewett, F.:** *Stakeholder-Dialog – Ethik der Digitalisierung*; Berkman Klein Center, Digital Asia Hub. Humboldt Institute for Internet and Society; Berlin; 06.10.2021; URL: <https://www.hiig.de/defining-digitalisation-defending-ethical-principles/>.

**Spitz, S., Züger, T. & Mastenbroek I.:** Technology is Never (Gender) Neutral Roundtable with Margrethe Vestager (European Commission) Deutsche Vertretung der EU Kommission, Berlin, Germany; 08.05.2023

**Stolberg-Larsen, J.; Hewett, F. & Züger, T.:** *Explainable AI Research Clinic – Explainable AI* Humboldt Institute for Internet and Society. Coconat Gutshof, Bad Belzig; 08.09.2021; URL: <https://www.hiig.de/en/research-clinic-explainable-ai/>.

**Enno, S., Faßbender, J., Züger, T.:** *Der Mensch im digitalen Zeitalter*. Humboldt Institute for Internet and Society, Berlin, Germany. Co-Organised by: ZHAW; 05.05.2022.

**Züger, T.:** *Gemeinwohlorientiert = Gemeinnützig? Unternehmen und gemeinwohlorientierte KI*. re:publica 2021; online, Berlin; 21.05.2021.

**Züger, T.:** *Young activists & human rights in a digital age*. Fundamental Rights Forum; online, Warsaw; 11.10.2021.

**Züger, T.:** *Public Interest Infrastructure Panel*; Tech for Democracy Days of Action, Auswärtiges Amt Dänemark; 10.11.2021.

**Züger, T.:** *Wie vermitteln wir Wissen über KI an die Bevölkerung?* Auftaktveranstaltung: Zentrum für vertrauenswürdige Künstliche Intelligenz. iRights.Lab; Berlin; 07.12.2021.

**Züger, T.:** *Public Round Table: Panel Discussion der Vienna Winter School „Taming the iMonster: Regulating digital platforms“*; Universität Wien; 24.02.2022.

**Züger, T.:** *Public Round Table: Panel Discussion der Vienna Winter School „Taming the iMonster: Regulating digital platforms“*; Universität Wien; 24.02.2022.

**Züger, T.:** *The Future of War: Drone Surveillance, Killer Robots & Lethal AI Weapons*. DNL26: The Kill Cloud: Networked Warfare, Drone & AI. Disruption Network Lab, Kunstquartier Bethanien, Berlin; 25.03.2022.

**Züger, T.:** *Science Talk "Converging Sciences"*. BMBF-Fachworkshop zur Integrierten Forschung. Forschungscluster Integrierte Forschung, Online, Nürnberg; 14.07.2022.

**Züger, T.:** *Künstliche Intelligenz + Gemeinwohl = ? Die dima und du! Dein Zukunftsassistent lädt ... ein!*. JUKUZ - Jugendkulturzentrum Aschaffenburg, Aschaffenburg; 28.10.2022.

**Züger, T.:** *Zukunft der Mobilität: Wenn Algorithmen unsere Wege beeinflussen*. 125 Jahre HSG: Kompass für die Zukunft. Universität St. Gallen. Universität St. Gallen, St. Gallen, Schweiz: 30.11.2023

**Züger, T.:** *Digitalpolitik für alle: Lunch Talk zum Gemeinwohl*. Podiumsdiskussion: "Digitalpolitik für alle: Lunch Talk zum Gemeinwohl". Heinrich-Böll-Stiftung - Bundesstiftung Berlin. Heinrich-Böll-Stiftung, Berlin, Deutschland: 08.11.2023

**Züger, T.:** *Strengthening Civic Engagement and Democracy*. I2SC Kick Off Event. Interdisciplinary Institute for Societal Computing (I2SC). Saarland University, Saarbrücken, Deutschland: 07.09.2023

**Züger, T.:** *Wie als Gesellschaft die Kontrolle über KI behalten? Ansätze aus der aktuellen Forschung*. KIDD @ work. KIDD – KI im Dienste der Diversität. Ludwig-Erhard-Haus, Berlin, Deutschland: 31.03.2023

**Züger, T.:** *G7 Hiroshima AI Process – Selbstverpflichtung als Schlüssel zur Akzeptanz*. KI-Round Table. Bundesministerium für Digitales und Verkehr (BMDV), Berlin, Deutschland: 25.09.2023

