INL/CON-23-73228-Revision-0

Integrating Conversational Artificial Intelligence Systems into Science Gateways

July 2023

Brandon Samuel Biggs Jr

Changing the World's Energy Future

Idaho National Laboratory

INL is a U.S. Department of Energy National Laboratory operated by Battelle Energy Alliance, LLC

INL/CON-23-73228-Revision-0

Integrating Conversational Artificial Intelligence Systems into Science Gateways

Brandon Samuel Biggs Jr

July 2023

Idaho National Laboratory
Idaho Falls, Idaho 83415

http://www.inl.gov

Prepared for the
U.S. Department of Energy
Under DOE Idaho Operations Office
Contract DE-AC07-05ID14517

July 24, 2023

Brandon Biggs
INL/CON-23-73228

Integrating Conversational Artificial Intelligence Systems into Science Gateways

Battelle Energy Alliance manages INL for the
U.S. Department of Energy's Office of Nuclear Energy

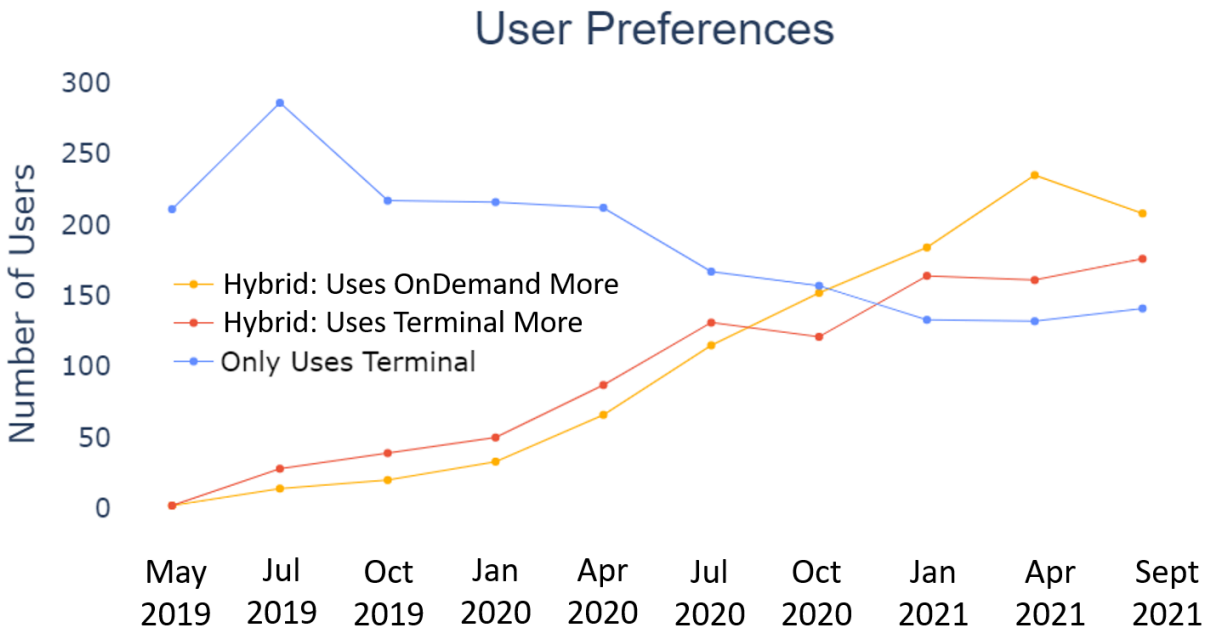
Idaho National Laboratory

Overview

- Background
- Science Gateways
- Conversational Systems
- Benefits
- Challenges
- Considerations

Background

- Scientific computing resources have historically been accessed via the terminal, but trends seem to be changing towards using science gateways.
- Two main science gateway categories:
  - Data
  - Compute*



Quantifying the Impact of Advanced Web Platforms on High Performance Computing Usage
https://dl.acm.org/doi/10.1145/3491418.3530758

