Report on the First LiPS Conference

H Ryott Glayzer

February 2024

Lab Assistant,
Schnee Group
SD Mines

LiPS:

The first Large Language Models in Physics Symposium, held at DESY in Hamburg, Germany.

Basis for LiPS

Large Language Models have shown impressive capabilities in recent times, regarding text and code generation. LiPS discussed a number of different topics in LLM research, including:

Basis for LiPS

Large Language Models have shown impressive capabilities in recent times, regarding text and code generation. LiPS discussed a number of different topics in LLM research, including:

- Direct use of LLMs in research

Basis for LiPS

Large Language Models have shown impressive capabilities in recent times, regarding text and code generation. LiPS discussed a number of different topics in LLM research, including:

- Direct use of LLMs in research
- Development of LLMs for fundamental physics

Basis for LiPS

Large Language Models have shown impressive capabilities in recent times, regarding text and code generation. LiPS discussed a number of different topics in LLM research, including:

- Direct use of LLMs in research
- Development of LLMs for fundamental physics
- Development of LLM techniques for physics

Basis for LiPS

Large Language Models have shown impressive capabilities in recent times, regarding text and code generation. LiPS discussed a number of different topics in LLM research, including:

- Direct use of LLMs in research

- Development of LLMs for fundamental physics

- Development of LLM techniques for physics

- AI ethics and reliability

Basis for LiPS

Large Language Models have shown impressive capabilities in recent times, regarding text and code generation. LiPS discussed a number of different topics in LLM research, including:

- Direct use of LLMs in research
- Development of LLMs for fundamental physics
- Development of LLM techniques for physics
- AI ethics and reliability
- Implicit LLM 'understanding' of physics

Basis for LiPS

Large Language Models have shown impressive capabilities in recent times, regarding text and code generation. LiPS discussed a number of different topics in LLM research, including:

- Direct use of LLMs in research

- Development of LLMs for fundamental physics

- Development of LLM techniques for physics

- AI ethics and reliability

- Implicit LLM 'understanding' of physics

In this presentation, we will go over a few topics I feel that the group will find interesting.

- Older Research $\rightarrow$ PDF Scans

Extracting Measurements from Legacy Publications

- Older Research $\rightarrow$ PDF Scans
- PDFs can be hard to parse

Extracting Measurements from Legacy Publications

- Older Research $\rightarrow$ PDF Scans

- PDFs can be hard to parse

- PDFs suck as a format.

Extracting Measurements from Legacy Publications

- Older Research $\rightarrow$ PDF Scans

- PDFs can be hard to parse

- PDFs suck as a format.

- OCR used to parse PDF

Extracting Measurements from Legacy Publications

- Older Research $\rightarrow$ PDF Scans

- PDFs can be hard to parse

- PDFs suck as a format.

- OCR used to parse PDF

- Retrieval Augmented Generation is used to make all this information available to LLMs

Extracting Measurements from Legacy Publications

- Older Research $\rightarrow$ PDF Scans

- PDFs can be hard to parse

- PDFs suck as a format.

- OCR used to parse PDF

- Retrieval Augmented Generation is used to make all this information available to LLMs

