

Karl Oskar Magnus Holm

# LLMS - The Death of GIS Analysis?

An Investigation into Using Large Language Models  
to Make GIS Analysis Simpler, Faster, and More  
Accessible

Specialization Project in Computer Science and Geomatics, June 2024

Supervisor: Hongchao Fan

External supervisors from Norkart: Alexander Salveson Nossun, Arild Nomeland and  
Rune Aasgaard

Department of Geomatics

Faculty of Engineering

Norwegian University of Science and Technology



## Abstract

## Sammendrag

## Preface

Karl Oskar Magnus Holm  
Trondheim, 30th October 2023

# Contents

<b>Abstract</b>	<b>i</b>
<b>Sammendrag</b>	<b>ii</b>
<b>Preface</b>	<b>iii</b>
<b>List of Figures</b>	<b>vii</b>
<b>List of Tables</b>	<b>viii</b>
<b>1. Introduction</b>	<b>1</b>
1.1. Background and Motivation . . . . .	1
1.2. Goals and Research Questions . . . . .	1
1.3. Research Method . . . . .	2
1.3.1. Literature Study . . . . .	2
1.4. Contributions . . . . .	2
1.5. Thesis Structure . . . . .	2
<b>2. Theory</b>	<b>3</b>
2.1. Large Language Models . . . . .	3
2.1.1. Benchmarking for Large Language Models . . . . .	3
HumanEval . . . . .	3
Multitask Language Understanding . . . . .	3
2.1.2. The GPT Family . . . . .	3
2.1.3. Google Products . . . . .	5
The BERT Family . . . . .	5
PaLM 2 . . . . .	5
Codey . . . . .	5
2.1.4. Meta Products . . . . .	5
LLama 2 . . . . .	5
Code LLama . . . . .	5
2.1.5. Honourable Mentions . . . . .	5
Falcon 180B . . . . .	5
Mistral . . . . .	5
Vicuna . . . . .	5
2.2. LLM providers . . . . .	5
2.2.1. OpenAI . . . . .	5

## Contents

2.2.2.	Microsoft Azure . . . . .	5
2.2.3.	Google Cloud . . . . .	5
2.2.4.	Amazon Web Services (AWS) . . . . .	5
2.2.5.	Anthropic . . . . .	5
2.3.	Geospatial Standards . . . . .	5
2.3.1.	International Standardization Work . . . . .	5
	OGC Standards . . . . .	5
	STAC Api Standard . . . . .	6
2.3.2.	Norwegian Standardization Work . . . . .	6
	SOSI . . . . .	7
	Geovekst . . . . .	7
	Norge digitalt . . . . .	7
2.4.	User Groups . . . . .	8
2.4.1.	General Public . . . . .	9
2.4.2.	GIS Professionals . . . . .	9
2.4.3.	City Planners . . . . .	9
2.4.4.	Business Analysts . . . . .	9
2.4.5.	Academics and Researchers . . . . .	9
2.4.6.	Emergency Services . . . . .	9
2.5.	Regulatory Bodies and Privacy Concerns . . . . .	9
2.5.1.	Local Regulations . . . . .	9
2.5.2.	The European Union (EU) . . . . .	9
2.6.	Related Work . . . . .	9
2.6.1.	GIS with LLMs . . . . .	9
2.6.2.	Retrieval Augmented Generation . . . . .	10
	LangChain . . . . .	10
	Microsoft Semantic Kernel . . . . .	10
	AutoGPT . . . . .	10
2.6.3.	Prompt Engineering and Planning Strategies . . . . .	10
<b>3.</b>	<b>Experiments and Results</b>	<b>12</b>
3.1.	Experimental Plan . . . . .	12
3.1.1.	Determining the Potential of LLM-based GIS analysis (RQ 1) . . .	12
3.1.2.	Testing ChatGPT's Ability to Perform Overlay Analysis using OGC API - Features (RQ 2) . . . . .	12
3.1.3.	Testing ChatGPT's Ability to To Use External Tools (RQ 2) . . .	12
3.2.	Experimental Setup . . . . .	12
3.3.	Experimental Results . . . . .	12
<b>4.</b>	<b>Discussion</b>	<b>13</b>
4.1.	Evaluation . . . . .	13
4.2.	Discussion . . . . .	13

## Contents

<b>5. Conclusion and Future Work</b>	<b>14</b>
5.1. Contributions . . . . .	14
5.2. Future Work . . . . .	14
5.2.1. Test regime . . . . .	14
5.2.2. Framework for Planning, Acting, and Reasoning . . . . .	15
5.2.3. Embeddings . . . . .	15
5.2.4. Fine-Tuning . . . . .	15
<b>Acronyms</b>	<b>16</b>
<b>Bibliography</b>	<b>18</b>
<b>Appendices</b>	<b>20</b>
<b>A. Task Description from Norkart</b>	<b>21</b>
<b>B. API Schemas</b>	<b>23</b>
B.1. OGC API - Features . . . . .	23
B.2. STAC API . . . . .	24
<b>C. Code Examples</b>	<b>25</b>