**Züger, T.:** *Das Dateninstitut für Deutschland: Wie kommt das Gemeinwohl in die Umsetzung?* Digital-Gipfel 2023 BMWK & BMDV. Volkshaus Jena, Jena, Deutschland: 20.11.2023

**Züger, T.:** *Applications, Ethics, and Futures of Artificial Intelligence*. International Futures (IF) 2023 Managing Global Governance (MGG). Auswärtiges Amt, Berlin, Deutschland: 11.10.2023

**Züger, T.:** *Studie zu den konsumsteigernden Wirkungen von Plattformen CO:DINA*. Online, Berlin, Deutschland: 23.03.2023

**Züger, T.:** *Digitales Europa 2030*. Seed Workshop Alfred Herrhausen Gesellschaft. Online, Berlin, Deutschland: 18.04.2023

**Züger, T.:** *Stitching Democracy* Federal Agency for Civic Education (Bundeszentrale für politische Bildung/bpb). Bundeszentrale für politische Bildung, Berlin, Germany: 25.09.2024

**Züger, T.:** *Digitalinfrastrukturpolitik Workshop: "Digitalinfrastrukturpolitik"*. Wikimedia Deutschland e. V.. Wikimedia Deutschland, Berlin, Germany: 25.09.2024

**Züger, T.:** *Meeting: AI "partnerships"*. Gesellschaft für Freiheitsrechte e.V. (GFF). Gesellschaft für Freiheitsrechte, Berlin, Germany: 12.09.2024

**Züger, T.:** *Roundtable: KI-Innovationen und der Wirtschaftsstandort Deutschland*. Das Progressive Zentrum. Räumlichkeiten des Progressiven Zentrums, Berlin, Germany: 10.09.2024

**Züger, T.:** *Multistakeholder-Dialog: „Künstliche Intelligenz im Schnittfeld von Wissenschaft und*

Öffentlichkeit". Wissenschaft im Dialog (WiD). Humboldt Institut für Internet und Gesellschaft, Berlin, Germany: 28.08.2024

**Züger, T.:** *Workshop: Principles and Benchmarks for Public Interest AI*. Digital Public Goods Alliance. Online, New York, USA: 16.08.2024

**Züger, T.:** *AI Palace 2024*. Metagov. Bueckeburg Palace, Bückeberg, Germany: 15.07.2024

**Züger, T.:** *Focus group: satellite image use for citizen-based monitoring*. Einstein Center Digital Future. Robert-Koch-Forum, Berlin, Germany: 11.07.2024

**Züger, T. & Nenno, S.:** *KI-Roundtable: KI und Desinformation*. Bundesministerium für Digitales und Verkehr. Bundesministerium für Digitales und Verkehr, Berlin, Germany: 24.05.2024

### Veröffentlichungen in digitalen Medien

**Borgesius, F. Z., & Asghari, H.:** Data breaches: Does the GDPR help? *Digital society blog*; 08. September 2023; URL: <https://www.hiig.de/en/data-breach-notification/>

**Fassbender, J.** in: Hauck, M.: Gemeinwohl. Wie man KI dazu bringt, etwas Gutes zu tun. *Süddeutsche Zeitung*; 12. Juli 2023; URL: <https://www.sueddeutsche.de/projekte/artikel/wirtschaft/kuenstliche-intelligenz-gemeinwohl-mozilla-e997292/?reduced=true>

**Fassbender, J.:** Why explainable AI needs such a thing as Society. *Digital society blog*; 17. Februar 2022; URL: <https://www.hiig.de/en/explainable-ai/>.

**Hewett, F., & Nenno, S.:** How to identify bias in Natural Language Processing. *Digital society blog*; 29. November 2021; URL: <https://www.hiig.de/en/bias-in-natural-language-processing/>.

**Hewett, F., & Stede, M.:** *Lexica corpus*. 2021; URL: <https://zenodo.org/record/5196030#.YmAJX5JBzq1>

**Hewett, F. & Asghari, H.:** Lowering the barriers: Accessible language and "Leichte Sprache" on the German Web. *Digital society blog*; 02. Mai 2023; URL: <https://www.hiig.de/en/lowering-the-barriers-accessible-language-and-leichte-sprache-on-the-german-web/>

**Nenno, S.:** Siri's evil sister. When the Dutch public service steals your data. *Digital society blog*; 3. Mai 2021; URL: <https://www.hiig.de/en/siris-evil-sister-when-the-dutch-public-service-steals-your-data/>.