- RAG allows LLMs to access extra information and avoid hallucinations
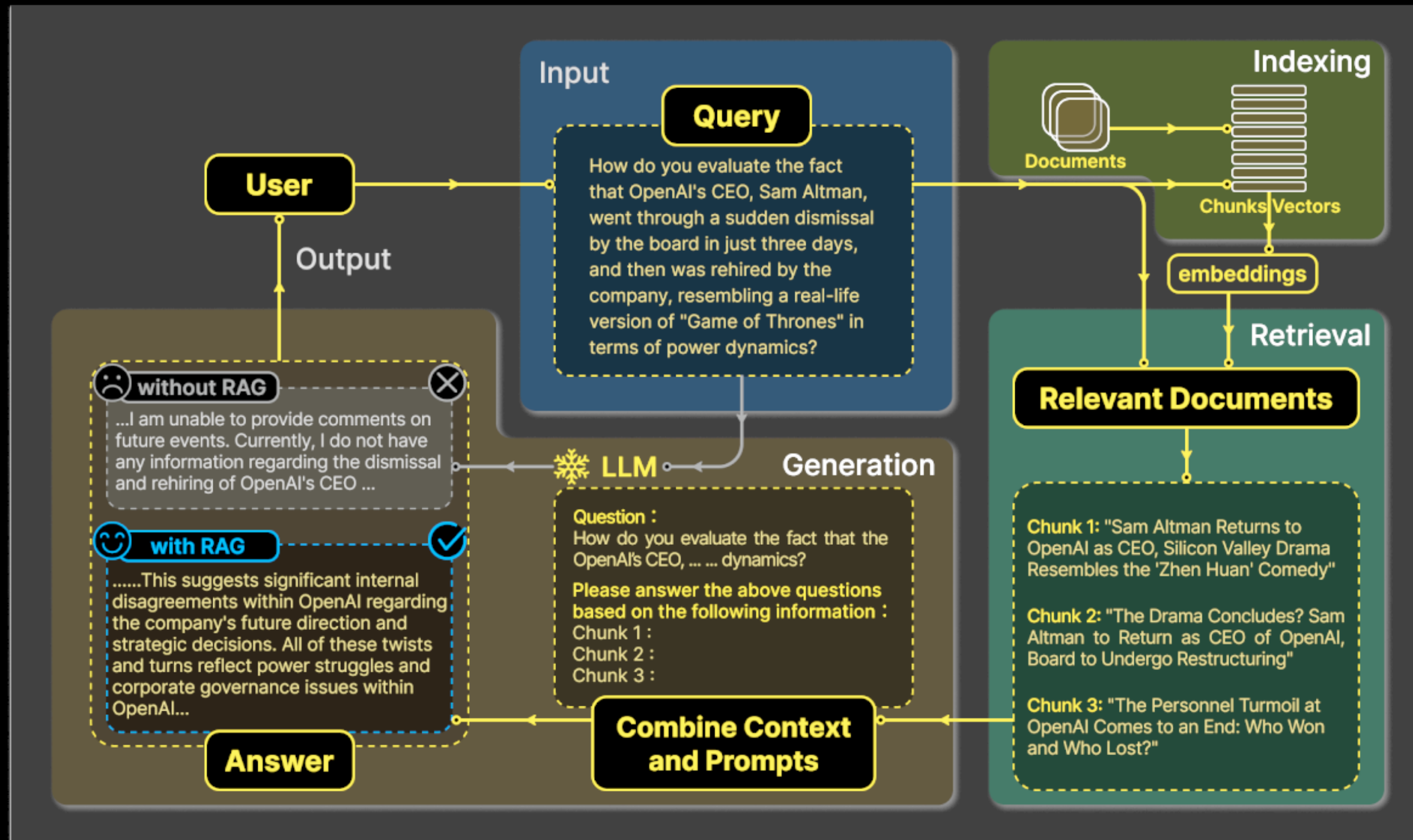
Retrieval Augmented Generation (RAG) [Lew+20]

Figure taken from [Gao+24]

Implications of this Talk

- Allows LLMs to avoid hallucinations

Implications of this Talk

- Allows LLMs to avoid hallucinations

- Allows LLMs to be more specialized

Implications of this Talk

- Allows LLMs to avoid hallucinations

- Allows LLMs to be more specialized

- Lets LLM assistants access information specific to a group

Implications of this Talk

- Allows LLMs to avoid hallucinations

- Allows LLMs to be more specialized

- Lets LLM assistants access information specific to a group

- e.g. an AI assistant for a group to access information scanned in experiment notes

ChATLAS: an AI Assistant for the ATLAS Collaboration

ATLAS Collaborators created an AI assistant using RAG to access information stored in their many different wikis, code docs, gitlab repos, meeting agendas, publications, and other information spots.

ChATLAS: an AI Assistant for the ATLAS Collaboration

ATLAS Collaborators created an AI assistant using RAG to access information stored in their many different wikis, code docs, gitlab repos, meeting agendas, publications, and other information spots.

- This has allowed them to access and organize a lot of information that would take years for grad and undergrad students to organize, allowing young scientists to participate more readily in research.

