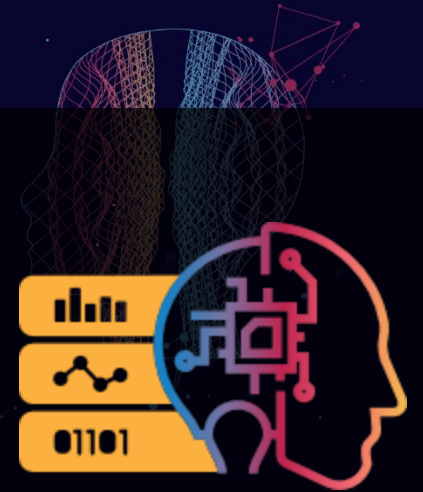




Demonstrating the cutting-edge capabilities of LLMs within the constraints of a secure, sand-boxed environment

Demitri Maestas



Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.



84.904
90.006

47.492

MM HY
[59.875 04.578]
L - 33 | F.965 - 5693

RT UR
[56.065 74.950]
H - 85 | 8594 - 9053



9.994

77.174

1.832

8.832



Our Goal: Bring LLM & NLP capabilities as a service to Sandia and its missions **as quickly as possible** while carefully upholding software, data, and compute policies

Potential Use Cases

Mission	Operational
Trend Analysis & Document Clustering	Knowledge Transfer
Code Generation w/Custom Grammars	Onboarding
Visualization and Virtual Reality	Resume Evaluation
Custom Applications & Pipelines	





The Story So Far

84.904

90.006



ChatGPT is released

SNL has labs-wide town hall about ChatGPT

The open source LLM/NLP space blossoms

Assortment of scoped projects and demonstrations shown at SNL

Funded, unified, structured LLM/NLP effort(s) begin

November 2022

January 2023

Early to Mid 2023

May – December 2023

January – March 2024





Sandia's challenges

High consequence decisions in National Security

Applying rigorous policies regarding software procurement, creation, and utilization.

Ensure tools remain useful in varied computational configurations.

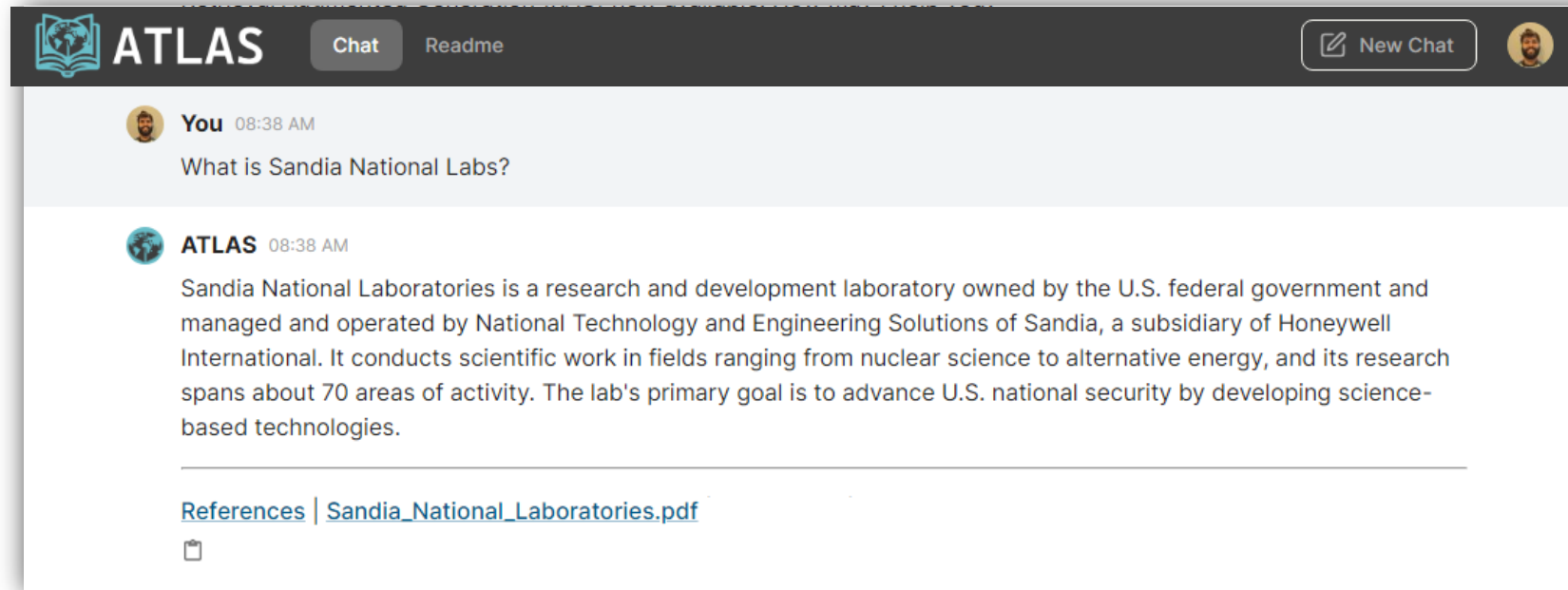
Limited ability to:

- Use commercial tools
- Contract out the work
- Share data amongst systems and people





ATLAS: A self-contained Retrieval Augmented Generation (RAG) web application powered with modular components



Shirty: A suite of modular NLP technology leveraging state-of-the-art concepts made by Sandians for Sandians.

 ATLAS

extract



embeddings



section

cross
-encodersvector
database

LLM



Network Infrastructure

Self-Contained Application

- **Modularity:** The flexibility to leverage the most current technology. Separation of concerns.
- **Customization:** Rapidly generate a solution best fit for the need *on that system*
- **Policy:** Built with SNL's software policies as a staple feature
- **Portability:** Self-Contained. Take it anywhere!



Sandia's challenges

High consequence decisions in National Security

Applying rigorous policies regarding software procurement, creation, and utilization.

Ensure tools remain useful in varied computational configurations.

Limited ability to:

- Use commercial tools
- Contract out the work
- Share data amongst systems and people

The ATLAS solutions

Equip the user with validation methods

Build with Sandia's posture as a staple component

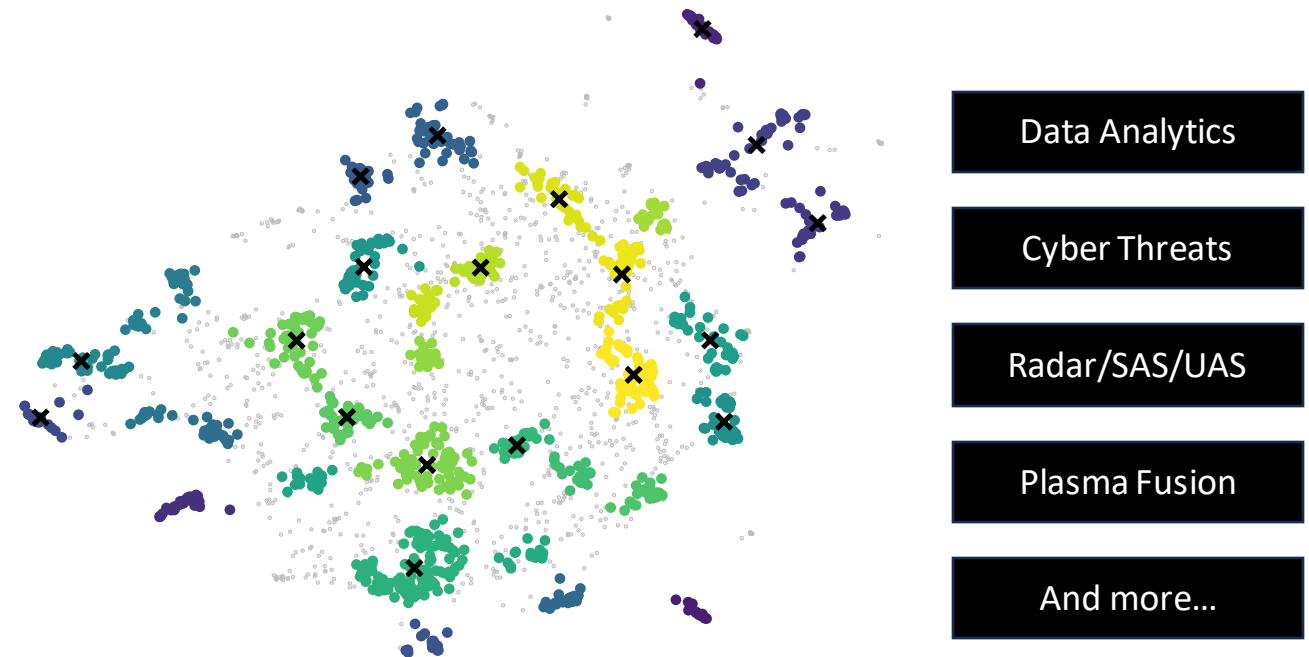
Design to:

- Be self-contained and standalone
- Leverage open-source
- Rapidly bring state of the art to SNL
- Maintain strict authorization constraints
- Be modular



The Goal: Uncover groups of topics via document summary, enabling topical clustering and tracking of scientific document trends.

