HARNESSING LANGUAGE MODELS FOR INVESTMENT WRITING AND CONTENT MARKETING

Investment writing is an important part of marketing investment products. Through ongoing engagement with clients, asset managers show that they value their business and constantly strive to deliver on the performance promise. Content also conveys expertise which is crucial to attract new clients. However, writing regular updates is often a tedious task that can distract investment staff from their main responsibilities.

The increasing popularity of machine learning and artificial intelligence raises the question of how machines can support the investment communication process.

This article outlines how to make use of the latest technology to make life easier for portfolio managers, product specialists, investment writers and relationship managers who often contribute to communications.

digipal

COMBINING COMPUTER SCIENCE AND LINGUISTICS

A branch of artificial intelligence known as Natural Language Processing (NLP) enables computers to understand text and spoken words. The goal is to integrate machine learning and statistics into computational linguistics to enable computers to understand human language. While NLP is about analyzing text, Natural language Generation (NLG) is about turning structured data into text.

In the quest to match human language breadth and capability, the size of language models has dramatically increased in recent years.

The Jumbo model launched by AI21 in September 2021 with its 178 billion parameters just slightly tops the competing OpenAi GPT-3 model.

Building such extensive models requires sheer computing power and huge amounts of example language to train the machines on. Such language models are often referred to as generative pre-trained transformers (GPT). With further advances in the field, it is expected that the largest models will soon exceed 200 billion parameters.

Thanks to their linguistic capabilities, these algorithms accept commands in the format of everyday language.

At the intersection of computer science and linguistics. NLG is a subset of NLP.
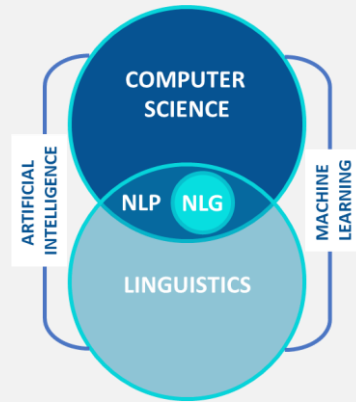
Language Models by launch date and size (in billion parameters)

digipal
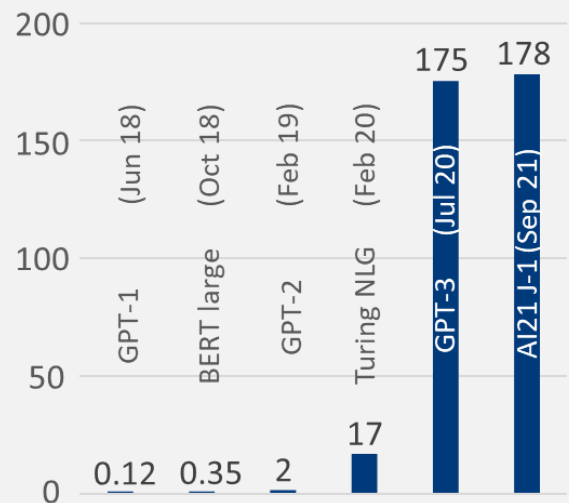
PROGRAMS THAT CAN GUESS THE NEXT WORD

Source: Pexels.

At the core of NLG lies the task of predicting the next word(s) following a seed phrase. We could ask to complete a statement starting with "Today is… ".

There is no deterministic answer. Possible variations depend on inputs such as desired length and creativity.

A language model has knowledge due to its previous training. It obtains the language logic from a text corpus, i.e. from a large and structured set of texts (usually from the Internet).

▶ "Today is the first day of the rest of your life."

▶ "Today is the day I am going to make a change."

▶ "Today is a good day to be alive. I'm alive and it's a good day."

digipal

Based on its 'experience', it can then formulate sentences and paragraphs. The output evolves as the knowledge base grows.

Note that the model cannot reason or argue purposefully. It will simply output valid statements from within the multitude of plausible statements.

To illustrate this point, let us examine more real-life cases based on the Davinci of GPT-3 Model. The command statement entered in the OpenAI playground is on the left of this page, and the response given by the model is on the right.

Command: "Outline the **main** asset classes of the financial markets"

This is certainly a valid response, although not the only possible answer.
The response can be very sensitive to the exact prompt. Let's change the prompt slightly:

"Outline **possible** asset classes of the financial markets"

One may wonder why exactly these five asset classes were selected. Why Currencies and not for example Private Equity?

