M5

Who pays for privacy?

Measuring and reporting subgroup-level utility in private synthetic data for diabetes and cardiovascular patients.

Master's · 30 credits | Statistics · Public health | Suits: KI (Public Health, Epidemiology) · KTH (ML)

The question

Under a fixed privacy budget, does synthetic data serve some cardiometabolic patients worse than others, such as younger people with type 2 diabetes, women with heart disease, or older patients with several conditions? How should that be measured and shown to the people relying on the data?

Why it matters

Published work shows that privacy noise can land hardest on small subgroups, and most medical studies do not report subgroup results at all. Regulators and the EU AI Act increasingly ask for subgroup evidence. Measuring it honestly is a design requirement for release documentation.

What you would do

01 Define subgroups among diabetes and cardiovascular patients in public seed data (age, sex, region, comorbidity) and a set of subgroup-level fidelity and inference measures.

02 Measure subgroup gaps across synthetic releases at several privacy budgets.

03 Relate the gaps to subgroup size and to the downstream tasks affected.

04 Propose a subgroup section for a release certificate.

What you would produce

- A subgroup-utility benchmark on public data
- Analysis of which groups lose most, and when
- A reporting template for subgroup results

Good fit if you have

Statistics, R or Python, interest in equity and fairness in health research.

Data and access

Public seed data (CMS DE-SynPUF) and synthetic releases from day one. Work on real patient records only inside a partner's approved secure environment, with ethics approval. We hold no Swedish patient data.

Timing and place

Spring or autumn 2027, 20 weeks / 30 credits. Stockholm (Solna) or hybrid.

Supervision

Gaurav Raj Chattarki, Founder, Synthetix Health, together with your academic supervisor.

Questions or a proposal of your own? gaurav@synthetixhealth.com · synthetixhealth.com/careers#thesis

All projects run under a confidentiality and IP agreement agreed before work begins. We review drafts only for confidential information, never for results. Negative results are publishable.