Natural Language Programming with FinTech Applications

Version 1.0

HKU Fintech Group
HKU Business School
March 2024

- Artificial Intelligence: Natural Language Programming.
- Artificial Intelligence: Logical Thinking and Problem Solving.
- Application of Artificial Intelligence: Financial Technology.

The reports about <u>the Q* model breakthrough</u> that you all recently made, what's going on there?

最近大家关于Q*模型突破的报道，到底是怎么回事？

Altman: No particular comment on that unfortunate leak. But what we have been saying — two weeks ago, what we are saying today, what we've been saying a year ago, what we were saying earlier on — is that we expect progress in this technology to continue to be rapid, and also that we expect to continue to work very hard to figure out how to make it safe and beneficial. That's why we got up every day before. That's why we will get up every day in the future. I think we have been extraordinarily consistent on that.

Altman：对于这次不幸的泄密事件没有特别评论。但我们一直在说——两周前、今天、一年前、早些时候——我们预计这项技术将继续快速进步，而且我们希望继续努力找出如何使其安全和有益的方法。这就是我们之前每天起床的原因。这就是为什么我们以后每天都会起床。我认为我们在这一点上一直非常一致。

Without commenting on any specific thing or project or whatever, we believe that progress is research. You can always hit a wall, but we expect that progress will continue to be significant. And we want to engage with the world about that and figure out how to make this as good as we possibly can.

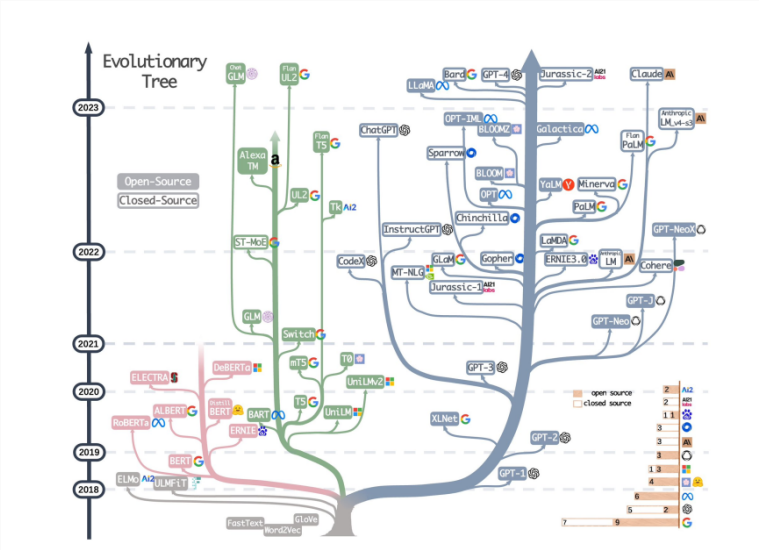在不评论任何具体事物或项目或其他任何东西的情况下，我们相信进步就是研究。你总是会碰壁。但我们预计进展将继续显着。我们希望与世界就此进行交流，并找出如何尽可能做到最好的方法。
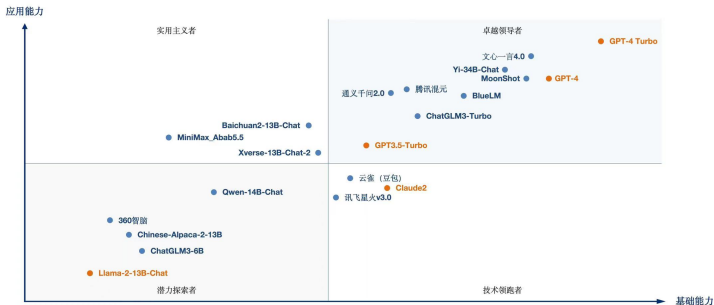
量子位

3

- 1. Natural Language Programming as a New Programming Approach.

Evolutionary Tree

- Dialogue Chatbot? Customer Service Tool? Chat Assistant? Question-Answering Bot?
- Efficient Writing Tool? Generates PPT, Word files?
- Latest perspective: Natural Language Programming (Explanation to follow).

