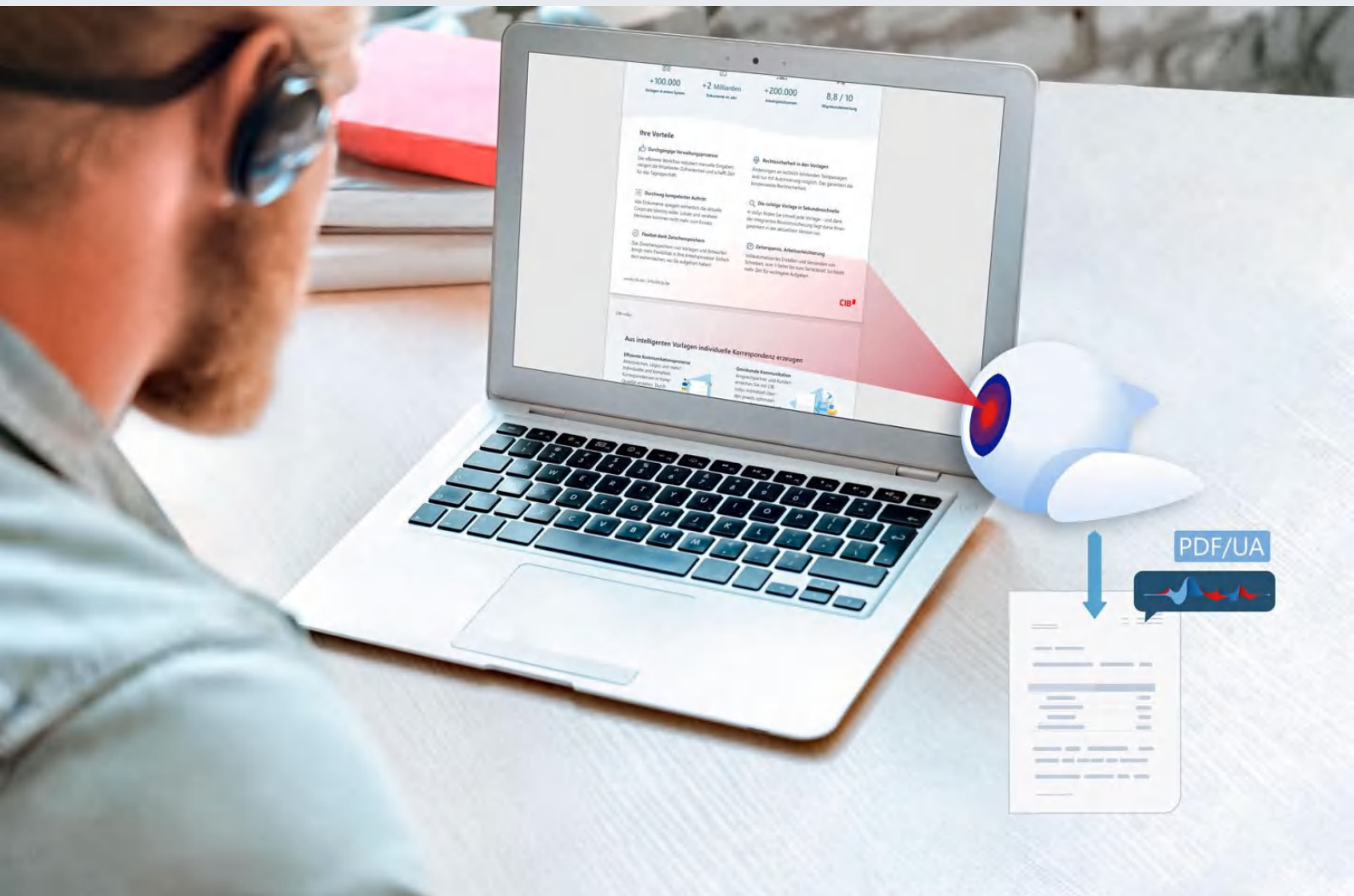


RIDMI

Lesen und Interpretation von Dokumenten durch Maschinenintelligenz



RIDMI

Lesen und Interpretation von Dokumenten durch Maschinenintelligenz

Förderkennzeichen 01IS23004

Laufzeit des Vorhabens: 01.04.2023 – 31.03.2025

Projektträger: Deutsches Zentrum für Luft- und Raumfahrt e. V. (DLR)

Projektpartner:

- CIB AI labs GmbH
- Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS

Projektleitung: Tammo Wüsthoff (CIB AI labs GmbH)

Das diesem Bericht zugrunde liegende Vorhaben wurde mit Mitteln des Bundesministeriums für Forschung, Technologie und Raumfahrt (BMFTR) unter dem Förderkennzeichen 01IS23004 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei der Autorin/beim Autor.

ZUSAMMENFASSUNG	5
DIE WICHTIGSTEN POSITIONEN DES ZAHLENMÄßIGEN NACHWEISES	6
Gesamtüberblick über die Mittelverwendung	6
CIB AI labs GmbH	6
Fraunhofer IAIS	6
NOTWENDIGKEIT UND ANGEMESSENHEIT DER GELEISTETEN PROJEKTARBEITEN	7
Beiträge der CIB AI labs GmbH	7
AP1 – Spezifikation	7
AP2 – Layout-Datensätze	8
AP3 – Transformer für Layout	9
AP4 – Qualitätsmessung für Layout	10
AP5 – Fußzeilen	11
AP6 – Tabellenstruktur	12
AP10 – Projektmanagement	13
Beiträge des Fraunhofer IAIS	13
AP3 – Transformer für Layout	13
AP6 – Tabellenstruktur	14
AP7 – Bilder-Datensätze	15
AP8 – Ersatztexte für Bilder	16
AP9 – Qualitätsmessung für Ersatztexte	16
AP10 – Projektmanagement	17
Technische und organisatorische Herausforderungen	17
Bewertung der Angemessenheit	18
VORAUSSICHTLICHER NUTZEN UND VERWERTBARKEIT DES ERGEBNISSES, AUCH KONKRETE PLANUNG FÜR DIE NÄHERE ZUKUNFT, IM SINNE DES FORTGESCHRIEBENEN VERWERTUNGSPLANS	19
Wirtschaftlicher Nutzen	19
Wissenschaftlicher Nutzen	19
Gesellschaftlicher Nutzen	20
Verwertungsperspektive und Zeitrahmen	20

Verwertungsstrategie.....	20
Risiken und Herausforderungen.....	20
Bereits erzielte wirtschaftliche Erfolge (ROI).....	21
DER WÄHREND DER DURCHFÜHRUNG DES VORHABENS DEM ZUWENDUNGSEMPFÄNGER BEKANNT GEWORDENE FORTSCHRITT AUF DEM GEBIET DES VORHABENS BEI ANDEREN STELLEN	22
ERFOLGTE ODER GEPLANTE VERÖFFENTLICHUNGEN DES ERGEBNISSES NACH NR. 5 DER NKBF	23
Wissenschaftliche Veröffentlichungen	23
Technische Veröffentlichungen.....	23
Öffentlichkeitsarbeit	23
Geplante Veröffentlichungen.....	23
Verwertungsrelevante Veröffentlichungen	24

Zusammenfassung

Das Verbundprojekt RIDMI („Lesen und Interpretation von Dokumenten durch Maschinenintelligenz“) verfolgte das Ziel, die automatisierte Analyse komplexer Dokumente durch den Einsatz moderner KI-Methoden grundlegend zu verbessern. Im Fokus stand die Entwicklung multimodaler Verfahren, die sowohl visuelle als auch semantische Informationen verarbeiten können, um Dokumente strukturell zu verstehen, Inhalte zu extrahieren und barrierefreie Ausgabeformate zu erzeugen.

