THEMEN ABSCHLUSSARBEITEN

Prof. Dr. Martin Steinebach Lehrstuhl Technische Universität Darmstadt

Martin Steinebach,
martin.steinebach@sit.fraunhofer.de,Telefon: 06151/869 349

This Testset does not Exist

- Im Sinne von "this person does not exist" (https://thisxdoesnotexist.com/)
- Netz trainieren mit CC0 Inhalten
- Netz so gestalten, dass daraus reproduzierbare Inhalte extrahiert werden. Also per Key/seed Zufallsgenerator steuern, der dann n Beispiele vom Punkt "seed" ausgehend produziert
- Die so generierten Inhalte werden als Testset verwendet:

  Robuste Hashverfahren, Wasserzeichen, …

- Untersuchen
  - Sind die Ergebnisse gleich/ähnlich zu den CC0 Inhalten?
  - Ähnlich zu bekannten Testsets?
    - https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.225.4857&rep=rep1&type=pdf
- Ziel wäre ein System, mit dem beliebige Mengen von Testmaterial unabhängig von Urheberrechten erstellt werden kann

Martin Steinebach,
martin.steinebach@sit.fraunhofer.de,Telefon: 06151/869 349

SYNTHESE VON REALISTISCHEN TRANSAKTIONSDATEN

- Finanzdaten sind hoch sensibel und die Finanzindustrie ist sehr stark reguliert bzw. Verstöße werden mit hohen Geldstrafen belegt. Die Weitergabe personenbezogener Daten bedarf der Zustimmung des entsprechenden Individuums, was dazu führt, dass es keine öffentlich zugänglichen Datensätze gibt.
- Für Forschungszwecke stehen deshalb i.d.R. nur künstlich erzeugte Datensätze zur Verfügung (z.B. PAYSIM)
- In der Thesis soll
  - Ein Konzept entwickelt werden, wie realistische Transaktionsdaten generiert werden können
  - Ein entsprechender Transaktionsdaten erzeugt
  - Und der Nutzen des generierten Datensatzes evaluiert werden.

- Ein guter Ausgangspunkt für dieses Thema ist z.B. https://dl.acm.org/doi/10.1145/3383455.3422554

Martin Steinebach, martin.steinebach@sit.fraunhofer.de,Telefon: 06151/869 349

GENERIERUNG VON PRIVATHEIT ERHALTENDEN TRAININGSDATEN

- Daten sind das neue Öl. Und je feingranularer die Daten, desto mehr lässt sich aus ihnen heraus lesen.
- Die Weitergabe von Daten bzgl. Einzelpersonen scheitert jedoch i.d.R. an Aspekten des Datenschutzes.
- Die Generierung von synthetischen Daten stellt somit eine attraktive Lösung für dieses Problem dar
- In der Thesis sollen
  - aktuelle State of the Art Ansätze im Bereich privacy-preserving synthetic data generation aufgearbeitet werden.
  - Diese dann gegeneinander evaluiert und
  - die resultierten synthetischen Datensätze anhand von Kennzahlen evaluiert werden, sowohl auf ihren Nutzen wie auch, ob sie wirklich privacy-preserving sind.
  - Ein guter Ausgangspunkt für dieses Thema ist z.B. https://arxiv.org/abs/2112.09238

Martin Steinebach, martin.steinebach@sit.fraunhofer.de, Telefon: 06151/869 349

Bei Interesse an obigen Themen ist eine formlose Kontaktaufnahme ausreichend. Auch bei Interesse an Abschlussarbeiten zu Forensik, Mediensicherheit, OSINT, ML-Sicherheit, NLP u.ä. kann sich gerne jeder melden, wir schauen dann, was wir anbieten können. Idealer Weise werden dabei gleich Erfahrungen in der jeweiligen Thematik, allgemeine Erfahrungen in der Programmierung und ein aktueller Notenspiegel mitgeliefert.

Kontakt:

Martin Steinebach

Rheinstrasse 75

64295 Darmstadt

Email: martin.steinebach@sit.fraunhofer.de

Telefon: 06151/869 349