МЕТОДЫ СНИЖЕНИЯ ГАЛЛЮЦИНАЦИЙ В БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЯХ

Оганян Роберт Владимирович

Внутренние и внешние галлюцинации

Внутренние

- The first Ebola vaccine was approved in 2021

- The first vaccine for Ebola was approved by the FDA in 2019.

Внешние

- Китай уже начал клинические испытания вакцины против COVID-19» ?

Данные

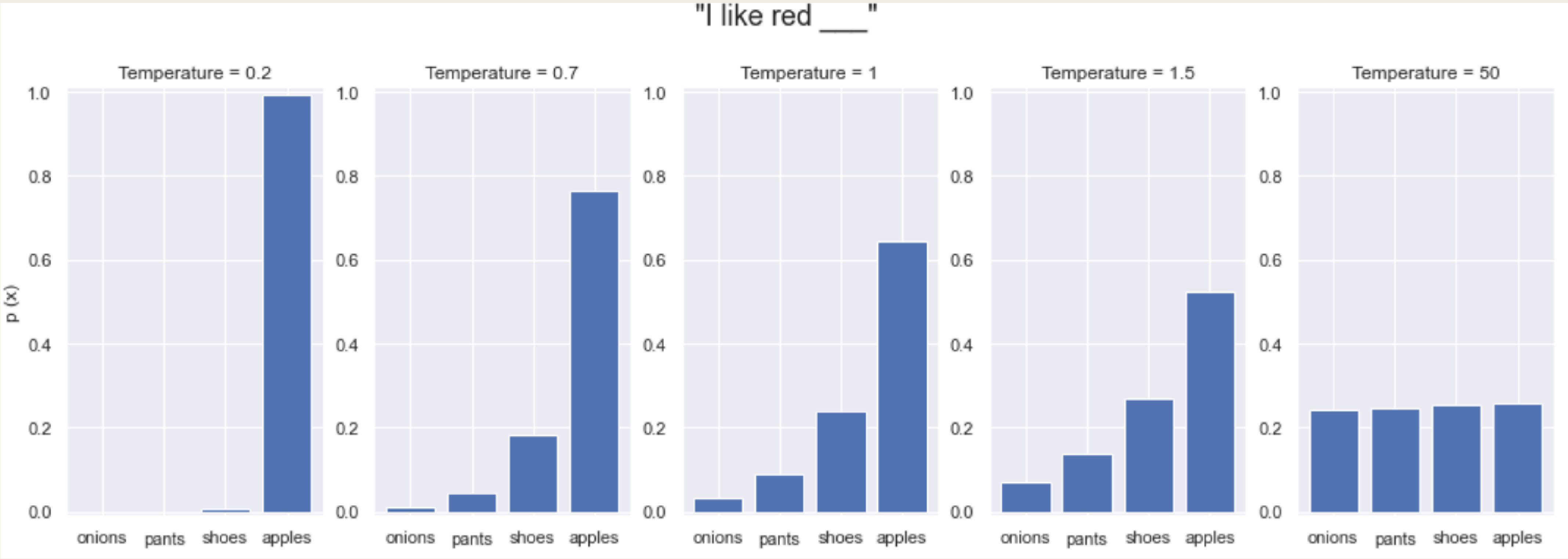
- Q: Translate the next paragraphs into Russian. Be as precise in terminology as possible. Everyone has the
  right to life, liberty and security of person.
  A (ground-truth):Каждый человек имеет право на жизнь, на свободу и на личную неприкосновенность.

- Q:Can you explain what is ology?
  A: Ology is the study of a particular subject or field of knowledge. It usually involves research and analysis to
  gain a deeper understanding of that subject

