

PriSyn Abschlussbericht - Kurzfassung

über die Ergebnisse des Kooperationspartners CISPÄ

Verbundprojekt: Repräsentative, synthetische Gesundheitsdaten mit starken
Privatsphärengarantien - PriSyn -



Saarbrücken, den 13.03.2026

CISPÄ Helmholtz Zentrum für Informationssicherheit	Förderkennzeichen: 16KISA133
Vorhabensbezeichnung: Verbundprojekt: Repräsentative, synthetische Gesundheitsdaten mit starken Privatsphärengarantien - PriSyn - Teilvorhaben: Vertrauenswürdige Generative Modelle	
Laufzeit des Vorhabens: 01.08.2023 – 14.12.2025	
Berichtszeitraum: 01.08.2023 – 14.12.2025	

Aufgabe/Zielsetzung

Zu den Kernbeiträgen gehört die umfassende Bewertung der Privatheits- und Anonymitätsrisiken generativer Modelle, die auf hochsensiblen biomedizinischen Daten trainiert werden. Dazu werden bestehende Privatheitsangriffe an den biomedizinischen Bereich angepasst, neuartige Bewertungsmethoden – etwa Dataset-Rekonstruktionsangriffe – entwickelt und strenge Privatheitsmetriken etabliert. Ziel ist die Integration dieser Werkzeuge in eine modulare Softwareplattform, mit der Modelleigner Privatheitsrisiken vor der Bereitstellung evaluieren können.

Ein weiterer zentraler Beitrag ist die Entwicklung differential-privater generativer Modelle für Gesundheitsdaten. Dazu werden bestehende Ansätze in einem gemeinsamen Rahmenwerk vereint, Domänenwissen zur Bewältigung kleiner Stichproben genutzt, die Repräsentation von Minderheitengruppen verbessert und verteilte Lernmethoden auf ihre effiziente und praktikable Einsetzbarkeit geprüft.

Wissenschaftlicher und technischer Stand

Differential-private generative Modelle sind ein vielversprechender Ansatz zur Synthese virtueller biomedizinischer Kohorten und zum datenschutzgerechten Austausch von Gesundheitsdaten. Sie bleiben jedoch anfälliger für fortgeschrittene Privatheitsangriffe. Bestehende Angriffe wie Membership Inference und Attribute Inference wurden vor allem für Bild- und Sprachdaten entwickelt; hochdimensionale biomedizinische Daten sind bislang kaum untersucht. Zudem liefern bestehende Bewertungsmethoden für generative Modelle oft nur geringe Aussagekraft, sodass die Risiken des Austauschs synthetischer biomedizinischer Daten leicht unterschätzt werden.

Zugleich zeigen Fortschritte im Deep Generative Modeling und in der Differential Privacy das Potenzial, synthetische biomedizinische Daten mit formalen Privatheitsgarantien zu erzeugen. Die Anwendung auf Gesundheitsdaten bleibt jedoch schwierig, da biomedizinische Datensätze meist hochdimensional, heterogen und klein sind. Dies reduziert die Wiedergabetreue und den Nutzen der Modelle. Hinzu kommt, dass bestehende Verfahren Verzerrungen aus den Trainingsdaten übernehmen und Minderheitengruppen oft unzureichend repräsentieren. Außerdem sind aktuelle Trainingsverfahren überwiegend auf zentralisierte Datensätze ausgelegt und passen daher nur begrenzt zur verteilten Struktur klinischer Daten.

Das Vorhaben adressiert diese Grenzen durch vereinheitlichte Trainingsrahmenwerke für private generative Modelle, die Einbeziehung von Domänenwissen und aufgabenspezifischen Zielen zur Verbesserung von Daten-Utility und Fairness sowie durch verteilte Lernschemata für reale Gesundheitsumgebungen.

Verlauf des Projekts

Unter der Leitung von CISPA wurden die AP 2 und AP 3 sowie die zahlreichen Kollaborationspunkte erfolgreich innerhalb des Projektzyklus von PriSyn implementiert.

Anpassung und Verbesserung bestehender Angriffe: Das Team passte Membership Inference und Attribute Inference Angriffe erfolgreich an biomedizinische generative Modelle an. Dazu wurden Baseline-Methoden zur Stichprobenrekonstruktion durch Adversarial Training Loss erweitert und Dimensionsreduktionstechniken zur Verarbeitung hochdimensionaler genomischer Daten untersucht.

Erster Dataset-Rekonstruktionsangriff: Es wurde der erste Angriff entwickelt, der darauf abzielt, den gesamten Trainingsdatensatz eines Zielmodells zu rekonstruieren. Dazu wurden Modellstichproben gesammelt, ein Surrogat-Generatives Modell aufgebaut und Stichprobenauswahlstrategien basierend auf Membership Inference eingesetzt. Ergänzend wurden strenge Privatheitsmetriken definiert, die sowohl Stichprobenunterschiede als auch Verteilungsähnlichkeit erfassen.

Generator-Doctor: Alle Bewertungsmethoden wurden in die modulare Softwareplattform Generator-Doctor integriert. Die Plattform ermöglicht das einfache Einbinden neuer Datensätze und Modelle sowie einen standardisierten Vergleich von Inferenzangriffen und Verteidigungsmechanismen.

Eine vereinheitlichte Sicht auf den Designraum für private Generatoren: Durch die Analyse verschiedener Ansätze privater generativer Modelle identifizierten wir Konfigurationen, die besonders für hochdimensionale biomedizinische Daten geeignet sind. Insbesondere verbessert die Integration einer kausalen Struktur in den Generator für Einzelzell-RNA-seq-Daten die Bewahrung biologischer Abhängigkeiten zwischen Genen und Zellpopulationen und führt zu besserer Datenqualität unter Differential Privacy.

Nutzung von Domänenwissen und aufgabenspezialisierten Generatoren: Biologisches Vorwissen wurde integriert, indem Generatoren an Genregulatorische Netzwerke (GRNs) konditioniert wurden. Zur automatischen Gewinnung solcher Netzwerke entwickelten wir eine Methode, die Large Language Models (LLMs) nutzt, um regulatorische Beziehungen aus biomedizinischer Literatur und Wissensdatenbanken zu extrahieren.

Neuartige Lernziele zur Einbeziehung von Minderheiten: Durch den Vergleich realer und synthetischer Einzelzell-RNA-seq-Daten zeigten wir, dass **kausale generative Modelle** einen **größeren Anteil seltener Zelltypen** bewahren und damit Minderheitenpopulationen besser repräsentieren.

Zusammenarbeit mit anderen Forschungseinrichtungen

Unter der Führung von CISPA beruhte der Erfolg stark auf der Zusammenarbeit mit den PriSyn-Partnern:

QuantPi brachte Expertise in Bewertungsmetriken und erklärbarer KI ein. Gemeinsam wurden Privatheitsangriffe evaluiert und Metriken zur Bewertung von Datenrekonstruktionsangriffen entwickelt. Zudem bewertete QuantPi die generierten Einzelzellendaten hinsichtlich Mode Collapse, halluzinierter Zellpopulationen und Diversität.

DZNE stellte hoch-sensible biomedizinische Datensätze und klinische Metadaten bereit und führte die biologische Validierung der generierten Daten durch. Dadurch konnten die entwickelten Methoden unter realistischen biomedizinischen Bedingungen getestet werden.

HPE unterstützte das verteilte Lernsetup durch Bereitstellung der Swarm-Learning-Infrastruktur. **AMD** unterstützte die Implementierung der Codebasis und Modellquantisierung für die Ausführung auf FPGA-Systemen.