**Nenno, S.:** How to find the right survey tool: problems and proposals. *Elephant in the Lab*; 8. Dezember 2021; URL: <https://elephantinthelab.org/how-to-find-the-right-survey-tool-problems-and-proposals/>.

**Nenno, S., & Richter, C.:** Sustainable AI – Wie nachhaltig ist Künstliche Intelligenz wirklich? *Digital society blog*; 29. Juni 2022; URL: <https://www.hiig.de/nachhaltige-ki/>

**Nenno, S.:** Inside Hugging Face. *Digital society blog*; 27. Juli 2023; URL: <https://www.hiig.de/inside-hugging-face/>

**Nenno, S.:** Wie können KI-Modelle energiesparender werden? Ein Überblick über aktuelle Trends. *te.ma*; 29. August 2023; URL: <https://te.ma/art/i3kt57/nenno-ki-modelle-energieverbrauch-reduzieren/>

**Züger, T.:** „KI stellt und nicht nur vor technische Fragen, sondern auch vor gesellschaftliche“. Interview mit #ki\_berlin; 13. März 2020, URL: <https://ki-berlin.de/blog/article/interview-dr-theresa-zueger-leiterin-des-ai-society-lab>.

**Züger, T.:** *Mit Open Mind und Open Source – das demokratische Potential gemeinwohlorientierter KI.* Interview mit Prototype Fund; 19. Juli 2021; URL: <https://prototypefund.de/interview-gemeinwohlorientierte-ki/>.

**Züger, T.:** *Künstliche Intelligenz und Frauen.* Interview mit D64 Zentrum für Digitalen Fortschritt; 5. Oktober 2021; URL: <https://d-64.org/interview-zueger-ki-frauen/>.

**Züger, T.:** „Keine handelnde Person, sondern eine Maschine“. Interview mit Sina Metz für *Wissenschaftskommunikation.de*, 20. Juni 2022; URL: <https://www.wissenschaftskommunikation.de/keine-handelnde-person-sondern-eine-maschine-59151/>

**Züger, T.:** CSRD, ein Löwe oder Papiertiger? Über Nachhaltigkeitsstandards künstlicher Intelligenz. *Digital society blog*; 5. Juli 2022; URL: <https://www.hiig.de/standards-nachhaltiger-ki/>

**Züger, T.:** Lernen mit Algorithmen. "Technologische Infrastrukturen kreieren Abhängigkeiten". Interview mit Lorena Jaume-Palás und Nina Heinrich für *bpb.de*, 18. Juli 2022; URL: <https://www.bpb.de/lernen/digitale-bildung/werkstatt/510774/technologische-infrastrukturen-kreieren-abhaengigkeiten/>

**Züger, T.:** *Verantwortung tragen – das können nur Menschen.* Interview mit zikzak; 01. Oktober 2022, URL: <https://zikzak.ch/preview/ER69-YSD2>

**Züger, T.:** Dietrich, K.: Wie künstliche Intelligenz der Gesellschaft dienen kann. *deutschlandfunkkultur*; 15. Januar 2023; URL: <https://www.deutschlandfunkkultur.de/gemeinwohlorientierte-ki-bessere-beratung-dank-lernen-der-software-100.html>

**Züger, T.:** Breitband Podcast: *KI muss nicht dem Kapital dienen.* *deutschlandfunkkultur*; 20. Mai 2023; URL: <https://www.deutschlandfunkkultur.de/ki-kapitalismus-katalysator-oder-gemeinwohlorientierte>

-technologie-dlf-kultur-8beec82e-100.html

**Züger, T., & Pothmann, D.:** The AI Transparency Cycle. *Digital society blog*; 24. März 2023; URL: <https://www.hiig.de/en/ai-transparency/>

#### Open Talks (Diskursives Format zum Wissenstransfer)

**Open Talk mit Anna Mätzener:** Automatic decision making systems, 20.03.2022, URL: <https://www.youtube.com/watch?v=i0675a1DpLk>.

**Open Talk mit Julia Gundlach:** Algorithms for the common good, 10.02.2022, URL: <https://www.youtube.com/watch?v=3sTnKflpwBY>.