# List of Figures

- 2.1. Actor map for stakeholders, providers, and other groups and organizations that could have some relevance to an autonomous LLM-based GIS-agent. 4



# List of Tables

1.1. Literature study search results. . . . .	2
2.1. Categories and frequency of ChapGPT usage (Skjuve et al., 2023, March, pp. 16–17). . . . .	9

# 1. Introduction

## 1.1. Background and Motivation

The field of [Large Language Models \(LLMs\)](#) is an emerging one. [Fan et al. \(2023, October, p. 2\)](#) found that the proportion of papers about [LLMs](#)<sup>1</sup> to arXiv has skyrocketed since 2020, with a six-times increase in percent points from 2022 to 2023. They write that prompt engineering has been extensively used as a way to improve code generation ([Fan et al., 2023, October, p. 7](#))

## 1.2. Goals and Research Questions

The overarching goal of this specialization project is to investigate how [Large Language Model \(LLM\)](#) can be utilized to make [GIS](#) analysis simpler, faster, and more accessible. As exemplified in the task description provided by Norkart (see [appendix A](#)), such a system should be able to create a meaningful response to a query such as:

"Find all buildings within a 100-meter belt that are above 100 meters above sea level and have docks."

The task then is to investigate how modern language models can be used to perform classical [GIS](#) analyses using standard [GIS](#) technologies like PostGIS and geospatial data catalogues adhering to [OGC](#) or [STAC](#) standards. Based on the task description I have constructed three research questions that I will attempt to answer in this specialization project report:

1. What is the potential of LLM-based GIS analysis?
2. How can OGC API Features be used in an overlay analysis using ChatGPT-4?
3. How can we give ChatGPT-4 access to external tools?

Researching the potential of using [LLMs](#) in GIS could uncover new methodologies for spatial analysis, predictive modeling, and decision-making. The aim with [RQ 1](#) would be to assess the capabilities and limitations of integrating machine learning algorithms with [Geographic information systems \(GISs\)](#), and also touch on how a mature [LLM](#)-based [GIS](#) could impact the daily work of (human) [GIS](#) professionals.

---

<sup>1</sup>Including articles whose title or abstract includes "LLM", "Large Language Model", or "GPT".

## 1. Introduction

Column1	arXiv	Google Scholar	Web of Science	PapersWithCode
Autonomous AI	2342	234	2	342
GIS	2342	234	2	342
Conditional random field	2342	234	2	342
Geospatial standardization	2342	234	2	342
LLM	2342	234	2	342
GPT	2342	234	2	342
BERT	2342	234	2	342
Transfer learning	2342	234	2	342
Foundation model	2342	234	2	342
Embedding	2342	234	2	342
Fine-tuning	2342	234	2	342
Prompt engineering	2342	234	2	342
Retrieval-augmented generation	2342	234	2	342

Table 1.1.: Literature study search results.

**RQ 2** focuses on the feasibility of using the **OGC API** - Features Standads in a typical overlay analysis within a conversational **AI** context like ChatGPT-4, having the users be able to express themselves using natural language queries. An answer to **RQ 2** would describe how **OGC APIs** can be called and manipulated in a flexible manner during a conversation to perform spatial queries or analyses, like intersecting layers or filtering features based on certain criteria.

**RQ 3** delves into the technical and ethical considerations of expanding ChatGPT-4's capabilities through integration with external tools, such as **GIS** software or data analytics platforms. It would explore options for secure and efficient data exchange, and assess the implications of such access in terms of data privacy and user consent.

## 1.3. Research Method

### 1.3.1. Literature Study

I set out by deciding on some search terms, and used a "snowball sampling" technique to find relevant articles. [Table 1.1](#) shows the number of search results on various sites.

## 1.4. Contributions

## 1.5. Thesis Structure

Fix  
table  
data  
and styl-  
ing

## 2. Theory

Chapter 2 of this specialization project will talk about the leading technologies in the field of [Large Language Model \(LLM\)](#), which in itself is a subfield of [Natural Language Processing \(NLP\)](#). [section 2.1](#) will go over the leading [LLM](#) models, their strengths and weaknesses, and briefly mention differences in model architecture. [Section 2.2](#) will name the most prominent providers of [LLM](#) services.

[Figure 2.1](#) shows an actor map which includes stakeholders, providers, and other groups and organizations that could have some relevance to an autonomous LLM-based GIS-agent. Following subsections will go over the different groups and explain why they are included in the actor map.

### 2.1. LLMs

#### 2.1.1. Benchmarking for Large Language Models

There are several ways of benchmarking [Large Language Models](#) but this section will focus on [HumanEval](#) and [Multitask Language Understanding \(MMLU\)](#).

##### **HumanEval**

[HumanEval](#) is a dataset of handwritten problems used to measure functional correctness for synthesizing programs for docstrings ([Chen et al., 2021, July](#), pp. 2–4).

##### **Multitask Language Understanding**

First introduced by [Hendrycks et al. \(2021, January\)](#) [Multitask Language Understanding](#) is a way of testing a [Large Language Model](#)'s multitask accuracy, covering 57 tasks including mathematics, computer science, and others.