Изменение температуры

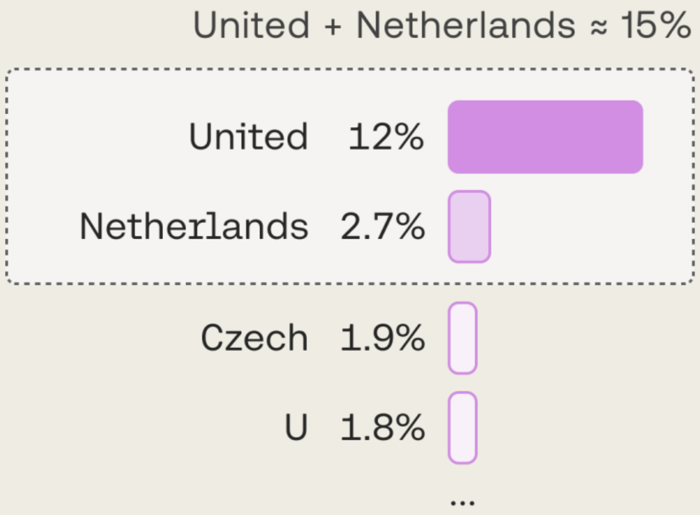
$$p_i' = \frac{\exp\left(\frac{p_i}{T}\right)}{\sum_j \exp\left(\frac{p_j}{T}\right)}$$

Изменение температуры

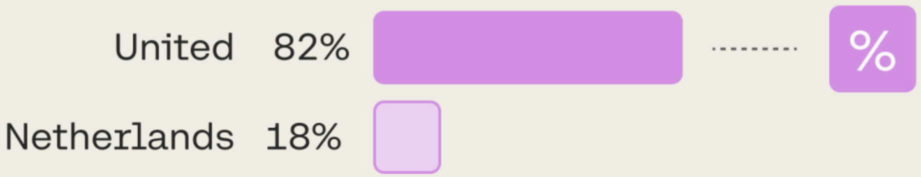


Factual-nucleus sampling

1. Consider only the top tokens whose likelihoods add up to 15%. Ignore all others.



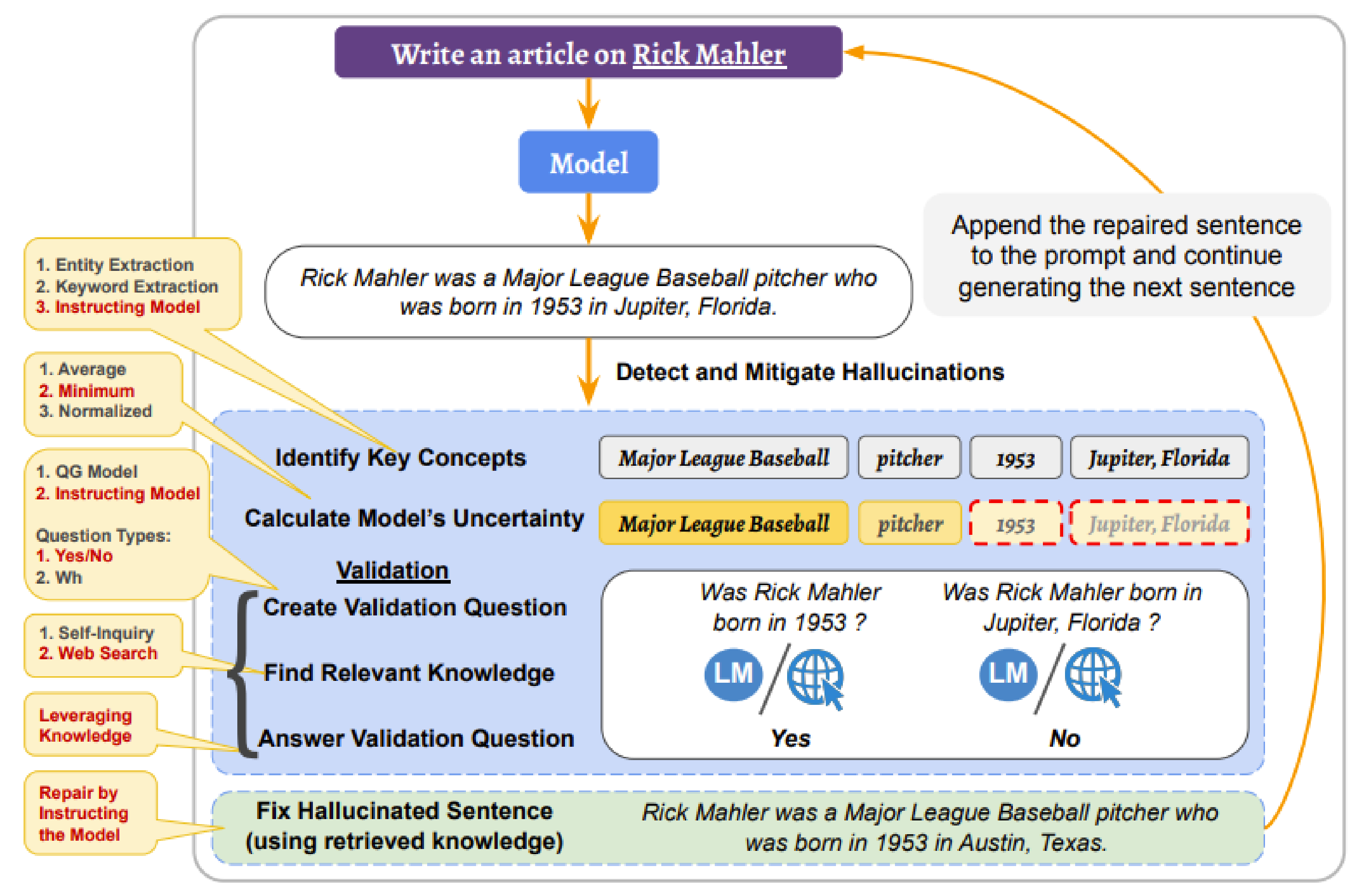
2. Sample from them based on their likelihood scores.