Ausgangspunkt war die Beobachtung, dass viele bestehende Systeme zur Dokumentenanalyse an den vielfältigen Layouts und Formaten realer Geschäftsdokumente scheitern. RIDMI adressierte diese Herausforderung durch den Aufbau robuster Modelle zu Layout-, Struktur- und Bildverständnis, insbesondere für deutschsprachige Dokumente. Dabei kamen moderne Transformer-Architekturen wie LayoutLMv3, DETR, Florence-2 und Gemma-2 zum Einsatz, ergänzt durch eigens entwickelte Autoencoder-Modelle (DVAE, GVAE) zur Generierung visueller Embeddings.

Ein zentrales Element des Projekts war die Erstellung und Aufbereitung eines umfangreichen, qualitativ hochwertigen Dokumentenkorpus. Über einen eigens entwickelten Web-Crawler wurden mehr als 1,5 Millionen Dokumente aus dem öffentlichen Sektor gesammelt. Ergänzt durch interne Quellen und manuelle Recherchen entstand ein einzigartiger deutschsprachiger Datensatz, der die Grundlage für das Training und die Evaluation der Modelle bildete. Die Annotation erfolgte teils automatisiert, teils über die Plattform CIB crowdsource, wodurch auch Aspekte wie Lesereihenfolge und semantische Struktur erfasst wurden.

Im Projektverlauf wurden mehrere technologische Meilensteine erreicht: eine leistungsfähige Tabellenerkennung (TATR), eine Logo- und Bildinhaltsanalyse, sowie die Integration von Large Language Models (LLMs) zur semantischen Texterkennung. Diese Komponenten wurden in die CIB-Produkte doXiview und doXisafe integriert und ermöglichen dort bereits die prototypische Erstellung barrierefreier PDF-Dokumente.

Die Zusammenarbeit mit dem Projektpartner Fraunhofer IAIS war geprägt von enger Abstimmung, gemeinsamer Spezifikation von Datenformaten und der Integration von Modellen in produktnahe Anwendungen. Die entwickelten Technologien wurden kontinuierlich evaluiert und in realen Anwendungsszenarien getestet.

RIDMI konnte seine Ziele in weiten Teilen erreichen. Die entwickelten Komponenten sind einsatzbereit, erste Kundenlösungen wurden realisiert, und es bestehen konkrete Pläne zur wirtschaftlichen Verwertung. Die Ergebnisse bilden zudem eine solide Grundlage für zukünftige Forschung und Produktentwicklung im Bereich der intelligenten Dokumentverarbeitung. Besonders hervorzuheben ist das Potenzial zur weiteren Verbesserung der Barrierefreiheit, zur Automatisierung von Geschäftsprozessen und zur Erschließung neuer Anwendungsfelder im öffentlichen und privaten Sektor.

Die wichtigsten Positionen des zahlenmäßigen Nachweises

Das Projekt RIDMI („Lesen und Interpretation von Dokumenten durch Maschinenintelligenz“) wurde im Rahmen des Förderprogramms KMU-innovativ als Verbundvorhaben mit den Partnern **CIB AI Labs GmbH** und **Fraunhofer IAIS** durchgeführt. Die bewilligten Fördermittel wurden im Zeitraum vom 01.04.2023 bis 31.03.2025 eingesetzt. Im Folgenden wird die Verwendung der Mittel im Verhältnis zur ursprünglichen Planung dargestellt und bewertet.

GESAMTÜBERBLICK ÜBER DIE MITTELVERWENDUNG

Die im Zuwendungsbescheid bewilligten Mittel wurden planmäßig verwendet. Die Projektlaufzeit wurde vollständig ausgeschöpft. Eine Verlängerung oder Aufstockung der Mittel war nicht erforderlich. Die Mittelverwendung erfolgte im Einklang mit dem Finanzierungsplan, wobei einzelne Positionen im Detail angepasst wurden, um auf operative Erfordernisse zu reagieren.

CIB AI LABS GMBH

Personalkosten

Der größte Teil der Mittel bei CIB entfiel auf Personalkosten. Diese wurden vollständig für projektbezogene Aufgaben eingesetzt. Der geplante Personaleinsatz konnte realisiert werden, wobei es zu **geringfügigen Verschiebungen zwischen den Arbeitspaketen** kam. Insbesondere die Entwicklung und Evaluation des **diskreten visuellen Autoencoders (DVAE)** sowie die **Aufbereitung und Annotation großer Dokumentdatensätze** erwiesen sich als aufwändiger als ursprünglich angenommen. Der Mehraufwand wurde durch interne Umverteilung kompensiert.

Sachkosten und Infrastruktur

Für das Training großer Sprach- und Layoutmodelle wurde geeignete Hardware beschafft. Die Anschaffung erfolgte im Rahmen der geplanten Investitionen. Die Infrastruktur wurde erfolgreich für das Training von **LayoutLMv3, DVAE** und weiteren Modellen eingesetzt. Weitere Sachkosten entfielen auf Softwarelizenzen, Datenmanagement und Visualisierungstools.

Externe Leistungen

Ein Teil der Annotationstätigkeiten wurde über die Plattform **CIB crowdsource** organisiert. Die Plattform wurde intern betrieben, sodass keine externen Dienstleister beauftragt werden mussten. Die Nutzung der Plattform ermöglichte eine kosteneffiziente und skalierbare Erstellung von Groundtruth-Daten.

Projektlaufzeit und Organisation

Die Projektlaufzeit wurde vollständig genutzt. Die Projektorganisation erfolgte über regelmäßige Abstimmungen mit dem Partner Fraunhofer IAIS. Die Koordination und das Projektmanagement lagen im geplanten Aufwand.

FRAUNHOFER IAIS

Personalkosten

Die Personalmittel bei Fraunhofer IAIS wurden im Wesentlichen wie geplant eingesetzt.

Notwendigkeit und Angemessenheit der geleisteten Projektarbeiten

Das Projekt RIDMI war in zehn Arbeitspakete (APs) gegliedert, die logisch aufeinander aufbauten: von der Spezifikation (AP1) über die Datenaufbereitung (AP2), Modellentwicklung (AP3–AP8), Qualitätssicherung (AP4, AP9) bis hin zum Projektmanagement (AP10). Die Arbeitspakete wurden arbeitsteilig zwischen den Partnern CIB AI labs GmbH und Fraunhofer IAIS durchgeführt. Die CIB übernahm primär die Entwicklung und Integration in produktnahe Anwendungen, während Fraunhofer IAIS sich auf die wissenschaftlich-technische Modellforschung konzentrierte.

BEITRÄGE DER CIB AI LABS GMBH

AP1 – Spezifikation

Zielsetzung: Definition der technischen und organisatorischen Schnittstellen.

Durchführung: Die Spezifikation wurde im Kooperationsvertrag geregelt.

Bewertung: Die Spezifikation war notwendig und wurde effizient umgesetzt.

Im Arbeitspaket „**Spezifikation**“ sind einige besonders hervorzuhebende Ergebnisse entstanden, die für den Projekterfolg von zentraler Bedeutung waren:

1. Kooperationsvertrag als zentrales Steuerungsinstrument

Die Zusammenarbeit zwischen CIB AI labs GmbH und dem Fraunhofer IAIS wurde frühzeitig in einem **Kooperationsvertrag** geregelt. Dieser enthielt nicht nur organisatorische und rechtliche Rahmenbedingungen, sondern auch **technische Details zur Zusammenarbeit**, insbesondere zum **Datenaustausch**, zur **Verantwortlichkeitsverteilung** und zu **Schnittstellenanforderungen**. Diese vertragliche Grundlage war entscheidend für die reibungslose und zielgerichtete Projektumsetzung.

2. Verzicht auf separates Spezifikationsdokument zugunsten agiler Abstimmung

In beiderseitigem Einvernehmen wurde auf ein klassisches Spezifikationsdokument verzichtet. Stattdessen wurden die Anforderungen **iterativ und agil** in regelmäßigen technischen Abstimmungen konkretisiert. Diese Vorgehensweise erwies sich als effizient und praxisnah, insbesondere angesichts der dynamischen Entwicklungen im Bereich KI und Layoutanalyse.