Let's go a step further and request to
 "Outline **all possible** asset classes of the financial markets"

Once again, a very comprehensive list, although it's not clear if ETFs and derivatives should be treated as separate asset classes. As impressive as the answers are, the examples make it clear that the model has acquired its own logic in looking at the world.

"Language models cannot reason or argue purposefully."

▶ Response: "The main asset classes of the financial markets are stocks, bonds, and cash."
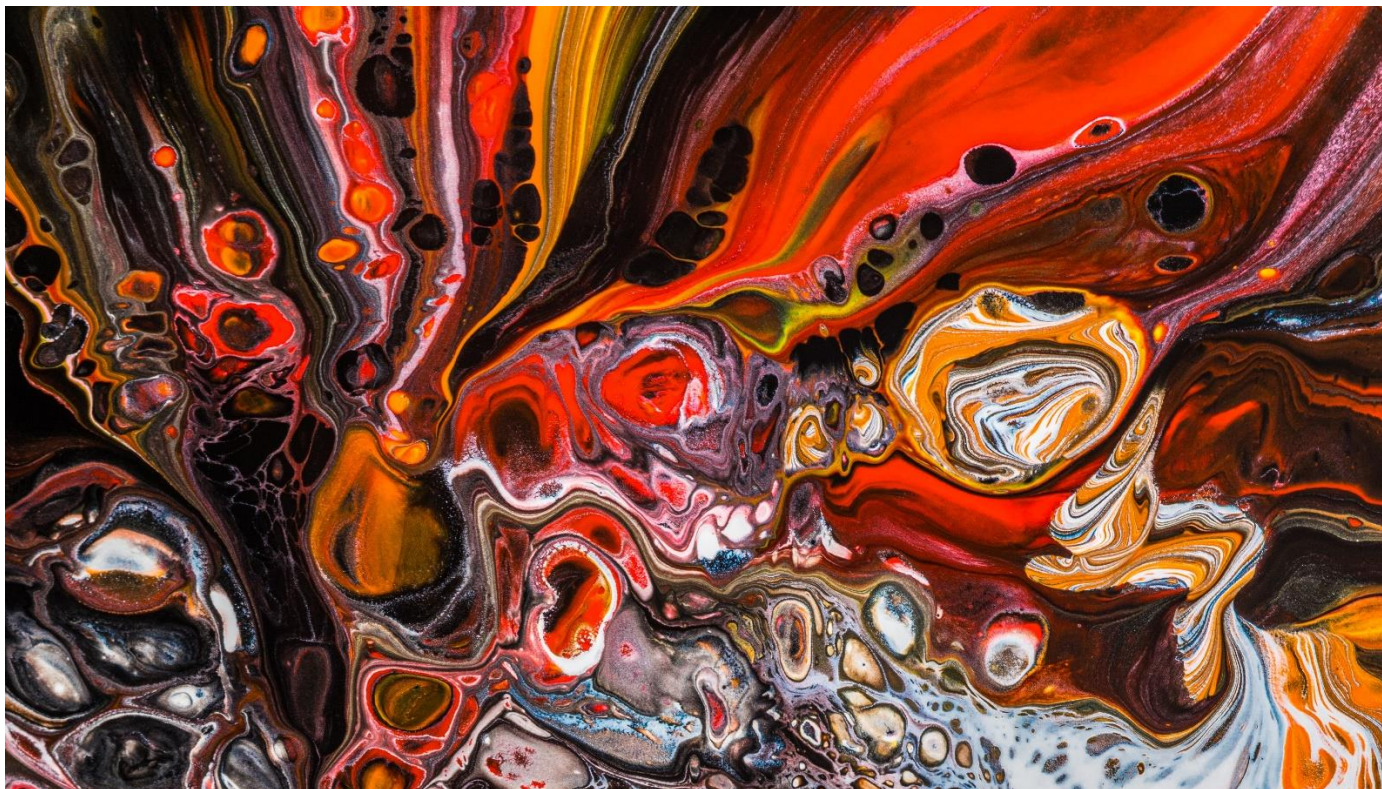
▶ This time the response comes in a nicely structured way:
- Stocks
- Bonds
- Commodities
- Currencies
- Real Estate

▶ GPT-3 responds:
- Cash and cash equivalents
- Bonds
- Stocks
- Mutual Funds
- Exchange-traded funds (ETFs)
- Hedge Funds
- Private Equity
- Real Estate
- Commodities
- Derivatives

digipal

RANDOMNESS AND CREATIVITY

Source: Pexels.