**Open Talk mit Stephan Dreyer:** Public service media & public value, 11.03.2022, URL: <https://www.youtube.com/watch?v=zkAuOq-vMEc>.

#### Webinare

**Pothmann, D., & Züger, T.:** *Webinar I: Mapping Public Interest AI*. 22.08.2023. Humboldt Institute for Internet and Society, Berlin, Germany (International)

**Pothmann, D., & Züger, T.:** *Webinar II: Mapping Public Interest AI*. 24.08.2023. Humboldt Institute for Internet and Society, Berlin, Germany (International)

Abschlusskonferenz (Teil von AP7): Die Konferenz diente der finalen Präsentation, Dissemination und Reflektion der Forschungsergebnisse mit Vertreter*innen aus der Zivilgesellschaft, Wissenschaft und Wirtschaft.	Q4/2023	Die Abschlusskonferenz wurde in Form von zwei Halbtages-Workshops auf einschlägigen Konferenzen, der KI 2023 und der AoiR 2023, durchgeführt (siehe Tabelle für Einzelheiten unter Organisierte Veranstaltungen). Dieses Format ermöglichte einen hohen Impact für die Präsentation und Diskussion der Forschungsergebnisse.
<b>Publikationen</b>		
Projektwebseite: Diese dient der Vorstellung des gesamten Forschungsvorhabens sowie der Dokumentation der entwickelten Prototypen und weiteren Projektergebnisse.	Launch Q4/2022	Online unter: <a href="https://www.hiig.de/project/public-interest-ai/">https://www.hiig.de/project/public-interest-ai/</a>
Wissenschaftliche Artikel, Veröffentlichung in relevanten und hochrangigen wissenschaftlichen Publikationsreihen <ul style="list-style-type: none"><li>je 4- 5 wiss. Paper von FGL und Post Doc</li></ul>	gesamte Projektlaufzeit	insgesamt bereits 22 Veröffentlichungen in einschlägigen wissenschaftlichen Publikationen (siehe Zwischenberichte 2021, 2022 und 2023), sowie 2 eingereichte, aber nicht bis September 2024 veröffentlichte Publikationen

- je 3-4 wiss. Paper pro Doktorand\*in, Veröffentlichung im Kontext von relevanten Fachtagungen

#### Wissenschaftliche Publikationen in alphabetischer Reihenfolge:

**Asghari, H., Fassbender, J., Züger, T. (2022):** *Gekommen, um zu bleiben. Zu der Frage, wie gemeinwohlorientierte KI wirtschaftlich tragfähig sein kann.* In: Zivilgesellschaft 4.0. Künstliche Intelligenz sozial, nachhaltig und partizipativ gestalten. Bundesministerium für Familie, Senioren, Frauen und Jugend (Hrsg.), Berlin, 38-40.

**Asghari, H. & Hewett, F. (2022).** HIIG at GermEval 2022: Best of Both Worlds Ensemble for Automatic Text Complexity Assessment. *Proceedings of the GermEval 2022 Workshop on Text Complexity Assessment of German Text*, 15–20.

**Asghari, H., Hewett, F., & Züger, T. (2023).** On the Prevalence of Leichte Sprache on the German Web. *ACM WebSci '23 Conference Proceedings*, 147-152. DOI: 10.1145/3578503.3583599

**Asghari, H., Birner, N., Burchardt, A., Dicks, D., Fassbender, J., Feldhus, N., Hewett, F., Hofmann, V., Kettemann, M. C., Schulz, W., Simon, J., Stolberg-Larsen, J., & Züger, T. (2022).** What to explain when explaining is difficult. An interdisciplinary primer on XAI and meaningful information in automated decision-making. *Humboldt Institute for Internet and Society*, 1-23. DOI: 10.5281/zenodo.6375784

**Asghari, H., Mahieu, R., Parsons, C., van Hoboken, J., Crete-Nishihata, M., Hilts, A., & Anstis, S. (2021).** *Measuring the Brussels Effect through Access Requests: Has the European General Data Protection Regulation Influenced the Data Protection Rights of Canadian Citizens?* In: *Journal of Information Policy*, 11, 301-349. DOI: 10.5325/jinfopoli.11.2021.0301.

**Asghari, H. & Nenno, S. (eingereicht August 2024).** Mechanistic Interpretability of Socio-Political Frames in Language Models: an Exploration. In: *Proceedings of the ECML/PKDD 2024 Workshop on Advances in Interpretable Machine Learning and Artificial Intelligence (AIMLAI)*.