3. Frühe Definition von Datenformaten und Schnittstellen

Bereits in der Spezifikationsphase wurden die **Datenformate für Trainings- und Testdaten** sowie die Schnittstellen zwischen den Softwarekomponenten definiert. Für Layoutannotationen kamen etablierte Formate wie hOCR und XML zum Einsatz. Diese frühe Festlegung ermöglichte eine parallele Entwicklung der Teilmodule und reduzierte spätere Integrationsaufwände erheblich.

Im Rahmen des Pretrainings von LayoutLMv3 wurde zusätzlich das TFRecord-Format eingeführt, um komplexe Dokumentrepräsentationen effizient zu speichern und zu verarbeiten. Die TFRecords enthalten multimodale Informationen, die für verschiedene Pretraining-Aufgaben genutzt werden:

Masked-Image-Modeling, Masked-Language-Modeling sowie ein Interaktionstask zwischen den Modalitäten (Text, Layout und Bild). Durch die klare Struktur der TFRecords konnten die Trainingspipelines frühzeitig aufgebaut und unabhängig von anderen Komponenten getestet werden.

4. Identifikation technischer Risiken und Priorisierung

Im Rahmen der Spezifikation wurden potenzielle Risiken frühzeitig identifiziert – etwa Verzögerungen beim Hardwareeinkauf oder Unsicherheiten bei der Datenverfügbarkeit. Diese Erkenntnisse führten zu einer **Priorisierung der Arbeitspakete** und zur **frühzeitigen Beschaffung kritischer Ressourcen**, was sich im weiteren Projektverlauf als vorausschauend erwies.

Diese Ergebnisse zeigen, dass die Spezifikationsphase nicht nur formale Grundlagen geschaffen hat, sondern auch **strategisch und operativ wichtige Entscheidungen** vorbereitet hat, die maßgeblich zum Projekterfolg beigetragen haben.

AP2 – Layout-Datensätze

Zielsetzung: Aufbau eines hochwertigen deutschsprachigen Datensatzes.

Durchführung: Sammlung von über 1,5 Mio. Dokumenten, Entwicklung von Extraktionstools, Einsatz von CIB doXiview und Crowdsourcing zur Annotation.

Bewertung: Die Arbeiten waren zentral für das Projektziel. Der Aufwand war hoch, aber gerechtfertigt.

Im Arbeitspaket **AP2 „Layout-Datensätze“** wurden im Projekt RIDMI zentrale Grundlagen für die spätere Modellentwicklung geschaffen. Hier sind die wichtigsten Ergebnisse und Leistungen im Überblick:

1. Aufbau eines einzigartigen deutschsprachigen Dokumentkorpus

- Sammlung von **über 1,5 Millionen Dokumenten (8 Millionen Dokumentseiten)** aus deutschen Kommunen und Universitäten mittels eines eigens entwickelten Web-Crawlers.
- Aufteilung in:
 - ca. **1,3 Mio. digital erzeugte PDF-Dateien** („digital born“)
 - ca. **200.000 gescannte PDF-Dateien**
 - ca. **8.000 DOCX-Dateien**

2. Entwicklung von Extraktionstools

Zur Vorbereitung der Trainingsdaten wurden drei spezialisierte Werkzeuge entwickelt:

1. Extraktionstool für Groundtruth-Daten aus PDF-Dateien

Dieses Tool ermöglicht die automatische Gewinnung von Layoutinformationen und Textinhalten aus digitalen sowie gescannten PDF-Dokumenten. Es bildet die Grundlage für die Erstellung strukturierter Trainingsdaten für verschiedene Trainingsphasen.

2. Tool zur Erstellung von TFRecords für das unsupervised Pretraining

Basierend auf dem spezifizierten TFRecord-Format werden multimodale Dokumentrepräsentationen aus digital-born Dokumenten und Scans erzeugt. Die Daten enthalten Informationen für Pretraining-Aufgaben wie Masked-Image-Modeling, Masked-Language-Modeling und Modalitätsinteraktionen. Das Tool erlaubt eine konsistente und skalierbare Vorbereitung großer Pretraining-Datensätze.

3. Finetuning-Tools zur Label-Extraktion aus digital-born Dokumenten

Für das supervised Finetuning wurden zwei Werkzeuge entwickelt:

- Eines extrahiert semantische Labels aus digital-born PDFs mit enthaltener Layoutstruktur.
- Ein weiteres Tool verarbeitet DOCX-Dateien und nutzt die bestehende Dokument-Output-Software von CIB zur strukturierten Auslesung von Elementen wie Überschriften, Absätzen und Tabellen.

3. OCR-gestützte Aufbereitung von Scans

- Verwendung der hauseigenen OCR-Engine **deepER** zur präzisen Erkennung von Wortkoordinaten in gescannten Dokumenten.
- Nutzung dieser Daten für das **unsupervised Pretraining** des diskreten visuellen Autoencoders (DVAE).
- Verbesserung der Bounding-Boxen für die Verwendung der Ergebnisse in Screenreadern, die besonders wichtig für Blinde und Sehbehinderte sind.

4. Crowdsourcing-basierte Annotation

- Anonymisierung und Aufbereitung ausgewählter Dokumente für **die CIB Crowdsourcing-Plattform**.
- Manuelle Annotation durch Nutzer zur Erfassung von:
 - **Lesereihenfolge**
 - **Layoutstrukturelementen** (z. B. Überschriften, Absätze, Tabellen)

5. Bereitstellung für Partner und Folge-APs

- Übergabe eines Teils der annotierten Dokumente an das **Fraunhofer IAIS** zur Weiterverarbeitung im Bereich Tabellenerkennung.
- Nutzung der aufbereiteten Daten für das Training des **Layout-Transformers** in AP3.

Gesamtbewertung zu AP2

Das Arbeitspaket AP2 war von zentraler Bedeutung für das gesamte Projekt. Die Qualität und Vielfalt der gesammelten und aufbereiteten Daten bilden die Grundlage für alle nachfolgenden KI-Modelle. Besonders hervorzuheben ist die **Fokussierung auf deutschsprachige Dokumente**, die in der internationalen Forschung bislang unterrepräsentiert sind. Damit wurde ein **Alleinstellungsmerkmal** für den deutschen Markt geschaffen.

AP3 – Transformer für Layout

Zielsetzung: Entwicklung eines multimodalen Layout-Transformers.

Durchführung: Evaluation von LayoutLMv3, Entwicklung eines eigenen GVAE, Optimierung mit Optuna, Integration in CIB flow.

Bewertung: Die Arbeiten waren technisch anspruchsvoll, aber notwendig und erfolgreich.

Im Arbeitspaket **AP3 „Transformer für Layout“** hat die **CIB AI labs GmbH** im Projekt RIDMI mehrere besonders bemerkenswerte Ergebnisse erzielt, die sowohl technologisch als auch praktisch relevant sind:

1. Entwicklung eines leistungsfähigen Layout-Transformers

- CIB hat ein **Transformermodell speziell für deutschsprachige Dokumente** trainiert, das in der Lage ist, komplexe Layoutstrukturen zu erkennen und in eine leserrichtige Reihenfolge zu bringen.
- Grundlage war die Architektur **LayoutLMv3**, die durch eigene Daten und Embeddings erweitert wurde.

2. Integration eines robusten Fallback-Mechanismus

- Für die Erkennung von Logos und wiederkehrenden Bildinhalten wurde **eine Fallbacklösung auf Basis von Feature-Matching** entwickelt.
- Diese Lösung verbessert die Layoutanalyse insbesondere bei **visuell komplexen oder unvollständigen Dokumenten**.

