Sachbericht zum Verwendungsnachweis

Verbundprojekt

EXPLAIN

Erklärbare Künstliche Intelligenz für die Industrie

Gefördert durch:

Bundesministerium
für Forschung, Technologie
und Raumfahrt

Förderhinweis:
Das diesem Bericht zugrunde liegende Vorhaben wurde mit Mitteln des Bundesministeriums für Forschung, Technologie und Raumfahrt (BMFTR) unter dem Förderkennzeichen 16IS22030A gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autorinnen und Autoren. Dieser Sachbericht wurde gemeinsam von allen Zuwendungsempfängern des Verbundvorhabens EXPLAIN erstellt und abgestimmt.

Gefördert durch:

Bundesministerium
für Forschung, Technologie
und Raumfahrt

EXPLAIN – Erklärbare Künstliche Intelligenz für die Industrie

Teil I – Kurzbericht

Zuwendungsempfänger:	Förderkennzeichen:
- ABB AG Forschungszentrum - Universität Hildesheim Software Systems Engineering - Eraneos Analytics Germany GmbH - Lausitz Energie Kraftwerke AG - Technische Universität Darmstadt Artificial Intelligence and Machine Learning Lab Assoziierter Partner: Schwermetall Halbzeugwerk GmbH & Co KG	**16IS22030A**
Koordination: ABB AG - Forschungszentrum Deutschland Dr. Ruben Hühnerbein Kallstadter Str. 1, 68309 Mannheim E-Mail: ruben.huehnerbein@de.abb.com	
Projektlaufzeit: 01.05.2022 - 30.04.2025	

Förderhinweis:
Das diesem Bericht zugrunde liegende Vorhaben wurde mit Mitteln des Bundesministeriums für Forschung, Technologie und Raumfahrt (BMFTR) unter dem Förderkennzeichen 16IS22030A gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autorinnen und Autoren. Der Teil I – Kurzbericht wurde gemeinsam von allen Zuwendungsempfängern des Verbundvorhabens EXPLAIN erstellt und abgestimmt.

Kurzbericht – EXPLAIN: Erklärbare Künstliche Intelligenz für die Industrie

Künstliche Intelligenz (KI) gewinnt in der Industrie zunehmend an Bedeutung etwa in der Energieerzeugung, Metallverarbeitung oder in der Prozessindustrie. Obwohl KI-Systeme leistungsfähig sind, gelten viele von ihnen weiterhin als „Black Boxes“: Ihre Entscheidungsprozesse sind für Anwenderinnen und Anwender schwer nachvollziehbar. Dies mindert Vertrauen, Akzeptanz und letztlich die praktische Nutzbarkeit von KI in sicherheitskritischen und komplexen industriellen Umgebungen.

Das Verbundprojekt *EXPLAIN – EXPLanatory interactive Artificial Intelligence for INdustry* hatte das Ziel, Künstliche Intelligenz transparent, nachvollziehbar und menschzentriert zu gestalten. Im Mittelpunkt stand die Entwicklung von Verfahren und Werkzeugen, die technische Erklärbarkeit mit nutzerorientiertem Design verbinden und so eine vertrauenswürdige Zusammenarbeit zwischen Menschen und KI ermöglichen. EXPLAIN leistete damit einen wesentlichen Beitrag zur verantwortungsvollen und sicheren Nutzung von KI in der Industrie.

Projektstruktur und internationale Zusammenarbeit

EXPLAIN wurde im Rahmen des Eureka-Clusters ITEA als internationales Verbundprojekt mit Partnern aus Deutschland, Schweden und den Niederlanden durchgeführt. Der deutsche Projektteil wurde vom Bundesministerium für Forschung, Technologie und Raumfahrt (BMFTR) gefördert und vereinte fünf Partner aus Industrie und Forschung.

Insgesamt arbeiteten 15 Organisationen aus drei Ländern zusammen, um erklärbare KI für unterschiedliche industrielle Domänen praktisch nutzbar zu machen. Für seine wissenschaftlich und technologisch herausragenden Ergebnisse wurde **EXPLAIN im Jahr 2025 mit dem ITEA Award of Excellence for Innovation ausgezeichnet** – als bestes europäisches Innovationsprojekt des ITEA-Programms.

Arbeiten und Ergebnisse

Im deutschen Teilvorhaben wurden drei industrielle Anwendungsfälle bearbeitet:

- Energiemarktvorhersage
- Unterstützungssysteme für den nachhaltigen Kraftwerksbetrieb
- Instandhaltungsplanung von Walzwerken

Die Forschung konzentrierte sich auf die Entwicklung erklärbarer maschineller Lernmodelle sowie interaktiver Werkzeuge, die Nutzerfeedback in den Lernprozess einbeziehen. Zentrale Ergebnisse waren unter anderem:

- Verfahren zur zeitreihenbasierten Anomalieerkennung und deren Erklärung mittels XAI-Methoden,
- ein Dashboard zur Unterstützung von Anlagenfahrern im Kraftwerksbetrieb, das Anomalien erkennt und erklärt,
- Feedback-Mechanismen, die Domänenwissen von Operatorinnen und Operatoren in die Modelle zurückführen,
- sowie eine MLOps-Architektur, die Erklärungen und Feedback in den gesamten Lebenszyklus von KI-Modellen integriert.

Die Ansätze wurden in enger Zusammenarbeit mit Endanwendern getestet, unter anderem im Kraftwerksbetrieb der LEAG und im Walzwerkskontext von Schwermetall. Die Erfahrungen zeigten, dass verständliche, interaktive Erklärungen das Vertrauen und die Effizienz im Umgang mit KI-Systemen signifikant steigern. Neben den technischen Ergebnissen entstand das öffentlich verfügbare EXPLAIN Guidebook *„Human-Centered Explainable AI for Process Industries“* (cf. https://itea4.org/news/explain-guidebook-human-centered-explainable-ai-for-process-industries.html), das praktische Leitlinien, Designprinzipien und Best-Practice-Beispiele für die Umsetzung erklärbarer KI in industriellen Umgebungen enthält.

Verwertung und Ausblick

Die entwickelten Methoden und Softwarekomponenten bilden eine tragfähige Grundlage für die Weiterentwicklung von vertrauenswürdiger und erklärbarer KI in der Industrie. Mehrere Partner führen die Arbeiten in nachfolgenden Forschungs- und Entwicklungsprojekten fort oder planen eine Integration der Konzepte in bestehende industrielle Plattformen. Die im Projekt gewonnenen Erkenntnisse tragen dazu bei, dass KI-Systeme künftig nicht nur leistungsfähig, sondern auch verständlich, überprüfbar und sicher eingesetzt werden können. Damit leistet EXPLAIN einen wichtigen Beitrag zur nachhaltigen und verantwortungsvollen Digitalisierung der Industrie.

Gefördert durch:
Bundesministerium für Forschung, Technologie und Raumfahrt

EXPLAIN – Erklärbare Künstliche Intelligenz für die Industrie

Teil II – Eingehende Darstellung

Zuwendungsempfänger:	Förderkennzeichen:
• ABB AG Forschungszentrum • Universität Hildesheim Software Systems Engineering • Eraneos Analytics Germany GmbH • Lausitz Energie Kraftwerke AG • Technische Universität Darmstadt Artificial Intelligence and Machine Learning Lab Assoziierter Partner: Schwermetall Halbzeugwerk GmbH & Co KG	**16IS22030A**
Koordination: ABB AG - Forschungszentrum Deutschland Dr. Ruben Hühnerbein Kallstadter Str. 1, 68309 Mannheim E-Mail: ruben.huehnerbein@de.abb.com	
Projektlaufzeit: 01.05.2022 - 30.04.2025	

Förderhinweis:
Das diesem Bericht zugrunde liegende Vorhaben wurde mit Mitteln des Bundesministeriums für Forschung, Technologie und Raumfahrt (BMFTR) unter dem Förderkennzeichen 16IS22030A gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autorinnen und Autoren. Der Teil II – Eingehende Darstellung wurde gemeinsam von allen Zuwendungsempfängern des Verbundvorhabens EXPLAIN erstellt und abgestimmt.

Kurzbeschreibung:

Das Verbundprojekt *EXPLAIN – EXPLanatory interactive Artificial Intelligence for INdustry* entwickelte Methoden und Werkzeuge, um Künstliche Intelligenz (KI) in industriellen Prozessen *erklärbar, transparent und menschzentriert* zu gestalten. Anhand von Anwendungsfällen aus der Energie-, Metall- und Prozessindustrie wurden neue Konzepte für Human-Centered Explainable AI (HCXAI) entworfen, prototypisch umgesetzt und gemeinsam mit Endanwendern validiert. EXPLAIN war Teil eines internationalen ITEA-Verbunds mit Partnern aus Deutschland, Schweden und den Niederlanden und bündelte die Expertise von 15 Industrie- und Forschungseinrichtungen.

Für seine wissenschaftlich und technologisch herausragenden Ergebnisse erhielt das Projekt 2025 den *ITEA Award of Excellence for Innovation* als bestes europäisches Innovationsprojekt des ITEA-Programms.

2025 Die deutschen Projektpartner des Verbundvorhabens EXPLAIN.

Struktur des Berichts / Übersicht der Arbeitspakete

AP 1 Anwendungsfälle, Benutzerziele, -bedürfnisse und -erwartungen & Anforderungen 3
AP 2 Datenerfassung und -aufbereitung 5
AP 3 Konventionelle KI-Modellierung 7
AP 4 Erklärungskomponente für industrielle KI 10
AP 5 Erklärungsbasiertes Feedback für industrielle KI 14
AP 6 Ende-zu-Ende *„Machine Learning“* System mit Erklärungen und Feedback 18
AP 7 Empirische Validierung und Pilotierung der Anwendungsszenarien 23
AP 8 Dissemination und Transfer 26
AP 9 Projektmanagement 31

Arbeitsplanung und Einbettung des Projekts

Die Vision des Projekts *EXPLAIN – EXPLanatory interactive Artificial Intelligence for INdustry* ist die Entwicklung eines erklärenden End-to-End-Workflows für Maschinelles Lernen, der es ermöglicht, das Fachwissen der wichtigsten Akteure im industriellen Ökosystem gezielt zu berücksichtigen und zu nutzen.

Ziel war es, Verfahren und Werkzeuge zu schaffen, mit denen Künstliche Intelligenz (KI) in industriellen Anwendungen *transparent, nachvollziehbar und menschzentriert* gestaltet werden kann. Damit leistet das Projekt einen Beitrag zur verantwortungsvollen Nutzung von KI in sicherheits- und prozesskritischen Umgebungen.

EXPLAIN war Teil des internationalen ITEA-Verbundsprojekts „EXPLAIN“ im Rahmen des Eureka-Clusters. Unter Koordination von ABB AG Forschungszentrum Deutschland und mit Partnern aus Schweden und den Niederlanden wurden insgesamt 15 Organisationen aus Industrie und Forschung eingebunden.
Der internationale Verbund ermöglichte den Austausch unterschiedlicher Perspektiven auf Erklärbarkeit, Interaktion und technische Integration von KI-Systemen.

Für seine wissenschaftlich und technologisch herausragenden Ergebnisse wurde EXPLAIN im Jahr 2025 mit dem *ITEA Award of Excellence for Innovation* ausgezeichnet – als bestes europäisches Innovationsprojekt des ITEA-Programms.

Im deutschen Teilvorhaben wurden drei zentrale industrielle Anwendungsfälle betrachtet:

- Energiemarkt Vorhersage
- Unterstützungssysteme für den nachhaltigen Kraftwerksbetrieb
- Instandhaltungsplanung von Walzwerken

AP 1 Anwendungsfälle, Benutzerziele, -bedürfnisse und -erwartungen & Anforderungen

Dieses Arbeitspaket umfasst die Erfassung der Ziele und Bedürfnisse der Endbenutzer in Bezug auf die beteiligten Prozesse und KI-Systeme. Diese werden in Anforderungen an die Erklärbarkeit übersetzt, die sich darauf auswirken, wie Datenerfassung und KI-Modellierung erfolgen, wie dem Benutzer Erklärungen präsentiert werden sollen und wie das System und das maschinell erlernte Modell das Feedback des Benutzers erfassen und verarbeiten können.

Die Tätigkeiten innerhalb des Arbeitspaketes bestehen in erster Linie in der Vorbereitung, Durchführung und Nachbereitung von Anwendungsfall- und Anforderungsworkshops unter Beteiligung des Personals der Anwendungspartner, sowie der Durchführung und Aufbereitung von Feldstudien, in denen die im Projekt produzierten Prototypen ausführlichen Nutzertests unterzogen werden.

Insgesamt konnte das Arbeitspaket wie geplant dazu beitragen, die Ziele, Bedürfnisse und Erwartungen der Endanwender systematisch zu erfassen und in Anforderungen an erklärbare KI zu übersetzen. Durch die Durchführung von Workshops, Interviews und Feldstudien mit den Anwendungspartnern wurden praxisnahe Erkenntnisse gewonnen, die direkt in die Entwicklung der Prototypen eingeflossen sind. Dabei zeigte sich insbesondere, wie wichtig eine klare Darstellung von Anomalien, die Berücksichtigung domänenspezifischen Fachwissens sowie Aspekte wie Schulung und Wissensmanagement sind. Auf dieser Basis konnten konkrete Anforderungen formuliert werden, die sowohl für die Gestaltung von Erklärungen als auch für die Einbindung von Nutzerfeedback in die Systeme entscheidend sind.

AP 1.1 – Klärung der Anforderungen bzgl. personenbezogener Daten und rechtlicher Rahmenbedingungen

Mitarbeitende Partner: ABB, ERANEOS, LEAG, SUH, TUDA

Dieses Arbeitspaket stellt sicher, dass die Erfassung und Verarbeitung von Daten im Rahmen des Projekts den rechtlichen und insbesondere datenschutzrechtlichen Rahmenbedingungen genügt. So ist für die einzelnen Datenquellen zu prüfen, inwieweit personenbezogene Daten betroffen sind oder betroffen sein können. Es werden geeignete Strategien entwickelt, die unzulässige Verknüpfung personenbezogener Daten zu unterbinden. Weitere rechtliche Rahmenbedingungen wie das Betriebsverfassungsgesetz oder Exportgesetze hinsichtlich des länderübergreifenden Datenexports, sind bei allen Aktivitäten durch alle Partner hinsichtlich der geplanten Arbeiten zu prüfen und ggf. geeignete Maßnahmen zu ergreifen, um die Rechtmäßigkeit der Projektarbeit sicherzustellen. ABB unterstützt die Konsortialpartner, wenn entsprechende Situationen in den Arbeitspaketen eintreten.

In diesem Arbeitspaket wurde sichergestellt, dass die Erfassung und Verarbeitung von Daten im Projekt den rechtlichen und insbesondere datenschutzrechtlichen Anforderungen entsprechen. Für die im Rahmen von Interviews erhobenen Daten wurden Einverständniserklärungen vorbereitet, den teilnehmenden Operatoren vorgelegt und von diesen unterschrieben. Auf diese Weise wurde die rechtmäßige Verarbeitung personenbezogener Daten gewährleistet.

AP 1.2 – Nutzer- und Anwendungsfallanalyse für einen nachhaltigen Kraftwerksbetrieb

Mitarbeitende Partner: ABB, ERANEOS, LEAG, SUH, TUDA

Fokus dieses Arbeitspaketes sind die Anwendungsfälle in den Anlagen der LEAG mit Ersatzbrennstoffverbrennung. Wie für das Gesamtpaket beschrieben, finden hier die Anwendungsfalls- und Anforderungs-Workshops sowie Feldstudien mit den entsprechenden Fachexperten und Endanwendern der LEAG statt. ABB unterstützt dieses Teilarbeitspaket des Arbeitspaket mit seiner Expertise und Erfahrung in der Vorbereitung, Durchführung und Nachbereitung von derartigen Workshops und Feldstudien. TUDA beteiligt sich an den Studien der für den Anwendungsfall relevanten technischen Anforderungen. SUH beteiligt sich an den Studien der für den Anwendungsfall relevanten technischen

Anforderungen. ERANEOS beteiligt sich an den Studien der für den Anwendungsfall relevanten technischen Anforderungen. ERANEOS beteiligt sich zudem an der Durchführung, Moderation und Aufbereitung der Workshops und Feldstudien.