**Asghari, H., Stolberg-Larsen, J., & Züger, T. (2022).** *Approximating Accessibility of Regions from Incomplete Volunteered Data.* In: CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI '22 Extended Abstracts), April 29-May 5, 2022, New Orleans, LA, USA. ACM, New York, NY, USA. DOI: 10.1145/3491101.3519706.

**Asghari, H., van Biemen, T., & Warnier, M. (2021).** *Amplifying Privacy: Scaling Up Transparency Research Through Delegated Access Requests.* In: 5th Workshop on Technology and Consumer Protection (ConPro '21). arXiv: 2106.06844

**Asghari, H., & Züger, T. (2024).** Es gibt keinen Algorithmus gegen Hass. In M. - S. Adeoso, E. Berendsen, L. Fischer & D. Schnabel (Hrsg.), *Code & Vorurteil. Über Künstliche Intelligenz*,

*Rassismus und Antisemitismus* (S. 49 - 60). Verbrecher Verlag Berlin.

**Fassbender, J. (2021).** *Particles of a whole: design patterns for transparent and auditable AI-systems.* In: Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings, 272-275. DOI: 10.1145/3460418.3479345.

**Fassbender, J., & Henderson, T. (2024).** Participatory to what end? Mapping motivations for participatory approaches in data-driven projects. *GoodIT '24: Proceedings of the 2024 International Conference on Information Technology for Social Good*, 301–305. DOI: 10.1145/3677525.3678675

**Fassbender, J., Kühnlein, I. & Henderson, T. (eingereicht September 2024).** Facing the ambiguities of participation in data-driven projects: A systematic literature review. *Data & Policy*.

**Hewett, F., Asghari, H., Stede, M. (2024).** Elaborative Simplification for German-Language Texts. *Proceedings of the 25th Annual Meeting of the Special Interest Group on Discourse and Dialogue*.

**Hewett, F., & Stede, M. (2021).** Automatically evaluating the conceptual complexity of German texts. In: Proceedings of the 17th Conference on Natural Language Processing (KONVENS 2021), 228–234.

**Hewett, F. & Stede, M. (2022).** Extractive Summarisation for German-language Data: A Text-level Approach with Discourse Features. *Proceedings of the 29th International Conference on Computational Linguistics (COLING)*, 756-765.

**Hewett, F. (2023).** APA-RST: A Text Simplification Corpus with RST Annotations. *Proceedings of the 4th Workshop on Computational Approaches to Discourse (CODI 2023)*, 173-179.

**Hewett, F. (2024).** Komplexitätsmerkmale auf der Textebene: Eine systematische Korpusanalyse. *Abstractband: "Einfach oder komplex? Befunde zur Passung geschriebener Sprache für verschiedene Zielgruppen"*, 1-4.

**Nenno, S. (2024).** Propositional Claim Detection: A task and dataset for the classification of claims to truth. *Journal of Computational Social Science*. DOI: 10.1007/s42001-024-00289-0

**Nenno, S. (2024).** Bootstrapping public entities. Domain-specific NER for public speakers. *Communication Methods and Measures*, 1-26. DOI: 10.1080/19312458.2024.2388098

**Nenno, S. (2024).** Is checkworthiness generalizable? Evaluating task and domain generalization of datasets for claim detection. *Neural Computing & Applications*, 36, 15165–15176. DOI: 10.1007/s00521-024-09896-4

**Nenno, S. (2024).** Potentials and Limitations of Active Learning for the Reduction of Energy Consumption during Model Training. *Weizenbaum Journal of the Digital Society*, 4(1), 1-29. DOI: 10.34669/WI.WJDS/4.1.3

**Züger, T., & Asghari, H. (2022).** AI for the public. How public interest theory shifts the discourse on AI. *AI & Society*. DOI: 10.1007/s00146-022-01480-5

**Züger, T., Kuper, F., Fassbender, J., Katzy-Reinshagen, A., Kühnlein, I. (2023).** Handling the hype: Implications of AI hype for public interest tech projects. *TATuP – Journal for Technology Assessment in Theory and Practice*, 32(3), 34-40. DOI: 10.14512/tatup.32.3.34