#### 2.1.2. The GPT Family

[Generative Pre-trained Transformers \(GPTs\)](#) are a type of [LLM](#) first introduced by [OpenAI](#) in 2018 ([Radford and Narasimhan, 2018](#)).

## 2. Theory

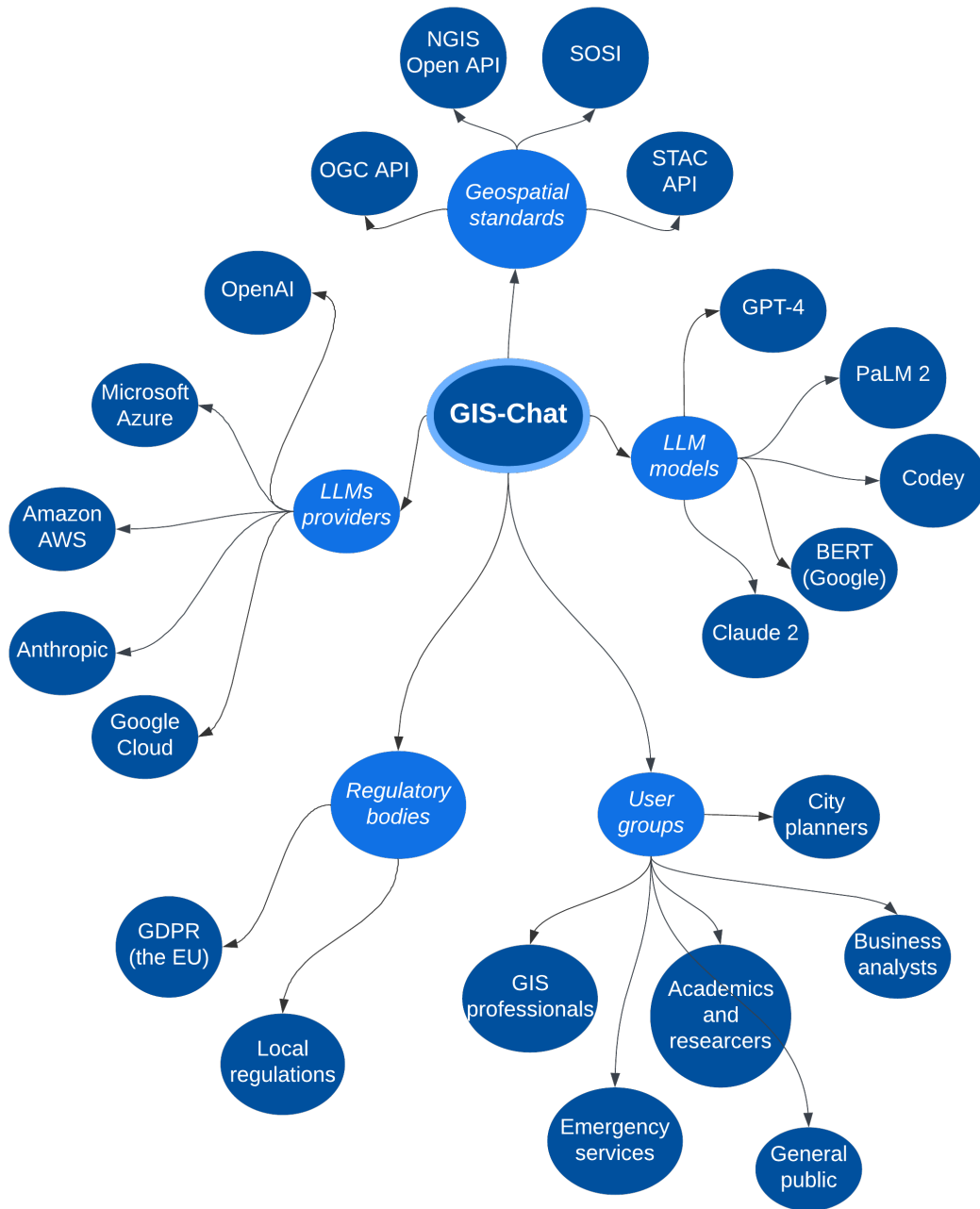


Figure 2.1.: Actor map for stakeholders, providers, and other groups and organizations that could have some relevance to an autonomous LLM-based GIS-agent.

## 2. Theory

### 2.1.3. Google Products

#### The **BERT** Family

[Bidirectional Encoder Representation from Transformers \(BERT\)](#) is a family of language models developed at Google ([Devlin et al., 2019, May](#)), first presented in 2018.

#### **PaLM 2**

#### Codey

### 2.1.4. Meta Products

Like [PaLM 2](#) The language models developed by Meta (previously Facebook) differ from the products from [OpenAI](#) in that Meta's LLama models are open-source.

#### **LLama 2**

#### Code LLama

### 2.1.5. Honourable Mentions

#### Falcon 180B

#### Mistral

#### Vicuna

## 2.2. [Large Language Model \(LLM\)](#) Providers

### 2.2.1. OpenAI

Being the most widely famous actor within the field of [LLMs](#), OpenAI has gained great influence through their vast portfolio.