- Source code written in a certain programming language.
- Source code compiled into binary machine language (compiler).
- Execution of the compiler (hardware).

```c
/*
 * This line basically imports the "stdio" header file, part of
 * the standard library. It provides input and output functionality
 * to the program.
 */
#include <stdio.h>

/*
 * Function (method) declaration. This outputs "Hello, world\n" to
 * standard output when invoked.
 */
void sayHello(void) {
    // printf() in C outputs the specified text (with optional
    // formatting options) when invoked.
    printf("Hello, world!\n");
}

/*
 * This is a "main function". The compiled program will run the code
 * defined here.
 */
int main(void)
{
    // Invoke the sayHello() function.
    sayHello();
    return 0;
}
```
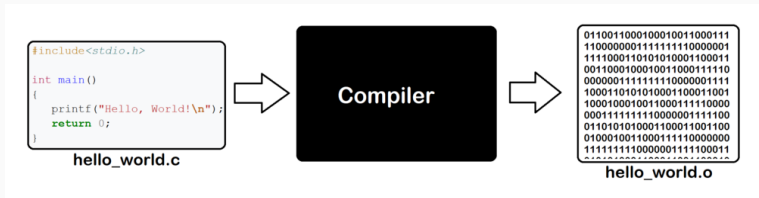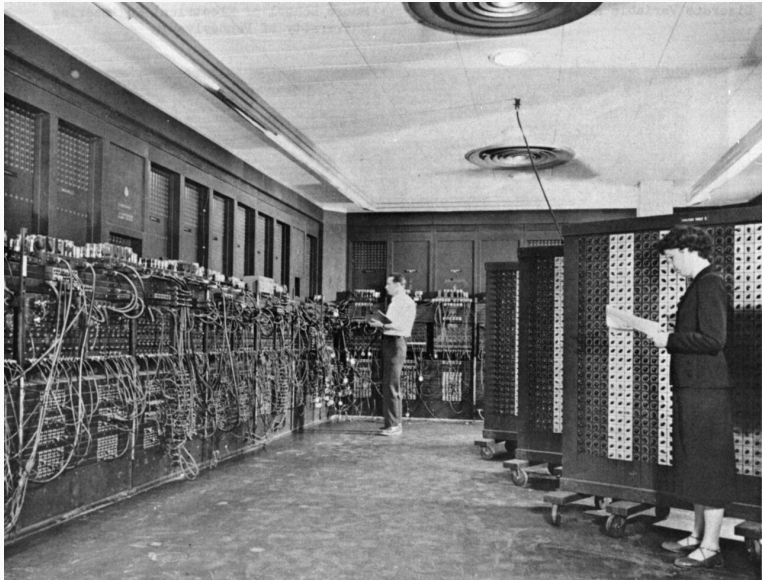
Evolution of Programming Languages

- Ancient Era (1940-1950): Programming in binary machine language (low programming efficiency, fast execution speed).
- Early Era (1960s): High-level programming languages like Fortran (1957), COBOL (1959), etc. Introduced compilers that could translate high-level languages into binary code.
- Mid Era (1970s): Pascal (1970), C (1972) - Dennis Ritchie, SQL (1978), etc.
- Late Era (1980-now): C++ (1980), Python (1990), R (1993), Java (1995), etc.
- Development Characteristics: The later programming languages have higher code efficiency but lower execution efficiency.
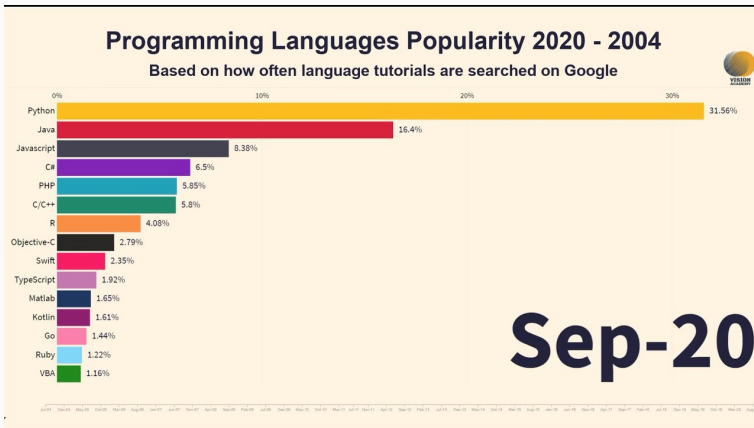
Disruption of Large Language Models: Natural Language Programming (Prompt)

- Similar to traditional programming, a prompt (code) based on large language models is used.
- This code is written in natural language, eliminating the need to be proficient in machine language.
- The code is compiled and executed by a compiler (large language model), producing an output.
- Characteristics: The code volume is significantly reduced (10-100 times less) compared to traditional programming like Python, and the execution efficiency is also greatly reduced. Particularly suitable for programming tasks related to natural language.

- In-context learning: to use examples for demonstration of tasks. In context learning is a useful trick to make LLM to align with your task.


```
Input: 2014-06-01
Output: !06!01!2014!
Input: 2007-12-13
Output: !12!13!2007!
Input: 2010-09-23
Output: !09!23!2010!
Input: 2005-07-23
Output: !07!23!2005!
```

in-context examples

test example

model completion

- Chain of Thoughts (CoT) is a trick to enhance the logical thinking ability of LLM. The key is to provide intermediate thinking steps to the LLM in prompt.

