

# 1. Darstellung der im Rahmen des Vorhabens durchgeführten Arbeiten

## 1.1 Gesamtziel des Verbundprojekts

Das Ziel des Verbundprojekts KI-AIM war die Entwicklung einer Plattform zur Kombination und Bewertung verschiedener Verfahren zur Anonymisierung und synthetischen Generierung medizinischer Daten. Die Plattform soll es ermöglichen, unterschiedliche Methoden zur Depersonalisierung sensibler Gesundheitsdaten flexibel zu kombinieren und deren Auswirkungen auf Datenschutz, Datenqualität und Nutzbarkeit systematisch zu bewerten.

Ein zentrales Anliegen des Projekts bestand darin, Domänenexpertinnen und -experten – etwa aus der Medizin, Medizininformatik oder dem Datenschutz – in die Lage zu versetzen, geeignete Verfahren zur Anonymisierung oder Synthetisierung medizinischer Daten auszuwählen und deren Eigenschaften nachvollziehbar zu bewerten. Dabei sollten sowohl Datenschutzanforderungen als auch Anforderungen an die Nutzbarkeit der Daten für Forschung und Entwicklung berücksichtigt werden.

Die DATATREE AG war im Verbundprojekt insbesondere für die Bearbeitung der datenschutzrechtlichen Fragestellungen sowie für Beiträge zur Verwertung der Projektergebnisse verantwortlich.

## 1.2 Beiträge im Arbeitspaket AP1 – Projektmanagement

Im Rahmen des Arbeitspakets AP1 beteiligte sich die DATATREE AG an den Aktivitäten zur Projektkoordination und -organisation. Dazu gehörten insbesondere die Teilnahme an regelmäßigen Projekttreffen, Telefonkonferenzen und Abstimmungsgesprächen mit den weiteren Projektpartnern.

Darüber hinaus wirkte die DATATREE AG an der Erstellung projektbezogener Dokumentationen sowie an der Erstellung der regelmäßigen Projektberichte mit. Hierzu zählten insbesondere Beiträge zu Sachstandsberichten sowie die Dokumentation der im Teilvorhaben erzielten Ergebnisse.

Die Aktivitäten im Projektmanagement dienten der Sicherstellung eines kontinuierlichen Informationsaustauschs zwischen den Projektpartnern sowie der Koordination der Arbeiten innerhalb des Verbundprojekts.

## 1.3 Beiträge im Arbeitspaket AP2 – Datenschutzrechtliche Aspekte

Der inhaltliche Schwerpunkt der Arbeiten der DATATREE AG lag im Arbeitspaket AP2 „Datenschutzrechtliche Aspekte“. Ziel dieses Arbeitspakets war es, datenschutzrechtliche Fragestellungen im Zusammenhang mit der Anonymisierung und synthetischen Generierung medizinischer Daten zu untersuchen und geeignete Bewertungsverfahren zu entwickeln.

Zu Beginn des Vorhabens wurden umfangreiche Recherche- und Analysearbeiten durchgeführt. Diese umfassten insbesondere eine Analyse der rechtlichen Anforderungen an anonymisierte Daten im Kontext der DSGVO sowie eine Auswertung einschlägiger Veröffentlichungen von Datenschutzaufsichtsbehörden und wissenschaftlicher Literatur.

Parallel hierzu wurden technische Ansätze zur Anonymisierung und synthetischen Datengenerierung analysiert. Dabei wurden verschiedene Verfahren hinsichtlich ihrer Eigenschaften, ihrer potenziellen Datenschutzrisiken sowie ihrer praktischen Nutzbarkeit untersucht.

Auf Grundlage dieser Analysen entwickelte die DATATREE AG ein Bewertungsverfahren zur Einschätzung des Grades der Anonymisierung von Datensätzen. Ziel dieses Ansatzes war es, mathematische und statistische Methoden zur Quantifizierung des Personenbezugs von Daten mit dem rechtlichen Verständnis des Personenbezugs gemäß DSGVO in Beziehung zu setzen.