Let's move from structuring a subject to generating explanations. This time we would like to know: "Explain equity performance in **August**"

While this seems plausible, it does sound a bit outdated. At the time of writing, it is mid-September 2022. Perhaps, we should have been more precise and indicated the exact month in the command.

"Explain equity performance in **August 2022**"

This response is legitimate. However, it becomes evident that GPT-3 relies more on common sense than the latest news. How about looking further back?

▶ "The equity market had a strong performance in August, with the S&P 500 index rising by over 5%. This was driven by a number of factors, including strong corporate earnings, positive economic data, and increasing optimism about a potential trade deal between the US and China."

▶ "The equity performance in August 2022 will be influenced by a number of factors, including the state of the global economy, the performance of the stock market, and political conditions."

digipal

"Explain equity performance in **August 2021**"

What an amazing paragon of knowledge this is! For a moment one can think that this model may soon replace all other data sources. However, a closer look reveals that August 2021 wasn't quite that great. The true historical returns were much lower.

That's an important take away at this point: Language models do not draw from a structured database, a compendium of facts, or an encyclopedia, even though some of these sources have been used for training purposes. Even if the formulated sentences sound completely plausible and are grammatically correct, one should not assume that the generated content is accurate.

Let's be clear. Generative pre-trained language models can do an amazing job when it comes to writing human-like texts. Entire articles on a given topic can be written automatically if there is sufficient room. The content will make sense from a language logic perspective.

However, it needs to be understood that there is a random element. Every time the command is executed another version is produced. Each version may have a different focus and a different storyline.

While this may work for blog posts, a poem or a song, a more intentional approach is required when it comes to writing reports about portfolios of investments.

▶ "The equity market had a strong performance in August, with the S&P 500 index gaining 6.4% and the Dow Jones Industrial Average gaining 5.5%. The Nasdaq Composite Index, which is heavily weighted towards technology stocks, had an even stronger performance gaining 8.2%."

"One should not assume that the generated content is accurate."

"Every time the task is repeated a random version is produced which may potentially have a new focus and a different storyline."

6

MAKE A LANGUAGE MODEL WORK WITHIN GUARD RAILS
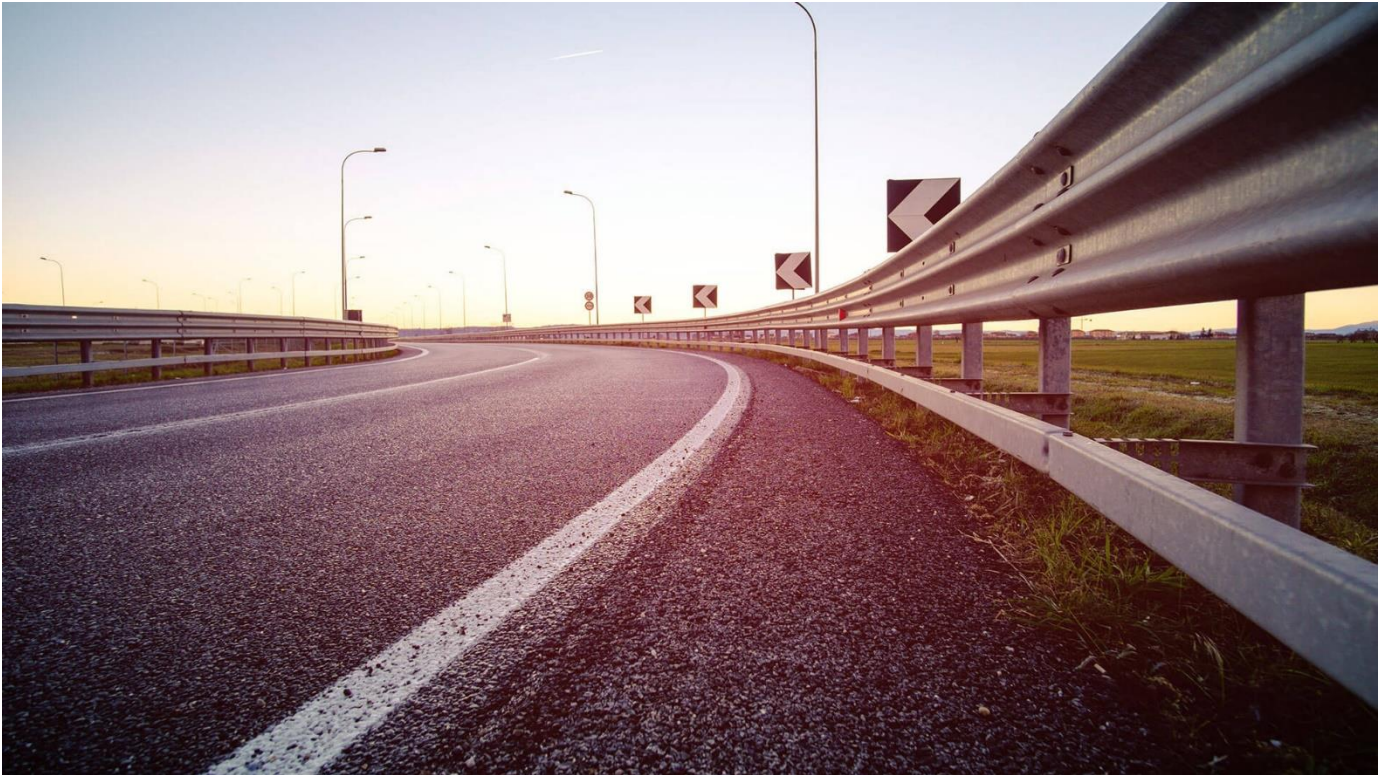
The most impressive results are achieved when the models have large degrees of freedom and can use their creative capability.[1]