Im Rahmen dieses Arbeitspakets standen die Anwendungsfälle in den Anlagen der LEAG mit Ersatzbrennstoffverbrennung im Fokus. Dazu wurden Workshops und Feldstudien mit Fachexperten und Endanwendern durchgeführt. ABB unterstützte mit seiner Erfahrung in der Konzeption, Durchführung und Auswertung dieser Aktivitäten und führte unter anderem einen Evaluationsworkshop im Trainingszentrum der LEAG in Lübbenau durch. Dabei konnten Operatoren mit unterschiedlichen Erfahrungsstufen befragt werden. Die Interviews bestätigten, dass die vorgestellte Lösung zur Erkennung und Erklärung von Anomalien im Anlagenbetrieb als hilfreiches Werkzeug wahrgenommen wird. Zugleich betonten die Befragten die Bedeutung der menschlichen Entscheidungshoheit: KI-Systeme sollen auf Anomalien hinweisen, aber keine autonomen Eingriffe vornehmen. ERANEOS untersuchte den Anwendungsfall der Day-Ahead-Strompreisprognosen aus Data-Science-Perspektive, berücksichtigte bestehende Erwartungen und Anforderungen und definierte auf dieser Basis geeignete Modelle und Zielvariablen. KI-Vorhersagemodelle wurden entwickelt, den LEAG-Mitarbeitern in Workshops vorgestellt und Feedback eingeholt. LEAG stellte den Projektpartnern einen umfassenden Überblick über die Arbeitssituation des Leitstandpersonals zur Verfügung, leitete mögliche Unterstützungsszenarien ab und bereitete auf dieser Grundlage ein Projekttreffen im Kraftwerk Boxberg vor. Zusätzliche Bedürfnisse, Erwartungen und Anforderungen wurden besprochen und dokumentiert und in weitere Modeliterationen eingearbeitet. SUH führte eine Feldstudie durch, in der Daten erhoben und Nutzerbedürfnisse für die relevanten Use Cases gesammelt wurden, und beteiligte sich zudem an der Analyse der technischen Anforderungen. TUDA brachte die in AP 4 und 5 entwickelten Erklärungs- und Feedback-Mechanismen für die Energiepreisvorhersage in die Diskussion ein, stellte diese dem Leitstandpersonal vor und leitete aus den Rückmeldungen Anforderungen für die Weiterentwicklung des Prototyps ab.

AP 1.3 – Nutzer- und Anwendungsfallanalyse Walzwerk

Mitarbeitende Partner: ABB, SUH, TUDA

[redacted]

[redacted]

[redacted]

[redacted]

AP 2 Datenerfassung und -aufbereitung

Dieses Arbeitspaket stellt die zur Bearbeitung der Anwendungsfälle und zur Methodenentwicklung notwendigen Daten bereit. Die Anwendungspartner extrahieren Daten aus ihren Betriebssystemen (z.B. Inspektionssysteme, Prozesshistorien, Kontrollsysteme) und generieren Daten mit Hilfe von realitätsnahen Prozesssimulationen. Sowohl extrahierte als auch generierte Daten müssen aufbereitet werden, indem relevante Metadaten und Labels für den Zweck des überwachten Maschinellen Lernens bereitgestellt werden. Die aufbereiteten Datensätze werden anderen Arbeitspaketen zur Entwicklung von ML-Modellen zur Verfügung gestellt.

AP 2.1 Erstellung von ML Datensatz für Energiemarktvorhersagen

Mitarbeitende Partner: ERANEOS, LEAG, TUDA

In diesem Arbeitspaket werden die notwendigen Datensätze für den Energiemarkt bereitgestellt. LEAG übergibt an die beteiligten Partner Datensätze, die das Verhalten des Energiemarktes und der wesentlichen Einflussparameter enthalten. In einen gemeinsamen Workshop soll ein gemeinschaftliches Verständnis über Begrifflichkeiten und Zusammenhänge auf dem deutschen Energiemarkt geschaffen werden. ERANEOS bereitet die Daten auf und identifiziert Features für die Modellierung. TUDA bietet den Industriepartnern Unterstützung bei der Definition des Datenmodells.

Im Rahmen des Anwendungsfalls „Day-Ahead Strompreisprognosen“ erfolgte eine umfassende Sichtung, Bewertung und Aufbereitung der relevanten Datenquellen. ERANEOS übernahm die Datenanalyse und -aufbereitung unter Einbezug verschiedener Datensätze, darunter EPEX-Spot-Marktpreise, Wetterdaten des Deutschen Wetterdienstes (DWD) sowie Lastgangdaten der ENTSO-E. Die LEAG stellte die erforderlichen Datengrundlagen zur Verfügung und unterstützte die Partner bei der fachlichen Einordnung der Datenquellen. TUDA brachte ihr methodisches Fachwissen in die Strukturierung und Konsolidierung des Datensatzes ein und arbeitete dabei eng mit Eraneos und LEAG zusammen. Durch die gemeinsame Bearbeitung entstand ein qualitativ hochwertiger Datensatz, der als Grundlage für die nachfolgenden Modellierungs- und Prognosearbeiten dient.

AP 2.2 – Erstellung von ML-Datensatz für den nachhaltigen Kraftwerksbetrieb

Mitarbeitende Partner: ABB, ERANEOS, LEAG, TUDA

Auf Grundlage des Workshops im AP 1.2 stellt LEAG in Absprache mit dem Partner einen Datensatz für die zu betrachtende Fragestellung zusammen und stellt bei Bedarf ein verfahrenstechnisches Modell für den thermischen Kraftwerksbetrieb zur Verfügung. ERANEOS bereitet die Daten auf und stellt ein Datenmodell zur Verfügung. Neben dem verfahrenstechnischen Modell steht bereits vor Projektbeginn ein Datensatz aus dem Projekt DIZPROVI 8 zur Verfügung. Aus diesem Datensatz sollen in AP 3 von ABB erprobte KI Methoden zur Anomaliedetektion und zur Modellierung von Nutzereingriffen angewandt werden. Sollten sich im Verlauf DIZPROVI geeignete Modelle zur vorausschauende Instandhaltung entwickelt werden, soll geprüft werden ob das DIZPROVI Vorgehen in EXPLAIN in den erklärenden ML Lebenszyklus integriert werden kann. TUDA bietet den Industriepartnern Unterstützung bei der Definition des Datenmodells.

Im Rahmen des AP 2.2 wurden umfangreiche Maßnahmen zur Erfassung, Aufbereitung und Analyse betrieblicher Kraftwerksdaten umgesetzt.

Zu Beginn exportierte die LEAG rund 1 Terabyte an Rohdaten aus den Archiven des Kraftwerks Boxberg Q. Diese Daten umfassten sämtliche Betriebs-, Mess- und Stellwerte sowie das Melde- und Alarmarchiv. In Kooperation mit ABB erfolgte die Konvertierung der Rohdaten in eine strukturierte und nutzerfreundliche Datenbank. Zur Sicherstellung einer effizienten Nutzung wurden Workshops durchgeführt, in denen der Aufbau des Datensatzes, die zugrunde liegenden Datenpunkte sowie das Kraftwerkskennzeichnungssystem (KKS) erläutert wurden.

Eraneos verantwortete die laufende Qualitätssicherung der bereitgestellten Daten. Im Verlauf des Projekts wurden die ursprünglichen Datenquellen aus Anwendungsfall 1 ersetzt, um eine kontinuierliche Datenverfügbarkeit zu gewährleisten. Darüber hinaus erfolgte eine systematische Überprüfung, Bewertung und Aufbereitung der Daten. Ergänzend wurden externe Einflussgrößen – wie Schulferien und Brückentage, gewichtet nach dem Bruttoinlandsprodukt der jeweiligen Bundesländer – als zusätzliche Inputfeatures integriert, um die Modellierungsqualität zu verbessern.

Die Technische Universität Darmstadt brachte ihr fachliches Know-how insbesondere in die Analyse der KKS-Struktur ein. Durch Vor-Ort-Workshops und begleitende Abstimmungen wurden Herausforderungen bei der eindeutigen Zuordnung der Daten im KKS-System identifiziert und Lösungsansätze entwickelt. Dieser Beitrag trug wesentlich zur erfolgreichen Orchestrierung und semantischen Konsistenz des Gesamtdatensatzes bei.

AP 2.3 - Erstellung von ML-Datensatz zur Betriebsoptimierung in Walzwerken

Mitarbeitende Partner: ABB, TUDA

[redacted]

AP 2.4 Erstellung von ML Datensatz zur Wartungsplanung in Walzwerken (ABB, TUDA)

Mitarbeitende Partner: ABB, TUDA

[redacted]

[redacted]

AP 3 Konventionelle KI-Modellierung

Unter Verwendung der Daten aus AP 2 werden in diesem Arbeitspaket ML-Basismodelle mit konventionellen maschinellen Lerntechniken ohne erklärende Elemente entwickelt. Die Projektpartner werden etablierte Whitebox- (z.B. Entscheidungsbäume) und Blackbox-Modelle (z.B. tiefe neuronale Netze) einsetzen, um die verschiedenen Anwendungsfälle zu bearbeiten. Auch wenn es bereits Modelle gibt, werden sie zu Vergleichszwecken neu trainiert, um sie mit den neuen Erklärungsmethoden und einem standardisierten Benchmarking-Datensatz zu vergleichen.

AP 3.1 Konventionelle ML Modelle für Energiemarktvorhersagen

Mitarbeitende Partner: ERANEOS, LEAG, TUDA, SUH

Um das Maschinelle Lernen um Erklärbarkeit zu erweitern, wird durch die LEAG eine Datenbasis so wie Beschreibungen zu möglichen Algorithmen bereitgestellt. Die aktuellen Modelle sind in der Umgebung Matlab erstellt. Diese eignen sich nur bedingt für das Forschungsvorhaben, da die Funktionen nur bedingt öffentlich zugänglich sind. Den Partnern ERANEOS und TUDA kommt die Aufgabe zu eine geeignete ML-Umgebung zu definieren und die Modelle in dieser umzusetzen (z.B. auf Basis von Python). Grundsätzlich kann die Vorhersage als Regressionsproblem modelliert werden. Aufgrund der Vielzahl von Einflüssen bieten sich klassische Whitebox-Methoden wie Entscheidungsbäume oder lineare Regression weniger als beispielsweise Ensemble-Methoden (Random Forests, Gradient Boosting) oder tiefe Neuronale Netze (inbesondere Recurrent Neural Networks mit Long-term-short memory oder Gated-Recurrent-Units oder Transformer-Architekturen) an. Geeignete Open Source Libraries wie Prophet werden auf Anwendbarkeit geprüft. Ergebnis dieses Arbeitspakets ist dementsprechend ein oder mehrere ML-Modelle, die sich grundsätzlich für die Energiemarktvorhersage eignen. SUH erhebt Informationen aus den Arbeiten im Arbeitspaket zur Planung der Integration des ML-Modells in das Ende-zu-Ende System ML in Arbeitspaket 6.

Im Rahmen des Arbeitspakets 3.1 wurden auf Grundlage der in Arbeitspaket 2.1 erzeugten Daten verschiedene KI-Modelle zur Energiepreisvorhersage entwickelt, getestet und wissenschaftlich ausgewertet. LEAG stellte dabei die Ausgangsdaten bereit und führte Workshops durch, in denen bestehende Modellierungsansätze und aktuelle wissenschaftliche Erkenntnisse diskutiert wurden. Auf dieser Basis wurden zunächst zehn Modellklassen identifiziert, von denen nach Evaluierung acht weiterverfolgt und optimiert wurden. Neben der Modellierung der Strompreisprognosen wurden auch Vorverarbeitungsschritte untersucht, insbesondere die Wirkung von Parametervariationen sowie unterschiedlichen Transformationen und Normalisierungen der Daten.

Die TUDA evaluierte verschiedene Modellfamilien für multivariate Zeitreihenprognosen, darunter moderne Transformer-Modelle wie PatchTST, Timer-XL und Chronos sowie MLP-basierte Ansätze wie TiDE. In enger Abstimmung mit LEAG und ERANEOS zeigte sich, dass klassische neuronale Netze – insbesondere Long Short Term Memory (LSTM) – weiterhin hohe Prognosegüte aufweisen. Aufbauend darauf wurde die neu vorgestellte xLSTM-Architektur erfolgreich für Zeitreihendaten adaptiert (xLSTM-Mixer) und auf Benchmark-Datensätzen sowie auf LEAG-Daten angewendet. Parallel wurden hybride Verfahren

entwickelt, die symbolische und neuronale Methoden kombinieren, wie etwa das AnswerSetNetwork, das interpretierbare Logikkomponenten in neuronale Modelle integriert.

ERANEOS entwickelte ein reproduzierbares Softwarepaket, das die getesteten KI-Modelle für die Strompreisprognose integriert und mehrere Marktszenarien abbildet. Zusätzlich wurde ein alternatives Modell auf Basis der Gaussian-Process-Regression konzipiert und erfolgreich validiert. Die Ergebnisse der gemeinsamen Forschung wurden in wissenschaftlichen Publikationen und auf internationalen Fachkonferenzen vorgestellt, darunter das International Symposium on Forecasting 2024 in Dijon. Die SUH begleitete den Modellierungsprozess durch die Entwicklung von Prototypen für ausgewählte Anwendungsfälle. Sie leitete aus den praktischen Erkenntnissen Anforderungen für die Gestaltung der MLOps-Architektur ab, die in die technische Weiterentwicklung des Projekts einflossen. Durch die enge Zusammenarbeit aller Partner entstand ein umfassendes, validiertes Modellportfolio zur Energiepreisprognose, das sowohl wissenschaftliche als auch praktische Anwendungsperspektiven eröffnet.

AP 3.2 Unterstützungssystem für den nachhaltigen Kraftwerksbetrieb

Mitarbeitende Partner: ABB, ERANEOS, LEAG, TUDA, SUH

Die Modelle des Maschinellen Lernens in diesem Arbeitspaket setzen in erster Linie auf den analogen Signaldaten aus dem Kraftwerk und/oder einem kraftwerkstechnischen Modell auf. Die Aufgabe der Modelle kann die Detektion oder Vorhersage abnormaler Anlagenzustände sein oder die Fortführung (Regression) eines Schadensmodells abhängig von der geplanten Betriebsstrategie. Neben klassischen Methoden wie PCA oder Regression mit Regressorauswahl bieten sich hier insbesondere wieder Recurrent Neural Networks mit Long-term-short memory, Gated-Recurrent-Units oder Transformer-Architekturen an. ABB und ERANEOS werden in diesem Arbeitspaket ihre Expertise hinsichtlich vorausschauender Wartung und Anomaliedetektion einbringen und entsprechende Modelle auf Basis der Kraftwerksdaten erstellen. LEAG stellt für die Umsetzung der Modelle Expertenwissen über den Prozess und beteiligt sich bei der Bewertung der Ergebnisse. TUDA hilft bei der Auswahl und Anpassung des ML-Modells. SUH erhebt Informationen aus den Arbeiten im Arbeitspaket zur Planung der Integration des ML-Modells in das Ende-zu-Ende System ML in Arbeitspaket 6.

Im Arbeitspaket 3.2 wurde die Entwicklung eines KI-basierten Unterstützungssystems zur Anomalieerkennung im Kraftwerksbetrieb fortgeführt und in enger Zusammenarbeit zwischen ABB, LEAG und TUDA weiterentwickelt. Ziel war die frühzeitige Identifikation von Betriebsabweichungen, um Störungen zu vermeiden und die Anlagenverfügbarkeit zu erhöhen. Das von ABB entwickelte Dashboard basiert auf einem Autoencoder-Neuronalen Netzwerk, das Anomalien in Zeitreihendaten aus dem Kraftwerk Boxberg erkennt. Ergänzend wurden verschiedene Methoden der erklärbaren Künstlichen Intelligenz (XAI) integriert, um die Ursachen erkannter Abweichungen transparent darzustellen. Die Visualisierung erfolgt über eine Python-Dash-Oberfläche, die eine intuitive Nutzung im Anlagenbetrieb ermöglicht. Erste Evaluierungen mit Anlagenfahrern aus zwei LEAG-Kraftwerken bestätigten den praktischen Mehrwert des Systems insbesondere hinsichtlich der Verständlichkeit der erklärenden Komponenten. Die LEAG unterstützte das Projekt durch die Bereitstellung umfangreicher Betriebsdaten und eine detaillierte Anforderungsanalyse. In Abstimmung mit dem ABB-Forschungszentrum wurden geeignete Zeitreihen ausgewählt und Forschungsszenarien definiert. Zudem erfolgten Gespräche mit Fachpersonal vor Ort, deren Ergebnisse in Showcases überführt wurden, die als Grundlage für die weitere Systementwicklung dienten. Ergänzend entwickelte die TUDA den „Alarm Prediction Transformer“ (APT), ein multimodales Transformer-Modell zur frühzeitigen Vorhersage von Alarmen im Kraftwerksbetrieb. Das Modell nutzt kombinierte Sensor- und Ereignisdaten, um drohende Alarme sowie deren Typen vorherzusagen und damit eine präventive Problembehebung zu ermöglichen.

