

Agenda

11. OJA Forum

18. & 19. November 2026

Wir weisen darauf hin, dass im Rahmen dieser Veranstaltung Fotos für öffentliche sowie nicht-öffentliche Zwecke gemacht werden und auf eine evtl. Nutzung dieser auf projekteigenen Social Media-Websites wie z. B. X, Blog, LinkedIn, Instagram. Sollten Sie damit nicht einverstanden sein, teilen Sie uns das bitte vorab oder zu Beginn der Veranstaltung mit. Vielen Dank!

Veranstaltungsort:

**Bertelsmann Stiftung
Werderscher Markt 6
10117 Berlin**

Kontakt:

Jana Fingerhut
Telefon: +49 5241 81-81393
E-Mail: jana.fingerhut@bertelsmann-stiftung.de

Stefan Winnige
Telefon: +49 228 107 1408
E-Mail: winnige@bibb.de

1. Tag | Mittwoch, 18. November 2026

11:45 Uhr **Ankunft**

12:00 Uhr **Mittagessen**

13:00 Uhr **Begrüßung**

Session 1 | Datenbanken

13:15 Uhr **Reproduzierbare LLM-Pipelines für die Analyse von Online-Stellenanzeigen**

Simon Buchert & Dr. Uwe Remer – Fraunhofer-
Informationszentrum Raum und Bau IRB

Sprachmodelle unterscheiden sich stark in Größe, Geschwindigkeit und Preis, doch welches Modell braucht welcher Analyseschritt wirklich? Wir stellen eine modulare Pipeline für Zoning, Kompetenzextraktion und Taxonomie-Matching vor und benchmarken pro Schritt verschiedene Modelle entlang von Genauigkeit, Inferenzzeit und Kosten. Zusätzlich zeigen wir das zugrundeliegende Datenmodell und die Verknüpfung zu Taxonomien wie ISCO, ESCO und KldB. Mitnehmen lässt sich ein Bewertungsrahmen, der sich auf eigene Pipelines übertragen lässt.

13:45 Uhr **Von Rohdaten zum Jobmonitor: Aufbau einer Dateninfrastruktur für Online-Stellenanzeigen**

Rahkakavee Baskaran – &effect data solutions GmbH

Dieser Vortrag zeigt, wie aus täglich rund 250.000 neuen Online-Stellenanzeigen die Zahlen im Jobmonitor entstehen. Vorgestellt wird der Weg der Daten von der Rohanzeige über Bereinigung und 15 orchestrierte Machine-Learning-Modelle bis zur fertigen Auswertung, aggregiert entlang verschiedener Konzepte wie Berufsgruppen oder Kompetenzen. Der Vortrag gibt einen Einblick darin, welche Herausforderungen etwa bei Datenqualität, Skalierung und Aktualität entstehen und gelöst werden, wenn aus einer einmaligen Analyse eine laufende Anwendung mit mittlerweile über 600 Millionen erfassten Stellenausschreibungen wird.

14:15 Uhr **ETL-Prozesse für Stellenanzeigen-Daten: Unser Workflow im BIBB**
Kristine Hein – Bundesinstitut für Berufsbildung (BIBB)

Der Vortrag gibt einen Einblick in unseren hausinternen Workflow zur Verarbeitung und Haltung von Stellenanzeigen-Daten. Gezeigt wird der Weg vom regelmäßigen Abruf beim Dienstleister über ETL-Prozesse, Bereinigung und Aufbereitung bis zur Aktualisierung der Datenbank. Dabei geht es insbesondere um die Prozesse zur Aktualisierung der Daten und das Einspielen ins DataWarehouse sowie um automatisierte Tests und Validierungen, mit denen wir die Qualität und Konsistenz der Datenbasis sicherstellen. Des weiteren gehen wir auf die Automatisierung der Prozess Pipelines zum Updaten der Rohdaten ein, sowie die Orchestrierung der DataWarehouse Prozesse und ihrer Abhängigkeiten und das Monitoring im laufenden Betrieb.