It is different when the task is to work toward a specific goal or conclusion. It then helps to think about the text generation process in steps.

One way to break it down is to segment into document planning, microplanning and realization.

The challenge becomes to effectively deploy GPT while retaining control over essential aspects of the generated text.
A skeleton is needed for the structure and the main arguments.

Three Stages of the NLG Process

Source: Sciforce.

"Provide a skeleton for the structure and the main arguments."

---

[1] To get a sense of the creative capabilities try out commands like "List interesting topics about financial market regulation" or "Provide a detailed synopsis about the effectiveness of financial market regulation" in OpenAI's playground.

digipal

Collecting information about a topic, analyzing relevant data, drawing conclusions and finding explanatory factors are typical elements at the 'content planning' stage.
Subsequently, the boundaries for the realization phase can be defined.

We could seize control by phrasing the command as follows:

"Write a single sentence with the following information:
- Equities were up by 2.3%
- Bonds were down by 1.7%
- Commodities moved sideways"

The resulting sentence is not only grammatically correct, it also incorporates the figures provided, which could be variable input from the analytics steps. One needs to control for the aspect of randomness. When the same command is repeated, different versions may be returned.

Within such a structured framework, language models can perform very highly valuable tasks such as:
- Convert specifications to real text
- Combine relevant sentences to summarize a topic
- Translate text into other languages
- Adjust text for predefined word choices
- Transform the style of a text

Thoughtfully combining data analytics with language models like GPT can steer the text generation precisely towards the desired outcome, while harnessing the capacity of language models at different stages. This way the required accuracy, personalization and scalability can all be achieved.

"Set the boundaries for the text realization task."

▶ "Assets were mixed with equities up by 2.3% and bonds down by 1.7%, while commodities neither rose nor fell significantly."

"Break down the text generation process into smaller parts that are then handled by well-suited models."

"Thoughtfully combining data analytics with GPT can steer the text generation precisely towards the desired outcome."

digipal

KEY TAKE-AWAYS AND CONCLUSION

Source: Shutterstock.

As demonstrated with the examples in this article, when producing factual reports, one should take note of the following realities:

- Language models cannot reason or argue purposefully
- One should not assume that the generated content is true
- Every time a command is repeated a random version is produced with a new focus and a different storyline

There are ways to address these circumstances:

- Provide a skeleton for the structure and the main arguments

- Set the boundaries for the text realization task

- Break down the text generation process into smaller parts that are then handled by specialized models

By Daniel Rudis, PhD, Co-Founder and CEO of digipal AG. September 2022.

Pre-trained language models are tremendously powerful. In some practical applications, the challenge is to 'tame the beast'.

One can use the models' immense capabilities within the constraints imposed by an analytical framework and a well-designed document structure, to provide the right explanations for a predetermined conclusion.

This in turn enables significant efficiency gains, for example, in investment writing and content marketing.

Digipal is a pioneer in automating content in the financial industry.

investment content
generated automatically

Testimonial
"Thanks to Digipal, we are in a position to provide well-founded information about our portfolios at the beginning of the month. This increase in productivity and efficiency allows us to spend more time on our investment activities and client services."

To learn more about our product offering, please get in touch:

digipal AG
8004 Zurich
info@digipal.ai
+41 43 508 76 35
www.digipal.ai