Durch die enge Kooperation der Partner entstand ein integriertes System, das datengetriebene Anomalieerkennung und Alarmprognose verbindet und einen wesentlichen Beitrag zur intelligenten, effizienten und sicheren Betriebsführung moderner Kraftwerksanlagen leistet.

AP 3.3 Modelle zur Betriebsoptimierung in Walzwerken

Mitarbeitende Partner: ABB, SUH, TUDA

Bei der Betriebsoptimierung in Walzwerken steht die Verknüpfung der im Walzprozess gesammelten Signaldaten mit Qualitäts- und Energiekennzahlen im Vordergrund. Dabei sind insbesondere die Signaldaten von besonderer Bedeutung. Zur Unterstützung des Anlagenpersonals können sowohl Klassifikations- (liegt ein Problem vor ja oder nein) als auch Regressions-Modelle (bspw. Bestimmung der besten Eingabeparameter) eingesetzt werden. Wegen des sequentiellen Charakter der Daten bieten sich auch hier vor allem Recurrent Neural Networks mit Long-term-short memory oder Gated-Recurrent-Units oder Transformer-Architekturen an. Neben den Signaldaten ist auch die Analyse von Videodaten von Bedeutung. Die Aufgabe besteht darin, Qualitätsproblemen im Walzwerk manuellen Eingriffen im Gießereiprozess zuzuordnen. Dadurch kann die eigentliche Ursache genau zeitlich bestimmt werden, wodurch eine verfahrenstechnische Analyse möglich wird. Zur Verarbeitung der Bilddaten bieten sich vor allem Convolutional Neural Networks an, um so von den Arbeiten des niederländischen Konsortium auf dem Gebiet der Bilddatenanalyse zu profitieren. Aufbauend auf Erfahrungen aus der chemischen Industrie, insbesondere der Modellierung von Batchprozessen, wird ABB Modelle zur Verknüpfung von Signal und Energie- und Qualitätskennzahlenarbeiten sowie ein erstes Modell zur Bilddatenanalyse aus der Gießerei erstellen. TUDA bringt ihre Expertise mit Deep-Learning-Methoden für Bild- und Videodaten ein. SUH erhebt Informationen aus den Arbeiten im Arbeitspaket zur Erstellung eines integrativen Modellierungsansatzes und zur Abstimmung mit dem existierenden Ansätzen zur Integration des ML-Modells in das Ende-zu-Ende System ML in Arbeitspaket 6.

Das Arbeitspaket wurde 2023 eingestellt. Grund hierfür waren unerwartete technische Hürden bei der Extraktion und Nutzung der Defektdaten aus einem proprietären Inspektionssystem. Insbesondere die fehlende Verfügbarkeit gelabelter Bilddaten sowie die Komplexität des proprietären Datenformats verhinderten eine zufriedenstellende Umsetzung mit den verfügbaren Mitteln. In Abstimmung mit den Projektpartnern wurde der Fokus daher auf das Thema „Wartungsplanung in Walzwerken“ verlagert. Die Entscheidung sowie die Hintergründe wurden bereits im Zwischenbericht 2023 dokumentiert.

AP 3.4 Modelle zur Wartungsplanung in Walzwerken

Mitarbeitende Partner: ABB, SUH, TUDA

Im Anwendungsfall zur Wartungsplanung geht es primär um das Aufdecken wiederkehrender, langfristiger Muster. Da die Dokumentation der Wartungsfälle im wesentlich als Freitext dokumentiert wird, sind hier Modelle für das Natural Language Processing einzusetzen. Einen möglichen Ansatz um wiederkehrende Muster aufzudecken stellt das Topic Modelling dar, das in den letzten Jahren zunehmend mit tief lernenden Modellen wie Variational Auto Encodern verknüpft wurde.

ABB soll auf Basis der Daten ein Topic Modell erstellen, dass die Aufdeckung von Mustern über längere Zeiträume ermöglicht. TUDA bringt ihr Fachwissen mit NLP-Methoden ein. SUH erhebt Informationen aus den Arbeiten im Arbeitspaket zur Planung der Integration des ML-Modells in das Ende-zu-Ende System ML in Arbeitspaket 6.

In dem AP 3.4 war ursprünglich vorgesehen, die in Freitext vorliegende Wartungsdokumentation mit Methoden des Natural Language Processing (NLP) zu analysieren und mithilfe von Topic Modelling wiederkehrende Muster über längere Zeiträume sichtbar zu machen. ABB sollte hierzu ein Modell entwickeln, TUDA ihr Fachwissen im Bereich NLP einbringen, während SUH die Integration in das Ende-zu-Ende-ML-System vorbereiten sollte. Erste Ansätze zum Topic Modelling wurden umgesetzt und lieferten wertvolle Einblicke.

In den begleitenden Nutzerinterviews zeigte sich jedoch, dass eine interaktive Unterstützung bei der

Informationssuche und Wissensnutzung von den Anwendern als deutlich hilfreicher eingeschätzt wurde als rein analytische Verfahren. Parallel dazu veränderten sich die Erwartungen durch die breite öffentliche Bekanntheit von ChatGPT, das erst nach Projektstart verfügbar wurde und Chatbots gesellschaftlich sichtbar machte.

Vor diesem Hintergrund verlagerte sich der Schwerpunkt der Arbeiten 2024 auf die Entwicklung eines domänenspezifischen Chatbot-Demonstrators zur Unterstützung der Wartungsplanung. Die technische Umsetzung basiert auf einem Retrieval-Augmented-Generation-Ansatz (RAG). Aufgrund rechtlicher Unsicherheiten beim Umgang mit proprietären Daten (vgl. AP 2.4) wurde der Demonstrator auf Grundlage öffentlich zugänglicher ABB-Produktinformationen realisiert. Als Plattform kam in Abstimmung mit dem Projektpartner eine Cloud-Lösung auf Basis von Microsoft Azure OpenAI zum Einsatz. Die Anwendung unterstützt verschiedene Dateiformate (z. B. Word, PowerPoint, PDF) und bietet damit eine flexible Grundlage zur Integration domänenspezifischen Wissens.

AP 4 Erklärungskomponente für industrielle KI

Das Arbeitspaket zielt darauf ab, generische Erklärungskomponenten zu entwickeln, die für die verschiedenen EXPLAIN-Anwendungsfälle verwendet werden können. Dazu gehören sowohl die Entwicklung von Erklärungsmethoden und inhärent erklärbaren Modellen als auch die Visualisierung und Präsentation der Erklärungen. Die Entwicklung dieser generischen Erklärungskomponenten wird iterativ mit zunehmender Reife über die drei Iterationen von PoCs und Piloten zum Testen (AP 7) entwickelt. Die mentalen Modelle der Endnutzer in den verschiedenen Anwendungsfällen (AP 1), die Eigenschaften der Datensätze in den verschiedenen Anwendungsfällen (AP 2) und die konventionellen KI-Modelle, die in AP 3 entwickelt wurden, werden den Ausgangspunkt für die Entwicklung der Erklärungskomponenten bilden. Während der Iterationszyklen reifen die Erklärungsmechanismen und Visualisierungen in ihrer Detailtreue und entwickeln sich von generischen zu anwendungsfallspezifischen Lösungen, mit dem Ziel einen High-Fidelity-Prototyp zu erstellen. Darüber hinaus ist das Arbeitspaket 4 eng mit derm Arbeitspaket 5 zu verknüpfen, um Lösungen zu finden und Mechanismen für erklärungsbasiertes Feedback unter Einbeziehung der Endnutzer zu erforschen. Die Ergebnisse des Arbeitspakets werden einer anwendungsbezogenen Bewertung mit dem Domänenexperten in AP 7 unterzogen. Die wahrgenommene Nützlichkeit und die Zeit, die man braucht, um die Erklärungen zu verstehen, werden die wichtigsten Messgrößen sein. Andere Metriken wie Stabilität, Trennbarkeit, Konsistenz und Ähnlichkeit zwischen Erklärungen werden vor der anwendungsbezogenen Bewertung berücksichtigt. Die Laufzeitanforderungen für die Erstellung von Erklärungen sind ein weiteres wichtiges Kriterium.

Das Arbeitspaket AP 4 wurde erfolgreich abgeschlossen. Der Schwerpunkt lag auf der finalen Entwicklung und prototypischen Implementierung generischer Erklärungskomponenten zur Erhöhung der Transparenz und Verständlichkeit der KI-Modelle in den verschiedenen industriellen Anwendungsfällen.

Im Rahmen von AP4 und AP5 wurde eine wöchentliche Workshopserie der TU Darmstadt geschaffen, um die Gemeinsamkeiten der einzelnen Use Cases herauszuarbeiten, diese zu generalisieren und neue Anwendungspotentiale zu entdecken. Ergebnis der Serie war unter anderem eine Übersicht aller im Projekt eingesetzten Erklärer und Feedback-Mechanismen und darüber, welche praktischen Annahmen gelten müssen, damit diese erfolgreich im industriellen Kontext eingesetzt werden können. Die Übersicht ist im EXPLAIN *Guidebook - Human-Centered Explainable AI for Process Industries*[1] der Öffentlichkeit zur Verfügung gestellt worden.

[1] Download: https://itea4.org/news/explain-guidebook-human-centered-explainable-ai-for-process-industries.html

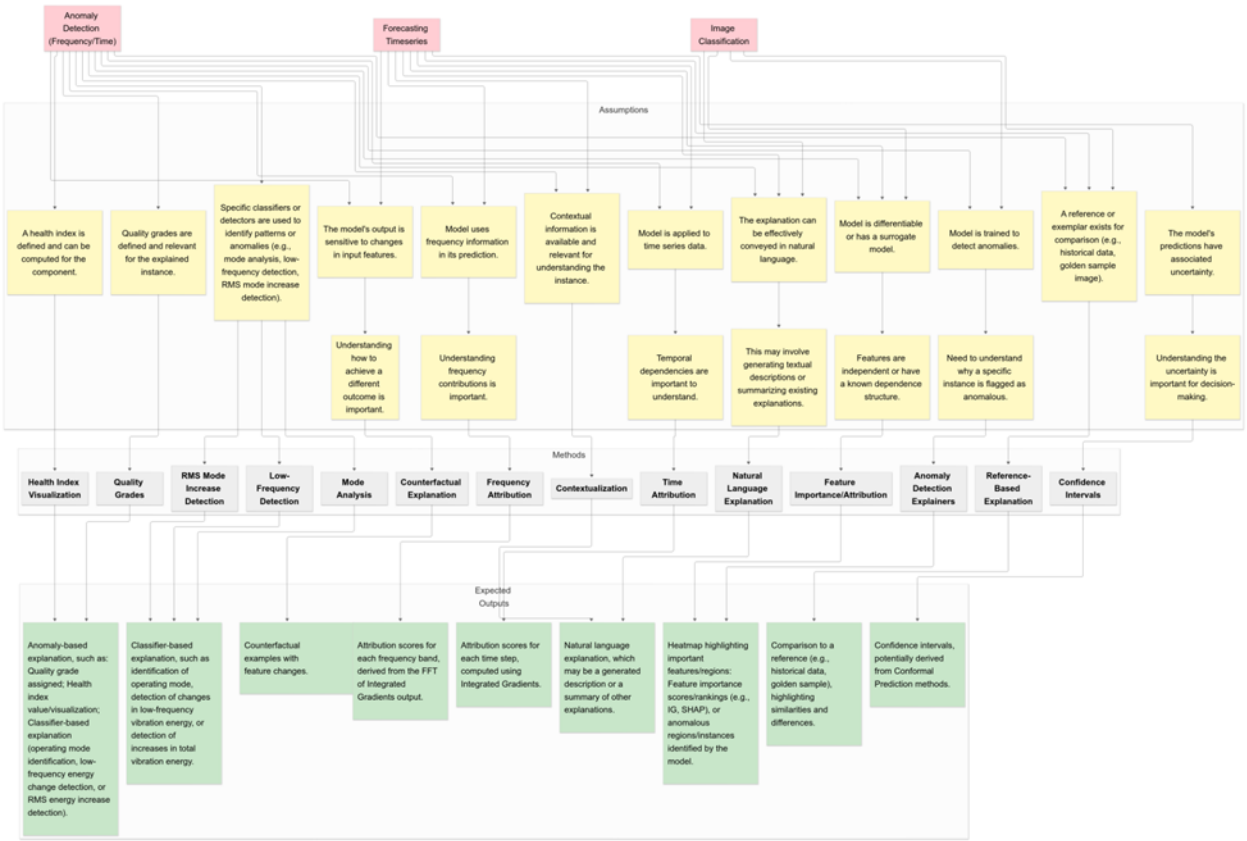


Abbildung 1 Zuordnung von Industrieproblemen und Datentypen zu Erklärungsmechanismen – von oben nach unten werden Anwendungsfälle (z. B. Anomalieerkennung, Prognose, Bildklassifikation), zentrale Annahmen (u. a. Zeitreihenbezug, verfügbare Kontextinformationen, Möglichkeit natürlicher Sprachbeschreibungen, Umgang mit Unsicherheit) sowie die daraus abgeleiteten Erklärungsmechanismen und erwarteten Ergebnisse abgebildet. Die erwarteten Outputs reichen von Gesundheits-/Qualitätsindikatoren und frequenzbezogenen Beiträgen über textuelle Zusammenfassungen und visuelle Hervorhebungen (z. B. Heatmaps) bis hin zu Referenzvergleichen und Konfidenzintervallen.

AP 4.1 Auswahl & Design von Erklärungsmechanismen

Mitarbeitende Partner: (ABB, LEAG, SUH, TUDA)

In diesem Arbeitspaket werden Erklärungsmechanismen aus dem momentanen Stand der Technik eingehender untersucht und hinsichtlich der Benutzerziele, -bedürfnisse und -erwartungen bewertet. Ausgehend von den identifizierten Erklärungsmechanismen werden zunächst mehrere Optionen als Mock-ups für eine frühe Bewertung durch die Benutzer erstellt. Hier kommen auch Wizard-of-Oz Ansätze in Frage. Basierend auf dem Feedback werden erste Prototypen auf Basis bspw. von Jupyter Notebooks umgesetzt. Schrittweise wird das Design der Erklärungsmechanismen gemeinsam mit dem Fortschritt in den übrigen Unterarbeitspakten über den Projektverlauf hinweg verbessert. ABB wird die Erkenntnisse aus AP 1 in dieses Arbeitspaket tragen und darüber hinaus die Mock-ups and Prototypen der Benutzerschnittstellen umsetzen. TUDA führt einen Überblick über den Stand der Technik bei Erklärungsmethoden durch und schlägt Ansätze vor, die für die Anwendungsfälle geeignet sind. LEAG beteiligt sich an der Validierung der Ergebnisse aus Sicht der Nutzer. SUH beteiligt sich an der Erhebung des Stand der Technik bei Erklärungsmethoden und evaluiert die Möglichkeit der Integration der Erklärungsmethoden in das Ende-zu-Ende System ML im Arbeitspaket 6.

Die TU Darmstadt führte eine systematische Analyse der verfügbaren Erklärungsverfahren durch und bewertete deren Eignung für industrielle Anwendungen. Dabei wurden insbesondere Verfahren wie LIME, SHAP und gradientenbasierte Methoden auf Zeitreihendaten angewandt. Darauf aufbauend wurden neue Verfahren zur Erklärbarkeit von Zeitreihenmodellen, darunter Feature Attribution, Time Interval Attribution und Time-Frequency Attribution, entwickelt und in Prototypen umgesetzt. Durch Workshops und wissenschaftliche Begleitung wurden die Ansätze kontinuierlich evaluiert und verbessert.