14:45 Pause

Session 2 | Ausbildung

15:15 Uhr **Generative KI und das Angebot von Ausbildungsplätzen: Eine Analyse mit deutschen Vakanzdaten**
Michael Stops – Institut für Arbeitsmarkt- und Berufsforschung (IAB)

Der rasante technologische Wandel verändert die Nachfrage von Unternehmen nach bestimmten Kompetenzen, wobei die Verbreitung generativer künstlicher Intelligenz (GenAI) diesen Prozess voraussichtlich beschleunigen wird. Diese Studie untersucht, wie sich GenAI auf die Nachfrage von Betrieben nach Auszubildenden in Deutschland auswirkt. Unter Verwendung von Daten zu offenen Ausbildungsstellen nutzen wir die weltweite Veröffentlichung von ChatGPT-3.5 im November 2022 als exogenen technologischen Schock. Wir stellen fest, dass die Nachfrage nach Auszubildenden in Berufen mit hoher GenAI-Exposition im Vergleich zu weniger exponierten Berufen signifikant zurückgeht. Wir zeigen, dass dieser Rückgang wahrscheinlich nicht auf allgemeine Trends der Arbeitskräftenachfrage oder Engpässe beim Arbeitskräfteangebot zurückzuführen ist. Unsere Ergebnisse deuten darauf hin, dass GenAI bereits die Beschäftigungsmöglichkeiten für Berufseinsteiger verändert, was potenzielle Auswirkungen auf die Ungleichheit zwischen den Generationen und die künftige Entwicklung der Erwerbsbevölkerung haben könnte.

15:45 Uhr **„Ausbildung im Fokus“ – Verknüpfung von Ausbildungsstellenanzeigen und einer Befragung der inserierenden Betriebe**

Stefan Winnige – Bundesinstitut für Berufsbildung (BIBB)

Die Befragung „Ausbildung im Fokus“ kombiniert Informationen aus Ausbildungsstellenanzeigen mit einer Befragung der inserierenden Betriebe. Nach einer Pilotstudie in Nordrhein-Westfalen in 2025 wird die Erhebung im

Herbst 2026 erstmals bundesweit durchgeführt. Der Vortrag stellt zunächst praktische und methodische Erfahrungen bei der Verknüpfung der Daten sowie der Nutzung der in den Stellenanzeigen verfügbaren E-Mail-Adressen für die Befragung vor. Im zweiten Teil werden Ergebnisse der Pilotstudie präsentiert. Im Fokus steht der zusätzliche Erkenntnisgewinn aus der Verknüpfung beider Datenquellen, insbesondere zur Identifikation von hemmenden und fördernden Faktoren bei der Besetzung von Ausbildungsstellen.

16:15 Uhr Sprachanforderungen in Jobprofilen – Auswertung von Stellenanzeigen für Hochschulabsolvent*innen in Deutschland
Nicolas Reum – Centrum für Hochschulentwicklung (CHE)

Welche Sprachkenntnisse erwarten Arbeitgeber von Hochschulabsolvent*innen? Dieser Frage geht der Beitrag auf Grundlage des Jobmonitors der Bertelsmann Stiftung nach. Analysiert werden Stellenanzeigen in Deutschland, die sich ausdrücklich an Bewerber*innen mit Bachelor- oder Masterabschluss beziehungsweise Promotion richten. Die Daten ermöglichen sowohl bundesweite als auch regionale, studienbereichs- und berufsbezogene Auswertungen. Dabei zeigt sich, dass Sprachkenntnisse für Hochschulabsolvent*innen verbreitet zum Anforderungsprofil gehören. In rund jeder zweiten untersuchten Stellenanzeige werden sie ausdrücklich genannt. Dabei dominieren zwei Sprachen: Deutsch und Englisch. Wie häufig Arbeitgeber Englisch- oder Deutschkenntnisse verlangen, unterscheidet sich aber erheblich nach Beruf und Studienfach.

16:45 Uhr Pause

Session 3 | Poster Session

17:15 Uhr Poster Pitch

Aufstiege ermöglichen - Kompetenzprofile mit dem Jobmonitor vergleichen

Hauke Janson / Gunvald Herdin – Bertelsmann Stiftung

Buchvorstellung: Skilled - How to Master a Changing World of Work

Fabian Stephany – University of Oxford

BSt-Destatis Studie: Ein mehrstufiges Verfahren zur Berechnung von Konjunkturindizes aus Online-Stellenanzeigen

Dr. Martina Rengers – Statistisches Bundesamt (Destatis)

Titel

Vanessa Weeke – Bertelsmann Stiftung

- 17:25 Uhr **Betrachtung der Poster und Diskussion**
- 18:30 Uhr **Ende erster Tag**
- 19:30 Uhr **Gemeinsames Abendessen – Ideen: Marienkäfer Pizzeria, Italo Fritzen**

2. Tag | Donnerstag, 19. November 2026

- 08:45 Uhr **Ankunft**
- 09:00 Uhr **Begrüßung**

Session 4 | Einblicke in die Praxis

- 09:05 Uhr **Titel**
Christina Langer – Stepstone

Die Teaser
- 09:35 Uhr **Titel**
Nina Frühling – Deutsche Bahn (angefragt)

