ChatGPT, Foundation Models, and Geo-Foundation Models

Dr. Gengchen Mai

[2022 - Now] Assistant Professor at Department of Geography, University of Georgia

https://gengchenmai.github.io/
Large Language Model

![Graph showing the growth of large language models from 2017 to 2024. Key models include:
- ELMo (94M)
- BERT-Large (340M)
- GPT-2 (1.5B)
- T5 (11B)
- GPT-3 (175B)
- Megatron-LM (8.3B)
- Turing-NLG (17.2B)
- Megatron-Turing NLG (530B)
- InstructGPT
- ChatGPT
- GPT-4

Image from Huggingface (https://huggingface.co/blog/large-language-models)
Foundation Models

**Task-Specific Models**
- Training **specific** models for **specific** tasks
  - Question Answering Models
  - Machine Translation Models
  - Common Sensing Reasoning Models
  - Reading Comprehension Models
  - Natural Language Inference Models
  - Image Classification Models
  - Text-to-image Generation Models
  - Image Editing Models

**Foundation Models**
- A large task-agnostic pre-trained model which can be adapted via **fine-tuning** or **few-shot/zero-shot** learning on a wide range of domains. (Bommasani et al., 2021)

**GPT-3** (Brown et al., 2020)
- **Few-shot Adaptation**

**DALL·E 2** (Ramesh et al., 2022)
- **Zero-shot Transfer**

**Various NLP Tasks**
- Closed Book Question Answering
- Machine Translation
- Common Sense Reasoning
- Reading Comprehension
- Natural Language Inference
- …

**Various CV Tasks**
- Text-to-image generation
- Image Completion
- Image Editing
- Style Transfer
- …

**Paradigm shift**
The Success of Artificial General Intelligence

Natural Language Processing

Stanford Alpaca

Stanford Alpaca

ChatGPT/GPT-4 (OpenAI 2023)

Computer Vision

Imagen (Saharia et al. 2022)

Segment Anything (Kirillov et al, 2023)

Reinforcement Learning

Gato (Reed et al. 2022)

Signal Processing

Whisper (Radford et al. 2022)
AGI on Geospatial Problems

How do the existing cutting-edge foundation models perform when compared with the state-of-the-art fully supervised task-specific models on various geospatial tasks?
A Multimodal FM for GeoAI

**Vision:** a multimodal FM for GeoAI that use their geospatial relationships as alignments among different data modalities.