ABB brachte Erkenntnisse aus den Nutzeranalysen in AP 1 ein und konzentrierte sich auf die Entwicklung benutzerfreundlicher Schnittstellen für erklärbare KI-Systeme. Im Rahmen des LEAG-Anomalieerkennungs-Dashboards entwickelte ABB mehrere Erklärungsarten – darunter kontrafaktische Erklärungen, Merkmalswichtigkeiten, Kontextualisierungen und Erklärungen in natürlicher Sprache – und validierte diese gemeinsam mit Endanwendern. Die iterative Rückkopplung mit den Nutzern führte zu einer verbesserten Verständlichkeit und Akzeptanz der KI-Modelle.

LEAG war federführend bei der anwendungsnahen Evaluierung der Mock-ups und Prototypen. In mehreren Workshops mit Anlagenfahrern wurden die entwickelten Erklärungskomponenten auf Verständlichkeit, Relevanz und Praxistauglichkeit getestet. Das Feedback aus diesen Sitzungen wurde an ABB und TUDA weitergegeben und floss direkt in die Weiterentwicklung der Komponenten ein.

SUH beteiligte sich an der Analyse des Stands der Technik und evaluierte die Integrationsmöglichkeiten der entwickelten Erklärungsmechanismen in das End-to-End-System (AP 6). Zudem unterstützte SUH die Diskussionen zu methodischen Standards und zur langfristigen Nutzbarkeit erklärbarer Modelle im industriellen Umfeld.

AP 4.2 Konzeption und Implementierung robuster Erklärungsalgorithmen

Mitarbeitende Partner: TUDA

Unter Berücksichtigung der Ergebnisse aus AP 4.1 werden die ausgewählten Algorithmen implementiert oder modifiziert, um die Eingabedaten aus den industriellen Anwendungsfällen zu verarbeiten und Algorithmen zu identifizieren, die in allen Anwendungsfällen robust funktionieren. Der Schwerpunkt dieser Aufgabe liegt auf statischen Erklärungen, d.h. auf Methoden, die eine feste Erklärung für eine gegebene Datenprobe liefern. Neben solchen peturbations-basierten Erklärungsansätzen kommen aber auch andere Ansätze, wie beispielsweise Surrogate-Modelle, Analyse der Aktivierung in neuronalen Netzen oder Netzwerkarchitekturen mit Erweiterungen für Erklärbarkeit in Frage. Ein wesentlicher Aspekt dieses Arbeitspakets wird die Entwicklung von Methoden zur Peturbation von Eingabedaten sein, um künstliche Daten als Eingabe für Erklärungsmethoden (wie bspw. LIME) oder in AP 5 als Trainingsdaten für das Explanatory Interactive Learning zu generieren.

TU Darmstadt übernahm die Implementierung der entwickelten Zeitreihen-Erklärungsverfahren. Dabei wurden perturbationsbasierte und surrogatebasierte Ansätze kombiniert, um eine hohe Robustheit der Ergebnisse zu gewährleisten. Sie stellte den Quellcode allen Projektpartnern zur Verfügung, richtete eine interne Plattform für den Austausch und die Wiederverwendung von Code ein und unterstützte die Partner bei der technischen Umsetzung durch Workshops und direkte Beratung. Zusätzlich wurden Verfahren zur künstlichen Datengenerierung entwickelt, die zur Schulung erklärungsbasierter Modelle in AP 5 beitrugen.

AP 4.3 Begleitende Test und Algorithmen Evaluation

Mitarbeitende Partner: ABB, LEAG, SUH, TUDA

Begleitend zur Implementierung werden die Erklärungsalgorithmen als Ergänzung zur Nutzerevaluation in AP 6 Tests und Bewertungen mittels rechnerischer Metriken unterzogen. Das Ziel dieser begleitenden Tests ist es, Algorithmen einerseits frühzeitig bzgl. ihrer Stabilität, Konsistenz und Wiedergabetreue und andererseits nach ihrer Laufzeit bewerten zu können.

ABB unterstützt hier mit der Domänen- und Datenexpertise in der Konstruktion geeigneter Testdatensätze. LEAG beteiligt sich an den Tests, um die Verständlichkeit und den Nutzen der Ergebnisse für den

Anwender zu gewährleisten. TUDA bietet Unterstützung bei der Evaluierung der entwickelten Algorithmen. SUH evaluiert die Möglichkeit der Integration der entwickelten Algorithmen in das Ende-zu-Ende System ML im Arbeitspaket 6.

ABB unterstützte die begleitenden Tests durch die Bereitstellung von Domänen- und Datenexpertise und konstruierte geeignete Testdatensätze für die Bewertung der Erklärungskomponenten. Die Ergebnisse ergänzten die von den Nutzerinterviews erhobenen Erkenntnisse und trugen dazu bei, die Zuverlässigkeit und Robustheit der entwickelten Ansätze abzusichern.

LEAG führte Nutzertests und Interviews mit Key Usern durch, um Erkenntnisse über Akzeptanz, Verständnis und Benutzerverhalten zu gewinnen. Diese Erkenntnisse wurden ausgewertet und dokumentiert und wurden wiederum verwendet für die Weiterentwicklung der Erklärungsmechanismen.

Die Universität Hildesheim entwickelte eine Architektur und implementierte darauf basierend einen Prototyp für eine End-to-End-Daten- und MLOps-Infrastruktur. Dabei wurden Anforderungen an eine nahtlose Integrierung von Erklär-Methoden berücksichtigt, um diese für Modelltests, -überwachung, -optimierung und -Auditierung einsetzen zu können.

Die Algorithmen wurden im Rahmen wissenschaftlicher Publikationen auf frei verfügbaren Benchmarks evaluiert. Darüber hinaus beteiligte sich die TU Darmstadt aktiv an der Zusammenarbeit mit Key Usern und führte Interviews durch, um Einblicke in Nutzungsmuster, Verständlichkeit und Akzeptanz zu gewinnen. Die daraus abgeleiteten Erkenntnisse wurden anschließend gemeinsam mit den Partnern diskutiert und ausgewertet.

AP 4.4 Implementierung der Erklärungskomponente

Mitarbeitende Partner: ABB, LEAG, SUH, TUDA

Die implementierten Algorithmen sind für die Integration in die späteren Prototypen in geeigneter Weise zu kapseln und bereitzustellen, sodass sie über die Ende-zu-Ende ML Plattform aus AP 6 in den ML-Lebenszyklus eingebunden werden können. ABB unterstützt diese Arbeiten um eine einfache Integration in die Zielsysteme in AP 6 sicherzustellen. SUH unterstützt diese Arbeiten um die problemlose Integration der Algorithmen in das Ende-zu-Ende System ML in AP 6 sicherzustellen.

TU Darmstadt stellte den Quellcode der entwickelten Erklärungsverfahren bereit, richtete eine gemeinsame Code-Plattform ein und unterstützte die Partner bei der Integration der Komponenten in das ML-System. Durch Workshops und technische Beratung wurde sichergestellt, dass die Methoden effizient in die Prototypen eingebunden werden konnten.

ABB wirkte daran mit, die entwickelten Erklärungskomponenten so aufzubereiten, dass eine spätere Einbindung in die Ende-zu-Ende-ML-Plattform (AP 6) möglich ist. Damit wurde die Grundlage geschaffen, ausgewählte Funktionen in den ML-Lebenszyklus und in nachfolgende Prototypen zu übertragen. Im

[redacted]

LEAG testete die integrierten Komponenten in der Praxis, sammelte Rückmeldungen von Endanwendern und bewertete die Verständlichkeit und Nützlichkeit der Erklärungen. Diese Erkenntnisse wurden zur weiteren Feinabstimmung der Visualisierung und Benutzerführung verwendet.

Die Universität Hildesheim übernahm die technische Validierung der Integration in das End-to-End-System. Sie stellte sicher, dass die Erklärungsverfahren nahtlos in bestehende MLOps-Prozesse eingebettet werden konnten, und begleitete den Übergang in AP 6 durch Qualitätssicherung und Dokumentation.

AP 5 Erklärungsbasiertes Feedback für industrielle KI

Das Arbeitspaket konzentriert sich auf die Entwicklung einer Feedback-Komponente für die in AP 3 erstellten ML-Modelle. Das Design der Feedback-Komponente unterstützt einen iterativen interaktiven Trainingsprozess, der mehrere Schritte umfasst. Zunächst wird eine Teilmenge der Trainingsdaten ausgewählt, und dem Benutzer werden die Vorhersagen des Modells für diese Teilmenge sowie die von der in AP 4 entwickelten Erklärungskomponente erstellten Erklärungen präsentiert. Zweitens wird über eine verständliche und benutzerfreundliche Schnittstelle das Feedback der Benutzer eingeholt, wobei sowohl die Korrektheit der Vorhersagen als auch die Gültigkeit der Erklärungen bewertet wird, wobei ein Gleichgewicht zwischen dem vom Benutzer geforderten Aufwand und der Fülle des Feedbacks hergestellt wird. Drittens wird das erhaltene Feedback in den Trainingsprozess einbezogen, um das Modell dazu anzuleiten, sich auf relevante Merkmale und Muster in den Daten zu stützen, was sowohl zu genaueren Vorhersagen als auch zu vernünftigeren Erklärungen führen sollte. Die Verbesserung, die durch das iterative interaktive Lernen gegenüber dem traditionellen Ansatz erzielt wird, kann durch den Vergleich konventioneller Leistungskennzahlen (F1-Score, Konfusionsmatrizen usw.) mit den Vorhersagen des Modells sowie der Übereinstimmung der Erklärungen mit dem Nutzerfeedback gemessen werden, um die Robustheit der Ergebnisse zu gewährleisten. Im Laufe der Entwicklung können auf der Grundlage von Zwischenvalidierungsergebnissen und qualitativen Rückmeldungen der Nutzer neue Effizienz- und Genauigkeitsmaße festgelegt werden.

Im Arbeitspaket 5 wurde ein interaktives Feedbacksystem entwickelt, das es Domänenexperten ermöglicht, Modellvorhersagen und deren Erklärungen zu bewerten. Dieses Feedback fließt in den Trainingsprozess ein und verbessert so Vorhersagequalität und Erklärbarkeit. Die Mechanismen sind für die Integration in die End-to-End-ML-Plattform aus AP 6 vorbereitet.

Die TU Darmstadt war für die Organisation und Leitung dieses Arbeitspakets verantwortlich. Im Rahmen dessen wurden Workshops durchgeführt, um Designentwürfe für die einzelnen Use Cases zu entwickeln, iterativ zu verfeinern und den Wissensaustausch zwischen den Use Cases zu fördern. Darüber hinaus fanden, wie in Arbeitspaket 4 beschrieben, Workshops zur Generalisierung der Erklärverfahren sowie zur Gestaltung von Feedback-Mechanismen statt.

Die Universität Hildesheim überprüfte die Entwurfsziele hinsichtlich ihrer Integration in den ML-Lebenszyklus und gab im Rahmen der initialen Architektur technische Empfehlungen zur Einbindung der Feedback-Mechanismen in das End-to-End-System.

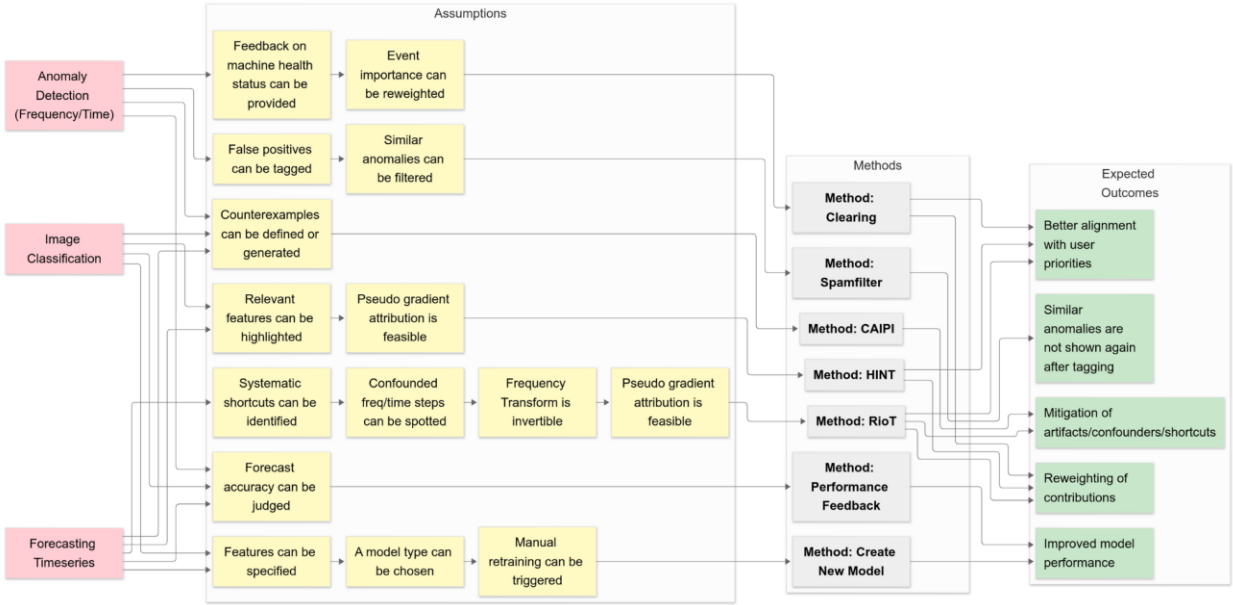


Abbildung 1 Zuordnung von Industrieproblemen und Datentypen zu Feedback-Mechanismen – von links nach rechts werden Annahmen, unsere Methoden und die erwarteten Outputs dargestellt. Die Darstellung zeigt, wie Feedback aus den Anwendungsfällen über diese Annahmen in passende Vorgehensweisen überführt wird, um Nutzerprioritäten besser zu treffen, unerwünschte Effekte zu mindern und die Modellleistung zu verbessern.

AP 5.1 – Design von Feedbackmechanismen und Prozessen

Mitarbeitende Partner: ABB, LEAG, SUH, TUDA

Das Konzept des erklärungsbasierten Feedbacks sieht vor, dass Domänenexperten im Trainingsprozess Rückmeldung mithilfe von Erklärungen einbringen. In diesem Arbeitspaket wird untersucht, in welcher Art und Weise diese Rückmeldung erbracht werden kann. Im einfachsten Fall kann dies über das Ablehnen oder Akzeptieren einer Erklärung geschehen oder durch die Angabe eines Gegenbeispiels. Ausgehend von den Datentypen, die in den Anwendungsfällen vorliegen, werden in diesem Arbeitspaket die Nutzerinteraktionen entworfen und in einem ersten Schritt durch Mock-Ups für eine erste Validierung durch die Domänenexperten aufbereitet. Im weiteren Verlauf des Arbeitspakets werden zunehmend dynamischere Prototypen entwickelt, beispielsweise von Implementierungen in Jupyter Notebooks bis zu vollwertigen web-basierten Nutzerschnittstellen.

ABB wird die Erkenntnisse aus AP 1 in dieses Arbeitspaket tragen und darüber hinaus die Mock-ups and Prototypen der Benutzerschnittstellen umsetzen. LEAG und ABB werden spezifische Anforderungen aus ihren Anwendungsdomänen in das Arbeitspaket einbringen. TUDA bietet Beratung und Unterstützung aus technischer Sicht. SUH bietet Beratung und Unterstützung aus technischer Sicht, insbesondere mit Hinblick auf die softwaretechnische Integration der Prototypen in das die Ende-zu-Ende System ML in AP 6.

ABB brachte die im Rahmen von Arbeitspaket 1 gewonnenen Erkenntnisse aus den Nutzeranalysen ein und arbeitete an der Konzeption benutzerzentrierter Feedbackprozesse. Gemeinsam mit den Projektpartnern wurde eine Workshop-Reihe initiiert, die eine enge Verzahnung der Arbeitspakete 4, 5 und 6 sicherstellte. Im Zuge dieser Zusammenarbeit entstanden erste Mock-ups und interaktive Prototypen, die gemeinsam mit Domänenexpertinnen und -experten diskutiert und iterativ weiterentwickelt wurden.

Für die von der LEAG betreuten Use Cases wurden Workshops durchgeführt, in denen verschiedene Feedback-Komponenten evaluiert und zu testende Methoden erarbeitet wurden. Besonderes Augenmerk lag dabei auf der Gestaltung intuitiver Eingabemöglichkeiten für die Bedienenden.

Die Universität Hildesheim unterstützte die Integration der entwickelten Komponenten in die übergreifende Referenzarchitektur.