### 2.2.2. Microsoft Azure

### 2.2.3. Google Cloud

### 2.2.4. [Amazon Web Services](#)

### 2.2.5. Anthropic

## 2.3. Geospatial Standards

### 2.3.1. International Standardization Work

#### [OGC](#) Api Standard

The [Open Geospatial Consortium \(OGC\)](#) API Standards serve as the glue in the field of [Geographic information system \(GIS\)](#), paving the way for interoperability and data

## 2. Theory

exchange between diverse systems. Leveraging common web protocols like [HTML](#) and supporting multiple data formats including [JSON](#), [GML](#), and [HTML](#). The [OGC API](#) standard provides a modular architecture consisting of a core specification and various extensions. This modularity allows for flexibility, enabling users to customize their services according to specific needs. According to their webpages, they provide 80 different standards, each for a specific geospatial purpose. Notable examples are 3D Tiles, CityGML, GeoTiff, and [OGC API - Features](#) ([OGC, 2023, October](#)).

[OGC API](#) Standards function as modern replacements to older standards like [WMS](#) and [WFS](#), and presents an evolved and more adaptable framework for spatial data operations, setting the stage for future innovations in the [GIS](#) domain.

### **STAC Api Standard**

The [SpatioTemporal Asset Catalog \(STAC\) API](#) is a standardized way to expose collections of spatial temporal data for online search and discovery. Built upon a [JSON](#) core, it aims to be a uniform and flexible environment from which developers can customize the API infrastructure to their domain. [STAC API](#) provides a powerful query language that allows users to search by various parameters like time, location, and keywords, making widely applicable. The [STAC](#) community has also defined specification in order to remove the complexity associated with having to create unique pipelines when consuming different spatial-temporary collection. The significance of the [STAC API](#) lies in its ability to democratize access to large volumes of geospatial data. By offering a common standard for data cataloguing and discovery, it reduces the barriers that often exist due to incompatible data formats. Developers or [GIS](#) professionals can take advantage of this through built-in tooling in QGIS, a desktop [GIS](#) for viewing, editing, and analysing spatial data, or through third-party packages in the Python and R programming languages. The API is also accessible through the command line interface when using [GDAL](#) ([STAC Tutorials n.d.](#)).

As [OGC](#) board member Chris Holmes puts it: "The [STAC API](#) implements and extends the [OGC API](#) — Features standard, and our shared goal is for [STAC API](#) to become a full [OGC](#) standard." ([Holmes, 2021, January](#)).

### **2.3.2. Norwegian Standardization Work**

Geospatial standardization work has been on the agenda of Norwegian governing powers for decades and have materialized in frameworks/collaborations like [Geovekst](#) and [Norge digitalt](#), as well as the [SOSI](#) file format. [Subsection 2.3.2](#) will delve into the work that has been done and what is expected for the future. The reasoning for the conclusion of this section is that the constraints of this specialization project is set by the Norwegian borders, and thus it is important to be aware of the standards that apply.

## 2. Theory

### SOSI

Samordnet Opplegg for Stedfestet Informasjon (SOSI) is a Norwegian file format for storing and exchanging geospatial data. It was first introduced in 1987 and has since approached international standards, the most important arenas currently being ISO/TC 211 and OGC (Mardal et al., 2015). SOSI is the adopted Norwegian standard for creating and delivering digital geographic data, administered by the Norwegian Mapping Authority (Statens kartverk) (Mæhlum and Rød, 2023).

In a SOSI dataset, terrain points, lines, and polygons are represented by their coordinates and classified into various object types according to the SOSI object catalog standard. However, there are few GIS systems that can read SOSI data directly, so data in SOSI format usually needs to be converted to another GIS-readable data format (Mæhlum and Rød, 2023).

### Geovekst

Geovekst is a collaborative initiative in Norway aimed at collecting, managing, and distributing geospatial information. Established in 1992, it is a partnership between national, regional, and local government bodies, as well as several private companies. Geovekst's primary focus is on creating a comprehensive, standardized geographical database for Norway that is easily accessible and updated regularly. It has played a vital role in various planning and development projects across the country, from urban planning to environmental conservation.

Unlike other geospatial initiatives, Geovekst emphasizes shared responsibilities and costs among its partners. This cooperative model ensures consistent data quality and efficient use of resources. It utilizes a variety of data sources, including aerial photographs, laser scans, and mapping, making it a rich resource for both public and private sectors. Moreover, its open-access policy allows for wider dissemination of geospatial information, thus encouraging innovation and informed decision-making across multiple disciplines.

### Norge digitalt

Established in 2005, Norge Digitalt is a more recent framework compared to Geovekst and is the name of Norway's national spatial data infrastructure. Norge Digitalt primarily involves governmental bodies (national, regional, and municipal), but also educational and research institutions and companies with responsibilities on a nation-wide scale; examples include Telenor and local and regional energy companies (Norge Digitalt, 2023, p. 6). Norge Digitalt aims to coordinate and streamline all geospatial activities in Norway, making it easier for users to discover, access, and use spatial data.

