

Large Language Models for Isotope Separation

Logan Endes, Department of Computer Science, Rochester Institute of Technology, Rochester, NY, 14623

Carlos Soto, Computer Science Initiative, Brookhaven National Laboratory, Upton, NY, 11973

Jasmine Hatcher-Lamarre, Collider-Accelerator Department, Brookhaven National Laboratory, Upton, NY, 11973

Abstract

The Brookhaven National Laboratory Medical Isotope Research and Production group researches and produces experimental isotopes that are used in nuclear medicine using the Brookhaven Linac Isotope Producer. Before producing isotopes, scientists must go through the process of analyzing research papers to identify ideal separation conditions of these isotopes. This study examines the use of open-source, frozen, large language models (LLMs), such as Galactica, MPT, and LLaMA-2 to extract information on separation conditions of isotopes from scientific literature using prompt tuning and prompt engineering. To test these models, we will be engineering prompts with different techniques to find which are most effective at extracting information from research papers. We found the best approach is Chain of Thought prompting with LLaMA-2-13b. This will support research of automated chemical extraction using artificial intelligence. As a result of this summer, I have developed research skills and have a grander understanding of prompt tuning LLMs.

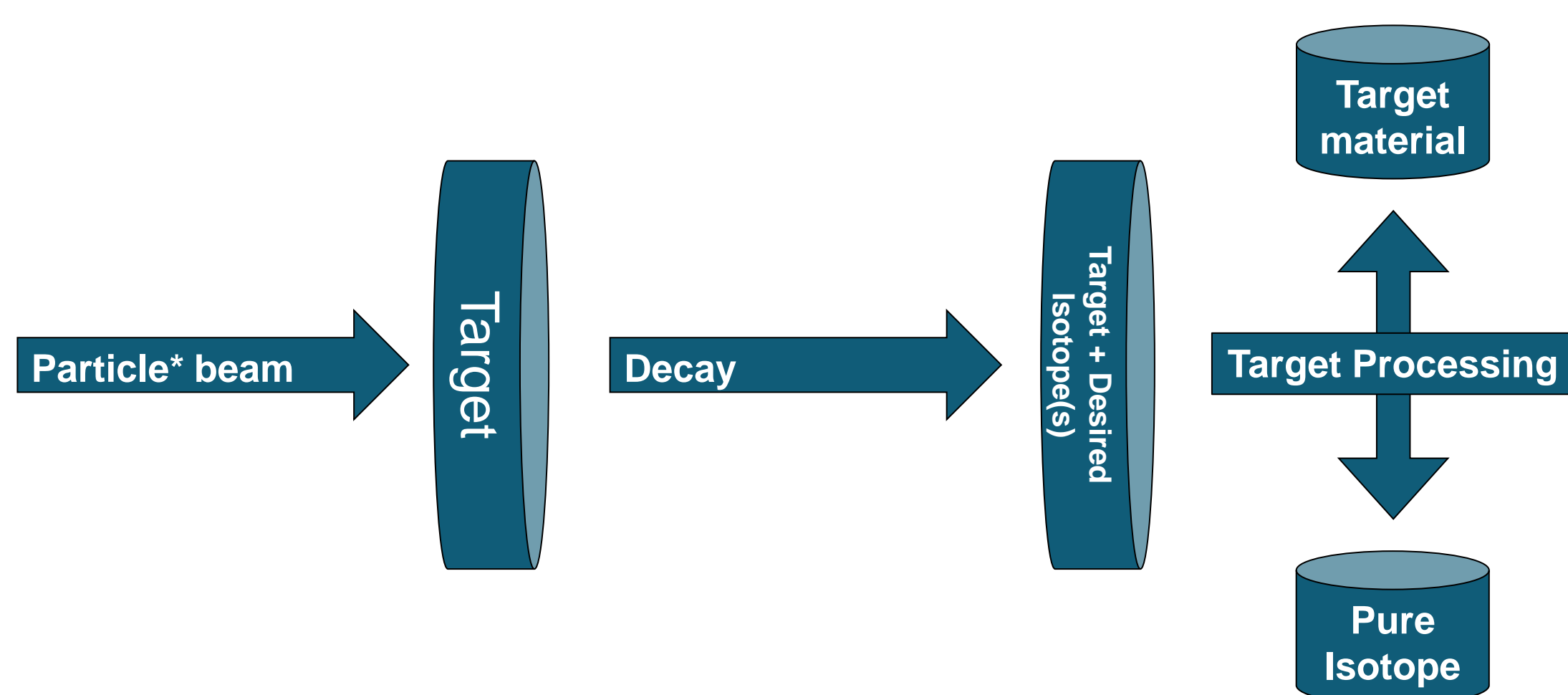
Introduction / Motivation

- BNL produces **Actinium-225** and other isotopes that are used for **imaging and treatment** of diseases
- Production of isotopes is a **time-consuming** process where much of the time is spent identifying **chemical separation** methodologies from previously published scientific literature

The use of a **large language model** to interact with text from research papers may **accelerate research** by extracting critical information for **chemical separations**.