Die in Arbeitspaket 4 entwickelten Erklärverfahren für Zeitreihen bildeten die Grundlage für die Feedback-Mechanismen im Rahmen des Explanatory Interactive Learning Frameworks. Diese Mechanismen wurden gemeinsam mit den Partnern konzipiert und an die unterschiedlichen Modalitäten angepasst. Dabei wurde je nach Anwendungsfall auf bestehende Algorithmen zurückgegriffen oder neue Ansätze modelliert.

AP 5.2 – Konzeption und Implementierung von Feedback-Mechanismen

Mitarbeitende Partner: ABB, LEAG, SUH, TUDA

Das Feedback von Domänenexperten muss im Trainingsprozess verarbeitet werden. Mit den Methoden aus den Vorarbeiten erfolgt dies über Gegenbeispiele, die aus dem Feedback der Domänenexperten abgeleitet werden. Dabei sind vor allem zwei Elemente auf die industrielle Anwendung zu untersuchen und

anzupassen: Die Query-Strategie und die Generierung von Gegenbeispielen aus dem Expertenfeedback. Die Query-Strategie entscheidet für welche Trainingsbeispiele das System das Feedback des Experten einholt. Damit hat diese Komponente einen entscheidenden Einfluss auf die Labeleffizienz und damit auf den Aufwand den die Domänenexperten in das Trainieren des Modells investieren müssen. Die Generierung von Gegenbeispielen entscheidet wie effektiv und korrekt das ML-Model durch das Feedback lernt. Zur Generierung kommen einerseits pertubations-basierte Ansätze in Frage (wie bspw. in LIME zur Erklärung genutzt) und andererseits generative Modelle wie Variational AutoEncoder oder Generativ-Adverserial-Networks. Eine wesentliche Herausforderung bei der Generierung wird darin bestehen, dass die generierten Beispiele aus verfahrenstechnischer Sicht ausreichend realistisch sind.

ABB und LEAG unterstützen dieses Arbeitspaket punktuell, um die Generierung von Gegenbeispielen aus der Anwendungsperspektive zu unterstützen. TUDA entwickelt und implementiert Abfragestrategien sowie Modellrevisionsmethoden mit Hilfe von Gegenbeispielen und Anpassung der Verlustfunktion. SUH bietet Beratung und Unterstützung aus technischer Sicht, insbesondere mit Hinblick auf die softwaretechnische Integration des Feedback-Mechanismus in das die Ende-zu-Ende System ML in AP 6.

Im Anwendungsfall *„Nachhaltiger Kraftwerksbetrieb“* (LEAG) entwickelte ABB einen Mechanismus, mit dem Anlagenfahrerinnen und -fahrer Modellvorhersagen bewerten können. Grundlage war ein Label-Propagation-Ansatz auf Basis siamesischer neuronaler Netze: Nutzerinnen und Nutzer konnten Vorhersagen als korrekt oder fehlerhaft markieren, woraufhin das System dieses Feedback auf ähnliche Fälle übertrug. Auf diese Weise ließen sich künftig falsch-positive Alarme automatisch herausfiltern.

Für die Entwicklung der Feedback-Mechanismen sichtete die LEAG die zugrunde liegenden Daten und wählte repräsentative Zustände aus, um den Trainingsprozess zu beschleunigen. Die Implementierung erfolgte in enger Abstimmung mit den Projektpartnern, um Algorithmen sowie die Darstellung von Abfragen und Ergebnissen optimal an die Bedürfnisse der Endnutzenden (z. B. Leitstandsfahrerinnen, Dispatcher) anzupassen.

Die Universität Hildesheim führte Interviews durch, nahm an Workshops und Besprechungen teil, sammelte Anforderungen und stellte diese den Projektpartnern bereit. Darüber hinaus übernahm SUH die technische Beratung und Unterstützung, insbesondere im Hinblick auf die Softwareintegration der Feedback-Mechanismen in die MLOps-Architektur des Arbeitspakets 6.

Die in Arbeitspaket 4 entwickelten Erklärerverfahren für Zeitreihen wurden in Arbeitspaket 5 genutzt, um einen Prototyp für eine interaktive Energiepreisvorhersage zu entwickeln. Im ersten Schritt können Nutzerinnen und Nutzer sich Modellvorhersagen visualisieren und erklären lassen. Anschließend besteht die Möglichkeit, gezieltes Feedback zu geben – z.B. wenn sich das Modell auf Faktoren stützt, die aus Sicht der Operatoren für die aktuelle Vorhersage nicht relevant sind. Über das Interface kann dieses Feedback an das System übermittelt werden. Hierfür stehen die von der TU Darmstadt entwickelte Methode RioT sowie HINT und Performance Rating zur Verfügung, mit denen das Feedback in das Modell zurückgeführt werden kann.

AP 5.3 – Tests und Algorithmen Evaluation

Mitarbeitende Partner: ABB, LEAG, TUDA

Neben der Validierung mit den Nutzen in AP 7 sollen die entwickelten Algorithmen bereits frühzeitig auf Basis von geeigneten Kennzahlen getestet werden. Neben dem Vergleich der konventionellen Metriken des Maschinellen Lernens (bspw. Treffergenauigkeit und F1-Score für Klassifikation oder Fehlerquadrate für Regression) spielt für das interaktive Lernen mit Feedback auch Labeleffizienz eine entscheidende Rolle. Hier ist die Frage, wie viel Feedback Domänenexperten geben müssen, um eine bestimmte Leistung

des ML-Modells zu erreichen. Daneben spielen Laufzeitbetrachtungen eine wichtige Rolle, um die Vision eines interaktiven Lernprozess erreichen zu können. Für den Aspekt der Datengenerierung wird diese Betrachtung um stark vereinfachte Nutzertest ergänzt, in denen die Domänenexperten den Realitätsgrad und die Entsprechung zum Feedback interaktiv bewerten.

Die TU Darmstadt bewertet die Genauigkeit und Effizienz des interaktiven Lernalgorithmus und deren Vergleich mit den Basismodellen. ABB und LEAG unterstützen hier vor allem die Bewertung der künstlich generierten Daten.

ABB unterstützte die begleitende Evaluation der entwickelten Mechanismen. Für das Anomalie-Dashboard wurden erste Tests mit Anlagenfahrern durchgeführt. Die Evaluation erfolgte anhand eines semi-strukturierten Fragebogens und ergänzte die rechnerischen Metriken um wertvolle Nutzerperspektiven. Parallel dazu untersuchte ABB algorithmische Konzepte zur Feedbackverarbeitung, insbesondere im Kontext von Few-Shot Learning mit siamesischen Netzwerken, um auch bei begrenztem Feedback Modellverbesserungen zu ermöglichen.

Die Key User der LEAG testeten und bewerteten die Ergebnisse in mehreren Iterationen. Das gesammelte Feedback zu den Algorithmen wurde direkt in die Weiterentwicklung der KI integriert und für nachfolgende Arbeitsschritte genutzt. Die Ergebnisse aus den Nutzerinterviews wurden gemeinsam mit allen Partnern ausgewertet und diskutiert. Sowohl positives als auch kritisches Feedback wurde systematisch aufbereitet und für künftige Entwicklungsschritte dokumentiert.

Die entwickelten Feedback-Mechanismen wurden von der TU Darmstadt evaluiert und wissenschaftlich veröffentlicht. Darüber hinaus wurden die Prototypen den Energie-Dispatcherinnen und -Dispatchern im Rahmen von Interviews vorgestellt, um zusätzliche Anforderungen zu erheben. Die gewonnenen Erkenntnisse wurden anschließend den Projektpartnern zur Verfügung gestellt.

AP 5.4 – Implementierung eines interaktiven Feedback-basierten Lernsystems

Mitarbeitende Partner: ABB, SUH, TUDA

Die implementierten Algorithmen sind für die Integration in die späteren Prototypen in geeigneter Weise zu kapseln und bereitzustellen, sodass sie über die Ende-zu-Ende ML Platform aus AP 6 in den ML Lebenszyklus eingebunden werden können.

ABB unterstützt diese Arbeiten um eine einfache Integration in die Zielsysteme in AP 7 sicherzustellen. SUH integriert Funktionalitäten zur Unterstützung der implementierten Algorithmen und deren Integration in das Ende-zu-Ende System ML aus AP 6.

[redacted]

Die Universität Hildesheim integrierte Funktionalitäten der entwickelten Partnerlösungen, um die implementierten Algorithmen zu unterstützen und deren Einbindung in das End-to-End-ML-System des Arbeitspakets 6 sicherzustellen. Der Quellcode des in Arbeitspaket 5.2 beschriebenen Prototyps wurde

einschließlich Benutzeroberfläche, Erklärverfahren und Feedback-Mechanismen dem Konsortium zur Verfügung gestellt. Die zugrunde liegenden Mechanismen wurden zudem separat veröffentlicht.

AP 6 Ende-zu-Ende *„Machine Learning“* System mit Erklärungen und Feedback

Entwicklung einer durchgängigen Daten- und MLOps-Infrastruktur unter Berücksichtigung des Bedarfs an nahtlosen Erklärungsmethoden und Nutzung von Erklärungen für Modelltests, Überwachung, Verbesserung und Audits. In diesem Arbeitspaket wird eine durchgängige Daten- und MLOps-Infrastruktur, das Ende-zu-Ende System ML, entwickelt, die den Bedarf an nahtlosen Erklärungsmethoden berücksichtigt und Erklärungen für Modelltests, Überwachung, Verbesserung und Audits nutzt. Um dies zu erreichen, ist die erste Aufgabe des Arbeitspakets die Modellierung und Automatisierung von MLOps-Prozessen mit der Fähigkeit zur Optimierung von MLOps-Prozessen. In diesem Schritt wird auch die Integration der Funktionalität für die Orchestrierung von MLOps-Prozessen sowie die Orchestrierung/Integration von Modellen in diese Prozesse entwickelt. In diesen Schritten werden auch die in Arbeitspaket 4 entwickelte und bereitgestellte Erklärungskomponente. Der nächste Schritt ist dann die Entwicklung von Funktionalitäten zur Automatisierung von MLOps-Prozessen, wie z.B. die automatische Generierung sowie die Nutzung von Funktionalitäten zur Modellüberwachung und die Aktualisierung und Ersetzung von Modellen. Der wichtigste Input für diesen Schritt sind die in Arbeitspaket 4 bereitgestellten Ansätze für erklärendes Training, die in die Funktionalitäten zur Automatisierung von MLOps-Prozessen integriert werden. Weitere Funktionalitäten werden dann durch die Entwicklung integrierter automatischer Konsistenzprüfungen bereitgestellt, z.B. bei der Datenauswahl für die Erstellung von Modellen, sowie für die automatische Prüfung auf Modell- oder Konzeptdrift. Jede der Funktionalitäten wird getestet und überprüft, um eine nahtlose Integration der Erklärungskomponente und der erklärenden Trainingsansätze aus Arbeitspaket 4 und Arbeitspaket 5 sicherzustellen. Beim Testen der Funktionalitäten werden Metriken verwendet, wie z. B. die Zeit bis zum erneuten Einsatz eines neuen Modells, der Umfang der manuellen Schritte beim Einsatz und die Zeit bis zum Abruf von Trainingsdaten, um die Leistung der Funktionalitäten zu messen. Die Gesamtheit der oben genannten Funktionen wird dann in eine Ende-zu-Ende MLOps-Infrastruktur integriert, das Ende-zu-Ende System ML, die ihrerseits wiederum anhand von Metriken getestet und evaluiert werden wird.

Im Rahmen von Arbeitspaket 6 entwickelten SUH und ABB gemeinsam einen ersten Demonstrator für ein MLOps-System und leisteten wichtige Beiträge zur Konzeption und Implementierung einer praxistauglichen Architektur.

SUH implementierte im Demonstrator automatisierte Konsistenzprüfungen, beispielsweise zur Datenauswahl bei der Modellerstellung und zur automatischen Modellvalidierung. Darüber hinaus wurden zentrale MLOps-Funktionalitäten integriert, darunter ein Model Store, Data Store, Feature Store, Versionierung und weitere notwendige Komponenten, die den operativen Betrieb und die Nachvollziehbarkeit von Machine-Learning-Prozessen unterstützen.

ABB brachte Erfahrungen aus dem Aufbau einer MLOps-Architektur für die schwedischen Use Cases ein. Insbesondere bei Boliden konnte aufgrund der vorhandenen IT-Infrastruktur eine vollständige Integration der entwickelten Komponenten vorbereitet und initialisiert werden. Die dort etablierte Architektur ermöglicht den direkten Betrieb von ML-Anwendungen in der Zielumgebung sowie deren Anbindung an bestehende Systeme über standardisierte Schnittstellen. Diese praktischen Erfahrungen flossen iterativ in die Abstimmung mit SUH ein und trugen zur Entwicklung von Architekturmuster bei, die die Grundlage für die operative Integration im finalen Innovationszyklus (Arbeitspaket 7) bilden.

AP 6.1 Gesamtarchitektur für Ende-zu-Ende System ML

Mitarbeitende Partner: ABB, ERANEOS, LEAG, SUH

In diesem Unterarbeitspaket werden die funktionalen und nicht-funktionalen Anforderung an eine durchgängige Daten- und MLOps-Infrastruktur, das Ende-zu-Ende System ML, erhoben und eine Komponentenstruktur, einschließlich der Abhängigkeiten zwischen den Komponenten der Infrastruktur und dem dynamischen Verhalten der Komponenten, definiert.

ABB bringt vor allem die Anforderungen aus AP 1 in dieses Arbeitspaket und seine Erfahrungen in Softwarearchitektur und Maschinellem Lernen beratend ein. SUH wird die funktionalen und nicht-funktionalen Anforderungen sammeln und dokumentieren und zur Erstellung einer ersten Software-Architektur beitragen. Aufbauend auf den erfassten Anforderungen wird SUH zur Realisierung einer ersten Infrastrukturvorlage beitragen. Die Vorlage wird containerisierte Komponenten und ihre Interaktionen beschreiben und es den Partnern ermöglichen, die Infrastruktur in Cloud-Plattformen oder vor Ort einzusetzen. ERANEOS bringt die technischen Anforderungen aus dem Anwendungsfall Energiemarktvorhersagen in das Arbeitspaket ein. Die LEAG definiert Anforderungen aus Sicht eines KRITIS Unternehmens an die Architektur.

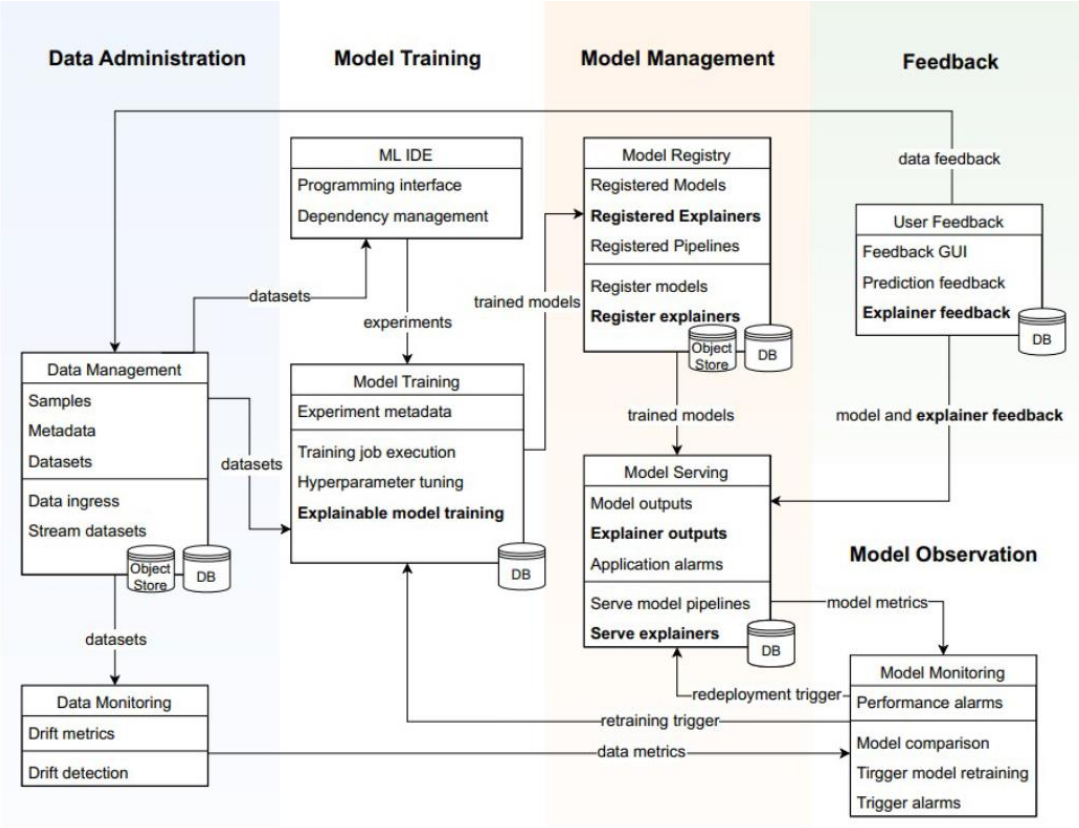
Im Rahmen von Arbeitspaket 6 leisteten ABB, LEAG und SUH wesentliche Beiträge zur Entwicklung und Implementierung einer MLOps-Architektur, die den spezifischen Anforderungen der industriellen Anwendungsfälle gerecht wird.