One key feature of Norge Digitalt is its focus on international standards and interoperability. While Geovekst is primarily a national initiative, Norge Digitalt aims to integrate Norway's geospatial data with that of other European countries. It supports a wide range of data formats and follows international standards, including those set by the Open Geospatial Consortium (OGC). The framework also provides various tools and



## 2. Theory

services, like metadata catalogues and web services, making it a comprehensive solution for geospatial data management and distribution in Norway.

### 2.4. User Groups

There are several user groups that could take advantage of an AI-based agent with general geographic and GIS knowledge. Questions can span from simple retrieval questions such as "How many people live in Trondheim" and "How long is the drive from Oslo to Bergen?", to more complicated questions that require problem-solving abilities and reasoning. While it is difficult to obtain dataset over common queries, Kumar (2023, May), creator of the chatbot app Pocket AI<sup>1</sup>, shared a dataset of ~13k user queries from his app along with classifications of these. Salient categories were:

- "task oriented" (23.1%)
- "informational" (20.2%)
- "social" (16.2%)
- "personal advice and self-improvement" (13.1%)

The main takeaway from these numbers is that the main motivation for use is productivity.

This aligns with the results of Skjuve et al. (2023, March) from their questionnaire-based study performed in late January 2023, about three months after its release (OpenAI, 2023, March). The goal with the study was to find out *why* people use ChatGPT. They found that most participants (55%) are motivated by productivity, and specifically applying it for routine tasks, information retrieval, text generation and writing support, and software development (Skjuve et al., 2023, March, pp. 17–21). Table 2.1 shows all categories and their frequencies. There were 197 samples in total, and more than one category could be assigned to each sample. It is worth noting that the study is likely to have included early adopters, and might therefore make the results less representative for the time at which this report is written (30th October 2023), now that use patterns have become more established (Skjuve et al., 2023, March, p. 37).

Given that the main reason people use conversational AI is for productivity, whether in a professional, academic, or personal context, such technology could be highly beneficial in a geospatial setting. 67 out of the 197 participants in Skjuve et al. (2023, March, p. 18) highlighted "ChatGPT's ability to understand complex queries" and that it is "efficient in alleviating the need to experiment with different phrasings of the query", as is often needed when 'Googling' for an answer to a specific question. This ease of information retrieval, along with its problem-solving abilities (Skjuve et al., 2023, March, p. 20), could also make conversational AIs highly relevant for geospatial purposes, and in the field of GIS. The following sections will elaborate on the potential user groups presented in Figure 2.1 that could benefit from such an artificial, and spatially aware, companion.

---

<sup>1</sup><https://github.com/varunon9/pocket-ai>

## 2. Theory

Table 2.1.: Categories and frequency of ChapGPT usage (Skjuve et al., 2023, March, pp. 16–17).

Category	% (n)
Productivity	55% (109)
Novelty	51% (101)
Fun and amusement	20% (41)
Creative work	18% (34)
Social interaction and support	9% (18)
Other	7% (15)

### 2.4.1. General Public

### 2.4.2. GIS Professionals

### 2.4.3. City Planners

### 2.4.4. Business Analysts

### 2.4.5. Academics and Researchers

### 2.4.6. Emergency Services

## 2.5. Regulatory Bodies and Privacy Concerns

### 2.5.1. Local Regulations

### 2.5.2. The [European Union](#)

## 2.6. Related Work

### 2.6.1. [GIS](#) with [LLMs](#)

[Roberts et al. \(2023, May\)](#) investigated extent of GPT-4’s geospatial awareness through a set of case studies with increasing difficulties, starting with general factual tasks and finishing with complex questions such as generating country outlines and travel networks.

[Li and Ning \(2023, May\)](#) states that “autonomous [GIS](#) will need to achieve five autonomous goals: self-generating, self-organizing, self-verifying, self-executing, and self-growing.”, and provide a “divide-and-conquer”-based method to address some of these goals. Furthermore, they propose a simple trial-and-error approach to addressing the self-verifying goal. They also highlight need of a memory system in a mature [LLM](#)-based [GIS](#) system, referring to the use of vector databases in autonomous agents like AutoGPT ([Richard, 2023, October](#)). Even with its shortages, the solution that ([Li and Ning, 2023, May](#)) provide, called [LLM-Geo](#), is able to solve provide good solutions in various case studies by providing executable assemblies in a Python environment when provided with URLs to relevant data sets, along with a user-specified query.

## 2. Theory

Zhang et al. (2023, July) uses the LangChain framework (Chase, 2022, October) in order to combine different GIS tools in a sequence in order to solve different sub-goals, and focuses on using the semantic understanding and reasoning abilities of LLMs like (e.g., ChatGPT) to call externally defined tools, employing the LLM as an agent or controller. The authors take great inspiration from the AutoGPT framework (Richard, 2023, October). The externally defined tools are described (manually) by its name and description. Said description contains information about the input parameters and output types of the tools/functions. Tools are defined for geospatial data collection, data processing and analysis, and data visualization. The effectiveness of the system is showcased in four case studies.