3. Erweiterung multimodaler Embeddings

- Aufbauend auf den eigens entwickelten **GVAE (Gaussian Variational Autoencoder)** und **DVAE (Discrete Variational Autoencoder)** wurden **visuelle Embeddings codiert**, um die Modelleistung weiter zu steigern.
- Verschiedene Fusionsstrategien der VAE-Embeddings und semantischer Embeddings (transformerbasiert und klassisch) wurden evaluiert.

4. Fehlerklassifikation und Qualitätsmessung

- Es wurden gezielte Tests zur **Fehlerklassifikation** durchgeführt, um Schwächen im Modellverhalten zu identifizieren.
- Die Ergebnisse flossen in die Weiterentwicklung der Architektur und die Verbesserung der Trainingsdaten ein.

Bewertung

Die CIB hat im AP3 nicht nur ein funktionierendes Transformermodell geliefert, sondern auch **mehrere innovative Komponenten** entwickelt, die über den Stand der Technik hinausgehen. Besonders hervorzuheben ist die **praxisnahe Ausrichtung** der Entwicklungen: Die Modelle wurden so konzipiert, dass sie sich **in bestehende CIB-Produkte integrieren lassen** (z. B. CIB flow).

AP4 – Qualitätsmessung für Layout

Zielsetzung: Entwicklung von Metriken zur Bewertung der Layoutanalyse.

Durchführung: Entwicklung eines Bewertungsprogramms, Zusammenarbeit mit geburtsblinden Informatikern.

Bewertung: Die Metriken waren essenziell für die Validierung. Der Aufwand war angemessen.

1. Entwicklung eines Bewertungsprogramms

- CIB hat ein **eigenes Programm zur Qualitätsmessung der Transformerausgaben** entwickelt.
- Dieses Tool ermöglicht den **Vergleich zwischen den vom Modell erzeugten Layoutstrukturen und annotierten Sollzuständen** (Groundtruth).

2. Definition spezialisierter Metriken

- Es wurden **spezialisierte Metriken zur Bewertung der Layoutqualität** definiert, die über klassische Genauigkeitsmaße hinausgehen.
- Die Metriken berücksichtigen z. B. **Lesereihenfolge, Segmentierungstreue und semantische Strukturierung**.

3. Einbindung von Betroffenen zur Validierung

- In Zusammenarbeit mit **geburtsblinden Informatikern** wurden zusätzliche Qualitätseinschätzungen eingeholt.
- Diese Perspektive war besonders wertvoll, um die **praktische Nutzbarkeit der Layoutanalyse für barrierefreie Dokumente** zu bewerten.

4. Visualisierung der Ergebnisse

- Die Ergebnisse der Qualitätsmessung wurden **visuell aufbereitet**, um Unterschiede zwischen Modelloutput und Groundtruth nachvollziehbar zu machen.
- Dies erleichtert die iterative Verbesserung der Modelle in AP3.

Bewertung

Auch wenn AP4 im Vergleich zu anderen Arbeitspaketen weniger umfangreich war, war es **methodisch und strategisch bedeutsam**. Die entwickelten Metriken und das Bewertungsprogramm bilden die **Grundlage für die objektive Evaluation** der Layoutanalyse und damit für **die Zertifizierbarkeit und Verwertbarkeit** der Ergebnisse im Kontext barrierefreier PDF-Erstellung.

AP5 – Fußzeilen

Zielsetzung: Erkennung und Zuordnung von Fußnoten.

Durchführung: Aufgrund von Priorisierungen wurde dieses AP nicht vollständig umgesetzt. Stattdessen wurden Annotationstools entwickelt und LLMs zur Annotation getestet.

Bewertung: Die Umpriorisierung war nachvollziehbar und im Sinne des Projekterfolgs.

Im Arbeitspaket **AP5 „Fußzeilen“** hat die CIB AI labs GmbH zwar **nicht die ursprünglich geplante vollständige Umsetzung** realisiert, aber dennoch **wertvolle Vorarbeiten und strategisch sinnvolle Umpriorisierungen** vorgenommen. Hier ist eine Zusammenfassung der wichtigsten Punkte:

1. Strategische Umpriorisierung

- Aufgrund der begrenzten Projektlaufzeit und der hohen Komplexität der Fußnotenerkennung wurde entschieden, den Fokus auf die **Verbesserung der allgemeinen Layoutanalyse** zu legen.

- Diese Entscheidung wurde getroffen, um die **Kernziele des Projekts (PDF-UA-Konformität, Barrierefreiheit)** effizienter zu erreichen.

2. Annotation und Datenaufbereitung

- Es wurden **Annotationswerkzeuge** entwickelt, mit denen Fußnoten und Referenzen in Dokumenten markiert werden können.
- Ein Teil der Datensätze wurde bereits **annotiert**, um eine spätere Modellierung zu ermöglichen.

3. Evaluation von LLMs zur Annotation

- Es wurde untersucht, inwieweit sich **Large Language Models (LLMs)** zur automatisierten Annotation von Fußnoten eignen.
- Die Ergebnisse waren **vielversprechend**, insbesondere im Hinblick auf eine mögliche spätere Umsetzung mit geringem manuellem Aufwand.

4. Crowdsourcing-basierte Annotation

- Als kostengünstige Alternative zur LLM-Nutzung wurde die **CIB Crowdsourcing-Plattform** eingesetzt, um Fußnoten manuell annotieren zu lassen.

Bewertung

Obwohl das AP5 nicht vollständig abgeschlossen wurde, sind die **geleisteten Vorarbeiten wertvoll**:

- Die entwickelten Tools und Erkenntnisse schaffen eine **gute Ausgangsbasis für eine spätere Umsetzung**.
- Die Entscheidung zur Umpriorisierung **war sachlich begründet und im Sinne des Projekterfolgs**.
- Die Nutzung von LLMs und Crowdsourcing zeigt, dass **innovative und flexible Wege** zur Zielerreichung eingeschlagen wurden.

AP6 – Tabellenstruktur

Zielsetzung: Erkennung und semantische Interpretation von Tabellen.

Durchführung: Einsatz des TATR-Modells, Entwicklung einer Schemaextraktion auf Word2Vec-Basis, Integration in bestehende Software.

Bewertung: Die Arbeiten waren erfolgreich und führten zu einem direkt verwertbaren Ergebnis.

Im Arbeitspaket **AP6 „Tabellenstruktur“** hat die **CIB AI labs GmbH** gemeinsam mit dem Fraunhofer IAIS ein besonders praxisrelevantes und erfolgreich abgeschlossenes Teilprojekt realisiert:

1. Erkennung und Umwandlung von Tabellenstrukturen

- CIB hat ein Verfahren zur **automatischen Erkennung von Tabellenstrukturen** in PDF-Dokumenten implementiert.
- Die erkannten Tabelleninhalte werden in **leserichtige Fließtexte** umgewandelt, was insbesondere für Screenreader und barrierefreie Darstellung entscheidend ist.

2. Semantische Auflösung von Tabellenschemas

- Es wurde eine Methode zur **semantischen Interpretation von Tabellen** entwickelt, die auf **Word2Vec-basierten Embeddings** aufbaut.
- Dadurch können **inhaltliche Zusammenhänge innerhalb von Tabellen** besser erkannt und für die Weiterverarbeitung nutzbar gemacht werden.

3. Integration in bestehende Softwarelösungen

- Die entwickelten Modelle und Verfahren wurden **in bestehende CIB-Produkte integriert**, insbesondere in den Dokumentbetrachter **CIB doXiview**.
- Damit ist die Technologie **direkt einsatzfähig** und kann in realen Anwendungsfällen genutzt werden.

4. Zusammenarbeit mit Fraunhofer IAIS

- Die CIB hat eng mit dem Fraunhofer IAIS kooperiert, das sich auf die Erkennung und Strukturierung der Tabellen konzentrierte, während CIB die **Integration und Anwendung** in der Praxis übernahm.

