

## Bachelor/Master Project

### Synthetic data: a critical analysis of approaches and tools

**Term:** Winter Term 2022/2023

**Language:** English

#### Motivation

The concept of synthetic data, i.e., data algorithmically generated within a synthesis process approximating original data, is gaining on popularity. It is even prognosed that the application of synthetic data will replace application of “real” data, especially in the various artificial intelligence-based undertakings. However, although the concept of the synthetic data sounds very promising, it is definitely not a ‘silver bullet’ that solves all problems in the data science/machine learning field.

#### Description

The main aim of this bachelor/master project is to critically analyze existing approaches and tools to generate synthetic data. To this aim the students should:

- (1) Make themselves familiar with basic ideas of synthetic data and its generation, including main assumptions of this domain.
- (2) Conduct market analysis of the existing approaches and tools in this area.
- (3) Define an evaluation framework for the needs of comparison of existing tools/approaches, as well as, depending on the agreed scope of the project, also a scenario together with a relevant data set to be used in the practical evaluation part.
- (4) Draw conclusions and formulate the recommendations for the further development.

#### Expected Outcomes

A report pointing to the conducted state of the art analysis (cf. point (1) and (2)), requirements and rationale for the evaluation framework and scenario, as well as detailed description of the evaluation results. In addition, a final presentation of the project results is expected.

**Institut für Informatik und  
Wirtschaftsinformatik (ICB)**

**Lehrstuhl für  
Wirtschaftsinformatik und  
Unternehmensmodellierung**

**Dr. Monika Kaczmarek-Heß**

Tel.: 0201 / 183 - 4330

[monika.kaczmarek-hess@uni-due.de](mailto:monika.kaczmarek-hess@uni-due.de)

**Pierre Maier**

Tel.: 0201 / 183 - 4563

[pierre.maier@uni-due.de](mailto:pierre.maier@uni-due.de)

R09 R04 H41

Universitätsstraße 9

45127 Essen

[www.umo.wiwi.uni-due.de](http://www.umo.wiwi.uni-due.de)

## Introductory Literature

- El Emam K, Mosquera L, Hoptroff R (2020) Practical Synthetic Data Generation. O'Reilly: Sebastopol, CA
- Figueira A, Vaz B (2022) Survey on Synthetic Data Generation, Evaluation Methods and GANs. *Mathematics* 10(15): 2733
- Assefa SA, Dervovic D, Mahfouz M, Tillman RE, Reddy P, Veloso M (2020) Generating Synthetic Data in Finance: Opportunities, Challenges and Pitfalls. *ICAIF '20: Proceedings of the First ACM International Conference of AI in Finance*

## Application Procedure:

Please apply via email to the supervisors. Please attach a short letter of motivation (approximately one A4 page) and a recent transcript of record ('Leistungsnachweis'). You can apply individually or in a group of **2-6 participants** (in this case each person should still send a separate e-mail, however point to the other members of the group).

**Application deadline:** 14 October 2022, 23:59 h