### 2.6.2. Retrieval Augmented Generation

Retrieval Augmented Generation (RAG) is tightly interwoven with explainable AI, being a framework for retrieving facts from an external knowledge base to allow a LLM-based agent access to accurate up-to-date information (Martineau, 2023, August). A common problem when working with language models, especially those designed to be general-purpose, is hallucination; that is, when the model provides an answer that is completely wrong but in a very convincing manner. While progress is being made with newer models even the better ones, like GPT-4, gives an incorrect answer about 1 out of 5 times, and even worse for certain categories of queries (for instance 'code' and 'business') (OpenAI, 2023, March, p. 10). Retrieval Augmented Generation can help mitigate this problem.

#### LangChain

LangChain (Chase, 2022, October) is an open-source project that provides tooling that can be used to create autonomous AI agents. It is designed to help with prompt management and optimization, creating chains of calls to LLMs, data augmentation, autonomous agent creation, and memory-related tasks.

#### Microsoft Semantic Kernel

Microsoft Semantic Kernel is an SDK that functions as the brain of an autonomous agent and provides connectors to models and memory, and connects to triggers and actions.

#### AutoGPT

Richard (2023, October) will try to split a task into subtasks and use the internet and other tools in an automatic loop to solve the task/subtasks.

### 2.6.3. Prompt Engineering and Planning Strategies

Zhou et al. (2023, October) introduces a framework called Language Agent Tree Search (LATS) "that synergizes the capabilities of LLMs in planning, acting, and reasoning."

## 2. *Theory*

As of writing (October 30th, 2023), the [LATS](#) framework is the highest scoring model on the [HumanEval](#) benchmark.

# 3. Experiments and Results

## 3.1. Experimental Plan

The experiments conducted for this report have been planned to answer the research questions described in [section 1.2](#).

### 3.1.1. Determining the Potential of LLM-based GIS analysis ([RQ 1](#))

[RQ 1](#) differs from [RQ 2](#) and [RQ 3](#) in that it is more open-ended. The tests will try to display the abilities of [GPT-4](#) for geospatial tasks, without providing it with any context or external tools. The tests are inspired by the work of [Roberts et al. \(2023, May\)](#) (see [subsection 2.6.1](#)), who did experiments with increasing difficulty on [GPT-4](#) to characterize what [GPT-4](#) knows about the geographical world, highlighting both capabilities and limitations.

### 3.1.2. Testing ChatGPT's Ability to Perform Overlay Analysis using OGC API - Features ([RQ 2](#))

Experiments on [RQ 2](#) will require three different elements:

1. Input data in correspondence with the [OGC API - Features](#) specification
2. Prompts to ChatGPT-4 using the "Code Interpreter" beta feature
3. Gold standard/expected output for the given combination of input data and prompts

### 3.1.3. Testing ChatGPT's Ability to Use External Tools ([RQ 3](#))

Experiments on [RQ 3](#) will test different tools and techniques intended to give an LLM access to up-to-date information and external tools (see [subsection 2.6.2](#) on [Retrieval Augmented Generation](#)). The main focus will be on how one can use a tool like [LangChain](#) hook up a conversational [AI](#) with sophisticated functionality found in various [GIS](#) software.

## 3.2. Experimental Setup

## 3.3. Experimental Results

## 4. Discussion

### 4.1. Evaluation

### 4.2. Discussion

Yiu et al. (2023) calls AI technologies like Large Language Models "powerful imitation engines" but claim that they are unable to innovate and that significant advancements in AI development is required for them to be able to learn the in the same way human children do.

# 5. Conclusion and Future Work

## 5.1. Contributions

## 5.2. Future Work

### 5.2.1. Test regime

In order to test the feasibility of different language models to serve as the brain of an autonomous GIS agent, a testing regime should be developed. In the examples of autonomous GIS agents described in the literature study of this report (see [subsection 2.6.1](#)), results have generally been presented in the form of case studies ([Li and Ning, 2023, May](#); [Zhang et al., 2023, July](#)). This type of qualitative testing is entirely appropriate to showcase the possibilities of the technologies but may be insufficient when comparing performance of different systems. In the latter case a quantitative approach would probably be preferable.

One idea is to create a test dataset which consists of inputs and corresponding desired outputs of typical GIS tasks. Inputs would in this case be natural language queries inputted by a mock user, and the output would be what you would expect a GIS professional to return when given the same tasks/queries. Inputs should reflect the varying level of GIS knowledge in the different user groups (see [section 2.4](#)). Outputs could be files with typical geospatial extensions (.shp, .geojson, .sosi, etc.), or they could adhere to API schemas specified by geospatial standards (see [section 2.3](#)).

While the inputs should be fairly simple to construct there are several questions to be answered in regard to the outputs:

- How does one evaluate the accuracy of the output?
- How should the AI agent respond when the user does not specify an output file format?
- How does one evaluate the usefulness of outputs to questions that should not return geospatial files, e.g. answers to general questions about geo-related subjects?

These are questions outside the scope of this specialization project. They will, however, be pursued in my master thesis.