Bewertung

Das Arbeitspaket AP6 war ein **vollständig abgeschlossenes und erfolgreiches Teilprojekt**. Die Ergebnisse sind **technisch ausgereift, wirtschaftlich verwertbar** und **sofort einsetzbar**. Besonders hervorzuheben ist die **Verbindung von KI-basierter Analyse mit konkreter Produktintegration**, was den Transfer von Forschung in die Anwendung beispielhaft demonstriert.

AP10 – Projektmanagement

Zielsetzung: Koordination und Steuerung des Projekts.

Durchführung: Regelmäßige Abstimmungen, Berichterstattung, Budgetkontrolle.

Bewertung: Das Projektmanagement war effektiv und im geplanten Budgetrahmen.

Das Projektmanagement wurde durch die CIB AI labs GmbH durchgeführt und umfasste die Koordination der Arbeitspakete, die Abstimmung mit dem Projektpartner Fraunhofer IAIS sowie die fristgerechte Erstellung der Berichte und Nachweise. Die Projektsteuerung verlief planmäßig, ohne besondere Vorkommnisse oder Abweichungen. Das Arbeitspaket wurde im vorgesehenen Budgetrahmen umgesetzt und hat die erfolgreiche Durchführung des Gesamtprojekts zuverlässig unterstützt.

BEITRÄGE DES FRAUNHOFER IAIS

AP3 – Transformer für Layout

Problemstellung „Lesereihenfolge“: Textdokumente lassen sich mit diversen Tools (PDF Parser oder OCR) in Blöcke aus Text mit zugehörigen Koordinaten umwandeln. Diese Blöcke sind aber sehr häufig nicht der korrekten Lesereihenfolge. Wird dieser Text nun durch einen Screenreader präsentiert oder von einem KI-System weiterverarbeitet, so entstehen erheblich Verständnisschwierigkeiten, sowohl für Mensch als auch für Maschine.

Zielsetzung: Bestimmung der korrekten Sortierung solcher Textblöcke.

Durchführung:

Die folgenden Ansätze wurden entwickelt und evaluiert:

- Große Sprachmodelle als Next-Sentence-Predictor
 - Aufgrund der typischerweise schlechten Datenlage (insbesondere von gelabelten Daten), wurde zunächst ein Ansatz entwickelt und erprobt, der komplett ohne Daten auskommt. Hierfür wurde das inhärente Sprachverständnis von großen Sprachmodellen (LLMs) verwendet. Leider hat sich die Qualität dieses Ansatzes als unzureichend und für manche Dokumententypen (z.B. Briefe) als ungeeignet erwiesen.
- Lesereihenfolge mit Dokumenten-Encoder
- Ein weiterer Ansatz war der Versuch das Lesereihenfolgenproblem auf verschiedene Weisen mittels dem Dokumenten-Encoder LiLT (Language-independent Layout Transformer) zu modellieren und durch ein Fine-tuning von diesem zu lösen. Auch hier hat die anschließende Evaluation der drei verschiedenen Modellierungen nicht die erhoffte Güte bewiesen.
- DocLLM für Lesereihenfolge
 - Parallel zum Projekt wurde beim Fraunhofer IAIS die Modellarchitektur „DocLLM“ für deutsche Sprache nachgebaut. Diese wurde noch im Rahmen des Projekts für den Lesereihenfolgen-Usecase finegetuned. Dieser Ansatz konnte aber nicht mehr innerhalb der Projektlaufzeit evaluiert werden und wird über den Projektverlauf hinaus weiterentwickelt.
- Evaluation neuer Technologien
 - Es wurden unter anderem folgende Technologien von dritter Seite während der Projektlaufzeit veröffentlicht, die für den Lesereihenfolge-Usecase evaluiert wurden:
 - Nougat Modell
 - GOT-OCR
 - Docling Library
 - Die Docling Library wurde nach der Evaluation erfolgreich in angepasster Form in weiteren Projekten eingesetzt.

AP6 – Tabellenstruktur

Zielsetzung: Die Tabellenstruktur-Erkennung setzt sich aus zwei Schritten zusammen. Zunächst werden Tabellen in einem gegebenen Dokument detektiert und lokalisiert. Anschließend wird deren Inhalt in einem digital weiter verarbeitbaren Format extrahiert.

Durchführung:

Für die Detektion von Tabellen wurde Finetuning Code für DETR (DEtection TRansformer) entwickelt. Dieser wurde auf öffentlichen Daten erprobt.

Anschließend wurden für die Extraktion ebenfalls GOT-OCR und Docling evaluiert.

Beide Schritte zusammen wurden auf einem Testdatensatz evaluiert, der von dem Projektpartner CIB AI labs GmbH zur Verfügung gestellt wurde.

AP7 – Bilder-Datensätze

Zielsetzung: Erstellen eines Multilingualen Datensatz für detaillierte Bildunterschriften für szenische Bilder.

Durchführung:

- **Problemdefinition:** Die direkte Übersetzung kurzer Captions erzeugen semantische Unklarheiten. Diese Ausgangslage erfordert mehr Kontext, um präzise, szenisch passende Bildunterschriften in deutscher Sprache zu erhalten.
- **Datenbasis für Continuous Pre-training:** Zunächst wurde der Wikipedia-Image-Text Datensatz (>3 Mio. Bild-Text-Paare) genutzt und später durch CC12M (10 Mio. Bilder) ersetzt, da CC12M eine höhere visuelle und thematische Diversität bietet und so robustere Repräsentationen ermöglicht.
- **Fine-tuning Datensatz für Bildunterschriften:** Für die gezielte Anpassung werden COCO Captions (~500k), ImageParagraphs (~10k), Multi30k (~30k DE) und DOCCI eingesetzt. Diese Mischung deckt sowohl knappe als auch absatzlange Beschreibungen sowie mehrsprachige Varianten ab.
- **Recaptioning zur Disambiguierung und Detailtiefe:** Englische VLMs erzeugen zunächst detaillierte, kontextreiche Bildbeschreibungen, die anschließend übersetzt werden. Der zusätzliche Kontext hilft, lexikalische Ambiguitäten zuverlässig aufzulösen und die Szenen präziser zu erfassen.
- **Übersetzungspipeline:** Die Pipeline nutzt NLLB für hohe Übersetzungsqualität; unvollständige oder fehlerhafte Übersetzungen werden entfernt, um Konsistenz und Sprachklarheit sicherzustellen.
- **CLIP-basiertes Matching zur semantischen Verknüpfung:** Die Bilder werden zusätzlich über CLIP-Embeddings mit textbasierten Machine-Translation-Daten von CCMatrix verknüpft, um die multilingualen Fähigkeiten und multimodalen Repräsentationen zu verbessern.
- **Ergebnis:** Das Ergebnis ist ein Datensatz mit Bildern, passenden detaillierten Bildunterschriften und Übersetzungsdaten sowie eine wiederverwendbare Pipeline. Der Datensatz umfasst Inhalte über szenische Bilder hinaus. Die Pipeline ist in Abbildung 1 visualisiert.