ChATLAS: an AI Assistant for the ATLAS Collaboration

ATLAS Collaborators created an AI assistant using RAG to access information stored in their many different wikis, code docs, gitlab repos, meeting agendas, publications, and other information spots.

- This has allowed them to access and organize a lot of information that would take years for grad and undergrad students to organize, allowing young scientists to participate more readily in research.

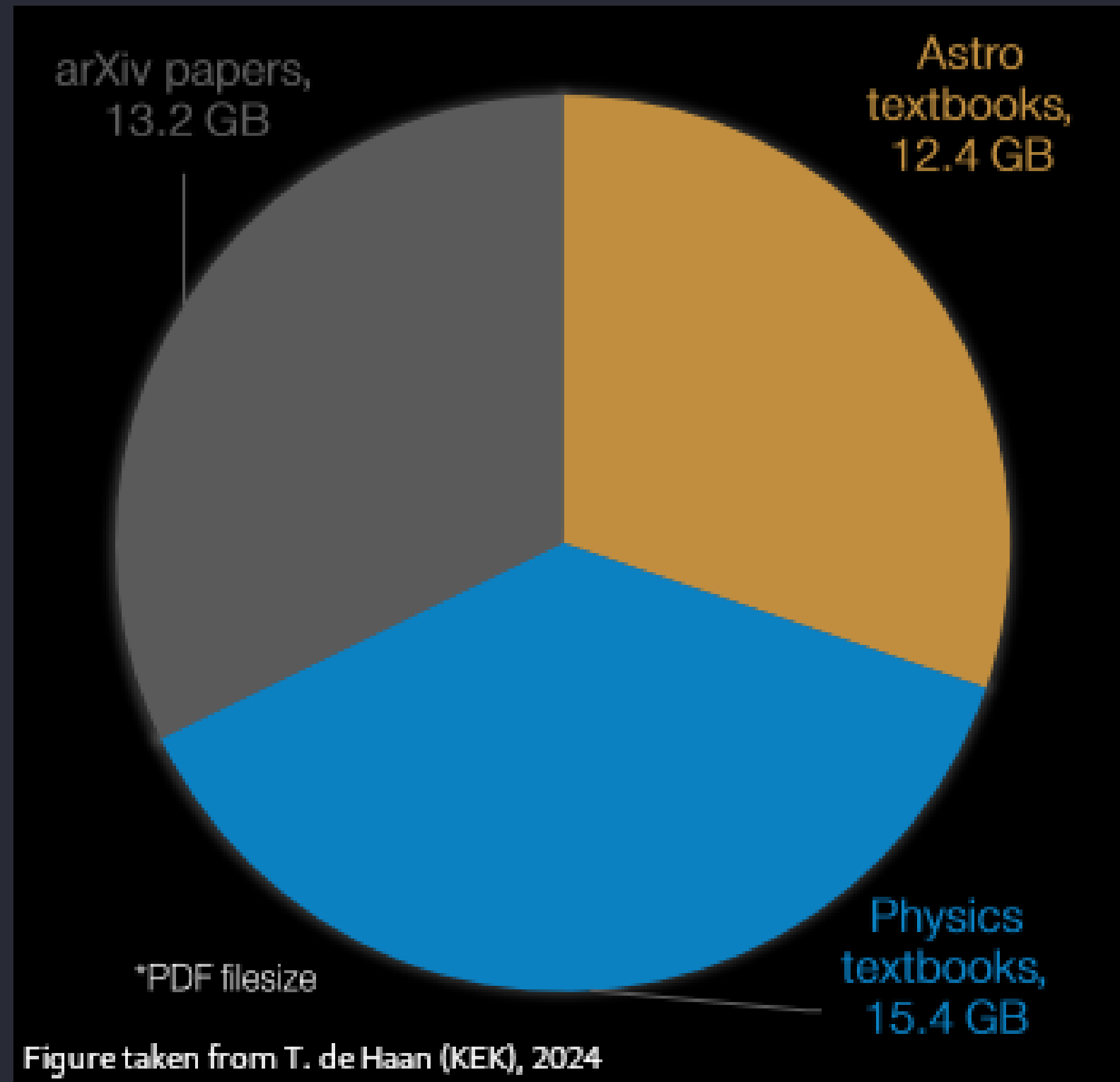- This AI assistant is hosted on their own servers.