ABB brachte Anforderungen aus den eigenen industriellen Use Cases in die Architekturarbeit ein und unterstützte SUH mit fundierter Expertise in den Bereichen Softwarearchitektur und Maschinelles Lernen. Ein besonderer Schwerpunkt lag auf den schwedischen Anwendungsfällen Södra und Boliden. Hier konnte ABB wertvolle Erfahrungen aus der industriellen Praxis einbringen, die maßgeblich in die Definition sowohl der funktionalen als auch der nicht-funktionalen Anforderungen einflossen.

LEAG ergänzte diese Arbeiten durch die Erhebung und Analyse von Anforderungen an ein Ende-zu-Ende-System. In Workshops wurden bestehende Erfahrungen eingebracht und zentrale Themen wie der sichere Umgang mit Daten aus kritischer Infrastruktur sowie die Anbindung an weiterverarbeitende Systeme identifiziert. Diese Erkenntnisse bildeten eine wichtige Grundlage für die Definition der System- und Sicherheitsanforderungen innerhalb der Gesamtarchitektur.

SUH war für die Erstellung der initialen Softwarearchitektur verantwortlich und dokumentierte die nicht-funktionalen Anforderungen. Auf dieser Basis entwickelte SUH ein Infrastruktur-Template, das containerisierte Komponenten und deren Interaktionen beschreibt und den Partnern eine Bereitstellung der Infrastruktur sowohl in Cloud-Umgebungen als auch on-premise ermöglicht. Ergänzend führte SUH Literaturrecherchen und Interviews durch, um gängige Standards, Werkzeuge und Praktiken im Bereich MLOps zu identifizieren. Die dabei gewonnenen Erkenntnisse wurden in die architektonische Ausgestaltung integriert. Darüber hinaus stellte SUH die identifizierten Herausforderungen von MLOps im industriellen Kontext auf der IN4PL-Konferenz vor und diskutierte diese mit Fachvertreterinnen und Fachvertretern aus Wissenschaft und Industrie.

Die von ABB, LEAG und SUH zusammengetragenen Anforderungen und Beiträge bildeten die Grundlage für die Beschreibung einer ersten MLOps-Architektur, die die spezifischen Herausforderungen und Bedarfe der industriellen Anwendungsfälle berücksichtigt. Diese Arbeiten trugen wesentlich zur Erreichung der Ziele von Arbeitspaket 6 und zur Vorbereitung der weiteren technischen Umsetzung in den folgenden Projektphasen bei.



Referenz MLOps-Architektur

Die MLOps-Architektur umfasst mehrere Kernkomponenten, die die Verwaltung, Nachvollziehbarkeit und Qualitätssicherung von KI-Modellen gewährleisten:

- **Data Management:** Sammlung, Annotation und Versionierung von Daten und Metadaten. Diese Komponente ermöglicht eine strukturierte Repräsentation und Bewertung von Datenquellen, wodurch die Wiederverwendbarkeit und Nachvollziehbarkeit von Trainings- und Testdaten verbessert wird.
- **Data Monitoring:** Überwachung der Modell- und XAI-Performance im Betrieb, Erkennung von *Concept Drift*sowie automatisierte Alarme und Schwellenwertprüfungen, um Modellabweichungen frühzeitig zu erkennen.
- **ML IDE:** Bereitstellung einer integrierten Entwicklungsumgebung für die Durchführung erster Experimente sowie als Grundlage für die Entwicklung neuer ML-Modelle und Pipelines.
- **Model Training:** Ausführung und Automatisierung von Trainingspipelines einschließlich Hyperparameter-Tuning. Diese Komponente unterstützt die systematische Erfassung und den Vergleich von Experimenten sowie die Auswahl optimaler Modelle.
- **Model Registry:** Verwaltung und Registrierung von trainierten Modellen, XAI-Methoden und zugehörigen Pipelines. Sie stellt sicher, dass Modelle eindeutig versioniert und reproduzierbar bereitgestellt werden können.
- **Model Serving:** Bereitstellung und Skalierung der registrierten Modelle in produktiven Umgebungen. Diese Komponente ermöglicht den robusten Betrieb von ML-Anwendungen direkt in industriellen IT-Systemen.
- **User Feedback:** Integration von Rückmeldungen der Endanwender zu Funktionalität und Qualität der Modelle. Diese Funktion erlaubt zudem das Labeln neuer Daten und die Visualisierung erklärbarer Modellentscheidungen, wodurch Vertrauen und Nutzbarkeit in der Praxis gesteigert werden.

- **Model Monitoring:** Kontinuierliche Bewertung der Modellgüte im produktiven Einsatz. Bei signifikanter Leistungsabnahme können automatisierte Prozesse für Retraining oder Modellanpassung ausgelöst werden.

Durch die Kombination dieser Komponenten wird ein vollständiger MLOps-Zyklus abgebildet, der sowohl technische Effizienz als auch Transparenz und Nachvollziehbarkeit fördert. Die Architektur ist modular aufgebaut, containerisiert und kann flexibel in Cloud- oder On-Premise-Umgebungen betrieben werden.

AP 6.2 Integration der Erklärungskomponente in das Ende-zu-Ende System

Mitarbeitende Partner: ABB, SUH

Die in AP 4.4 entwickelte Erklärungskomponente ist in den Prozessablauf und in die technische Infrastruktur des Ende-zu-Ende System ML zu integrieren. Hier ist insbesondere eine relevante Frage, wie die Entwicklung eines ML-Modells gemeinsam mit der Entwicklung der spezifischen Erklärungskomponente gesteuert und protokolliert werden kann und welche DevOps Prozesse notwendig sind, um sicherzustellen, dass immer eine zum aktuell verwendeten Modell passende Erklärungskomponente eingesetzt wird. ABB unterstützt SUH in der Integration um so die spätere Einbidung in die Zielsysteme in AP 7 sicherzustellen. SUH erweitert der Infrastrukturvorlage, um den besonderen Anforderungen bei der Erstellung und Bereitstellung der Erklärungskomponente gerecht zu werden und einen erklärenden ML-Workflow für die Trainingsphase von ML Modellen sowie einen erklärungsbasierten Evaluierungs- und Freigabeprozess für ML Modelle einzubeziehen.
Im Rahmen von Arbeitspaket 6.2 arbeiteten SUH und ABB gemeinsam an der Weiterentwicklung der MLOps-Architektur, insbesondere im Hinblick auf die Integration von Erklärbarkeitskomponenten (Explainability).

Im Rahmen von Arbeitspaket 6.2 arbeiteten SUH und ABB gemeinsam an der Weiterentwicklung der MLOps-Architektur, insbesondere im Hinblick auf die Integration von Erklärbarkeitskomponenten (Explainability).

SUH war maßgeblich an der architektonischen Konzeption beteiligt und entwickelte eine Architektur, die aufzeigt, wie erklärbare KI-Modelle in den bestehenden MLOps-Workflow integriert werden können. Dabei erfolgte ein enger Austausch mit den Projektpartnern, um sicherzustellen, dass die konzipierten Lösungen sowohl technisch umsetzbar als auch mit den Anforderungen der unterschiedlichen industriellen Anwendungsfälle kompatibel sind.

ABB beteiligte sich beratend an der Integration der in Arbeitspaket 4.4 entwickelten Erklärungskomponenten in den MLOps-Workflow. Der Schwerpunkt lag darauf, die spezifischen Anforderungen der ABB-Use-Cases zu berücksichtigen und sicherzustellen, dass die entwickelten Erklärungsfunktionen praxisnah in die Trainings- und Evaluierungsprozesse eingebettet werden können. Durch diese enge Zusammenarbeit wurde gewährleistet, dass die entwickelten Methoden nicht nur forschungsorientiert, sondern auch industriell anwendbar sind.

AP 6.3 Integration der Feedback Mechanismen in das Ende-zu-Ende System

Mitarbeitende Partner: ABB, SUH

Das in AP 5.4 entwickelte interaktive feedback-basierte Lernsystem ist in den Prozessablauf und technische Infrastruktur des Ende-zu-Ende System ML zu integrieren. Dabei ist die relevante Frage, wie der inkrementelle Lernprozess sowie die Zwischenergebnisse in der Plattform protokolliert und bereit gehalten werden können. Beim Einsatz und Feedback im Betrieb stellt sich außerdem die Frage, wie sowohl Modellausgaben als auch ein Feedbackschnittstelle so bereitgestellt werden kann, dass beides flexibel in die Arbeitsumgebung der Endnutzer integriert werden kann.

ABB unterstützt SUH in der Integration um so die spätere Einbidung in die Zielsysteme in AP 7 sicher zustellen. SUH erweitert der Infrastrukturvorlage, um den besonderen Anforderungen bei der Erstellung

und Bereitstellung der Feedback Mechanismen und deren Integration in die Arbeitsumgebung der Endbenutzer.

Im Rahmen von Arbeitspaket 6.3 arbeiteten SUH und ABB gemeinsam an der Integration von Feedback-Mechanismen in die bestehende MLOps-Architektur.

[redacted]

SUH aktualisierte auf Basis der ersten Iteration die Softwarearchitektur und erweiterte das bestehende Infrastruktur-Template. Dadurch konnten die spezifischen Anforderungen für die Erstellung und Bereitstellung der Feedback-Mechanismen sowie deren Integration in die Arbeitsumgebung der Endnutzer berücksichtigt werden. Die Zusammenarbeit von SUH und ABB gewährleistete, dass die Feedback-Mechanismen sowohl technisch robust als auch praxisnah einsetzbar sind.

AP 6.4 Daten- und Artefaktmanagement im Ende-zu-Ende System ML

Mitarbeitende Partner: ABB, Eraneos, SUH

Das Erstellen von Modellen mit Maschinellen Lernen erfordert das Management von Code und Daten. Im Fall vom interaktiven Lernen mit Erklärungen und Feedback haben die Daten und Artefakte einen stark unterschiedlichen technischen Charakter und relativ komplexe Schnittstellen zum Bereitstellen von interaktiven Erklärungen oder zur Eingaben von Feedback. In diesem Arbeitspaket werden für diese unterschiedliche Arten von Daten Artefakten unterschiedliche Mechanismen zur Bereitstellung sowie zur Konfiguration und Anpassung in das Ende-zu-Ende System ML integriert. Ein weiterer wesentlicher Aspekt dieses Arbeitspakets ist das Management der Daten, die zur Entwicklung der ML-Modelle verwendet werden sowie der Eingabedaten während des produktiven Betriebs. Im Unterschied zu konventionellen maschinellen Lernprozessen muss das Ende-zu-Ende System ML hier auch die Eingabe der Fachexperten berücksichtigen, protokollieren und zugänglich machen und außerdem zwischen künstlich generierten Daten und Daten aus dem echten Anlagenbetrieb unterscheiden können.

ABB unterstützt SUH in der Integration um so die spätere Einbidung in die Zielsysteme in AP 7 sicher zustellen. SUH integriert eine automatisierte Konsistenzprüfung, z.B. bei der Datenauswahl für die Modellerstellung, sowie für die automatische Prüfung von Modellen. SUH integriert Funktionalitäten zur Unterstützung der notwendigen Funktionalitäten für MLOps, wie z. B. einen Model Store, Data Store und einen Feature Store und ihre notwendigen Funktionen, wie z.B. Versionierung, Annotation und Schnittstellen zur Erklärungskomponente (Arbeitspaket 6.2) und zu den Feedback Mechanismen (Arbeitspaket 6.3) im Ende-zu-Ende System ML. ERANEOS bringt die technischen Anforderungen aus dem Anwendungsfall Energiemarktvorhersagen in das Arbeitspaket ein.

Im Rahmen von Arbeitspaket 6.4 entwickelten SUH und ABB gemeinsam einen ersten Demonstrator für ein MLOps-System und leisteten wichtige Beiträge zur Konzeption und Implementierung einer praxistauglichen Architektur.

SUH implementierte im Demonstrator automatisierte Konsistenzprüfungen, beispielsweise zur Datenauswahl bei der Modellerstellung und zur automatischen Modellvalidierung. Darüber hinaus wurden zentrale MLOps-Funktionalitäten integriert, darunter ein Model Store, Data Store, Feature Store,

Versionierung und weitere notwendige Komponenten, die den operativen Betrieb und die Nachvollziehbarkeit von Machine-Learning-Prozessen unterstützen.

ABB brachte Erfahrungen aus dem Aufbau einer MLOps-Architektur für die schwedischen Use Cases ein. Insbesondere bei Boliden konnte aufgrund der vorhandenen IT-Infrastruktur eine vollständige Integration der entwickelten Komponenten vorbereitet und initialisiert werden. Die dort etablierte Architektur ermöglicht den direkten Betrieb von ML-Anwendungen in der Zielumgebung sowie deren Anbindung an bestehende Systeme über standardisierte Schnittstellen. Diese praktischen Erfahrungen flossen iterativ in die Abstimmung mit SUH ein und trugen zur Entwicklung von Architekturmuster bei, die die Grundlage für die operative Integration im finalen Innovationszyklus (Arbeitspaket 7) bilden.

AP 7 Empirische Validierung und Pilotierung der Anwendungsszenarien

Das Ziel des Arbeitspakets ist es, die generischen Ergebnisse aus den Arbeitspaketen 4, 5 und 6 in die Anwendungsfälle und in die Werkzeuge und Softwareumgebungen der KI-Anbieter und KI-Betreiber zu übertragen. In jeder Iteration des Projekts wird dieses Arbeitspaket Artefakte liefern, die mit Nutzern getestet werden können. Mit jeder Bewertung wird die Wiedergabetreue der Artefakte erhöht. Am Ende der letzten Iteration werden die Prototypen in einer realistischen Laborumgebung oder im Pilotbetrieb getestet. Die Durchführung des Tests ist ebenfalls Teil dieses Arbeitspakets. Zum Test der Lösungen kommen einerseits Nutzertest mit einfachen statistischen Bewertung (z.B. Zustimmungsrate zum Ergebnis und Erklärung) andererseits Evaluationen für den gesamten Arbeitsprozess, beispielsweise Think Aloud Testing oder der System Usability Score Fragebogen, zum Einsatz.

Zu den spezfischen Arbeitsinhalten der Unterarbeitspakete gehört die Modifikation der Visualisierungen aus den Arbeitspaketen 4 und 5. Diese Arbeiten sind notwendig, um die generischen Konzepte auf die vorherrschenden Arbeitsumgebungen der Anwendungspartner anzupassen und um den domänenspezifischen Konventionen oder der von den Benutzern bevorzugten Visualisierung zu entsprechen. Es sind ggf. auch Schnittstellen zur Plattform für das interaktive Maschinelle Lernen mit Feedback zu schaffen, um die Ergebnisse aus AP 4, AP 5 und AP 6 in den Zielumgebungen nutzen zu können. Schließlich fallen auch die Vorbereitung, Durchführung und Nachbereitung der Nutzertest in dieses Arbeitspaket.

In der Struktur haben wir uns entschieden, die Unterarbeitspakete nach Anwendungsfall zu gruppieren, da auf diese Weise die Arbeitsteilung im Konsortium klarer wird.