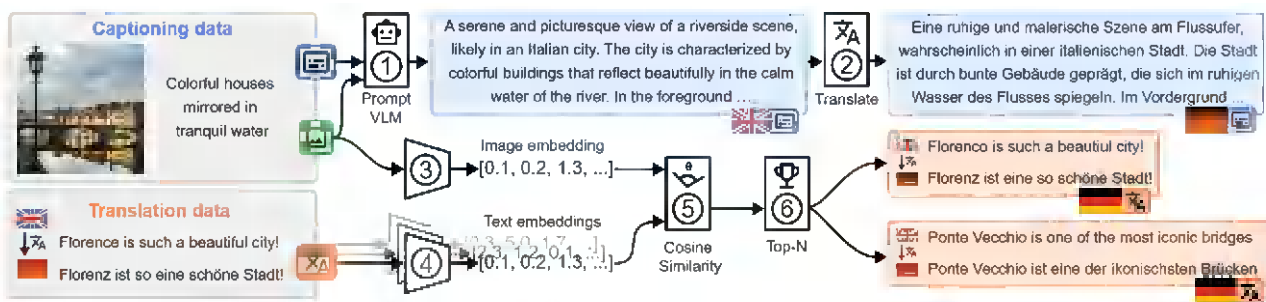


Abbildung 1: Pipeline zur Datensatzgenerierung. Eingabe ist ein Bilddatensatz mit kurzen Bildunterschriften sowie ein Übersetzungsdatensatz in Englisch und der Zielsprache Deutsch. (1) Bild und kurze Bildunterschrift werden einem VLM zugeführt, um eine detaillierte englische Beschreibung zu erzeugen, die (2) in die Zielsprache übersetzt wird. (3) Das Bild und (4) alle englischen Sätze des Übersetzungsdatensatzes werden in einen gemeinsamen Vektorraum eingebettet. (5) Die Cosine Similarity wird berechnet und (6) ein Top-N-Matching verknüpft die ähnlichsten Bilder und Übersetzungen.

Bewertung: Die Maßnahmen gehen über die Ziele hinaus: Der Datensatz enthält nicht nur szenische Bilder, sondern bietet vielfältige Inhalte und mehr Robustheit. Die verbesserte Übersetzungspipeline sorgt für

kontextreiche Beschreibungen, höhere Qualität und bessere Konsistenz. CLIP-Matching stärkt die Verbindung von Bild und Text. Das Ergebnis ist ein zuverlässiger, mehrsprachiger Datensatz.

AP8 – Ersatztexte für Bilder

Zielsetzung: Entwicklung eines KI-Modells zur Bildbeschreibung.

Durchführung:

- **Florence-2-basierte Modelllinie (0,4–11,2 Mrd.) für lange und komplexe Captions:** Florence-2-base bildet den robusten Kern für detailreiche, zusammenhängende Bildbeschreibungen; die Model-Palette reicht über Florence-2-large bis zur Kombination aus Florence-2-Encoder mit Gemma-2 als Decoder (bis 11,2 Mrd. Parameter), wodurch Ausgabequalität skalierbar ist und kleinere Varianten effizient auf Consumer-Hardware laufen.
- **Migration zum Gemma-2-Tokenizer für Multilingualität:** Zur Erweiterung der Sprachabdeckung wird auf den Gemma-2-Tokenizer umgestellt; der Embedding-Layer wird mittels Token-Übereinstimmungen bzw. Mapping angepasst, um vorhandenes Wissen im Modell zu erhalten.
- **Systematische Untersuchung von Skalierungsgesetzen:** Die Modelle zwischen 0,4–11,2 Mrd. Parametern werden unter variierenden Daten- und Compute-Budgets trainiert, um den Transfer von der Übersetzungsaufgabe zu der Captioning-Aufgabe zu untersuchen. Die Modelle erzeugen in Sprachen ohne Caption-Training sinnvolle Bildunterschriften auf Basis von Übersetzungsdaten, was generalisierende visuell-sprachliche Repräsentationen zeigt. Schlüsselfaktoren der Performance sind Multilingualität des vortrainierten Modells, Modellgröße und Anzahl gesehener Trainingsbeispiele.
- **Übertragbarkeit durch Feintuning:** Nach Feintuning werden kompetitive Ergebnisse in MMT (Multi30K, CoMMuTE), lexikalischer Disambiguierung (CoMMuTE) sowie Captioning (Multi30K, XM3600, COCO Karpathy) erreicht.

Bewertung: Die getroffenen Maßnahmen erfüllen nicht nur die Zielsetzung, sondern erweitern diese auch. Die skalierte Florence-2-Modelllinie ist in der Lage, unterschiedliche Anforderungen zu erfüllen, von effizienter Inferenz bis hin zu höchster Qualitätsausgabe. Zudem ermöglichen die durchgeführten Skalierungsstudien Zero-shot-Captioning in anderen Sprachen, ohne dass ein spezifisches Caption-Training in der Zielsprache erforderlich ist. Mit der belegten Übertragbarkeit auf MMT, Disambiguierung und Captioning ist das Ziel einer robusten und praxisnahen Bildbeschreibung erreicht worden.

AP9 – Qualitätsmessung für Ersatztexte

Zielsetzung: Entwicklung von Metriken zur Bewertung der Bildbeschreibungen.

Durchführung:

Ausgewählte Metriken: Für die automatische Bewertung wurden BLEU, CIDEr, ROUGE-L und CLIPScore genutzt. Außerdem wurden LLMs im Zusammenhang mit der CLAIR-Metrik für die Bewertung genutzt. SPICE und METEOR wurden verworfen, da ihre Annahmen im Deutschen zu inkonsistenten und wenig aussagekräftigen Scores führen. Die Leistung wurde auf Multi30k, COCO Captions, CoMMuTE und XM3600 geprüft, um deutsche Captions, multimodale Übersetzung und mehrsprachige Generalisierung abzudecken.

Bewertung: Die Maßnahmen erfüllen die Zielsetzung, indem sie ein praxistaugliches Metriken-Set etablieren und dessen Eignung über mehrere Benchmarks demonstrieren. Es wurde ein zuverlässiger Bewertungsrahmen ausgewählt, der die Qualität von Bildbeschreibungen konsistent und aussagekräftig erfasst.

AP10 – Projektmanagement

Zielsetzung: Unterstützung der Projektkoordination.

Durchführung: Teilnahme an Abstimmungen, wissenschaftliche Dokumentation.

Bewertung: Der Beitrag war angemessen und unterstützend.

TECHNISCHE UND ORGANISATORISCHE HERAUSFORDERUNGEN

Technische und organisatorische Herausforderungen

Einzelne Herausforderungen wurden bereits im Rahmen der jeweiligen Arbeitspakete beschrieben. Im Folgenden werden sie noch einmal **gebündelt und übergreifend dargestellt**, um einen kompakten Überblick über die zentralen technischen und organisatorischen Hürden im Projektverlauf zu geben.

1. Verzögerungen beim Hardwareeinkauf durch späte Vertragsunterzeichnung

Die vertragliche Abstimmung und die finale Unterzeichnung des Kooperationsvertrags erfolgten später als ursprünglich geplant. Dies hatte direkte Auswirkungen auf den zeitlichen Ablauf der Projektinitialisierung, insbesondere auf die Beschaffung der für das Training großer KI-Modelle notwendigen Hardware. Da die Auswahl und Inbetriebnahme von GPU-Servern ein kritischer Pfad für die Modellentwicklung war, führte diese Verzögerung zu einem verspäteten Aufbau der Trainingsinfrastruktur. Die Projektleitung reagierte mit einer Priorisierung vorbereitender Arbeiten, um den Zeitverlust zu kompensieren.

2. Hoher Aufwand bei der Annotation komplexer Layoutstrukturen

Die manuelle Annotation von Dokumenten zur Erfassung von Layoutstrukturelementen und Lesereihenfolgen erwies sich als deutlich aufwändiger als zunächst angenommen. Insbesondere bei komplexen Dokumenten mit Tabellen, Fußnoten und verschachtelten Layouts war eine präzise Annotation durch menschliche Annotatoren notwendig. Die CIB setzte hierfür ihre eigene Crowdsourcing-Plattform ein, musste jedoch feststellen, dass die Qualität der Annotation stark von der Schulung und Betreuung der Annotatoren abhing. Dies erforderte zusätzliche Ressourcen für Qualitätssicherung und Nachbearbeitung.