Standard Prompting

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The answer is 27. ✗

Chain-of-Thought Prompting

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. 5 + 6 = 11. The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had 23 - 20 = 3. They bought 6 more apples, so they have 3 + 6 = 9. The answer is 9. ✓

- Prompt:
  The odd numbers in this group add up to an even number: 4, 8, 9, 15, 12, 2, 1. A: Adding all the odd numbers (9, 15, 1) gives 25. The answer is False. The odd numbers in this group add up to an even number: 15, 32, 5, 13, 82, 7, 1. A:

- Output:
  Adding all the odd numbers (15, 5, 13, 7, 1) gives 41. The answer is False.

Prompt- 研报正文总结

· 作为一名资深的研究报告（研报）总结专家，您的任务是将研报内容精炼为 1 句话的摘要。这些句子将用于展示研报摘要，专注于 {field} 领域。请按照以下指导原则进行工作： 1. 精简总结：摘要仅包含 {sentence} 句话，且总字数不超过 50 字。 2. 数据重视：尽可能多地包含相关数据信息。 3. 投资者导向：内容应针对投资者，突出重要信息。接下来，我将提供一个例子，以帮助您更好地理解这项任务：

The first part: Provide Task Guidance

· 研报内容例子： {example_content}
· 对应的研报摘要例句： {example_summary}"

The second part: Use example as ICL

· 请注意，这些例子仅用于说明任务，并无实际内容意义。

The third part: to clarify and avoid confusion

在此基础上，您需要总结以下研报内容，形成一句关于 {field} 的摘要。请勿使用例子中的内容作为您的总结基础。如果无法从所给内容中提炼出关于 {field} 的摘要，也请勿使用示例内容进行总结。

需要总结的研报内容： {content}

Part 5: Extract Results as Output

Part 4: Plug in the data that we wish LLM to analyze.

18

- Compared to the over 99.99999999% accuracy of traditional compilers, the compilation accuracy (task execution correctness) of large language models like GPT-4 is much lower, possibly only around 90%, and even lower (around 50% or less) for many specific tasks.
- How to improve accuracy: Task decomposition, using natural language programming reasonably in combination with traditional machine language programming.
- Next 5-10 years: High-level language programming will evolve towards a combination of natural language and high-level language programming, eventually transitioning to pure natural language programming.
- Proficiency and practical experience are required to effectively use natural language programming. The barrier to programming may not decrease, but the efficiency of programming is greatly improved.

- 2. Artificial Intelligence: Logical Thinking and Problem Solving.

- Programming is a human-designed activity and a pre-programmed tool.
- Problem solving is an optimization process, which involves automatically finding the best solution. It often relies on mathematical algorithms and is implemented using computer languages.
- Problem solving is a special type of programming, where solvers can solve many complex problems that humans may not be able to solve quickly.

- The Q-function helps us solve problems, while the Q*-function helps us find the optimal solution to a problem.
- AlphaGo relies on the Q*-function, which is estimated through numerous Monte Carlo simulations, to solve specific problems.
- With Q*, we can find the optimal solution to a problem. However, can we estimate Q* in a general domain?

Despite recent advances, using AI to solve, or at least assist with solving, advanced mathematical problems remains an incredibly complicated and multifaceted challenge. It will be important to experiment with multiple approaches to this goal, and to benchmark the performance of each of them.

The AI-MO Prize promises to provide at least one such set of benchmarks which will help compare different AI problem solving strategies at a technical level, in a manner that will be accessible and appealing to the broader public.

Terence Tao
University of California, Los Angeles
Fields Medallist (2006)

陶哲轩的肯定👍

量子位

AI-MO Prize   About   Supporters   Get Involved

$10mn AI Mathematical Olympiad Prize Launches

XTX Markets is launching a new $10mn challenge fund, the **Artificial Intelligence Mathematical Olympiad Prize** (AI-MO Prize). The fund intends to spur the development of AI models that can reason mathematically, leading to the creation of a publicly-shared AI model capable of winning a gold medal in the International Mathematical Olympiad (IMO).

The grand prize of $5mn will be awarded to the first publicly-shared AI model to enter an AI-MO approved competition and perform at a standard equivalent to a gold medal in the IMO.

There will also be a series of progress prizes, totalling up to $5mn, for publicly-shared AI models that achieve key milestones towards the grand prize.