Im Arbeitspaket 7 wurden die in den vorherigen Arbeitspaketen entwickelten Konzepte, Prototypen und technischen Komponenten in realen Anwendungsszenarien erprobt und gemeinsam mit den Endanwendern evaluiert. Ziel war es, die Praxistauglichkeit der erklärbaren und feedbackbasierten KI-Ansätze in verschiedenen industriellen Kontexten zu überprüfen und deren Nutzen für den operativen Einsatz zu bewerten. Die Prototypen für die Energiemarktvorhersagen und den nachhaltigen Kraftwerksbetrieb wurden erfolgreich in die Arbeitsprozesse der LEAG integriert und in mehreren Evaluierungszyklen getestet. Dabei zeigte sich, dass erklärungsbasierte Visualisierungen und nutzerzentrierte Feedbackmechanismen wesentlich zur Vertrauensbildung und Akzeptanz von KI-Systemen beitragen. Die Ergebnisse wurden auf der ML4CPS-Konferenz 2024 vorgestellt und fließen in Konzepte zur zukünftigen industriellen Umsetzung ein. Der geplante Use Case zur Betriebsoptimierung in Walzwerken konnte aufgrund fehlender Daten nicht weitergeführt werden. Dagegen wurde im Anwendungsfall Wartungsplanung in Walzwerken ein domänenspezifischer Chatbot-Prototyp mit integrierten Erklärungs- und Feedbackfunktionen entwickelt und mit Servicetechnikern getestet. Erste Ergebnisse zeigen ein hohes Potenzial für den Einsatz generativer KI-Systeme im industriellen

Wissensmanagement. Insgesamt bestätigen die durchgeführten Tests und Pilotierungen die Praxistauglichkeit der entwickelten Ansätze und liefern wertvolle Erkenntnisse zur Gestaltung transparenter, nutzerzentrierter KI-Systeme in industriellen Umgebungen.

AP 7.1 Evaluierung im Anwendungsszenario Energiemarktvorhersagen

Mitarbeitende Partner: ABB, ERANEOS, LEAG, SUH, TUDA

In diesem Unterarbeitspaket werden die Prototypen für die Energiemarktvorhersagen den entsprechenden Mitarbeitern (Tradern, Disponenten) der LEAG verfügbar gemacht. Dabei wird in den frühen Projektphase mit einfachen Prototypen (Mock-Ups, Jupyter Notebooks) gearbeitet und über den Projektverlauf eine immer bessere Integration in den tatsächlichen Arbeitsprozess angestrebt. Dabei wird sowohl der Trainingsprozess als auch der produktive Einsatz der Vorhersagen mitsamt der Erklärungen getestet.

ABB unterstützt die Arbeiten mit der Expertise in Nutzer- und Usabilitytests. ERANEOS arbeitet an der Umsetzung der Prototypen der jeweiligen Iteration. ERANEOS nimmt zudem bei Bedarf Anpassungen an der Datenaufbereitung und den Modellen vor. TUDA übernimmt notwendige Anpassungen der Erklärungsmethoden und des interaktiven Lernalgorithmus. SUH integriert notwendige Anpassungen in die Infrastruktur des Ende-zu-Ende System ML. LEAG-Experten testen das Anwendungsszenario Energiemarktvorhersagen im Geschäftsbetrieb.

Im Rahmen der Evaluierung des Anwendungsszenarios Energiemarktvorhersagen wurde der entwickelte Prototyp mit integrierten Feedback- und Erklärungskomponenten in der Google Cloud Plattform produktiv gestellt und liefert seitdem täglich Prognosen. Verschiedene Abteilungen innerhalb der LEAG erhielten Zugriff auf das System, um dessen Nutzung im realen Arbeitsumfeld zu testen. Die Tests wurden in mehreren Evaluierungszyklen durchgeführt und durch kleinere Anpassungsiterationen an Datengrundlage und Darstellung begleitet.

SUH passte die Erklärungsvisualisierungen an die spezifischen Anforderungen des Use Cases an und überführte die notwendigen Änderungen in den Demonstrator des Ende-zu-Ende-ML-Systems. Die TU Darmstadt stellte die in den Arbeitspaketen 4 und 5 entwickelten Prototypen in Interviews den Key Usern vor, um zusätzliche Anforderungen zu erfassen – insbesondere hinsichtlich der Effizienz der Anwendung bei begrenzter Nutzungszeit. ABB unterstützte diese Arbeiten mit seiner Expertise in der Konzeption und Durchführung von Nutzer- und Usability Tests.

AP 7.2 Evaluierung im Anwendungsszenario zur Unterstützung des nachhaltigen Kraftwerkbetriebs

Mitarbeitende Partner: ABB, ERANEOS, LEAG, SUH, TUDA

In diesem Unterarbeitspaket werden die Prototypen für die Analyse und Unterstützung des Kraftwerkbetriebs den entsprechenden Mitarbeitern (Anlagenfahrer, Wartungspersonal) der LEAG verfügbar gemacht. Dabei wird in den frühen Projektphase mit einfachen Prototypen (Mock-Ups, Jupyter Notebooks) gearbeitet und über den Projektverlauf eine immer bessere Integration in den tatsächlichen Arbeitsprozess angestrebt Dabei wird sowohl der Trainingsprozess als auch der produktive Einsatz der Vorhersagen mitsamt der Erklärungen getestet.

ABB unterstützt die Arbeiten mit der Expertise in Nutzer- und Usabilitytests. ERANEOS nimmt bei Bedarf Anpassungen an der Datenaufbereitung und den Modellen vor. TUDA übernimmt notwendige Anpassungen der Erklärungsmethoden und des interaktiven Lernalgorithmus. SUH integriert notwendige Anpassungen in die Infrastruktur des Ende-zu-Ende System ML.

Im Rahmen der Evaluierung des Anwendungsszenarios zur Unterstützung des nachhaltigen Kraftwerksbetriebs wurden die im Jahr 2023 entwickelten Erklärungskomponenten und Prototypen getestet und ausgewertet. Ergebnisse der Untersuchungen wurden im März 2024 auf der ML4CPS-Konferenz in Berlin vorgestellt. Das von den Partnern entwickelte System zur frühzeitigen Erkennung von Anomalien im Kraftwerksbetrieb wurde als sekundäres System am Kraftwerkssimulator der LEAG unter realistischen Bedingungen mit Leitstandsfahrern getestet. Dabei wurde direktes Feedback gesammelt, das in Form mehrerer Verbesserungszyklen in die Weiterentwicklung des Dashboards einfloss. Die gewonnenen Erkenntnisse bilden die Grundlage für ein Konzept zur künftigen Implementierung des Systems im realen Kraftwerksbetrieb am Standort Boxberg Block Q.

ABB führte im September 2023 einen Evaluationsworkshop mit Anlagenfahrern des LEAG-Kraftwerks durch, um die Nützlichkeit der Lösung zur Erkennung und Erklärung abnormaler Anlagensituationen zu bewerten. Fünf Operatoren mit unterschiedlichem Erfahrungsniveau – von Auszubildenden bis zu langjährigen Mitarbeitenden – nahmen an der Studie teil. Die Interviews zeigten, dass die Lösung als hilfreiches Werkzeug zur Eingrenzung von Anlagenstörungen wahrgenommen wird. Zugleich wurde betont, dass KI-Systeme den Operator lediglich unterstützen, aber keine autonomen Eingriffe vornehmen sollten. Besonders geschätzt wurden die klaren Erklärungen und Visualisierungen, die das Vertrauen in die Technologie stärken.

SUH begleitete die technische Integration der Erklärungskomponenten in die Infrastruktur des Ende-zu-Ende-ML-Systems und stellte die Kompatibilität mit den bestehenden Schnittstellen sicher. Die TU Darmstadt übernahm eine beratende Rolle und brachte ihre Expertise zu erklärungsbasierten Methoden in Workshops und Abstimmungstreffen ein.

AP 7.3 Evaluierung im Anwendungsszenario zur Betriebsoptimierung in Walzwerken

Mitarbeitende Partner: ABB, SUH, TUDA

[redacted]

AP 7.4 Evaluierung im Anwendungsszenario zur Wartungsplanung in Walzwerken

Mitarbeitende Partner: ABB, SUH, TUDA

[redacted]

[redacted]

AP 8 Dissemination und Transfer

AP 8.1 Zusammenarbeit mit dem Internationalem Konsortium und anderen Dritten

Mitarbeitende Partner: ABB, Eraneos, LEAG, SUH, TUDA

Das Vorhaben ist im Rahmen des Eureka Cluster AI Calls in ein internationales Konsortium mit Partnern in Schweden und den Niederlanden eingebunden. Dadurch ergeben sich vielfältige positive Effekte und Synergien, vor allem die breitere Absicherung der Ergebnisse durch die Vielzahl an Anwendungsfällen im internationlen Konsortium. Hervorzuheben ist die starke Kompetenz der akademischen Partner im Schweden im Kontext Mensch-KI Interaktion sowie die Ergänzenden Schwerpunkt mit einen Fokus auf Anwendungen auf Basis von Prozess- und Signaldaten in Deutschland und Schweden sowie Bilddaten in den Niederlanden. Hier bestehen große Chancen von den Schwerpunkten der anderen Konsortien zu lernen. Dies setzt natürlich neben den pflichtmäßigen internationalen Review-Treffen auch noch weitere Treffen auf Arbeitsebene und entsprechende Dokumentation der Ergebnisse voraus.

Die internationale Zusammenarbeit bildete einen zentralen Bestandteil von Arbeitspaket 8 und trug maßgeblich zur erfolgreichen Durchführung und Weiterentwicklung der Projektaktivitäten bei. Der Austausch zwischen den nationalen und internationalen Konsortialpartnern ermöglichte eine enge Abstimmung technischer, wissenschaftlicher und organisatorischer Aspekte und unterstützte die Integration zusätzlicher industrieller Anwendungsfälle.

Ein besonderer Schwerpunkt lag auf der Kooperation mit dem schwedischen Teilkonsortium. Die zusätzlichen Use Cases von *Södra* und *Boliden* wurden in enger Abstimmung mit den schwedischen Partnern sowie den internationalen Beteiligten weiterentwickelt. Durch den kontinuierlichen Austausch zwischen ABB Deutschland, ABB Schweden und den jeweiligen Industriepartnern konnten die Anforderungen und technischen Lösungsansätze erfolgreich harmonisiert und in die gemeinsame MLOps-Architektur integriert werden.

Zur Förderung des internationalen Wissensaustauschs wurden im Berichtszeitraum mehrere Konsortialtreffen und Workshops durchgeführt:

- *Kick-off-Meeting in Ladenburg, Deutschland:* Projektinitialisierung und Festlegung der gemeinsamen Arbeitsstrukturen.

- *Konsortiumstreffen in Västerås, Schweden:* Planung und Durchführung des internationalen Konsortiumstreffens organisiert von ABB Deutschland und ABB Schweden.
- *ITEA Review Meeting in Eindhoven, Niederlande (Prodrive):* Erfolgreiche Koordination und Durchführung des jährlichen *ITEA Review Meetings* mit allen Partnern des internationalen Verbundprojekts in Eindhoven, Niederlande.
- *Konsortiumstreffen in Lübbenau, Deutschland:* Abschluss des Projekts mit dem finalen Face-to-Face Konsortiumstreffen bei der LEAG in Lübbenau, das zugleich als Vorbereitung des finalen ITEA Review Meeting diente.

Ein weiterer wesentlicher Aspekt der internationalen Zusammenarbeit war die *gemeinsame Durchführung der Workshop-Reihen 4 und 5* im Rahmen von Arbeitspaket 4 und 5, organisiert durch die *Technische Universität Darmstadt*. Diese Workshops dienten dem bilateralen Austausch von Erfahrungen zur Implementierung und Anwendung der im Projekt entwickelten Erklärungs- und Feedbackmechanismen. Dabei wurden sowohl algorithmische Konzepte als auch Fragen der Benutzerinteraktion und Oberflächengestaltung behandelt. Die aktive Teilnahme mehrerer Partner – unter anderem ABB, SUH, LEAG und Vertreter des schwedischen Konsortiums – trug dazu bei, methodische Entwicklungen mit praktischen Anforderungen aus der industriellen Anwendung zu verbinden.

Insgesamt zeigte sich die enge internationale Zusammenarbeit als entscheidend für den erfolgreichen Wissenstransfer innerhalb des Konsortiums. Durch den regelmäßigen Austausch von Erkenntnissen, die gemeinsame Durchführung von Workshops und die iterative Abstimmung technischer Komponenten konnte ein hoher Grad an Kohärenz und Interoperabilität zwischen den nationalen Teilprojekten erreicht werden. Damit wurde eine zentrale Voraussetzung für die spätere operative Integration der entwickelten Lösungen geschaffen und die internationale Sichtbarkeit des Projekts nachhaltig gestärkt.

AP 8.2 Publikationen

Mitarbeitende Partner: ABB, Eraneos, LEAG, SUH, TUDA

Publikationen und Konferenzen:

- Dix M., Koltermann J., Mieck S., Pastler B. & Kloepper B. (2024). XAI for anomaly analysis by power plant operators – a case and user study. In Machine Learning for Cyber-Physical Systems (ML4CPS), Berlin, Germany, März 2024.

- Dix M., Koltermann J., Mieck S., Petersen H., Taege S. & Koltermann J. J. (2024). Von geheimnisvoll zu verstehbar – der Schlüssel zu einem vertrauensvollen KI-Einsatz in der Leitwarte. Vortrag auf dem 56. Kraftwerkstechnischen Kolloquium, Dresden, Deutschland, Oktober 2024

- Dix M., Koltermann J. J., Mieck S., Petersen H., Taege S. & Anjanappa G. (2024). Using Siamese neural networks for the open-set recognition of anomalies detected in industrial time-series data. In Proceedings of the IEEE 29th International Conference on Emerging Technologies and Factory Automation (ETFA 2024), Padova, Italy, September 2024

- Manca G., Kunze F. C. & Fay A. (2024). A novel process plant alarm dataset and methodology for alarm data generation. In Proceedings of the IEEE 29th International Conference on Emerging Technologies and Factory Automation (ETFA 2024), Padova, Italy, September 2024

- Zhang Y., Brorsson E., Methnani L., Zohrevandi E., Bhattacharya N., Darnell A. & Tammia R. (2024). Human-centered Explainable-AI: an empirical study in process industry. In Proceedings of the 15th AHFE International Conference on Artificial Intelligence and Social Computing, San Francisco, USA, Juli 2024

- Brorsson E., Zhang Y., Zohrevandi E., Bhattacharya N., Theodorou A., Darnell A., Tammia R. & Van Driel W. (2024). Evaluating explainability of time-series models: a user-centered approach in industrial settings. In Proceedings of the 15th AHFE International Conference on Artificial Intelligence and Social Computing, San Francisco, USA, Juli 2024

- Zohrevandi E., Bhattacharya N., Tammia R., Zhang Y., Darnell A., Bang M. & Kollegen (2025). Design and visualisation of time-series-based explanation mechanisms for industrial AI applications. In Proceedings of the 8th International Congress and Workshop on Industrial AI and eMaintenance (IAI 2025), Luleå, Schweden

- Kraus M., Divo F., Dhami D. S. & Kersting K. (2025). xLSTM-Mixer: multivariate time-series forecasting by mixing via scalar memories. In Proceedings of the Neural Information Processing Systems (NeurIPS 2025), Vancouver, Kanada

- Kraus M., Steinmann D., Wüst A., Kokozinski A. & Kersting K. (2025). Right on Time: revising time-series models by constraining their explanations. European Conference on Machine Learning and Knowledge Discovery in Databases (ECML PKDD 2025), Porto, Portugal

- Kraus M., Steinmann D., Wüst A., Kokozinski A. & Kersting K. (2024). Right on Time: revising time-series models by constraining their explanations. Poster im NeurIPS 2024-Workshop „Interpretable AI: Past, Present and Future“ (NeurIPSW 2024), Vancouver, Kanada

- Steinmann D., Divo F., Kraus M., Wüst A., Struppek L., Friedrich F. & Kersting K. (2024). Navigating shortcuts, spurious correlations and confounders: from origins via detection to mitigation. arXiv preprint arXiv:2412.05152.

- Skryagin A., Ochs D., Deibert P., Kohaut S., Dhami D. S. & Kersting K. (2024). Answer Set Networks: Casting Answer-set Programming into Deep Learning. arXiv preprint arXiv:2412.14814.

- Manca G. & Fay A. (2024). Convolutional kernel-based classification of industrial alarm floods. Data-Centric Engineering, 5, e30.

- Manca G., Kunze F. C. & Fay A. (2024). Addressing Uncertainty in Online Alarm Flood Classification Using Conformal Prediction. IEEE Access, 12, 165626–165652.

- Manca G., Kunze F. C., Brorsson E. & Fay A. (2023). Dynamic Causal Analysis with Operator-Centric Visualization for Managing Industrial Alarm Floods. In Proceedings of the IEEE 2nd Industrial Electronics Society Annual On-Line Conference (ONCON 2023), Online, Februar 2024.