## *5. Conclusion and Future Work*

### **5.2.2. Framework for Planning, Acting, and Reasoning**

### **5.2.3. Embeddings**

### **5.2.4. Fine-Tuning**



# Acronyms

**AI** Artificial Intelligence.

**API** Application Programming Interface.

**AWS** Amazon Web Services.

**BERT** Bidirectional Encoder Representation from Transformers.

**EU** European Union.

**GDAL** Geospatial Data Abstraction Library.

**GIS** Geographic information system.

**GML** Geography Markup Language.

**GPT** Generative Pre-trained Transformer.

**HTML** HyperText Markup Language.

**HTTP** Hypertext Transfer Protocol.

**ISO** International Organization for Standardization.

**JSON** JavaScript Object Notation.

**LLM** Large Language Model.

**NLP** Natural Language Processing.

**OGC** Open Geospatial Consortium.

**RAG** Retrieval Augmented Generation.

**SDK** Software Development Kit.

**SOSI** Samordnet Opplegg for Stedfestet Informasjon.

**SQL** Structured Query Language.

## *Acronyms*

**STAC** SpatioTemporal Asset Catalog.

**TC** Technical committee.

**WFS** Web Feature Service.

**WMS** Web Map Service.

**XML** Extensible Markup Language.

# Bibliography

- Chase, H. (2022, October). LangChain. Retrieved October 5, 2023, from <https://github.com/langchain-ai/langchain>
- Chen, M., Tworek, J., Jun, H., Yuan, Q., Pinto, H. P. d. O., Kaplan, J., Edwards, H., Burda, Y., Joseph, N., Brockman, G., Ray, A., Puri, R., Krueger, G., Petrov, M., Khlaaf, H., Sastry, G., Mishkin, P., Chan, B., Gray, S., Ryder, N., Pavlov, M., Power, A., Kaiser, L., Bavarian, M., Winter, C., Tillet, P., Such, F. P., Cummings, D., Plappert, M., Chantzis, F., Barnes, E., Herbert-Voss, A., Guss, W. H., Nichol, A., Paino, A., Tezak, N., Tang, J., Babuschkin, I., Balaji, S., Jain, S., Saunders, W., Hesse, C., Carr, A. N., Leike, J., Achiam, J., Misra, V., Morikawa, E., Radford, A., Knight, M., Brundage, M., Murati, M., Mayer, K., Welinder, P., McGrew, B., Amodei, D., McCandlish, S., Sutskever, I., & Zaremba, W. (2021, July). Evaluating Large Language Models Trained on Code.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019, May). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.
- Fan, A., Gokkaya, B., Harman, M., Lyubarskiy, M., Sengupta, S., Yoo, S., & Zhang, J. M. (2023, October). Large Language Models for Software Engineering: Survey and Open Problems.
- Hendrycks, D., Burns, C., Basart, S., Zou, A., Mazeika, M., Song, D., & Steinhardt, J. (2021, January). Measuring Massive Multitask Language Understanding.
- Holmes, C. (2021, January). SpatioTemporal Asset Catalogs and the Open Geospatial Consortium. Retrieved October 23, 2023, from <https://medium.com/radiant-earth-insights/spatiotemporal-asset-catalogs-and-the-open-geospatial-consortium-659538dce5c7>
- Kumar, V. (2023, May). What are people asking to ChatGPT? Retrieved October 25, 2023, from <https://varunon9.medium.com/what-are-people-asking-to-chatgpt-f5f324a6cc27>
- Li, Z., & Ning, H. (2023, May). Autonomous GIS: The next-generation AI-powered GIS.
- Mæhlum, L., & Rød, J. K. (2023). SOSI. *Store norske leksikon*. Retrieved October 23, 2023, from <https://snl.no/SOSI>

## Bibliography

- Mardal, G., Borreb, M., Christensen, L., Jetlund, K., Ryghaug, P., & Hokstad, I. (2015). Nasjonal strategi for videreutvikling av SOSI.
- Martineau, K. (2023, August). What is retrieval-augmented generation? Retrieved October 6, 2023, from <https://research.ibm.com/blog/retrieval-augmented-generation-RAG>
- Norge Digitalt. (2023). *Generelle vilkår for Norge Digitalt-samarbeidet* (tech. rep.). Retrieved October 23, 2023, from <https://www.geonorge.no/globalassets/geonorge2/avtaler-og-bilag-norge-digitalt/generelle-vilkar.pdf>
- OGC. (2023, October). OGC Standards. Retrieved October 23, 2023, from <https://www.ogc.org/standards/>
- OpenAI. (2023, March). GPT-4 Technical Report.
- Radford, A., & Narasimhan, K. (2018). Improving Language Understanding by Generative Pre-Training. Retrieved October 9, 2023, from <https://www.semanticscholar.org/paper/Improving-Language-Understanding-by-Generative-Radford-Narasimhan/cd18800a0fe0b668a1cc19f2ec95b5003d0a5035>
- Richard, T. B. (2023, October). AutoGPT: The heart of the open-source agent ecosystem. Retrieved October 5, 2023, from <https://github.com/Significant-Gravitas/AutoGPT>
- Roberts, J., Lüddecke, T., Das, S., Han, K., & Albanie, S. (2023, May). GPT4GEO: How a Language Model Sees the World’s Geography.
- Skjuve, M., Bae Brandtzaeg, P., & Følstad, A. (2023, March). Why People Use ChatGPT.
- STAC Tutorials. (n.d.). Retrieved October 23, 2023, from <https://stacspec.org/en/tutorials/>
- Yiu, E., Kosoy, E., & Gopnik, A. (2023). Transmission Versus Truth, Imitation Versus Innovation: What Children Can Do That Large Language and Language-and-Vision Models Cannot (Yet). *Perspectives on Psychological Science*, 17456916231201401.
- Zhang, Y., Wei, C., Wu, S., He, Z., & Yu, W. (2023, July). GeoGPT: Understanding and Processing Geospatial Tasks through An Autonomous GPT.
- Zhou, A., Yan, K., Shlapentokh-Rothman, M., Wang, H., & Wang, Y.-X. (2023, October). Language Agent Tree Search Unifies Reasoning Acting and Planning in Language Models. Retrieved October 30, 2023, from <http://arxiv.org/abs/2310.04406>