**Züger, T. & Asghari, H. (2024).** Introduction to the special issue on AI systems for the public interest. *Internet Policy Review*, 13(3). DOI: 10.14763/ 2024.3.1802

**Züger, T. & Asghari, H. (2024).** AI systems for the public interest [Special issue]. *Internet Policy Review*, 13(3). DOI: DOI: 10.14763/ 2024.3.1802

**Arbeitspapiere:**

**Iglesias Keller, C. & Züger, T. (2023).** Promises and perils of democratic legitimacy in Social Media Councils. *M. C. Kettemann, J. Francke, C. Dinar, & L. Hinrichs (Eds.), Plattform://Democracy. Perspectives on Platform Power, Public Values and the Potential of Social Media Councils: Research Report Europe*, 36-42.

Drei Doktorarbeiten in Form von kumulativen Dissertationen, die jeweils bereits veröffentlichte Artikel inkludieren.	Q4/2024	Eine Doktorarbeit wurde eingereicht, die Verteidigung findet voraussichtlich in Q1/2025 statt. Zwei Doktorarbeiten sind kurz vor der Einreichung und werden planmäßig 2025 verteidigt.
Beiträge zu habilitationsäquivalenten Leistungen durch FGL und Post-Doc.	Q4/2024	Wie beschrieben wurde eine äquivalente Leistung zu den Monographien in einzelnen Veröffentlichungen angestrebt. Im Zeitraum der Forschungsgruppe konnten durch den FGL 4 Veröffentlichungen als Hauptautor und 5 Veröffentlichungen durch den Post-Doc als Hauptautor erbracht werden.
<b>Weitere Verwertungsformate</b>		
Demonstrator: geplant ist eine interaktive Applikation, die sowohl auf Konferenzen als auch online oder bei Events am HIIG als Veranschaulichung der Ergebnisse des Projekts dienen kann.	Entwicklung in 2021 / Nutzung ab 2022	Online unter: <a href="http://www.publicinterest.ai">www.publicinterest.ai</a> Auf der Website finden sich die Anwendungen: - der Simba Text Assistant - das Claimspotting Tool

		Weitere Projektergebnisse: - die Public Interest AI Project Map - ein Stakeholder Index - die Definition und Bedingungen für gemeinwohlorientierte KI
OER in Form einer Poster-Journey	2023	Entwicklung von Postern zum Thema Bias und Bias Mitigation als OER mit Irma Mastenbroek, Präsentation auf der langen Nacht der Wissenschaften 2024
	Q4/2024	Veröffentlichung von OER Material in Form von Zines und zwei Postern
Veröffentlichung von Datensätzen	Q1/2023	<p>Neeno, S. (2023). Propositional Claim Detection, (NLP Datensatz). Data File Version 1.0.0, <a href="https://doi.org/10.7802/2538">https://doi.org/10.7802/2538</a>. URL: <a href="https://search.gesis.org/research_data/SDN-10.7802-2538?doi=10.7802/2538">https://search.gesis.org/research_data/SDN-10.7802-2538?doi=10.7802/2538</a></p> <p>Hewett, F. (2023) APA-RST: A Text Simplification Corpus with RST Annotations. URL: <a href="https://github.com/fhewett/apa-rst">https://github.com/fhewett/apa-rst</a></p> <p>Hewett, F., Stede, M. (2021) Lexica corpus. A corpus for German text simplification. URL: <a href="https://zenodo.org/records/6319803">https://zenodo.org/records/6319803</a></p>
Unser Konzept gemeinwohlorientierter KI soll auch außerhalb der Forschungskreise bekannter gemacht werden. Dafür werden verstärkt auch Vorträge auf Veranstaltungen, die gezielt die Brücke zwischen Forschung und Praxis schlagen, angestrebt.	Q1/2024	DATA Festival "Getting real with AI" von BARC am 19. -20. März 2024 durch Theresa Züger und Sami Neeno, wo das Konzept und die Prototypen der Gruppe präsentiert wurden: <a href="https://barc.com/events/data-festival/">https://barc.com/events/data-festival/</a>
Folgeantrag: Zur Erweiterung der Forschung zu Public Interest AI ist im letzten Jahr des Forschungsprojekts ein Folgeantrag geplant, wodurch das erarbeitete Modell in internationalen Kontexten erprobt und erweitert werden soll. Infrage kommen hier vor allem europäische Förderlinien wie Horizon Europe.	Q2/2024	<p>erfolgreich eingereicht bei Volkswagen Stiftung: <a href="https://www.volkswagenstiftung.de/de/foerderung/foerderangebot/change-fellowships-and-research-groups">https://www.volkswagenstiftung.de/de/foerderung/foerderangebot/change-fellowships-and-research-groups</a></p> <p>Das Projekt entwickelt auf Basis unserer Kenntnisse zu gemeinwohlorientierter KI und weiteren Ansätzen eine Evaluationsmethode für gemeinwohlorientierte KI-Projekte.</p>
Fertigstellung von Prototypen als Showcases:	Q4/2024	