Ein besonderes Augenmerk lag darauf, komplexe technische Verfahren und deren Parametrisierung für unterschiedliche Zielgruppen verständlich aufzubereiten. Die entwickelten Bewertungsansätze sollten insbesondere Datenschutzbeauftragten, Ethikkommissionen sowie Aufsichtsbehörden eine nachvollziehbare Einordnung der eingesetzten Verfahren ermöglichen.

Im weiteren Projektverlauf wurde der entwickelte Bewertungsansatz zur Analyse der im Projekt entwickelten Anonymisierungs- und Synthetisierungsverfahren eingesetzt. Hierzu wurden unterschiedliche generative Modelle und Anonymisierungsverfahren untersucht und hinsichtlich ihrer Eigenschaften bewertet.

Darüber hinaus wurden Evaluationsansätze entwickelt, mit denen potenzielle Datenschutzrisiken synthetischer Datensätze untersucht werden können. Dabei wurden insbesondere mögliche Angriffsszenarien betrachtet, bei denen versucht wird, Informationen über Trainingsdaten aus generativen Modellen abzuleiten.

Die Arbeiten erfolgten in enger Abstimmung mit den technischen Arbeitspaketen des Verbundprojekts, insbesondere mit Blick auf die Bewertung der im Projekt entwickelten Anonymisierungs- und Synthetisierungsverfahren.

#### **1.4 Beiträge im Arbeitspaket AP7 – Dissemination und Verwertung**

Neben den inhaltlichen Arbeiten beteiligte sich die DATATREE AG auch an Aktivitäten zur Dissemination und Verwertung der Projektergebnisse.

Ein Schwerpunkt lag hierbei auf der Entwicklung eines Verwertungsplans für die im Projekt erzielten Ergebnisse. Ziel dieser Arbeiten war es, mögliche Nutzungsszenarien der Projektergebnisse sowohl aus wissenschaftlicher als auch aus wirtschaftlicher Perspektive zu analysieren.

Darüber hinaus wurde untersucht, in welchen Anwendungsfeldern die entwickelten Bewertungsansätze zukünftig eingesetzt werden können. Insbesondere im Kontext der zunehmenden Nutzung von KI-Verfahren im Gesundheitswesen ergibt sich ein wachsender Bedarf an Verfahren zur datenschutzkonformen Nutzung sensibler Gesundheitsdaten.

## **2. Wichtigste Positionen des zahlenmäßigen Nachweises**

Die im Projekt angefallenen Kosten entfielen überwiegend auf Personalaufwendungen für wissenschaftliche und technische Mitarbeitende der DATATREE AG. Diese waren insbesondere mit folgenden Tätigkeiten befasst:

- Durchführung rechtlicher und technischer Analysearbeiten
- Entwicklung eines Bewertungsansatzes für den Grad der Anonymisierung
- Analyse und Bewertung von Verfahren zur synthetischen Datengenerierung
- Durchführung von Evaluationsarbeiten zur Untersuchung möglicher Datenschutzrisiken
- Mitwirkung an Projektkoordination, Dokumentation und Berichtswesen
- Beiträge zu Disseminations- und Verwertungsaktivitäten

Die eingesetzten Mittel wurden entsprechend der ursprünglichen Projektplanung verwendet und dienten der Durchführung der vorgesehenen Projektarbeiten.

### **3. Notwendigkeit und Angemessenheit der geleisteten Projektarbeiten**

Die im Projekt durchgeführten Arbeiten waren erforderlich, um die im Verbundprojekt entwickelten Verfahren zur Anonymisierung und synthetischen Datengenerierung sowohl technisch als auch datenschutzrechtlich bewerten zu können.

Gerade im medizinischen Kontext bestehen besonders hohe Anforderungen an den Schutz personenbezogener Daten. Gleichzeitig besteht ein wachsender Bedarf, medizinische Daten für Forschungs- und Entwicklungszwecke nutzbar zu machen.

Die Entwicklung geeigneter Bewertungsverfahren zur Einordnung des Anonymisierungsgrades sowie zur Analyse möglicher Datenschutzrisiken stellt daher eine zentrale Voraussetzung für den praktischen Einsatz entsprechender Technologien dar.

Die im Projekt entwickelten Ansätze leisten einen Beitrag zur systematischen Bewertung solcher Verfahren und unterstützen damit die Entwicklung datenschutzkonformer Lösungen zur Nutzung medizinischer Daten.