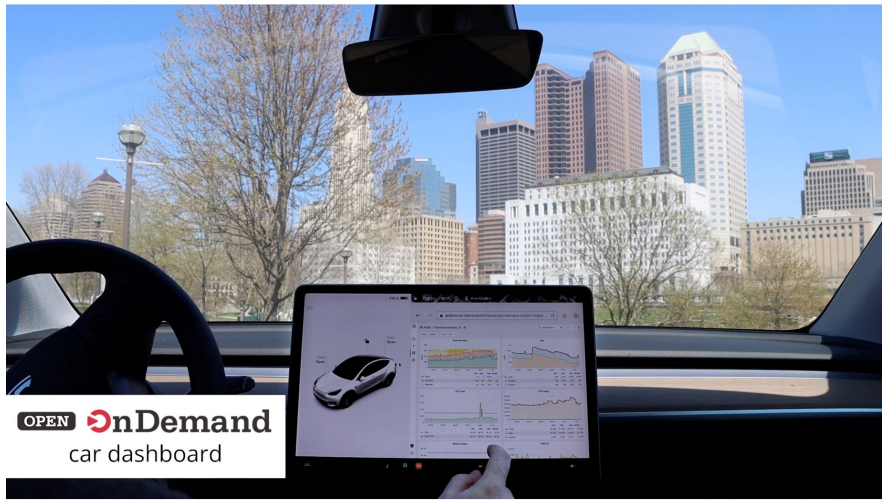
* This is what we're focused on

What is a science gateway?

- Web based portal to scientific or high performance computing (HPC) resources
- Provides access to resources and applications

Why would I want to use a science gateway?

- One-stop-shop for HPC
  - Visualization tools
  - IDEs
  - File management
- Lower the technical barrier
  - Fewer challenges submitting jobs
  - Tool access is streamlined
  - Most devices have a web browser
  - Abstract the challenges away from users so they have more time to do science



Open OnDemand Device Demonstration
https://vimeo.com/825953856/479872dc8b

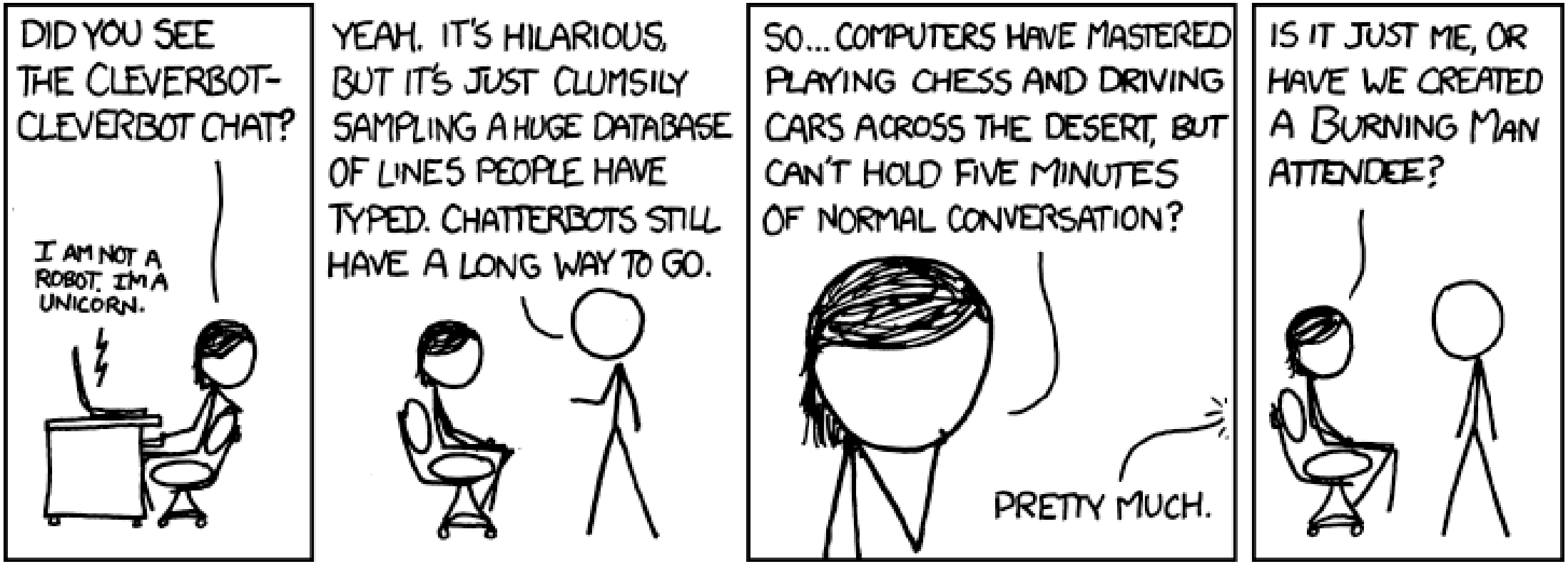
Computation Focused Science Gateways

- Apache Airavata
- Altair Access
- Jupyter Hub
- Open OnDemand

So HPC in a web browser.. What's next?