Clustering of DOE/OSTI articles on ML/AI. We identify clusters on data analytics, cyber threats, etc.

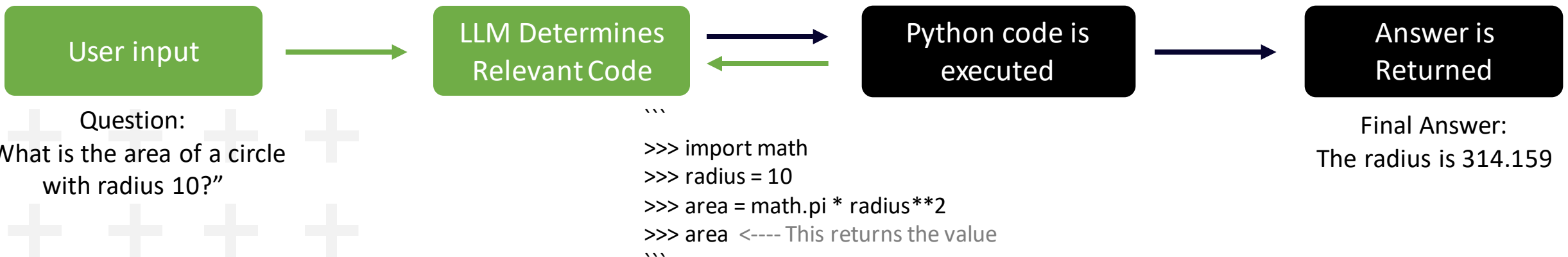


2,400 OSTI ML/AI SAND reports (2010-2023),
colored by topic





- **The Goal:** Use LLMs to drive generation and execution of code to **derive an accurate quantitative answer from a quantitative question**
- **Examples:**
- "What is the area of a circle with radius 10?" -> "314.1592653589793"
- "Can you make a csv named latlon.csv and fill it with 100 random longitude and latitude values?" -> "Done"
- "Does the polygon with verticies [(0,0),(1,2),(5,3)] intersect [(1,1),(1,2),(0,0)]?" -> "The polygons intersect"
- "Does the polygon with verticies [(0,0),(1,2),(5,3)] intersect [(10,1),(11,10),(11,11)]?" -> "The polygons don't intersect"



- **The Goal:** Interact and ask questions about VR Environments/Models in real-time using audible speech.



Gemini Scout Robot in the CAD2VR Application

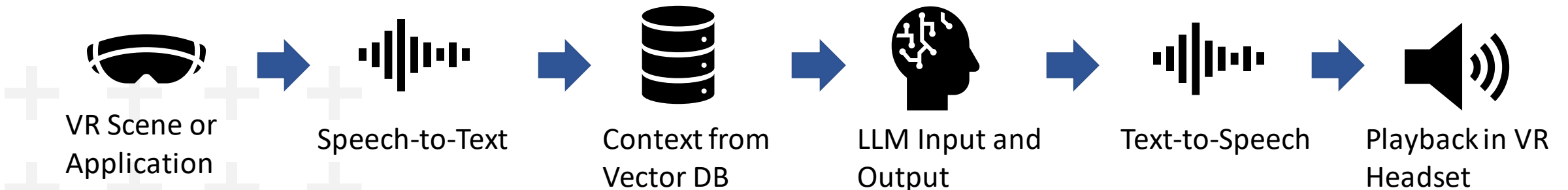
Example Dialogue

User: What am I looking at?

JARVIS: That is the Gemini Scout Mine Rescue Robot, a robotic assistant designed by Sandia robotics engineers to help with rescue planning and response during a mining accident.

User: How big is it?

JARVIS: It is less than four feet long and two feet tall and can navigate around tight corners and over safety hatches a foot high. It weighs approximately 190 Lbs.



Avoids the need to remove the VR headset at any point.



Summary:

Bringing LLM/NLP capability to SNL in a quick, modular, multi-use way while meeting SNL requirements & procedures.

This enabled rapid development of mission applications, some of which are noted here.

Questions?

Demitri Maestas

demmaes@sandia.gov

Other SNL LLM Projects:

- IRIS (Virtual Subject Matter Expert)
- Sandia AI Chat with Microsoft Azure
- Derivative Classification Assistance Portfolio

Thanks to these Sandia Staff:

- Ron Oldfield
- Tyler Ganter
- Danny Gomez
- Catherine Appleby
- Aniket Pant
- Brian Kessel

This work was funded with a mix of the Advanced Simulation and Computing Program (ASC) and Sandia's Nuclear Deterrence (ND) Programs