AstroLLaMA: An AI trained on Astrophysics Abstracts

AstroLLaMA is a 7B parameter LLM trained on Astronomy and Physics abstracts. It is able to answer more specialized questions in astrophysics without being super vague or hallucinating, unlike ChatGPT and other generalized LLMs.

You can play around with it here: AstroLLaMa on the Hugging Face

Cosmosage: an AI trained on Cosmology Papers

arXiv papers, 13.2 GB

Astro textbooks, 12.4 GB

Physics textbooks, 15.4 GB

*PDF filesize

Figure taken from T. de Haan (KEK), 2024

Cosmosage is a general-purpose assistant specialized in answering questions about cosmology. This project is fully open-sourced, and many improvements are planned. It is a 7B LLM Fine-Tuned on Cosmology Papers and textbooks, based on the Mistral-7B model, trained on 8k papers from the ArXiV.

Cosmosage on Helmholtz Blabador Server

I asked both ChatGPT and AstroLLaMA to generate an abstract for the silicone sealants analysis, specifically to compare the models.

Prompt:

Write an abstract for presentation at an undergraduate student research symposium, focusing on undergraduate analysis of data collected on the emanation of Radon-222 from two commercially available silicone sealants. The background of the measurements are to minimize backgrounds for the SuperCDMS dark matter experiment. The abstract should be no more than 300 words, and should be approachable for both specialists and non-specialists. There are two sealants involved, both of which have an emanation rate less than 0.01 milliBecquerel, with uncertainties on the order of 0.02 millibecquerel. Both sealant samples had a surface area of about 200 sq. cm and an average width of 0.7 cm. According to Bowles (2019), silicone has a Radon diffusion length of 2.6 +/- 0.1 mm.

Abstract generated by ChatGPT 3.5

ChatGPT 3.5

The emission of Radon-222 (222Rn) from materials is a significant concern in experiments aiming to detect dark matter, such as the Super Cryogenic Dark Matter Search (SuperCDMS). In this study, we investigate the emanation of 222Rn from two commercially available silicone sealants commonly used in experimental setups, with the objective of minimizing background signals in the SuperCDMS experiment. The two sealants under examination exhibited emanation rates below 0.01 milliBecquerel, with uncertainties on the order of 0.02 milliBecquerel. Each sealant sample, with a surface area of approximately 200 sq. cm and an average width of 0.7 cm, was analyzed to determine its 222Rn emanation characteristics.

Based on the work by Bowles (2019), which provides a diffusion length for Radon in silicone of 2.6 +/- 0.1 mm, we devised an experimental setup to measure the 222Rn emanation. Through careful data collection and analysis, we quantified the emanation rates and characterized the uncertainty associated with each sealant sample.

Our findings indicate that both sealants emit 222Rn at rates below the specified threshold for background radiation in the SuperCDMS experiment. The obtained emanation rates, coupled with the known diffusion properties of Radon in silicone, contribute valuable data for understanding and mitigating background signals in dark matter detection experiments. This research underscores the importance of thorough characterization of materials used in experimental setups to ensure the sensitivity and accuracy of sensitive scientific instruments.

Abstract Generated by AstroLLaMA

AstroLLaMA

In the quest for ultra-low background levels necessary for direct searches of weakly interacting massive particles (WIMPs), it becomes imperative to evaluate the potential sources of background noise in the experimental setup. Silicone sealants are commonly employed in many applications ranging from electronics manufacturing to medical devices, yet little attention had been paid towards evaluating their suitability for use in particle detection systems until now. This research presents measurements performed to quantify the amount of radon-222 emitted from commercially available silicone sealants, specifically those intended for use in the SuperCDMS detector system. Results indicate that the sampled products emit negligible amounts of radon-222, thus meeting stringent standards established for low background operation. These findings underscore the need for careful material screening measures early in the design phase of particle detection systems, particularly ones aimed at probing extremely rare interactions.