Factual-nucleus sampling

$$p_t = \max\{\omega, p \times \lambda^{t-1}\}$$

Detecting and Mitigating Hallucinations of LLMs by Validating Low-Confidence Generation



Rick Mahler was a Major League Baseball pitcher who was born in 1953 in Jupiter, Florida.

Append the repaired sentence to the prompt and continue generating the next sentence

Was Rick Mahler born in 1953 ?

Was Rick Mahler born in Jupiter, Florida ?

Rick Mahler was a Major League Baseball pitcher who was born in 1953 in Austin, Texas.

Запросы о значении терминов

- Шаг 1 - Написание промпта.
- Шаг 2 - Поиск релевантных знаний в интернете и создание соответствующего промпта

Запросы о значении терминов

Q: Can you explain what is billon?
A: Sure, a billion is a number equal to 1,000 million or 1,000,000,000 (one thousand million). It has 9 zeros after the initial digit...

Q: Here are some ground truth information about billon: Billon may refer to:
Billon (alloy), a metal alloy containing mostly copper or bronze with small quantity of silverBillon () is an alloy of a precious metal (most commonly silver, but also gold) with a majority base metal content (such as copper). It is used chiefly for making coins, medals, and token coins. The word comes from the French bille, which means "log". Any of several alloys of precious metals and base metals that are used (where legal) to make coins, medals etc. Using provided evidence above, Can you explain what is billon?
A: Certainly! According to the provided information, "billon" refers to an alloy used in the production of coins, medals, and token coins. This alloy is typically composed of a precious metal, most commonly silver but sometimes gold, with a majority base metal content such as copper. The purpose of using billon is to create a material that has some of the desirable properties of the precious metal while being more cost-effective due to the inclusion of a less expensive base metal.

Запросы о переводе

- Шаг 1 - Написание промпта.
- Шаг 2 - Выделение важных фраз (концепций).
- Шаг 3 - Поиск релевантных знаний в интернете и создание соответствующего промпта (а именно перевод ключевых фраз)

Запросы о переводе

Q: Translate the next paragraphs into Russian. Be as precise in terminology as possible. Everyone has the right to life, liberty and security of person
A: Каждый человек имеет право на жизнь, свободу и безопасность личности.

Q: Identify all the important keyphrases from the above sentence in english and return a comma separated list in quotes
A: right to life, liberty, security of person

Q: Here are some ground truth information about translations: right to life translates as право на жизнь; liberty translates as свобода; security of person translates as личную неприкосновенность. Based on provided evidence above, Translate the next paragraphs into Russian. Be as precise in terminology as possible. Everyone has the right to life, liberty and security of person
A: Каждый человек имеет право на жизнь, свободу и личную неприкосновенность

Результаты для терминов

	Верно	Не верно
GigaChat	14	6
ChatGPT	19	1
Claude	19	1
llama	6	14
GigaChat_stitch_in_time	14	6
ChatGPT__stitch_in_time	20	0
Claude_stitch_in_time	20	0
llama 3B_temperature_sampling	4	16
llama 3B_factual_nucleus_sampling	5	15
ChatGPT-temperature_sampling	20	0
ChatGpt-factual-nucleus_sampling	20	0