- While science gateways decrease the time-to-science that comes with using advanced systems, progress can still be made to make things easier..
- What about artificial intelligence?

Using AI for HPC instead of HPC for AI



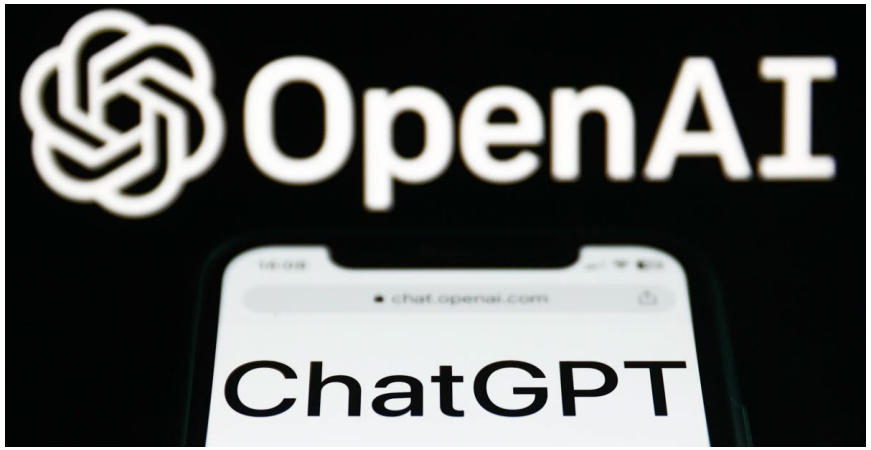
Conversational Systems

- Speech and text chat assistants
- Large Language Models (open/closed weights)
- Rule based with AI (Rasa, ‘skills’)

Hey Siri

Google Assistant

alexa



Meet Claude

A next-generation AI assistant for your tasks, no matter the scale

Bard

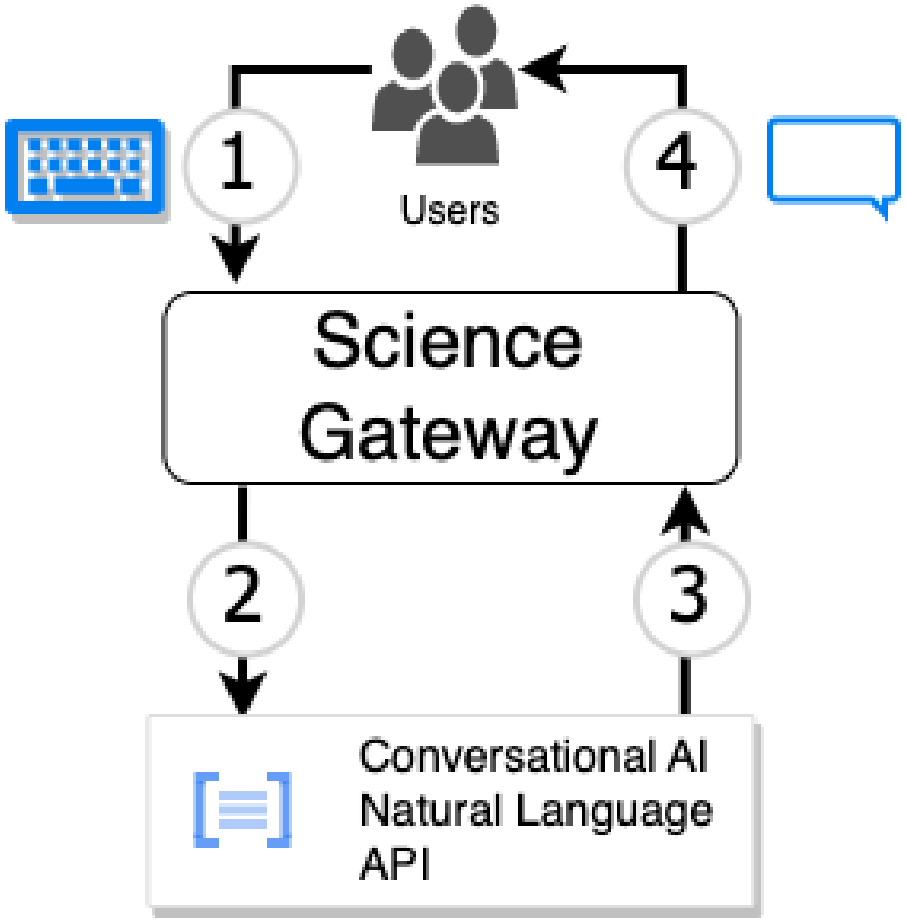
How might these conversational systems be used?