Die Teaser
- 10:05 Uhr **Anzeige inklusiv umgeschrieben – noch die gleiche Persönlichkeit gesucht? Eine HEXACO-Analyse KI-gestützten Rewritings von Stellenanzeigen**

PD Dr. Olena Linnyk – milch&zucker GmbH / Embrace.family / Justus-Liebig-Universität Gießen

Inklusive Sprache in Stellenanzeigen gilt als wirksamer Hebel, um mehr qualifizierte Bewerber:innen zu erreichen. Doch was passiert dabei mit der impliziten Botschaft der Anzeige – dem psychologischen Anforderungsprofil, das zwischen den Zeilen kommuniziert wird? Bleibt es beim Umschreiben erhalten

oder wird es verzerrt? Dieser Frage sind wir empirisch nachgegangen und haben mehrere Tausend Online-Stellenanzeigen automatisiert mit zwei KI-Systemen verarbeitet. Das erste, betterads, schreibt Anzeigen automatisch um und ersetzt agentische (männlich konnotierte) sowie absolutistische (Fixed-Mindset-)Formulierungen durch genderneutrale, kommunale bzw. Growth-Mindset-Sprache. Das zweite, von uns entwickelte System extrahiert aus dem Anzeigentext das gesuchte HEXACO-basierte Persönlichkeitsprofil; eine Vorstudie hat die Präzision dieser Extraktion validiert. Durch den Vergleich der extrahierten Profile vor und nach dem Umschreiben zeigen wir, wie stabil psychologische Anforderungsprofile unter inklusivem Rewriting bleiben – und was das für die automatisierte Anzeigenoptimierung in der Praxis bedeutet.

10:35 Uhr Pause

Session 5 | Wie verändert KI die OJA-Analysen

11:00 Uhr **Welche Ansätze für welche Aufgaben? Generative KI in der Stellenanzeigenanalyse**

Ann-Sophie Gnehm – UZH Zürich

Generative KI eröffnet neue Möglichkeiten für die Klassifikation und Informationsextraktion – von der Erzeugung und iterativen Verbesserung von Trainingsdaten bis hin zum direkten und agentischen Einsatz von Sprachmodellen. Anhand aktueller Projekte mit Schweizer Stellenanzeigen zeigt der Beitrag, wie LLMs mit strukturiertem Wissen, spezialisierten Modellen und menschlicher Expertise kombiniert werden können. Beispiele sind die Klassifikation von Berufen und Bildungsanforderungen sowie die Geokodierung von Arbeitsorten. Im Mittelpunkt steht dabei die Frage, welche Ansätze sich für welche Aufgaben eignen und wie sich die Rolle menschlicher Expertise verändert – von der Annotation über die methodische Konzeption bis zur Qualitätssicherung. Abschließend werden Herausforderungen für eine unabhängige Evaluation und die Validität der Ergebnisse diskutiert.

11:30 Uhr **LLM-basierte Extraktion strukturierter Anforderungen aus Online-Stellenanzeigen**

Kai Krüger – Bundesinstitut für Berufsbildung

Wie gut eignen sich Large Language Models für die Extraktion komplexer Anforderungsstrukturen aus Online-Stellenanzeigen? Der Vortrag stellt erste Experimente mit dem GOJA-Datensatz vor, in dem Anforderungen als strukturierte Kombinationen aus Entitäten und Relationen annotiert sind. Im Mittelpunkt stehen die Extraktionsqualität der LLMs, praktische Herausforderungen bei der Anwendung auf diese vergleichsweise komplexe Informationsstruktur sowie Überlegungen zur Inferenzzeit bei der Verarbeitung größerer Datensätze.

12:00 Uhr Zur Rolle von Datenannotation und kosteneffizientem Sampling in Zeiten von LLMs

Julia Romberg – GESIS – Leibniz-Institut für Sozialwissenschaften

Um maschinelle Lernmodelle auf spezifische Aufgaben auszurichten, werden sie häufig mithilfe (großer Mengen) annotierter Daten trainiert. Die Annotation von Daten ist jedoch zeitaufwendig und damit kostspielig. Mit dem Aufkommen leistungsstarker großer Sprachmodelle (LLMs) könnten sich diese Anforderungen jedoch verändern. In diesem Vortrag werden die Ergebnisse einer Umfrage innerhalb der Sprachverarbeitungs-Community vorgestellt, in der die Bedeutung von Datenannotation und kosteneffizientem Sampling in Zeiten von LLMs hinterfragt wird.

12:30 Uhr Mittagessen

14:00 Uhr Ende der Veranstaltung