Cosmosage_v2

The purpose of this study is to measure the emanation of Radon-222 from two commercially available silicone sealants, which are used in the Super Cryogenic Dark Matter Search (SuperCDMS) experiment. The emanation of radioactive elements from materials used in the experiment's setup can significantly impact the experiment's performance. In this study, we analyze the emanation of Radon-222 from the two silicone sealant samples, Sealant A and Sealant B, using a high-precision alpha spectrometer. Our results provide valuable information on the emanation of Radon-222 from these silicone sealants, which can help optimize the SuperCDMS experiment and other dark matter searches.

Mistral-7B

Mistral is a general-purpose 7B-Parameter LLM used as a control.

Mistral-7B-Instruct-v0.2

The Super Cryogenic Dark Matter Search (SuperCDMS) is an ongoing experiment aimed at detecting weakly interacting dark matter particles using cryogenic detectors. One of the main challenges for this experiment is minimizing the background radiation, particularly from radon-222 (Rn-222), which has a half-life of 3.8 days and a beta decay that can mimic the signal of dark matter particles. In this study, we measured the Rn-222 emanation rates from two commercially available silicone sealants, which are commonly used in the SuperCDMS experiment.

Two samples of each sealant, with surface areas of approximately 200 sq. cm and an average width of 0.7 cm, were sealed in custom-designed containers and left to equilibrium for over a month. The Rn-222 activity was measured using a high-purity germanium detector, and the data were analyzed to determine the emanation rates and uncertainties. Both sealants had Rn-222 emanation rates below 0.01 milliBecquerel, with uncertainties on the order of 0.02 millibecquerel.

According to Bowles (2019), silicone has a Radon diffusion length of 2.6 $\pm$ 0.1 mm. To assess the impact of the measured sealant emanation rates on the SuperCDMS experiment, we calculated the expected activity in the detector based on the sealant surface area and the diffusion length. Our results indicate that even with the highest measured emanation rate, the contribution of Rn-222 to the background is below the level that can be readily distinguished from the dark matter signal in the SuperCDMS experiment.

This study provides valuable information for the selection of low-background materials for the SuperCDMS experiment and demonstrates the capabilities of undergraduate students in conducting research that contributes to ongoing experiments in the field of particle physics. Further investigations into the relationship between sealant properties and Rn-222 emanation rates could lead to improved sealing materials with even lower backgrounds for dark matter detectors.

A local LLaMA^2 7B model on Bison

The following is the response from Simon, a local LLaMA^2 7B model I have trained, running on Bison.

A local LLaMA^2 7B model on Bison

The following is the response from Simon, a local LLaMA^2 7B model I have trained, running on Bison.

Simon

Abstract: The emanation of Radon-222 from commercially available silicone sealants is an important consideration in the design and operation of the SuperCDMS dark matter experiment. In this study, we present the results of undergraduate analysis of data collected on the emanation of Radon-222 from two commercially available silicone sealants. The purpose of these measurements is to minimize backgrounds for the SuperCDMS dark matter experiment.

The two sealant samples used in this study had an emanation rate less than 0.01 milliBecquerel, with uncertainties on the order of 0.02 millibecquerel. Both sealant samples had a surface area of about 200 sq. cm and an average width of 0.7 cm. According to Bowles (2019), silicone has a Radon diffusion length of 2.6 $\pm$ 0.1 mm, which is consistent with the results of this study.

The emanation rates from these two sealants were measured using a Geiger counter and a scintillator detector. The data collected was analyzed using a statistical analysis software to determine the background levels for each sealant sample. The results showedthat the background levels for both samples were significantly lower than the expected background levels based on the emanation rates of Radon-222 from silicone.

The findings of this study have important implications for the design and operation of the SuperCDMS dark matter experiment. By minimizing the background levels, we can improve the sensitivity and accuracy of the detector and increase the chances of detecting dark matter particles. The results of this study demonstrate the feasibility of using commercially available silicone sealants for the emanation of Radon-222 in the SuperCDMS experiment.