3. Schwierige Ableitung geeigneter Metriken aus der Fachliteratur

Für die Bewertung der Layoutanalyse war es notwendig, objektive Metriken zu definieren, die sowohl technische Genauigkeit als auch barrierefreie Nutzbarkeit abbilden. Die Recherche in der Fachliteratur ergab jedoch, dass bestehende Metriken entweder zu allgemein oder zu spezifisch für andere Anwendungsfälle waren. Die CIB entwickelte daher eigene Bewertungsmetriken in enger Abstimmung mit geburtsblinden Informatikern. Diese mussten technisch implementiert, getestet und mit realen Beispielen validiert werden – ein zusätzlicher methodischer Aufwand, der jedoch zu einem praxisnahen Bewertungssystem führte.

4. Dynamische Entwicklungen im Bereich LLMs und Foundation-Modelle

Während der Projektlaufzeit kam es zu rasanten Fortschritten im Bereich großer Sprachmodelle (LLMs) und multimodaler Foundation-Modelle. Diese Entwicklungen eröffneten neue Möglichkeiten, stellten das

Projektteam aber auch vor die Herausforderung, laufende Arbeiten regelmäßig zu evaluieren und ggf. anzupassen. So wurden z. B. neue Modelle wie GPT-4, Claude 3.7 oder Gemini 2.5 hinsichtlich ihrer Eignung für Bildbeschreibung und semantische Annotation getestet. Die Integration solcher Modelle erforderte zusätzliche technische Prüfungen, insbesondere im Hinblick auf Datenschutz, Lizenzierung und API-basierte Nutzung.

BEWERTUNG DER ANGEMESSENHEIT

Die geleisteten Arbeiten waren in Umfang und Tiefe angemessen, um die ambitionierten Projektziele zu erreichen. Die Aufteilung der Aufgaben zwischen den Partnern war sinnvoll und effizient. Trotz einzelner Umpriorisierungen wurden alle Kernziele erreicht oder vorbereitet. Die eingesetzten Ressourcen stehen in einem guten Verhältnis zu den erzielten Ergebnissen. Die Zusammenarbeit zwischen CIB und Fraunhofer IAIS war konstruktiv, zielgerichtet und synergetisch.

Voraussichtlicher Nutzen und Verwertbarkeit des Ergebnisses, auch konkrete Planung für die nähere Zukunft, im Sinne des fortgeschriebenen Verwertungsplans

WIRTSCHAFTLICHER NUTZEN

Die im Projekt RIDMI entwickelten Technologien werden direkt in bestehende und neue Produkte der CIB AI labs GmbH integriert. Dazu zählen insbesondere:

- **CIB ridmi:** Eine Layout-Generations-Komponente für die Herstellung von Barrierefreiheit und Annäherung an vollständiges PDF/UA.
- **CIB flow:** Ein Workflow-System für BPMN-basierte Geschäftsprozesse, bei CIB erweitert um die hochwertige Verarbeitung von Dokumenten in Ein- und Ausgang.
- **CIB doXiview:** Ein webbasierter Dokumentbetrachter, der bereits um die Layoutanalyse-Komponente erweitert wurde.
- **CIB doXisafe:** Eine Austauschplattform, die künftig barrierefreie PDF-Erstellung auf Basis der RIDMI-Komponenten ermöglichen wird.
- **CIB easyRead:** Ein Modul zur strukturierten Darstellung von Dokumentinhalten, das die Ergebnisse der Layout- und Tabellenanalyse nutzt.

Bereits während der Projektlaufzeit konnten erste Kunden für die Logoerkennung und die dokumentbasierte Klassifikation gewonnen werden. Die Nachfrage nach barrierefreier Dokumentverarbeitung steigt insbesondere im öffentlichen Sektor und bei Unternehmen mit hohem Dokumentenvolumen. Die RIDMI-Ergebnisse ermöglichen eine automatisierte Umsetzung gesetzlicher Anforderungen (z. B. Barrierefreiheitsstärkungsgesetz) und bieten damit einen klaren Wettbewerbsvorteil. Deswegen haben wir sehr viele Kundenanfragen aus dem Behördenumfeld.

Die im Rahmen der Entwicklungsarbeiten gewonnenen Kenntnisse, der erstellte Code, die gesammelten Daten und die entwickelten Modelle können für zukünftige Aktivitäten am Fraunhofer IAIS genutzt werden.

WISSENSCHAFTLICHER NUTZEN

Das Projekt hat zur Weiterentwicklung von Methoden im Bereich Document Understanding, multimodaler Embeddings und Layoutanalyse beigetragen. Die entwickelten Modelle (z. B. GVAE, Layout-Transformer) und die annotierten Datensätze bilden eine solide Grundlage für weitere Forschungsvorhaben. Die Zusammenarbeit mit dem Fraunhofer IAIS und die geplante Dissemination auf Fachveranstaltungen (z. B. Munich Datageeks, BMFTR-Fachtagung) stärken den wissenschaftlichen Austausch.

Darüber hinaus wurden neue Ansätze zur semantischen Interpretation von Tabellen und zur automatisierten Bildbeschreibung erprobt, die in zukünftigen Forschungsprojekten weiter vertieft werden können.

Im Rahmen des Projekts wurde außerdem eine Preprint-Publikation veröffentlicht, die sich mit automatisierten Bildbeschreibungen und dem Skalierungsgesetz im Kontext systematischer Generalisierung befasst:

Spravil, J., Houben, S., & Behnke, S. (2025). Florenz: Scaling Laws for Systematic Generalization in Vision-Language Models. arXiv preprint arXiv:2503.09443.

Ein wichtiger Aspekt besteht darin, den Source Code zur Erzeugung von Bildbeschreibungen nach der erfolgreichen Veröffentlichung der Preprint-Publikation auf einer geeigneten wissenschaftlichen Veranstaltung der Öffentlichkeit zur Verfügung zu stellen. Damit wird nicht nur die Nachnutzbarkeit der Ergebnisse für Forschung und Praxis gestärkt, sondern auch die Transparenz und Innovationskraft in der Community gefördert.

GESELLSCHAFTLICHER NUTZEN

RIDMI leistet einen direkten Beitrag zur digitalen Barrierefreiheit. Die entwickelten Technologien ermöglichen es, PDF-Dokumente automatisiert so aufzubereiten, dass sie auch für blinde und sehbehinderte Menschen zugänglich sind. Dies fördert die digitale Teilhabe und unterstützt die Umsetzung gesetzlicher Vorgaben auf Bundes- und EU-Ebene.

Die Zusammenarbeit mit dem Bayerischen Blinden- und Sehbehindertenbund (BBSB) und der Mediablis-Stelle zeigt, dass die Ergebnisse auch in der Bildung und im sozialen Bereich Anwendung finden können – etwa bei der barrierefreien Aufbereitung von Schulbüchern.

VERWERTUNGSPERSPEKTIVE UND ZEITRAHMEN

Kurzfristig (0–6 Monate nach Projektende): Integration der RIDMI-Komponenten in CIB-Produkte, Abschluss der Tests zur Bildbeschreibung, erste Pilotanwendungen bei Partnern.

Mittelfristig (6–24 Monate): Vermarktung der barrierefreien PDF-Erstellung als Standardmodul, Ausbau der Kundenbasis im öffentlichen Sektor, Weiterentwicklung der Layoutanalyse für Spezialformate.

Langfristig (ab 2027): Erweiterung der Lösung um dialogbasierte Funktionen („Chat with your Document“), semantische Suche und adaptive Dokumentnavigation.

VERWERTUNGSSTRATEGIE

Die Verwertung erfolgt primär durch interne Integration in die CIB-Produktpalette. Ergänzend ist eine Lizenzierung an Partner (z. B. AKDB, öffentliche Verwaltungen) vorgesehen. Die Marke „CIB RIDMI“ wird eingetragen, um die Sichtbarkeit der Lösung zu erhöhen.

