Invitation to the Oral Examination – Department CS

For the occasion of his examination for a Doctoral Degree,

Alexander Andreas Ziller

will present his dissertation entitled

Privacy-preserving Artificial Intelligence in Medicine

on **16th December 2024** at **10:00 am CET**

Attendance to the presentation is open to the public. The presentation will be in English.

The candidate, all members of the Examination Committee, and authorized examiners of the TUM School of CIT are invited to the presentation and subsequent oral examination.

The presentation and subsequent examination will take place online via zoom:
https://tum-conf.zoom-x.de/j/62562683910?pwd=WJEsWsfT1fNt46bHNJQOKRzOM71aXA.1

Meeting ID: 625 6268 3910
Passcode: 573991

and at **TranslaTUM** Einsteinstraße 25 (Bau 522)
81675 München, **Raumnummer 00.22.01** Johannes B. Ortner Forum, Großes Auditorium.

Examination committee:

Chair: **apl. Prof. Georg Groh**

First Examiner: **Prof. Daniel Rückert**

Second Examiner: **Prof. Jens Kleesiek, Universität Duisburg/ Essen**

Third Examiner: **Prof. Sotirios Tsaftaris, The University of Edinburgh**

Garching, **2nd** of **December 2024**

Abstract:

Artificial Intelligence (AI) has become paramount in many areas over the last decade. It has proven to be a valuable addition to medical workflows, where it can assist doctors in precise evaluations of patient conditions.
However, highly performant AI models crucially depend on large and diverse datasets. While these datasets are continuously generated in hospitals and medical institutions, they are inaccessible due to the risks of privacy infringements. The term Privacy-enhancing technologies (PETs) summarises the field of technical

approaches and algorithms, which aim to reunite AI training and the protection of its training data from unintended leakage.

In this dissertation, we investigate the use of PETs in the context of medical AI approaches. Specifically, we demonstrate a holistic workflow comprised of various PETs that provides protection from attackers while yielding highly performant AI models, even outperforming expert radiologists. The most important PET in this thesis, Differential Privacy (DP), provides mathematical bounds on the risks of information leakage. We analyse the computational overhead DP implementations impose on the training of AI models and provide an alternative which is competitive in runtime and generically compatible with most AI network architectures. Furthermore, we investigate the impact of using DP for medical AI training on the fairness and non-discrimination of subgroups.
Here, in contrast to prior work, we find that not the representation of subgroups in the training data is driving fairness impacts, but rather the difficulty of predicting the respective subgroup. In particular, we see that groups with a lower prediction performance in non-private AI models suffer further performance losses with increasing privacy guarantees. This may impact the way of assembling datasets for the training of privacy-preserving fair AI models.
Lastly, we analyse how an appropriate level of protection can be determined and find that, for many scenarios, typical privacy budgets are overly pessimistic. We show that by adapting the privacy budget to a concrete threat model, the negative impact of DP on the performance of AI models can be largely mitigated. With these contributions, we hope to advance the widespread breakthrough of technical and mathematical approaches to protecting patient privacy when training medical AI models.