<p>Anwendung für Übersetzung in einfache Sprache, Dokumentation dieser auf der Projektwebseite und wenn möglich als funktionierende Anwendung / funktionale Erweiterung in den jeweiligen Entwicklungskontexten</p> <p>Anwendung für Faktenchecker*innen zum effizienteren Filtervorgang von potentiellen Misinformationen</p>		<p>nach der Veröffentlichung eines Plug-Ins für Vereinfachung /Zusammenfassung (Simba Text Assistant) streben wir nun die Zusammenarbeit mit verschiedenen Praxispartner*innen an wie dem DRK und Einrichtungen der Diakonie</p> <p>Veröffentlichung des Claimspotting Tools als Monitor für potentielle Misinformationen auf Telegram. Zusammenarbeit findet bereits mit der Deutschen Welle (dw) statt. Darüber hinaus wird eine Zusammenarbeit mit weiteren Faktencheck Organisationen angestrebt (AR Faktenfinder, BR Faktenfuchs).</p>
<p>Interdisziplinäre Veröffentlichung</p>	<p>Q4/2024</p>	<p>Special Issue zum Thema "AI in the public Interest" im Internet Policy Review, einem Open Source Journal</p>
<p>Übertragbares Analyseverfahren zu Public Interest AI, das als Prüfung der Gemeinwohlorientierung auch zur Beurteilung von industriellen Projekten verwendet werden kann.</p> <p>Publikation dazu in geeigneten Kanälen, beispielsweise auf der Webseite, als Artikel in einer industrienahen Publikation oder in Form von Vorträgen auf industrie-relevanten Fachkonferenzen</p>	<p>Q1/2023</p>	<p>Veröffentlichung im Rahmen eines Sammelbandes des BMAS März 2023: <a href="https://www.bundesregierung.de/breg-de/suche/kuenstliche-intelligenz-2173100">https://www.bundesregierung.de/breg-de/suche/kuenstliche-intelligenz-2173100</a></p> <p>Des weiteren Vorträge auf industrienahen Konferenzen wie beispielsweise AI:NOW 2024 oder DATA Festival 2024</p>
<p>Digitaler Salon: Einmal monatlich stattfindende Veranstaltung mit Gäst*innen zu Fragen über vernetzte Gegenwart. In dieser Diskussionsreihe werden Internet-Phänomene beleuchtet sowie Vor- und Nachteile technischer Veränderungen erörtert.</p>	<p>Q3/2021</p>	<p>Thematische Einführung von Judith Faßbender beim Digitalen Salon zu "Cookies make you lose control" über verschiedene Cookie-Richtlinien, Datenschutzbestimmungen und Nutzungseinwilligungen</p>
	<p>Q3/2023</p>	<p>Thematische Einführung von Sami Nenno beim Digitalen Salon zu "Friede, Freude, Meinungsvielfalt? Was die neuen EU-Spielregeln (nicht) ändern und die neue-alte Rolle von Social-Media-Plattformen"</p> <p>Thematische Einführung von Theresa Züger beim Digitalen Salon zu "Frozen</p>

		Eye – Die Zukunft der Gesichtserkennung im öffentlichen Raum”
	Q1/2024	Thematische Einführung von Irina Kühnlein beim Digitalen Salon zu “ Digitaler Salon: Quanten statt Quaken – Steckt die Rettung des Klimas in KI und Supercomputern?”
	Q2/2024	Thematische Einführung von Freya Hewett beim Digitalen Salon zu “Wie bitte?” über einfache Sprache
Teilnahme an Veranstaltung der deutschen Vertretung der Europäischen Kommission	Q2/2023	Sarah Spitz, Theresa Züger und Irma Mastenbroek bei einem Roundtable zum Thema “Technology is Never (Gender) Neutral” mit Margrethe Vestager (Mitglied der Europäischen Kommission)
Etablierung des Themas Public Interest AI auf wissenschaftlichen Konferenzen	2023ff.	jährlicher Public Interest AI Workshop auf der German Conference on Artificial Intelligence. Gesellschaft für Informatik (GI)