Die Dissemination erfolgt über:

- Fachvorträge und Konferenzen
- Blogbeiträge, Web-Auftritt und Produktdemos
- Kundenpräsentationen und Schulungen

RISIKEN UND HERAUSFORDERUNGEN

Mögliche Herausforderungen bei der Verwertung betreffen:

- Datenschutz und Lizenzfragen (Kosten) bei der Nutzung externer LLMs
- Marktdurchdringung bei konservativen Zielgruppen, insbesondere Vertrieb gegenüber Behörden
- Technologische Weiterentwicklung durch große Anbieter

Diese Risiken werden durch eine modulare Architektur, eigene Trainingsdaten und enge Kundenbindung aktiv adressiert.

BEREITS ERZIELTE WIRTSCHAFTLICHE ERFOLGE (ROI)

Bereits während der Projektlaufzeit konnten erste Returns on Investment (ROI) erzielt werden:

- **Logoerkennung:** Die im Projekt entwickelte Komponente zur Erkennung wiederkehrender Bildinhalte (z. B. Logos) wurde bereits bei Kunden eingesetzt. Sie verbessert die Dokumentenklassifikation und ermöglicht automatisierte Prüfprozesse.
- **Dokumentenklassifikation mit Fingerabdruck (Spin-Off-des Förderprojektes):** Auf Basis der Layoutanalyse wurde ein individueller „Fingerprint“ für Dokumente entwickelt. Dieser wird bereits in Pilotanwendungen zur Betrugserkennung im Gesundheitswesen eingesetzt – ein klarer wirtschaftlicher und gesellschaftlicher Nutzen.
- **Posteingangs-Klassifikation:** Die im Projekt entwickelten multimodalen Embeddings wurden außerhalb des Projektrahmens in ein Standardprodukt zur Posteingangsverarbeitung überführt. Dieses Produkt ist bereits am Markt verfügbar und generiert Einnahmen.
- **Kundengewinnung durch Öffentlichkeitsarbeit:** Durch Blogbeiträge und Projektkommunikation wurden neue Kundenkontakte geknüpft, die sich für die Themen Barrierefreiheit, Dokumentenstrukturierung und KI-gestützte Analyse interessieren.

Der während der Durchführung des Vorhabens dem Zuwendungsempfänger bekannt gewordene Fortschritt auf dem Gebiet des Vorhabens bei anderen Stellen

Während der Projektlaufzeit wurden relevante Entwicklungen bei anderen Forschungseinrichtungen und Unternehmen beobachtet und bewertet. Diese Beobachtungen dienen sowohl der strategischen Einordnung des eigenen Projektansatzes als auch der Validierung der gewählten Methoden.

Im Bereich der Layoutanalyse wurden zwei neue Modelle bekannt: „**Nougat**“ und „**Document-Image-Transformer (DIT)**“. Beide Ansätze wurden von der CIB AI labs GmbH anhand von Beispieldaten getestet. Die Ergebnisse zeigten jedoch, dass diese Modelle nicht die erforderliche Qualität für den Einsatz im Projektkontext erreichten. Insbesondere wurde festgestellt, dass die Trainingsdaten dieser Modelle stark auf wissenschaftliche Artikel aus dem arXiv-Umfeld fokussiert waren, was zu einer eingeschränkten Generalisierbarkeit auf deutschsprachige Geschäftsdokumente führte. Für die Anforderungen des RIDMI-Projekts – insbesondere im Hinblick auf Barrierefreiheit und komplexe Layouts – waren diese Modelle daher nicht geeignet.

Im Bereich der multimodalen Sprachmodelle wurden während der Projektlaufzeit bedeutende Fortschritte erzielt. Modelle wie **GPT-4**, **Claude 3.7** und **Gemini 2.5 Pro** wurden hinsichtlich ihrer Eignung für Aufgaben wie Bildbeschreibung, semantische Annotation und Dokumentenverständnis evaluiert. Erste Tests zeigten vielversprechende Ergebnisse, insbesondere im Hinblick auf die Generierung von Bildersatztexten. Diese Entwicklungen wurden im Projektverlauf berücksichtigt und flossen in die Planung der weiteren Verwertung ein – insbesondere im Hinblick auf mögliche hybride Ansätze mit externen APIs oder lokal gehosteten Modellen.

Auch im Bereich der Tabellenstrukturerkennung wurden neue Modelle identifiziert, darunter **GeoLM** und weitere auf Transformer-Architekturen basierende Ansätze. Diese wurden als potenziell relevant für die Weiterentwicklung der RIDMI-Komponenten eingestuft und werden im Rahmen der Nachnutzung weiter beobachtet.

Insgesamt wurde der Stand der Technik kontinuierlich beobachtet und kritisch bewertet. Die im Projekt RIDMI entwickelten Lösungen zeichnen sich durch eine hohe Spezialisierung auf deutschsprachige, realweltliche Dokumente aus und bieten damit ein Alleinstellungsmerkmal gegenüber generischen, international trainierten Modellen.

Erfolgte oder geplante Veröffentlichungen des Ergebnisses nach Nr. 5 der NKBF

Im Verlauf des Projekts RIDMI wurden verschiedene Maßnahmen zur Verbreitung der Projektergebnisse initiiert oder vorbereitet. Diese Aktivitäten dienen sowohl der wissenschaftlichen und technischen Dissemination als auch der Unterstützung der wirtschaftlichen Verwertung.

WISSENSCHAFTLICHE VERÖFFENTLICHUNGEN

Zum Zeitpunkt des Projektabschlusses sind keine formalen wissenschaftlichen Publikationen (z. B. in Journals oder Konferenzbänden) erschienen. Es ist jedoch geplant, die im Projekt gewonnenen Erkenntnisse zur Bildbeschreibung in geeigneter Form zu veröffentlichen. Zielkonferenzen sind unter anderem:

- European Conference on Computer Vision (ECCV)
- AAAI Conference on Artificial Intelligence
- BMFTR-Fachtagungen

TECHNISCHE VERÖFFENTLICHUNGEN

Die CIB AI labs GmbH hat im Rahmen des Projekts mehrere technische Inhalte veröffentlicht, darunter:

- Blogbeiträge zur Anwendung der Layoutanalyse und zur Integration in CIB-Produkte
- Technische Demos im Rahmen von Kundenpräsentationen
- Interne Whitepapers zur Architektur des Layout-Transformers und zur Bildersatztexterstellung

Ein öffentlich zugängliches Repository ist derzeit nicht vorgesehen, da die entwickelten Komponenten in kommerziellen Produkten verwendet werden.

ÖFFENTLICHKEITSARBEIT

Zur Verbreitung der Projektergebnisse wurden folgende Maßnahmen ergriffen:

- Veröffentlichung von Projektfortschritten auf der CIB-Webseite
- Beiträge in thematisch passenden Fachblogs
- Präsentationen auf internen und externen Veranstaltungen, u. a. bei der AKDB, dem BBSB und der Stadt München
- Veranstaltung eines Meetups „Munich Datageeks“
<https://www.cib.de/ki-expertentreffen-munich-datageeks-cib/>

GEPLANTE VERÖFFENTLICHUNGEN

Für die Zeit nach Projektende sind weitere Veröffentlichungen geplant, darunter:

- Ein Fachartikel zur Kombination multimodaler Embeddings mit Layoutanalyse
- Ein Erfahrungsbericht zur Zusammenarbeit mit assoziierten Partnern im Bereich Barrierefreiheit
- Produktbroschüren zur barrierefreien PDF-Erstellung mit RIDMI-Komponenten

- Konferenzbeitrag zur Bildersatztexterzeugung und dessen Skalierungsgesetze für die systematische Generalisierung von Machine-Translation-Daten.

VERWERTUNGSRELEVANTE VERÖFFENTLICHUNGEN

Die im Projekt entwickelten Technologien werden in Produktdokumentationen, Marketingmaterialien und Schulungsunterlagen überführt. Diese dienen der Markteinführung und der Unterstützung von Vertriebspartnern.