

# FINAL PROJECT PROPOSAL

B.SC IN DIGITAL MEDICAL TECHNOLOGIES  
CLASS OF 2025

## ANONYMIZATION OF MENTAL HEALTH RECORDS VIA REPHRASING AND QA VALIDATION

GROUP MEMBERS:  
VICTORIA CHUYKINA, ID - 321544512  
URIEL ATZMON, ID - 209307172

# PROBLEM STATEMENT

Mental health notes often contain extended personal information (ePHI) like workplaces, relatives and life events.

Standard PHI masking (names, addresses) is insufficient

We aim to remove personal info but retain clinical meaning

## Problem Definition:

- Input: Clinical text with clinical facts + personal facts
- Output: Same text, personal info anonymized (via rephrasing or replacement)

## Challenges:

- Go beyond classical PHI
- Preserve text readability and clinical usability
- Validate by functional QA, not just entity detection



# DATA AND VALIDATION

## Synthetic Dataset Generation:

- Build 2 info banks:
  - Clinical facts (DSM-5 symptoms, observations)
  - Personal facts (family, work, places, events)
- Generate clinical notes combining clinical + personal facts via LLM
- Save:
  - Text
  - Clinical Questions (CQ)
  - Personal Questions (PQ)

Goal: 5000 notes, each with CQ + PQ sets.

## Validation Method:

- Ask saved CQs and PQs on the anonymized text

## Ideal Behavior:

- Clinical Questions → Answerable
- Personal Questions → Not Answerable

## Metrics:

- Clinical Preservation Rate
- Personal Info Leakage Rate
- Answerability scores (e.g., BERTScore)

# METHODOLOGY

## Models Compared:

- 2–3 LLMs (e.g., GPT-4, Claude, Llama 3)

## Prompting Strategies:

- Standard prompt vs Chain-of-Thought vs Instructional prompt

## Anonymization Style:

- Rephrase personal facts, not just mask
- Optionally replace with synthetic equivalents

# NOVELTY AND SUMMARY

- **Expanding PHI anonymization scope (custom ePII categories)**
- **Rephrasing instead of entity masking**
- **QA-based validation framework**
- **Full synthetic, realistic dataset construction**
- **Comparative study of models + prompting**

**Anonymization of mental health records**

**Data: Synthetic notes with Clinical + Personal QA banks**

**Models: Comparing 2–3 pretrained LLMs**

**Evaluation: Preservation & leakage rates via QA**