References: Bowles, J. (2019). The properties of silicone and its applications. Journal of Materials Science, 54(1), 1-13.

Comparison of Models

- GPT 3.5 is a 175B-parameter model, while the rest are only 7B models.

Comparison of Models

- GPT 3.5 is a 175B-parameter model, while the rest are only 7B models.

- The prominence of hallucination prevalent in the 7B models aligns with the usage of fine-tuning rather than RAGs.

Comparison of Models

- GPT 3.5 is a 175B-parameter model, while the rest are only 7B models.

- The prominence of hallucination prevalent in the 7B models aligns with the usage of fine-tuning rather than RAGs.

  - Incorrect or hallucinated answers are often closer to the training data than an answer like "I don't know"

Comparison of Models

- GPT 3.5 is a 175B-parameter model, while the rest are only 7B models.

- The prominence of hallucination prevalent in the 7B models aligns with the usage of fine-tuning rather than RAGs.

  - Incorrect or hallucinated answers are often closer to the training data than an answer like "I don't know"

- Cosmosage is working on a RAG pipeline which will be deployed once it is worked out.

Comparison of Models

- GPT 3.5 is a 175B-parameter model, while the rest are only 7B models.
- The prominence of hallucination prevalent in the 7B models aligns with the usage of fine-tuning rather than RAGs.
    - Incorrect or hallucinated answers are often closer to the training data than an answer like "I don't know"
- Cosmosage is working on a RAG pipeline which will be deployed once it is worked out.
- Simon, a general-purpose LLM I trained with 7B general parameters exhibits a great deal of hallucination, showcasing the specialization and fine-tuning of other models shown here

Comparison of Models

- GPT 3.5 is a 175B-parameter model, while the rest are only 7B models.
- The prominence of hallucination prevalent in the 7B models aligns with the usage of fine-tuning rather than RAGs.
  - Incorrect or hallucinated answers are often closer to the training data than an answer like "I don't know"
- Cosmosage is working on a RAG pipeline which will be deployed once it is worked out.
- Simon, a general-purpose LLM I trained with 7B general parameters exhibits a great deal of hallucination, showcasing the specialization and fine-tuning of other models shown here
- These models show promising progress for specialization of small models.

Astro-COLIBRI and NIMBUS

Astro-COLIBRI and NIMBUS are LLM-integrated suites for multi messenger astronomy, ie: astronomy & astroparticle physics. They allow both researchers and the general public to access information regarding detected events.

Astro-COLIBRI and NIMBUS

Astro-COLIBRI and NIMBUS are LLM-integrated suites for multi messenger astronomy, ie: astronomy & astroparticle physics. They allow both researchers and the general public to access information regarding detected events.

The LLMs employed look at the data that is available, and generate text to make it human-readable. These are both publicly available and we will take a look at them:

Astro-COLIBRI and NIMBUS

Astro-COLIBRI and NIMBUS are LLM-integrated suites for multi messenger astronomy, ie: astronomy & astroparticle physics. They allow both researchers and the general public to access information regarding detected events.

The LLMs employed look at the data that is available, and generate text to make it human-readable. These are both publicly available and we will take a look at them:

Astro-COLIBRI web interface

Astro-COLIBRI and NIMBUS

Astro-COLIBRI and NIMBUS are LLM-integrated suites for multi messenger astronomy, ie: astronomy & astroparticle physics. They allow both researchers and the general public to access information regarding detected events.

The LLMs employed look at the data that is available, and generate text to make it human-readable. These are both publicly available and we will take a look at them:

Astro-COLIBRI web interface

NIMBUS web interface

Conclusions

LiPS had talks by many others, including CERN AI assistants, LLMs for Accelerator Control, AI ethics, LLM understanding of symbolic math, and the use of LLMs for students and professors as a study assistant.

Conclusions

LiPS had talks by many others, including CERN AI assistants, LLMs for Accelerator Control, AI ethics, LLM understanding of symbolic math, and the use of LLMs for students and professors as a study assistant.

Altogether, the talks at LiPS set a path forward for the use of LLMs in different areas of physics and astronomy. I'm excited to see where things will go and how it can help researchers of the future.