M1

Red-teaming synthetic clinical data

Membership-inference, reconstruction and linkage attacks against certified synthetic releases.

Master's · 30 credits | Privacy · Security | Suits: KTH (Cybersecurity, ML) · SU (DSV)

The question

A synthetic release produced under ε-differential privacy comes with a mathematical ceiling on what any attacker can learn about a real patient. How close do real, well-resourced attacks get to that ceiling, and where does the gap come from?

Why it matters

Formal guarantees are only persuasive to a hospital or an ethics board when someone independent has tried hard to break them. Published attacks on synthetic data show that similarity scores can hide real leakage. An open, reproducible attack harness is the evidence the field is missing.

What you would do

01 Survey and implement the main attack families from the literature: shadow-model membership inference, distance-based attacks, attribute inference and record linkage.

02 Run them against public-seed releases (CMS DE-SynPUF) at several privacy budgets, with a clear threat model for each attack.

03 Compare measured attack advantage against the theoretical bound printed on the release certificate.

04 Where a partner's secure environment and ethics approval allow it, repeat the strongest attacks on releases made from real records, without data leaving the environment.

What you would produce

- An open-source attack harness with documented threat models
- A results report: where attacks come close to the bound, and where they do not
- A publishable paper; negative results count

Good fit if you have

Python, ML fundamentals, some probability. Interest in security or cryptography. Prior work with privacy attacks is a plus, not a requirement.

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.