Related Work

- "**Attention Is All You Need**"¹ proposed the **transformer** architecture of machine learning models, which caused a boom in the advancement of **large language models**
- Large language models (LLMs)** are machine learning algorithms designed to process text in a word-by-word generative mode
 - At large scales, LLMs demonstrate impressive information extraction and reasoning capabilities



Approach

LLMs are used as fixed ML models, with "learned" capabilities emerging through prompt manipulation and engineering

Of the many prompt engineering strategies, the following were selected:

- Zero-Shot Prompting**
- Few-Shot Prompting**
- Chain-of-Thought (CoT) Prompting²**
- Automatic Prompt Engineer³**

In order to determine the best language model, we tested 8 different models with varying parameters and configurations in tandem with text extracted from papers with different filtering techniques.

Name	Publisher	Parameters	Quantized?	Scientific?
Galactica-1.3b	Meta AI	1.3 billion	No	83%
Galactica-6.7b	Meta AI	6.7 billion	Yes	83%
LLaMA-2-7b	Meta AI	7 billion	Yes	--%
LLaMA-2-13b	Meta AI	13 billion	Yes	--%
MPT-7b-Instruct*	MosaicML	7 billion	Yes	5.2%

Table 1 depicts relevant models to our problem, their publisher, the parameters in the model, whether the model is quantized, and the percentage of training data which was scientific literature. LLaMA-2 did not report its data training corpus.

Results / Insights

After working with an existing **automatic prompt engineer²**, we decided that it would be best to follow other avenues before revisiting it. Ultimately, we concluded from research that **CoT** prompting would be best for extracting information.

Through testing, we concluded the following based on a benchmark of conclusions drawn by a human:

- Galactica-1.3b⁴ identified 37% with 77 generations per minute
- Galactica-6.7b identified 35% with 7 generations per minute
- LLaMA-2-7b⁵ identified 53% with 10 generations per minute
- LLaMA-2-13b identified 69% with 8 generations per minute

(These results come from generations made by the models on keyword-filtered data from 2 different radioisotope research papers. I went through these papers, and extracted important paragraphs with key features and drew conclusions based on those. The score is how accurate the model was compared to my conclusions).

Anion-exchange chromatography also isolates cobalt-5x from iron target material [94]. After dissolving the target in acid, the iron is oxidized into the 3+ state with hydrogen peroxide. Instead of loading the cobalt-5x on to the anion-exchange resin and washing away the target material, as was performed in the nickel/cobalt anion-exchange separation, cobalt and iron are retained on the column when loaded. The cobalt can then be eluted from the anion-exchange column in 4 M HCl. The iron target will remain on the column during the 4 M HCl wash, but can be recovered in low concentrations of HCl. The HCl eluant containing cobalt-5x can also be evaporated and the cobalt-5x reconstituted in alternative matrices without incurring significant losses [26,28,86,87].	Question: What is the target material in the above reaction? Answer: 59Fe, or Iron-59, must be the target. Score: 23.69
	Question: What acid is the target material dissolved in during the above reaction? Answer: 4 M HCl is the acid used to dissolve the target. Score: 22.98
	Question: What acid is used in the elution during the above reaction? Answer: 4 M HCl is used in the elution during the above reaction. Score: 30.67
	Question: What are the products of the above reaction? Answer: 59Fe and 59Co, or Cobalt-59, must be the products. Score: 26.66

Figure 2 shows human annotations (left) vs. AI generations (right) on a paragraph from "Production, Purification, and Applications of a Potential Theranostic Pair: Cobalt-55 and Cobalt-58m"⁶

Acknowledgements

This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI).

I would like to especially thank my mentor, Carlos Soto, and my PI, Jasmine Hatcher-Lamarre, for the great experience this summer, all the help, and the great insights that they have provided me.

Not export controlled.

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. ✅

Figure 1 shows Standard Prompting (left) vs. Chain-of-Thought prompting (right). Standard prompting gives an example (few-shot) without much reasoning whereas Chain-of-Thought embeds reasoning in examples given.