Результаты для перевода

	Верно	Не верно
GigaChat	7	12
ChatGPT	12	7
Claude	17	2
GigaChat_stitch_in_time	7	12
ChatGPT__stitch_in_time	19	0
Claude_stitch_in_time	19	0
ChatGPT-temperature_sampling	12	7
ChatGpt-factual-nucleus_sampling	9	10

Q: Can you explain what is ology?

llama: Ology is a word that is used to describe a particular type of blog. It is a portmanteau of "blog" and "ology".

Llama(temperature):The Wikipedia article is very good about explaining ology (link below):
*A ology (/ˈɑːlˌwɪdʒ/ or /ˈˌoʊlədʒ/) is a branch of study within a particular science or field of knowledge, or other type of science.
*In this article ology and scientific field are treated synonymously. A field of study is not a synonym of a branch of science; nor is a branch a synonym of an ology…

Llama (factual-nucleus): Ology is a word that means “the study of.” It is a noun that is used to describe a field of study

Q: Can you explain what is billon?

llama: The term billon is used to describe a coin that was struck in a small quantity. The term is used to describe coins struck in the 1st century BC and 1st century AD

Llama(temperature):(1) Billon is silver/base metal that was alloyed with a small amount of copper. It was considered to be of the finest silver for use in coins. (2) For much of ancient history the coins were lost until much later, when they discovered that silver/base metal that was of good color and appearance and had a distinctive shape was of more economic value to the owner.

Llama (factual-nucleus): Ology is a word that means “the study of.” It is a noun that is used to describe a field of study

Q: "Translate the next paragraphs into Russian. Be as precise in terminology as possible.Everyone has the right to an effective remedy by the competent national tribunals for acts violating the fundamental rights granted him by the constitution or by law."

Chatgpt_stitch_in_time: Каждому человеку принадлежит право на эффективное средство защиты перед компетентными национальными судами в случае совершения действий, нарушающих его основные права, предоставленные ему Конституцией или законом.

Q: Identify all the important keyphrases from the above sentence in english and return a comma separated list in quotes

Chatgpt_stitch_in_time: "right to an effective remedy", "competent national tribunals", "acts violating the fundamental rights"

Q: Here are some ground truth information about translations: right to an effective remedy translates as право на эффективную правовую защиту \ права на эффективную правовую защиту; Possible contexts for right to an effective remedy: It might be useful for the Committee to understand how the right to an effective remedy functioned in the case law of other treaty bodies.==Комитету было бы целесообразно разобраться, каким образом право на эффективную правовую защиту действует в прецедентном праве других договорных органов.

acts violating the fundamental rights translates as случаях нарушения его основных прав; Possible contexts for acts violating the fundamental rights: Everyone has the right to an effective remedy by the competent national tribunals for acts violating the fundamental rights granted him by the constitution or by law.==Каждый человек имеет право на эффективное восстановление в правах компетентными национальными судами в случаях нарушения его основных прав, предоставленных ему конституцией или законом..

Based on provided evidence above, Translate the next paragraphs into Russian. Be as precise in terminology as possible. Everyone has the right to an effective remedy by the competent national tribunals for acts violating the fundamental rights granted him by the constitution or by law.

Chatgpt_stitch_in_time : Каждому принадлежит право на эффективное восстановление в правах компетентными национальными судами в случаях нарушения его основных прав, предоставленных ему конституцией или законом.

Q: Translate…Everyone is entitled to a social and international order in which the rights and freedoms set forth in this Declaration can be fully realized.

ChatGPT: Каждый имеет право на социальный и международный порядок, в котором права и свободы, изложенные в этой Декларации, могут быть полностью реализованы.

ChatGPT_temperature: Каждому гарантировано социальное и международное устройство, в рамках которого права и свободы, изложенные в этой Декларации, могут быть полностью реализованы.

ChatGPT_factual_nucleus: Каждому человеку предоставляется социальный и международный порядок,в рамках которого права и свободы,изложенные в данной Декларации, могут быть полностью осуществлены.

Конец

- Спасибо за внимание