- Baskan D. E., Meyer D., Mieck S., Faubel L., Klöpper B., Strêm N., Wagner J. A. & Koltermann J. J. (2023). A Scenario-Based Model Comparison for Short-Term Day-Ahead Electricity Prices in Times of Economic and Political Tension. Algorithms, 16(4), 177.

- Faubel L., Woudsma T., Klöpper B., Eichelberger H., Buelow F., Schmid K., Ghezeljehmeidan A. G., Methnani L., Theodorou A. & Bång M. (2024). MLOps for Cyber-Physical Production Systems: Challenges and Solutions. IEEE Software, 42(1), 65–73.

- Faubel L. & Schmid K. (2024). A systematic analysis of MLOps features and platforms. WiPiEC Journal – Works in Progress in Embedded Computing Journal, 10(2), 97–103.

- Faubel L., Woudsma T., Methnani L., Ghezeljehmeidan A. G., Buelow F., Schmid K., van Driel W., Klöpper B., Theodorou A., Nosratinia M. & Bång M. (2024). A MLOps architecture for XAI in industrial applications. In Proceedings of the IEEE 29th International Conference on Emerging Technologies and Factory Automation (ETFA 2024), Padova, Italy.

- Faubel L. & Schmid K. (2024). MLOps: a multiple case study in Industry 4.0. in Proceedings of the IEEE 29th International Conference on Emerging Technologies and Factory Automation (ETFA 2024), Padova, Italy.

- Faubel L. & Schmid K. (2024). An MLOps platform comparison. Hildesheimer Informatik Berichte, 1/2024 (SSE 1/24/E)

AP 8.3 Multimediale Öffentlichkeitsarbeit und Mitarbeit in Gremien
Mitarbeitende Partner: ABB, ERANEOS, LEAG, SUH, TUDA

Die Öffentlichkeitsarbeit durch wissenschaftliche Publikationen wird ergänzt durch Broschüren, Presseveröffentlichungen und Social Media Aktivitäten. Diese Aktivitäten unterteilen sich in drei Phasen. In der ersten Phase (Monate 1-15) ist das Ziel Aufmerksamkeit für das EXPLAIN Projekt und seine Ziele, angestrebten Ergebnisse und den Forschungsansatz zu generieren. In der zweiten Phase (Monat 16-31) sollen die ersten Ergebnisse verbreitet und Rückmeldungen aus den jeweiligen Fach- und Anwendungsdomänen gesammelt werden. In der dritten Phase liegt der Schwerpunkt auf die Verbreitung der abschließenden Ergebnisse.

Die Disseminationsaktivitäten im Rahmen von Arbeitspaket 8 dienten der Sichtbarmachung der Projektergebnisse, der Kommunikation mit Fachöffentlichkeit und Industrie sowie der Förderung des Austauschs zwischen Forschung, Wirtschaft und Gesellschaft. Die Aktivitäten wurden in mehreren Phasen durchgeführt und umfassten Publikationen, Fachvorträge, Konferenzteilnahmen, Online-Kommunikation sowie öffentliche Veranstaltungen.

Phase 1 – Online-Kommunikation und internationale Sichtbarkeit

Zu Beginn des Berichtszeitraums lag der Schwerpunkt auf dem kontinuierlichen Ausbau der Projektpräsenz und der internationalen Sichtbarkeit. Die Projektwebsite (https://explain-project.eu/) wurde regelmäßig gepflegt und mit aktuellen Inhalten, Projektfortschritten und Veranstaltungshinweisen aktualisiert. Dadurch konnte eine transparente Kommunikation der Projektziele und -ergebnisse gegenüber Stakeholdern, Forschungspartnern und der interessierten Öffentlichkeit gewährleistet werden.

Im September 2024 nahm das Projektteam an den ITEA PO Days in Antwerpen teil, wo das EXPLAIN-Projekt vor einem internationalen Fachpublikum vorgestellt wurde. Diese Veranstaltung bot die Gelegenheit, Projektergebnisse mit anderen ITEA-Initiativen zu vernetzen und Synergien mit thematisch verwandten Projekten zu identifizieren.

Im November 2024 beteiligte sich ein Projektmitglied an einer Podiumsdiskussion beim AI Grid Summit 2024, bei der Fragen zu industrieller KI-Anwendung, Transparenz und Vertrauen in datengetriebene Systeme diskutiert wurden. Zudem wurde das EXPLAIN-Projekt im Rahmen einer Sitzung des Forschungsbeirats Industrie 4.0 (acatech) präsentiert, wodurch die Ergebnisse in den nationalen Diskurs zu industrieller Digitalisierung und KI-Standardisierung eingebracht werden konnten.

Phase 2 – Wissenschaftliche Dissemination und Fachaustausch

Die wissenschaftliche Verwertung der Projektergebnisse wurde durch mehrere Fachbeiträge und Konferenzpräsentationen unterstützt.

- *„Right on Time: Revising Time Series Models by Constraining their Explanations“* – präsentiert von Mitarbeitenden der Technischen Universität Darmstadt (TUDA) auf dem NeurIPS-Workshop „Interpretable AI: Past, Present and Future“ (Vancouver, Dezember 2024) sowie als Konferenzpaper auf der ECML PKDD (Porto, September 2025).
  Bezug: Erklärungsgeleitete Modellrevision für robuste Zeitreihenmodelle; Anknüpfung an Interactive-Learning-Workflows.
- *„xLSTM-Mixer: Multivariate Time Series Forecasting by Mixing via Scalar Memories“* – vorgestellt als Preprint auf NeurIPS (Dezember 2024) und wird vorgestellt als Konferenzpaper auf NeurIPS (San Diego, Dezember 2025).
  Bezug: Kernbeitrag effiziente Forecasting-/Predictive-Methode des Projekts für hochdimensionale industrielle Zeitreihen.

- *“XAI for anomaly analysis by power plant operators – a case and user study”*, präsentiert von Marcel Dix auf der ML4CPS in Berlin (21.–22. März 2024).

- *“Using Siamese Neural Networks for the Open Set Recognition of Anomalies Detected in Industrial Time Series Data”*, vorgestellt auf der IEEE ETFA 2024 in Padua (10.–13. September 2024).

- *„Von geheimnisvoll zu verstehbar – Der Schlüssel zu einem vertrauensvollen KI-Einsatz in der Leitwarte“*, präsentiert auf dem 56. Kraftwerkstechnischen Kolloquium in Dresden (8.–9. Oktober 2024).

Diese Beiträge trugen dazu bei, die wissenschaftlichen Ergebnisse zu erklärbarer KI (XAI), Anomalieerkennung und MLOps in industriellen Anwendungen in der Fachcommunity zu etablieren und zugleich den Transfer in praxisnahe Kontexte zu fördern.

Darüber hinaus beteiligten sich Projektpartner an mehreren *Workshop-Reihen*, organisiert im Rahmen der Arbeitspakete 4 und 5, bei denen methodische und gestalterische Aspekte von Erklärungs- und Feedbackmechanismen diskutiert wurden. Dadurch wurde der wissenschaftliche Austausch innerhalb des Konsortiums und mit externen Forschenden weiter gestärkt.

Phase 3 – Öffentlichkeitsarbeit und Auszeichnungen

Im abschließenden Projektjahr lag der Fokus auf der öffentlichen Darstellung der Projektergebnisse und der Anerkennung der geleisteten Arbeit im internationalen Kontext.

Mehrere Pressemitteilungen sowie Beiträge auf LinkedIn informierten über die Projekterfolge und Auszeichnungen. Besonders hervorzuheben ist die ITEA Award Verleihung im Rahmen der ITEA PO Days 2025 in Estoril (Portugal), bei der das EXPLAIN-Projekt für seine Innovationsleistung ausgezeichnet wurde. Darüber hinaus erhielt das Projekt eine Nominierung für den FREDDIE Award, bei dem insbesondere die Zusammenarbeit von LEAG und ABB im Bereich erklärbarer und vertrauenswürdiger KI gewürdigt wurde.

Durch diese Aktivitäten konnte die Bekanntheit des Projekts innerhalb der europäischen Forschungslandschaft und der Industrie erheblich gesteigert werden. Die Disseminationsmaßnahmen

haben wesentlich dazu beigetragen, das Bewusstsein für die Bedeutung von Transparenz, Nachvollziehbarkeit und Zuverlässigkeit in industriellen KI-Systemen zu schärfen und die Akzeptanz der entwickelten Lösungen über die Projektgrenzen hinaus zu fördern.

AP 9 Projektmanagement

AP 9.1 Organisation der Zusammenarbeit

Mitarbeitende Partner: ABB, ERANEOS, LEAG, SUH, TUDA

Im Rahmen der Konsortialleitung wird die Termin- und Ressourcenüberwachung vorgenommen. Zum Projektstart neben einem Kick-Off Treffen des internationalen Konsortiums auch eine Veranstaltung auf nationaler Ebene geplant, bei dem sich die Projektmitarbeiter der Partner persönlich kennenlernen. Diese Veranstaltung wird auch genutzt, um die Modalitäten der Zusammenarbeit zwischen den Partnern festzulegen und den Umgang mit Projektdokumenten, insbesondere Produktionsdaten der Anwendungspartner zu besprechen. Weiterhin soll der Projektträger die Gelegenheit erhalten, organisatorische-adminstrative Hinweise zur gewünschten Zusammenarbeit zwischen Projektpartnern und Projektträger zu geben.

Auf Arbeitsebene werden regelmäßige Arbeits- und Koordinationstreffen als Remote-Meetings angesetzt, an denen jeweils die für ein Thema oder Anwendungsfall relevanten Projektmitarbeiter teilnehmen. Um die Zusammenarbeit zwischen den Projektarbeiten zu stärken wird angestrebt, anhängig an die Konsortialtreffen weitere Arbeitstreffen durchzuführen, in denen bestimmte für den aktuellen Projektverlauf relevante Themen intensiver bearbeitet werden. Die Agenda solcher Treffen wird von den betroffenen Projektbearbeitern im Vorfeld der Konsortialtreffen vorbereitet und mit den übrigen Partnern geteilt.

Die Konsortialleitung organisiert Entwurf und Unterzeichnung eines Konsortialvertrags im Verbundprojekt. Dieser wird sich an den üblichen Standardverträgen orientieren. Er regelt u. a. die Geheimhaltung, den Umgang mit ins Projekt eingebrachtem und im Projekt entstehenden Daten und geistigen Eigentums. Die Konsortialleitung ist der erste Ansprechpartner bei Interessenskonflikten zwischen Projektpartnern und nimmt in dieser Funktion eine neutrale Position ein. Die Konsortialleitung wird von ABB wahrgenommen.

Im Rahmen der Konsortialleitung übernahm ABB die Koordination der Zusammenarbeit zwischen den Partnern sowie die übergeordnete Termin- und Ressourcenüberwachung. Das „*Project Collaboration Agreement (PCA)“* im Verbundprojekt wurde erfolgreich durch ABB koordiniert, abgestimmt und von allen Partnern unterzeichnet.

Zur Sicherstellung einer reibungslosen Abstimmung fanden über den gesamten Projektverlauf hinweg zweiwöchentliche Online-Konsortialmeetings statt, die den laufenden Informationsaustausch und die Koordination der Arbeitspakete gewährleisteten. Ergänzend dazu wurden drei Präsenztreffen des Gesamtkonsortiums organisiert: das Kick-off-Meeting 2022 am ABB-Forschungszentrum in Ladenburg, das Konsortiums Treffen 2024 bei den schwedischen Partnern in Västerås sowie das Abschlusstreffen 2025 bei der LEAG in Lübbenau. Diese Treffen dienten dem persönlichen Austausch, der technischen Abstimmung und der gemeinsamen Planung der nächsten Projektphasen.

AP 9.2 „Bereitstellung projektspezifischer IT-Services und Projekt-Website“

Mitarbeitende Partner: ABB

ABB richtet eine projektinterne zentrale Informations- und Dokumentenmanagementplattform in der Teamsplattform zur Unterstützung kooperativer Arbeit ein und betreibt diese. Die Plattform wird bspw. für gemeinsam erarbeitete Dokumente genutzt wie auch als Archiv für Belegexemplare von

Projektveröffentlichungen. Für gemeinsame Softwareentwicklungen wird ein Software Repository verwendet. ABB richtet eine Internetpräsenz für die Außendarstellung des Projekts ein und pflegt diese. Die Internetpräsenz stellt Informationen zu Projektzielen, der Rollen der Partner im Projekt und im Projekt entstandene Publikationen oder anderen Medien (wie freigegebene Videoclips von Demonstrationen) zur Verfügung. Des Weiteren werden Hinweise auf öffentliche Projektveranstaltungen gegeben.

ABB übernahm die Einrichtung und den Betrieb der projektinternen Informations- und Dokumentenmanagementplattform. Dafür wurden ein gemeinsamer Microsoft Teams-Kanal und ein SharePoint-System eingerichtet, die als zentrale Arbeitsumgebung für das Konsortium dienten. Über diese Plattformen erfolgten der Austausch und die Ablage aller Projektdokumente, einschließlich gemeinsamer Entwürfe und Protokolle.

Darüber hinaus richtete ABB die öffentliche Projektwebsite ein, die während der gesamten Laufzeit gepflegt und regelmäßig aktualisiert wurde. Die Website stellt die Projektziele, die Rollen der Partner sowie ausgewählte Ergebnisse und Publikationen vor und dient als zentrale Plattform für die Außendarstellung des Projekts. Ergänzend wurden über die Seite auch Informationen zu öffentlichen Veranstaltungen und Demonstratoren veröffentlicht.

AP 9.3 Berichtswesen im Verbundprojekt

Mitarbeitende Partner: ABB, Eraneos, LEAG, SUH, TUDA

Das projektspezifische Berichtswesen der einzelnen Verbundprojektpartner orientiert sich an den Vorgaben des Förderers bzw. des Projektträgers. Als Projektkoordinator stimmt das ABB Forschungszentrum das inhaltliche Berichtwesen mit den Projektpartnern ab und ist verantwortlich für die Schnittstelle zum Projektträger. Konsortialberichte werden von der Konsortialleitung auf der Grundlage der Berichte der Projektpartner erstellt.

Zudem wurden die Durchführung der Projekt-Reviews und die fristgerechte Erstellung der Deliverables im Verbundprojekt erfolgreich umgesetzt und von allen Partnern unterstützt. Die Koordination und Kommunikation zwischen den nationalen und internationalen Partnern verlief effizient und trug wesentlich zum erfolgreichen Projektfortschritt bei.

Als Projektkoordinator übernahm ABB die Leitung des projektweiten Berichtswesens und die Schnittstelle zum Projektträger. Dazu gehörten die Abstimmung der Inhalte mit den Partnern, die Koordination der Beiträge sowie die fristgerechte Erstellung und Einreichung der Konsortialberichte. ABB Deutschland leitete zudem erfolgreich die Durchführung der Projekt-Reviews und die termingerechte Erstellung der Deliverables, die von allen Partnern unterstützt wurden. Die enge und effiziente Kommunikation zwischen den nationalen und internationalen Partnern trug wesentlich zum erfolgreichen Projektfortschritt bei.

Ein wichtiger Meilenstein war die Überarbeitung und thematische Bündelung der ITEA-Deliverables in enger Abstimmung mit ITEA und den Konsortialpartnern. Diese Maßnahme, initiiert und koordiniert durch ABB, reduzierte den Dokumentationsaufwand und orientierte die Berichte stärker am tatsächlichen Fortschritt der Anwendungsszenarien. Sie wurde von allen Partnern begrüßt und half, im Jahr 2024 ein neues Momentum im Projektverlauf zu schaffen – ohne inhaltliche Kürzungen oder Ressourceneinschränkungen. Wesentliche Berichtspflichten wurden im Berichtszeitraum erfolgreich erfüllt, darunter das ITEA Review Meeting in Eindhoven (Mai 2024), der Project Progress Report (September 2024) sowie das finale Projektreview im April 2025, das mit einer starken Beteiligung der deutschen Partner erfolgreich abgeschlossen wurde.

Ein herausragendes Ergebnis war die Verleihung des *ITEA Award of Excellence 2025* in der Kategorie Innovation an das EXPLAIN-Projekt. Diese Auszeichnung würdigt die hervorragende internationale Zusammenarbeit und den praxisnahen Forschungsbeitrag zur Entwicklung erklärbarer KI für industrielle Anwendungen.