# Verslag BAP 25/03/2024

## Besproken

* Keuze Architectuur
  + Data analyse uitgevoerd
  + Mistral is Optimaal voor eerste poging
  + Later misschien Llama en Mixtral erbij doen
* Domein benchmark
  + FinEval
    - Multiple Choice voor snelle evaluatie
    - Chain of Thought voor correcter antwoord
    - Zero-shot voor intrinsieke vaardigheden te testen
    - Few-shot voor betere antwoorden
  + Chemical Tasks:
    - Benchmark verdelen in verschillende taken
    - Per taak ander formaat en andere evaluatie
    - 1 Standaard prompt voor alle taken
  + Onze Benchmark
    - 10 vragen letterlijk in tekst te vinden
    - 10 vragen afleiden uit de tekst
    - 5 vragen waar de tekst te weinig info over bevat om te antwoorden
    - 5 vragen die het model op het foute been zet
    - 5 vragen die helemaal niks met het domein te maken hebben
    - Allemaal multiple choice met chain of thought. Chain of Thought nog niet evalueren, potentieel BLUE score ervan in de toekomst
* Van de grond op methode valt weg, te weinig data en tijd
* Retrieval Augmented Generation
  + GPT-4, Gemini-ultra en Claude v3 proberen.

## TODO tegen …

* Automatische TestBench opzetten
* Walter Daems Contacteren
* Eerste deel van paper uitschrijven
* Documenten formatteren voor RAG en Finetuning
* Finetuning proberen op te zetten voor paasvakantie