Home / GenAI In-Depth

GenAI In-Depth: The Science and Capabilities of GenAI

The field of Generative AI has experienced exponential advancements in recent years, demonstrating remarkable progress across diverse modalities such as text, images, sound, and more. These advancements can be primarily attributed to the three main factors: **methods, data,** and **scale of computation.**

First, advances in deep learning techniques, specifically transformer architectures, have enabled more powerful and efficient modeling of complex relationships in the data. Second, the availability of vast and diverse datasets, encompassing extensive text corpora and video repositories, has significantly contributed to the quality and diversity of generated outputs. Third, the substantial growth in computational power has played a crucial role in enabling the training and deployment of increasingly complex generative models. Collectively, these technical factors have fueled the rapid evolution of GenAI, paving the way for breakthroughs in various domains of content generation.

Language Generation

Language generation builds on the strong foundation of language models, dating back to 1948. For instance, the GPT (Generative Pre-trained Transformer) model, including versions such as GPT-3, 3.5, 4, is a state-of-the-art language model developed by OpenAI. It relies on a transformer deep learning architecture, designed to process and generate natural language text. The architecture includes multiple layers of self-attention mechanisms, enabling it to capture the contextual relationships and dependencies between words and generate coherent and contextually relevant responses.

GPT models have been trained on an extensive corpus of diverse text data of more than 400 billion tokens, including books, articles, and web pages, using unsupervised learning techniques. The training process involves predicting the next word in a sentence based on the preceding context, enabling the model to learn grammar, semantics, and common language patterns. This pre-training phase equips GPT-3 with a broad understanding of human language and knowledge.

InstructGPT is a variant of the GPT (Generative Pre-trained Transformer) model developed by OpenAI. It shares the technical foundation of GPT, utilizing a transformer architecture and unsupervised learning. However, what sets InstructGPT apart is its specific training objective, which focuses on generating text conditioned on user instructions. During the pre-training phase, InstructGPT is trained to predict the next word in a sentence given both the preceding context and an additional instruction prompt.

This conditioning enables the model to generate text that adheres to specific guidelines provided by the user. By fine-tuning InstructGPT on custom datasets with specific instruction-based tasks, it can be tailored to perform a range of practical applications, such as generating code, writing essays, answering questions, or providing detailed instructions. The technical aspects of InstructGPT leverage the power of transformer-based architectures and instruction conditioning to generate contextually coherent and user-guided text outputs.

Multimodal Generation

Advances have also been made in the space of multimodal GenAI, which includes other modalities such as audio (including music), images, and video. For instance, CLIP is a transformer-based model that uses a dual-encoder architecture, and encodes both images and language. The image encoder processes images by passing them through a convolutional neural network, while the text encoder utilizes a transformer-based architecture to process textual descriptions. By performing this process jointly across images and text, CLIP learns to align their representations in a shared embedding space, which allows it to capture meaningful relationships between images and their associated textual descriptions. By maximizing the similarity between corresponding image-text pairs and minimizing the similarity between non-matching pairs, CLIP learns to understand the semantic connections between visual and textual data.

The resulting joint embedding space enables a range of applications, including zero-shot image classification, where CLIP can find the most likely textual description of an image, and zero-shot image retrieval, where it can find images that match a given text query.

Another example of a multimodal GenAI is DALL-E, which combines language models with an image encoder-decoder architecture. DALL-E has the ability to generate novel images based on textual prompts, showcasing impressive creativity in synthesizing unique and imaginative visuals. By learning a latent space representation of images, DALL-E can generate highly detailed and diverse images, transcending conventional image synthesis techniques.

Autonomous GenAI Systems

The first wave of widely adopted GenAI tools such as those introduced above are interactive or conversational agents that are dependent on human prompts to perform actions. More recently, applications that can plan and operate more autonomously have emerged:

Auto-GPT is a system for autonomous task execution, built on top of other GenAI models, such as GPT. Unlike interactive systems such as U-M GPT, ChatGPT, Auto-GPT operates without constant human input, by setting its own objectives to pursue, generating responses to prompts, and adapting its prompts recursively based on new information. It can autonomously perform several actions such as web searching, web form interactions, or API interactions. As an early self-driven system for task execution, Auto-GPT is an example of an AI system that can not only perform tasks determined by human users, but also define its own tasks, a step forward toward more complex AI-driven autonomy and problem solving.

Autonomous GenAI tools have the potential to revolutionize numerous technological and societal sectors. However, the deployment of unsupervised autonomous GenAI systems also introduces substantial technological challenges and societal risks, necessitating careful consideration and robust safeguards in their development.

Limitations and Risks

There are several risks and limitations that come with the use of GenAI. A few of these risks are listed below:

- **Unintended bias**
  GenAI models are trained on large language or vision datasets, which are often reflective of multiple societal biases. These biases can be propagated in the generated content, perpetuating and amplifying societal biases and prejudices.
- **Misinformation and disinformation**
  GenAI systems often produce misinformative statements and make unsupported claims. Their inability to provide a confidence level for the information they provide makes it difficult to determine when to trust these models. Further, these AI systems can be exploited to generate false or misleading information. The resulting misinformative content can have significant negative implications for trust, credibility, and the manipulation of public opinion.
- **Lack of transparency**
  Because of being particularly complex, it is often challenging to interpret and understand the current AI models. This lack of transparency can make it difficult to determine how the model arrives at its generated outputs, which also makes it hard to identify biases, errors, or potential ethical issues, and hinders accountability.
- **Privacy concerns**
  Very large amounts of data are typically needed to build AI systems, which can lead to a risk of privacy breaches and unauthorized use of personal information during data collection, storage, and utilization.
- **Intellectual property and copyright infringement**
  In a similar vein, the use of very large datasets, and the functioning of generative AI which often replicates patterns from the training data, can result in the unauthorized generation or replication of copyrighted material.
- **Ethical considerations**
  There is a vast set of ethical considerations surrounding GenAI. In addition to the risks mentioned before, other ethical implications include the appropriate use of AI-generated content, consent for data usage, potential impact on human creativity, impacts on labor markets and employment, and others.
- **Existential risk**
  A small yet growing fraction of technology leaders have articulated apprehensions that GenAI systems—and AI systems more broadly—represent a potential existential threat to society. This is predominantly attributable to the unpredictable

characteristics of expansive AI algorithms and the potential for these systems to autonomously optimize themselves, which could result in unanticipated and undesirable consequences. Irrespective of one's position within the spectrum of these concerns, there is unanimous consensus that AI systems necessitate the establishment of stringent safeguards during their construction. Furthermore, it is imperative that robust policies are instituted at various levels to mitigate the risks associated with these systems.

Information and Technology Services

- About ITS
- Safe Computing
- ITS Service Status
- Work at ITS

Stay Connected

Contact ITS

- Wolverine Access
- Office of the VPIT-CIO
- U-M Website Privacy Notice