# Appendices

# A. Task Description from Norkart

Oppgave med omfang som kan tilpassast både prosjekt og masteroppgave

## LLMs - GIS-analysens død

(kan justerast seinare)

### BAKGRUNN

Nyere modeller for kunstig intelligens har demonstrert spesielt gode evner til å kunne lære av store mengder ustrukturert og semi-strukturert informasjon. ChatGPT fra OpenAI tok verden med storm – og chat-baserte systemer florerer. Kan chat-baserte modeller skapes for å hente ut GIS-data effektivt? Norkart har en stor dataplattform hvor brukere utvikler mot API'er som i stor grad har GIS/Geografiske data i bunn. GeoNorge er en stor datakatalog hvor brukere slår opp, eller søker kategorisert for å finne data. QGIS, Python, PostGIS, FME og andre verktøy brukes ofte til å gjennomføre GIS-analyser – hvor en GIS-analytiker/data-scientist gjennomfører dette.

*«Finn alle bygninger innenfor 100-meters-belte som er over 100 kvm og har brygger»*

Er dette mulig å få til med dagens tilgjengelige chat-modeller?

### OPPGAVEBESKRIVELSE

Oppgaven har som hovedmål å undersøke hvordan nyere språkmodeller kan benyttes for å gjennomføre klassiske GIS-analyser ved å bruke standard GIS-teknologi som PostGIS/SQL og datakataloger (OGC API Records fks). Hva finnes av tilgjengelig chat-løsninger? Hvordan spesialtilpasse til GIS-anvendelser? Hvor presise kan en GIS-Chat bli?

Relevante delmål for oppgaven:

1. Kartlegge state-of-the-art
2. Utvikle proof-of-concepts
3. Analysere begrensninger og kvalitet

Oppgaven vil med fordel deles i prosjektoppgave og masteroppgave

- Prosjektoppgave
  - State-of-the-art: Ai-modeller og multi-modal maskinlæring
  - Innhente og utvikle datagrunnlag og API-tilgjengelighet
- Masteroppgave
  - Utvikle proof-of-concepts med tilgjengelige åpne modeller/teknologi
  - Gjennomføre eksperimenter for analyse av kvalitet

## *A. Task Description from Norkart*

Detaljert oppgavebeskrivelse utvikles i samarbeid med studenten.

### **ADMINISTRATIVT/VEILEDNING**

Ekstern veileder: (en eller flere)

Mathilde Ørstavik, Norkart

Rune Aasgaard, Norkart

Alexander Nossun, Norkart

Aktuelle vegleiarar og ansvarleg professor ve NTNU (den som har fagansvar nærast oppgåva):

Terje Midtbø (GIS, kartografi, visualisering)

Hongchao Fan (3D modellering, fotogrammetri, laser)

## B. API Schemas

### B.1. OGC API - Features

OGC API specification for a 'collection' object<sup>1</sup>:

---

```
type: object
required:
  - id
  - links
properties:
  id:
    description: identifier of the collection used, for example, in URIs
    type: string
    example: address
  title:
    description: human readable title of the collection
    type: string
    example: address
  description:
    description: a description of the features in the collection
    type: string
    example: An address.
  links:
    type: array
    items:
      $ref: link.yaml
  example:
    - href: http://data.example.com/buildings
      rel: item
    - href: http://example.com/concepts/buildings.html
      rel: describedby
      type: text/html
  extent:
    $ref: extent.yaml
  itemType:
    description: indicator about the type of the items in the collection (the default value is 'feature').
    type: string
    default: feature
  crs:
    description: the list of coordinate reference systems supported by the service
    type: array
    items:
      type: string
```

---

<sup>1</sup><https://schemas.opengis.net/ogcapi/features/part1/1.0/openapi/schemas/collection.yaml> (retrieved October 25, 2023)



## *B. API Schemas*

default :

– <http://www.opengis.net/def/crs/OGC/1.3/CRS84>

example:

– <http://www.opengis.net/def/crs/OGC/1.3/CRS84>

– <http://www.opengis.net/def/crs/EPSG/0/4326>

---

### **B.2. STAC API**

## C. Code Examples