### **4. Voraussichtlicher Nutzen und Verwertbarkeit der Ergebnisse**

Die im Projekt entwickelten Ergebnisse besitzen sowohl wissenschaftliche als auch wirtschaftliche Verwertungspotenziale.

Insbesondere die Generierung synthetischer Daten stellt einen vielversprechenden Ansatz dar, um sensible Gesundheitsdaten für Forschungs- und Entwicklungszwecke nutzbar zu machen, ohne reale Datensätze unmittelbar weitergeben zu müssen.

Für die DATATREE AG ergeben sich aus den Projektergebnissen insbesondere Möglichkeiten zur Weiterentwicklung des eigenen Dienstleistungsportfolios im Bereich Datenschutz und Informationssicherheit im Gesundheitswesen.

Die entwickelten Bewertungsansätze können beispielsweise bei der datenschutzrechtlichen Bewertung von KI-Anwendungen oder bei der Analyse von Anonymisierungsverfahren eingesetzt werden.

Darüber hinaus besteht ein wachsendes Marktpotenzial für Lösungen zur datenschutzkonformen Nutzung medizinischer Daten im Kontext von KI-Anwendungen. Vor dem Hintergrund der zunehmenden Digitalisierung des Gesundheitswesens sowie der steigenden Nutzung datengetriebener Verfahren wird erwartet, dass entsprechende Beratungs- und Analyseleistungen weiter an Bedeutung gewinnen.

Neben der Nutzung einzelner Projektergebnisse wird auch eine Verwertung des Gesamtvorhabens angestrebt. Hierzu gehört insbesondere die im Projekt entwickelte Plattform zur Kombination und Bewertung von Anonymisierungs- und Synthetisierungsverfahren, die perspektivisch als Open-Source-Software bereitgestellt werden soll.

### **5. Fortschritt auf dem Gebiet des Vorhabens bei anderen Stellen**

Während der Projektlaufzeit konnte beobachtet werden, dass synthetische Datengenerierung zunehmend als Ansatz zur datenschutzkonformen Nutzung sensibler Daten diskutiert wird. Insbesondere im Bereich der künstlichen Intelligenz und der medizinischen Forschung werden generative Modelle verstärkt eingesetzt, um synthetische Datensätze zu erzeugen.

Gleichzeitig wird auch die Diskussion über mögliche Datenschutzrisiken solcher Verfahren intensiver geführt.

Die im Projekt durchgeführten Arbeiten greifen diese Entwicklungen auf und leisten einen Beitrag zur systematischen Bewertung entsprechender Verfahren.

## **6. Veröffentlichungen**

Im Rahmen des Projekts wurden mehrere Ergebnisdokumente erstellt, in denen zentrale Aspekte der Projektarbeiten dokumentiert sind.

Ein wesentliches Ergebnis stellt eine wissenschaftliche Ausarbeitung zur Generierung und Bewertung synthetischer Gesundheitsdaten dar. In diesem Dokument werden unterschiedliche Verfahren zur synthetischen Datengenerierung beschrieben und hinsichtlich ihrer Eigenschaften sowie ihrer praktischen Nutzbarkeit analysiert.

Darüber hinaus wurde eine Untersuchung zur Bewertung von Datenschutzrisiken synthetischer Gesundheitsdaten durchgeführt. Diese Arbeit analysiert insbesondere mögliche Angriffsszenarien und bewertet, unter welchen Bedingungen synthetische Datensätze potenzielle Risiken für die Privatsphäre enthalten können.

Ergänzend wurde ein Bericht zum rechtlichen Rahmen für Datenverarbeitung und KI-Anwendungen im Gesundheitswesen erstellt. Dieser Bericht untersucht die regulatorischen Anforderungen an die Nutzung medizinischer Daten und ordnet die im Projekt betrachteten Verfahren in den rechtlichen Kontext ein.

Die drei Ergebnisdokumente bilden gemeinsam eine Grundlage für die Bewertung von Verfahren zur Anonymisierung und synthetischen Datengenerierung im medizinischen Kontext und leisten damit einen Beitrag zur weiteren wissenschaftlichen und praktischen Diskussion in diesem Themenfeld.