LIKOLA MISHECK

ЛИКОЛА МИШЕК
5130203/20101

Summary of Week 01: Introduction to AI Subfields

The first week of our AI course presented a foundational overview of the vast and dynamic landscape of Artificial Intelligence (AI). The introductory lecture emphasized how AI is an umbrella term that encompasses a wide array of subfields, each with its own methodologies, tools, and applications. AI is fundamentally concerned with creating machines and systems capable of performing tasks that would normally require human intelligence, such as decision-making, language processing, vision, and robotics.

Slide 5 from the course material illustrated many of these subfields, highlighting key topics like **Machine Learning**, **Natural Language Processing**, **Neural Networks**, **Large Language Models**, **AI Agents**, and **Reinforcement Learning**, among others. These subfields are not isolated but rather interconnected in a way that advances the overall capabilities of AI. For example, the fusion of machine learning and neural networks drives innovations in **Deep Learning**, while **Multimodal AI** integrates multiple forms of data (e.g., text, images, audio) into cohesive AI systems.

In today's world, the rapid progression of AI technology is reshaping industries from healthcare and finance to entertainment and education. The course thus opened with a high-level overview to ensure that we understand AI not just as a concept but as a multifaceted discipline with immense real-world impact.

A Closer Look at Large Language Models (LLMs)

One subfield that especially captured my attention is **Large Language Models (LLMs)**. The rise of LLMs has revolutionized how machines interact with human language. These models, often built using massive datasets, are capable of understanding, interpreting, and generating human-like text. **GPT (Generative Pre-trained Transformer)** is one of the most notable LLMs and exemplifies how these models are transforming natural language processing (NLP).

LLMs work by predicting the next word in a sequence based on the context provided by previous words. Their ability to generate coherent and contextually appropriate text makes them invaluable for a wide range of applications, from conversational agents to machine translation, content generation, and even advanced research tools.

How LLMs Work: Transformer Architecture

At the core of LLMs like GPT is the **Transformer architecture**, which has been a game-changer in NLP. Unlike traditional models such as Recurrent Neural Networks (RNNs) or Long Short-Term Memory (LSTM) networks, transformers use a mechanism called **self-attention**. This mechanism allows the model to focus on different parts of a sentence when generating or understanding language, giving it the ability to capture long-range dependencies more effectively. In simpler terms, transformers enable the model to understand the context better by focusing on the most relevant words or phrases, even if they are far apart in a sentence.

Additionally, the scalability of transformers allows them to be trained on enormous datasets. This has led to models with billions of parameters that can generate high-quality text, understand multiple languages, and even solve complex problems requiring reasoning and comprehension. The training of these models involves several computational challenges, requiring high-performance hardware and a sophisticated understanding of **gradient descent** algorithms, data augmentation, and parallelization techniques.

Impact of Large Language Models on the AI Industry

The success of LLMs has pushed the boundaries of what AI can achieve in terms of human-computer interaction. For instance, LLMs like GPT-3 can generate creative content, write code, summarize information, translate languages, and answer complex questions. They are also beginning to serve as the backbone for many AI-powered applications, such as virtual assistants, customer support bots, and even AI-based tutoring systems.

Despite their remarkable capabilities, LLMs are not without limitations. One significant challenge is the model's tendency to produce outputs that are factually incorrect or biased, as these models are often trained on vast and uncurated datasets from the internet. Additionally, there are ongoing debates about the ethical implications of deploying LLMs in sensitive areas, such as healthcare or journalism, where accuracy and reliability are paramount.

Previous Knowledge and Experience

I have had the opportunity to interact with LLMs in various capacities, primarily in the realm of content generation and automation. In my previous projects, I’ve worked with AI-driven tools that assist in writing marketing copy, generating reports, and automating communication for customer support. These experiences gave me a glimpse into the power and versatility of LLMs, particularly how they can produce human-like text and assist in streamlining tasks that would otherwise be time-consuming.

Although I have used pre-trained models for these tasks, my understanding of the inner workings of LLMs remains somewhat basic. I am familiar with concepts like **pre-training** and **fine-tuning**, where a model is initially trained on a large dataset and later fine-tuned on a specific, smaller dataset for a targeted application. However, I have not had the chance to delve deeply into the technical aspects, such as how the models are architected, trained, or optimized at a large scale.

Expectations for the Course

Through this course, I hope to gain a deeper and more structured understanding of how LLMs are built, trained, and applied. I am particularly keen on learning about:

1. **The mathematics behind transformers** – I expect to understand in detail how the attention mechanism works, how it improves over traditional neural networks, and how it allows for better scalability in NLP tasks.
2. **Challenges in training large models** – The sheer size of models like GPT-4 or GPT-5 requires a profound understanding of not just the algorithms but also the computational infrastructure. I want to explore what it takes to train models of this size and how problems like overfitting, data bias, and computational efficiency are tackled in the real world.
3. **Applications beyond text generation** – While I have experience with content generation, I am eager to explore how LLMs are used in fields such as healthcare for medical diagnostics, law for document analysis, and education for personalized learning systems. Additionally, I want to explore **multimodal applications**, where language models are combined with vision or auditory data to create more holistic AI systems.
4. **Ethical considerations** – With great power comes great responsibility. I look forward to discussing the ethical challenges of deploying LLMs, particularly around issues like bias, misinformation, and job displacement.
5. **Future advancements** – Finally, I am intrigued by the potential of LLMs to contribute to the broader AI goal of developing **Artificial General Intelligence (AGI)**. LLMs are already performing tasks that previously seemed reserved for human intelligence, but AGI requires models that can learn and reason across a wider array of contexts, domains, and tasks.

Conclusion

The first week of the course has set the stage for what promises to be an exciting and insightful journey into the world of AI. As I continue with this course, I look forward to not only expanding my technical knowledge of LLMs and transformers but also understanding the broader implications of AI in society. The rapid evolution of AI technology offers incredible opportunities for innovation, and I am eager to be part of this transformative field.