量子位

23

- 3. Fintech as an application to Natural Language Programming

Case Study: Automatic Generation of Research Report Abstracts using Natural Language Programming

- The financial industry produces a large number of research reports daily, covering macroeconomics, company fundamentals, investment recommendations, and more.
- Reading through an entire research report, often spanning several pages, is a time-consuming task.
- Can we use natural language programming to automatically generate precise summaries of the report's main content? Can we compare multiple reports horizontally? Can we analyze a company's past reports vertically?
- Similar work exists in reading scientific literature and generating abstracts, extracting and outlining novel or prose, and even completing primary school assignments.
- How can we refine and make the process more precise? Natural language programming.
- Demonstrating the effectiveness of the approach.
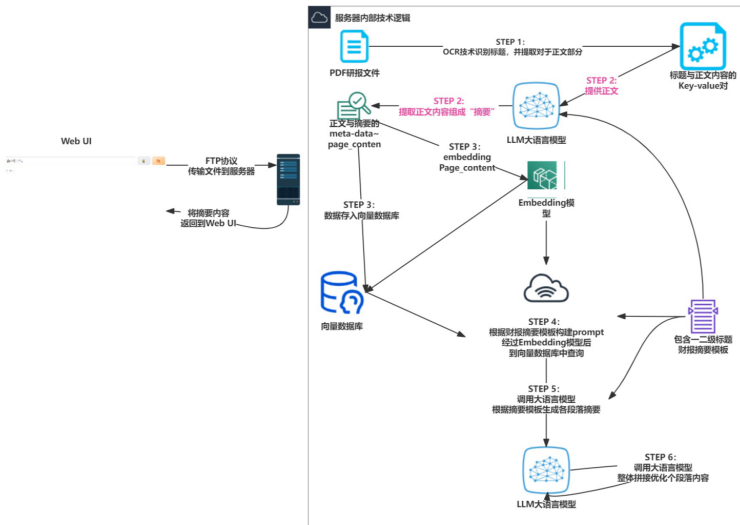
Prompt- 研报正文归类

· 从现在开始你是一个阅读和归纳助手，请帮忙判断某段内容属于以下类别中哪些部分。

· 你需要按照如下格式：

· Question: 美联储加息渐入末期有望催化金价。23 年 6 月美联储首次宣布暂停加息，23 年 7 月美国 CPI 同比上涨 3%。CME 预测年内或结束加息，明年存在降息预期。央行购金走高。22 年全球央行净购黄金 1081 吨，同比 +140%，利好金价上行。

· Output:" 该段落属于经济背景分析类别。"

· Question: 随着国内外局势愈发复杂、中美关系不确定性增强以及我国经济结构优化需求迫切，我国为加快国内电子测量行业快速发展推出一系列鼓励政策。8 月 1 日出版的第 15 期《求是》杂志发表中共中央总书记、国家主席、中央军委主席习近平的重要文章《加强基础研究实现高水平科技自立自强》。文章指出，建设基础研究高水平支撑平台。

· Output:" 该段落属于政治背景分析类别。"

· ......

· 回答不要超过 100 个字，回答的类别必须在如下各类别中：

· {second_label}

· 以上 Question 和 Output 的例子仅供格式参考，无实际的内容上的意义。接下来我会给你一段公司研究报告的段落，请帮忙梳理句子内容使其语义完整，按照上述格式给出这段内容属于哪些类别并总结。如果段落中有 "给予公司买入评级" 的字样，请归纳到投资建议类别中。用中文回答我。你要非常注意的是，请勿使用以上 Question 和 Output 的例子的内容进行总结。

· 好的，现在开始！

· Question: {Question}

Prompt- 研报正文总结

· 作为一名资深的研究报告（研报）总结专家，您的任务是将研报内容精炼为 1 句话的摘要。这些句子将用于展示研报摘要，专注于 {field} 领域。请按照以下指导原则进行工作： 1. 精简总结：摘要应仅包含 {sentence} 句话，且总字数不超过 50 字。 2. 数据重视：尽可能多地包含相关数据信息。 3. 投资者导向：内容应针对投资者，突出重要信息。接下来，我将提供一个例子，以帮助您更好地理解这项任务：

· 研报内容例子： {example_content}

· 对应的研报摘要例句： {example_summary}"

· 请注意，这些例子仅用于说明任务，并无实际内容意义。

· 在此基础上，您需要总结以下研报内容，形成一句关于 {field} 的摘要。请勿使用例子中的内容作为您的总结基础。如果无法从所给内容中提炼出关于 {field} 的摘要，也请勿使用示例内容进行总结。

· 需要总结的研报内容： {content}

Prompt- 段落优化

- 以下段落每句话是按照 {sub_topics} 顺序构成的，请你在不改变叙述顺序和大体意思的情况下，精简内容，去除掉一些冗余重复的叙述。比如这段话重复了相同的语义内容，精简这些语义内容，删减到只保留一处。比如这段话重复提到了一些相同的数据，检查这些重复的语句是否讲的是同一个内容，如果是，删减多余的内容，只保留一个。这很重要。需要处理的内容如下 :{content}

Thank You!