#### Wissenschaftliche und wirtschaftliche Anschlussfähigkeit

Eine wirtschaftliche und wissenschaftliche Anschlussfähigkeit an die Ergebnisse der Forschungsgruppe besteht auf mehreren Ebenen: Zum einen auf der Ebene möglicher anschließender Forschungsprojekte, zum anderen auf der Ebene der professionellen Weiterentwicklung der Forschenden. Tatsächlich wurde eine Anschlussförderung zur Weiterentwicklung des Forschungsfeldes mit einer Forschungsgruppe zu Beginn 2025 bereits bei der VolkswagenStiftung eingeworben. Des Weiteren hat die hervorragende Qualifikation, die die Forschenden im Rahmen der Forschungsgruppe erworben haben, sie dazu befähigt sehr erfolgreich nächste Karriereschritte zu gehen und Folgeanstellungen (beispielsweise als Post-Docs an der TU Berlin und wissenschaftlicher Mitarbeiter an der TU Dresden) zu finden.

#### Mögliche Folgeprojekte

Die Forschungsergebnisse der gesamten Forschungsgruppe bieten eine Grundlage für Folgeprojekte im europäischen Raum, die sowohl für die Wirtschaft, als auch für die politische und soziotechnische Forschung relevant sind.

Das entwickelte Konzept von gemeinwohlorientierter KI ist in verschiedensten Kontexten, wirtschaftlichen wie auch gemeinnützigen, anwendbar und kann somit zur Analyse und Beratung oder der Begleitforschung dienen. Besonders für KI-Anwendungen in der Verwaltung oder anderen öffentlichen Infrastrukturen wird der Anspruch, dem Gemeinwohl zu dienen, in den kommenden Jahren stetig an Relevanz gewinnen. Hierzu wurden bestehende Partnerschaften, z.B. mit der Deutscher Welle, ausgebaut,

um ggf. gemeinsame Projekte zu entwickeln.

Weiter ist es möglich und sinnvoll, die bestehenden Forschungserkenntnisse im Rahmen von Folgeprojekten im europäischen Raum zu validieren und an weiteren Fallbeispielen zu erproben. Die in den Projekten entwickelten Prototypen in den Bereichen der automatisierten Sprachverarbeitung (NLP) stellen auf technischer Ebene einen signifikanten Beitrag zum Stand der Forschung dar: Durch die Offenlegung der Ergebnisse können die erarbeiteten Modelle und Vorgehen in zukünftigen wissenschaftlichen Projekten weiterentwickelt werden, aber auch insbesondere durch europäische KMUs in Produkte integriert werden.

Konkret wird im erfolgreich eingeworbenen Folgeprojekt die Evaluation von gemeinwohlorientierten KI-Projekten weiter erforscht. Des Weiteren zielt ein weiteres beantragtes Projekt darauf ab, auf den Ergebnissen des AP 5 (Claimspotting) im Kontext der Erkennung von Misinformation aufzubauen. Die Rückmeldung durch den Fördergeber steht hier noch aus.

## **5. Fortschritte im Bereich Public Interest AI außerhalb der Forschungsgruppe**

Das Civic Coding Innovationsnetzwerk arbeitet weiterhin an der Förderung gemeinwohlorientierter KI und wurde unserer Kenntnis nach bis 2026 verlängert. International gewinnt das Thema ebenso an Relevanz, z. B. auf dem im Februar in Paris geplanten AI Action Summit, bei dem sich ein ganzer Programmtrack dem Thema Public Interest AI widmet. Theresa Züger konnte sich hier als Teil der Arbeitsgruppe für den AI Action Summit einbringen. Des Weiteren wurde wie erwähnt eine Folgeförderung bei der VW Stiftung über 5 Jahre in Form eines Change! Fellowships eingeworben, wodurch die Forschungsergebnisse der Gruppe als Basis dienen und von einem neuen Forschungsteam weiterentwickelt werden.

## **6. Veröffentlichungen der Ergebnisse**

Siehe Seite 14ff.