- General questions
  - “What HPC systems can I submit to?”
  - “What is the maximum amount of time my job can run for?”
- Fact gathering/troubleshooting issues
  - “My job died and I’m not sure why!”
  - AI: “Please provide the following pieces of information and we’ll create a support ticket for you.”
- Submitting jobs
  - “Create a jupyter notebook session for me.”
  - AI: “Sure, which system would you like to submit to and for how long do you want your job to run?”

Chat Interface AI Assistants

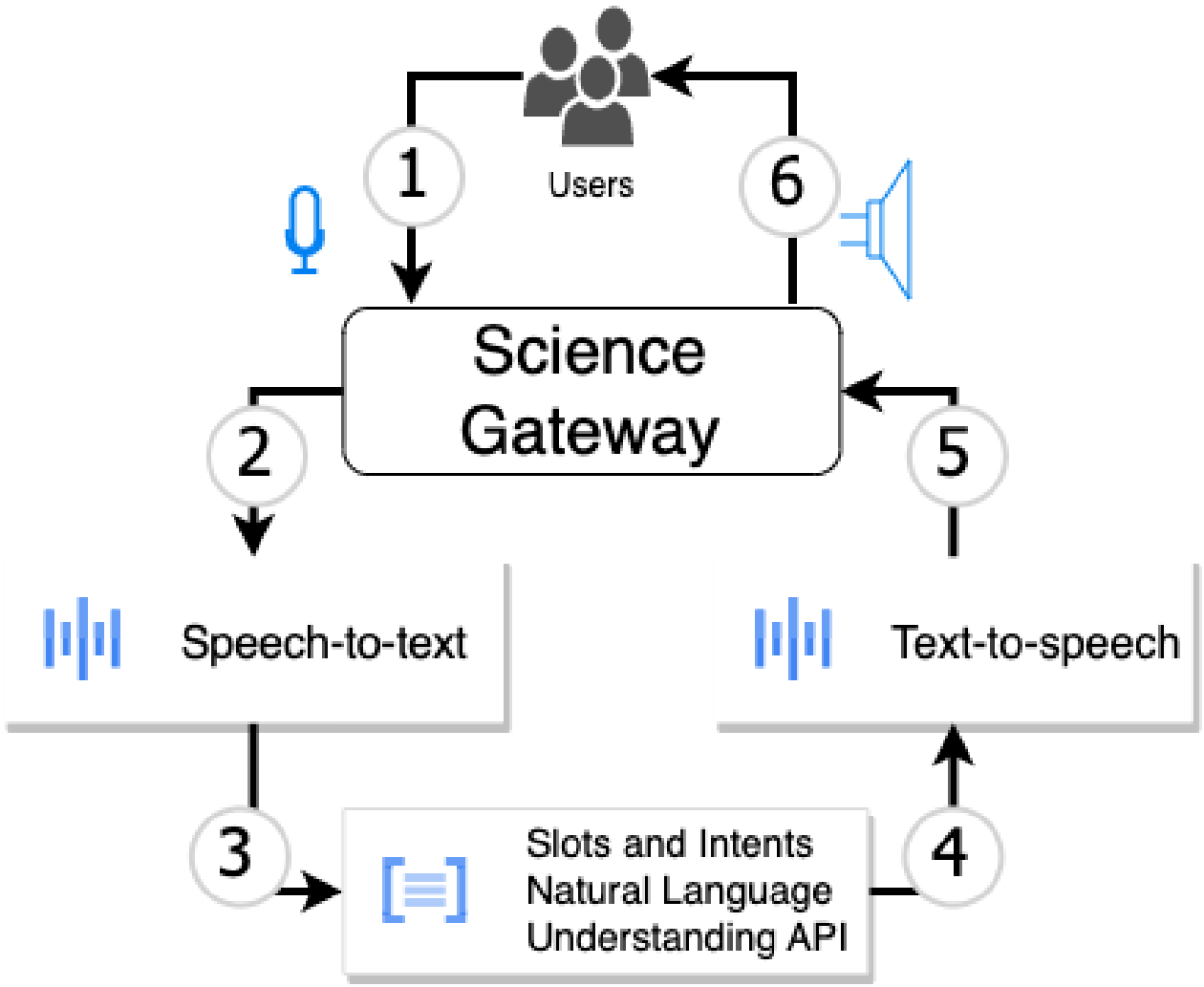
1. A user interacts with a chat interface integrated into the science gateway.
2. The user's message is sent to a conversational API.
3. The API generates a response and forwards it to the science gateway.
4. The generated response is displayed to the user in the chat interface.

