# Disposition – “LLMs – The Death of GIS Analysis”

## Frontpage (Deadline: October 9th)

* Title: “An investigation into using LLMs for GIS analysis”

## Abstract (Deadline December 8th)

* **1 page**
* 4-6 words per section
* Purpose (1/4)
* Methods (1/4)
* Results (1/4)
* Implications (1/4)

## Sammendrag (Deadline December 8th)

* Same as abstract, only in Norwegian (required)

## Preface (Deadline December 11th)

* **1 page**
* Context (result of subject TBA4560, etc.)
* Personal motivation: why did I select this task?
* Thanks to mom and dad, Norkart, supervisor, etc.

## Table of figures

## Table of tables

## Acronyms

## 1. Introduction (Deadline: October 9th)

### Motivation

* LLM is a booming field of research, etc.
* Big potential is GIS
  + Examples of usefulness
  + Who needs GIS analysis?

### Goals and Research Questions

* Goal/objective in a single sentence
  + Expand in text below
* Research questions (sub-goals)
  + Main challenges
  + Repeat them in the discussion
  + 3 RQs maybe?
  + Suggestions
    - How can OGC API Features be used in an overlay analysis using ChatGPT-4?
    - How can the main task be divided into more manageable sub-problems?
    - How can we give ChatGPT-4 access to external tools?
    - What is the potential of LLM-based GIS analysis?

### Research Method

* Theoretical
* Experiments
* Return to choices of research methods in the Discussion

### Report Structure

* Overview of what’s coming next

## 2. Theory (Deadline: October 30th)

* Actor map
  + Different language models
    - GPT-4 (OpenAI)
    - BERT (Google)
    - Claude 2 (Anthropic)
    - PaLM 2 (Google)
    - Codey (Google)
  + LLM Providers
    - OpenAI
    - Google Cloud (Vertex AI, etc.)
    - Microsoft Azure
    - Amazon AWS
    - Anthropic
  + User groups
    - GIS professionals
    - City planners
    - Academics and researchers
    - Emergency services
    - Business analysts
    - General public
  + API standards
    - OGC API
    - NGIS Open API
    - STAC API
  + Regulatory bodies
    - European Union (GDPR)
    - Local municipal regulations
  + Data sources
    - Government databases
    - Private databases
    - Crowdsources data
* GIS
  + How is classical GIS work performed?
  + Current work
    - Standardisation central
    - Norway
      * Geodataarbeid
      * Standardisation
      * Geovekst
      * FKB-data
        + SFKB
        + NVDB
      * Orthophoto
      * Laser data
    - International
      * Standardisation
      * OGC
      * STAC
* LLMs
  + What are they?
  + Use-cases
    - General use-cases
    - Not specific to GIS analysis
  + Different models
    - Models for specific purposes (trained to produce code, etc.)
  + Strengths/weaknesses
    - Prompting vs. fine-tuning
    - How does language used affect the results (Norwegian vs. English)
    - Access to up-to-date information
  + Fine-tuning
  + Embeddings
  + Prompt engineering
* Geospatial technology, API standards, etc.
  + State of the art in machine readable standards
  + Geonorge
  + NGIS Open API
  + OGC Records API
  + STAC API
  + Cloud native

## 4. Datasets

## 5. Architecture

## 6. Proofs of concept (Deadline November 6th)

### 6.1. Plan

* 2-3 different, small technical proofs-of-concepts/experiments. Examples:
  + Test what is possible without any modifications (much like Arild’s demonstration)
    - Chat GPT4 with code completion
    - Emphasise limitations in results chapter
  + Make a web application that takes a prompt, gets relevant datasets from data catalogues, and displays them on the client
    - No spatial queries (try to limit the task)
  + Make web application that connects to a PostGIS database, accepts natural language queries, converts them to SQL and retrieves result to client

### 6.2. Setup

* What are the “lab settings”
  + What provider is used?
    - ChatGPT4 API
    - Azure ML studio
* How were the experiments performed?

### 6.3. Results

* Literature that answers you research questions
* Present findings from experiments
* Elaborate

## 7. Evaluation and Discussion (Deadline December 4th)

### 7.1. Evaluation

* Strengths/weaknesses of results from experiments
* Comparison against traditional methods
* Financial aspects
  + Cost-effectiveness compared to traditional methods

### 7.2. Discussion

* Privacy/legal/regulatory concerns
* Integration challenges
  + Want to avoid too much hard-coding – trying to utilize the LLM’s flexibility
* Scalability
  + Handling more complex tasks
  + Combining multiple datasets
* Adaptability to different languages and regions
  + Focus on Norway at first?
  + Parallel tests using Norwegian and English
    - Same queries, different results?
* Data security
  + Can we mindlessly give ChatGPT access to cadastral data?
  + Azure ML Studio better option in terms of data security?
* Ethical concerns
  + Unintentional bias or inaccuracies

## 8. Conclusion and Future Work (Deadline December 6th)

### 8.1. Contributions

### 8.2. Future Work

* How to improve weaknesses found in the results

## Bibliography (Deadline December 8th)

* List of cited sources

## Appendix (Deadline December 8th)

* Supplementary material
* Additional data
* Detailed information