The user can choose to repeat the process or end the conversation.



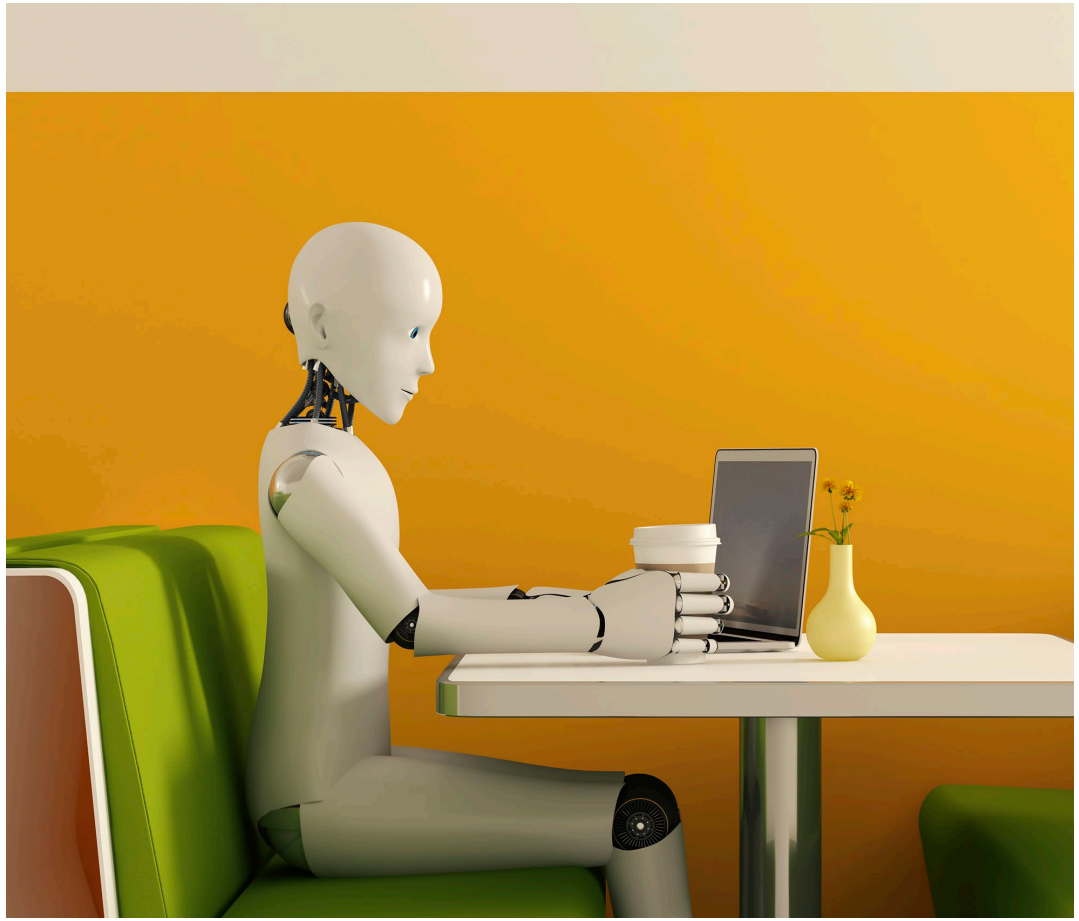
Voice Interface AI Assistants for Job Requests

1. A user interacts with a voice interface integrated into the science gateway.
2. The audio is sent to a server API to be transcribed via a speech-to-text model
3. The transcribed text is sent through a natural language understanding (NLU) model
4. The NLU model interprets the request and sends a text response to a text-to-speech (TTS) model.
5. The TTS model creates an audio file that is sent back to the science gateway.
6. The science gateway plays the generated audio for the user to hear and respond to.



Benefits

- AI can assist users by reducing the barrier to HPC while shifting time to more productive efforts
  - Searching documentation
  - Saves HPC staff and researchers' time by getting answers faster



Challenges of integrating AI

- Interacting with various systems
  - Different schedulers, many different codes, versions, other tools
- Various jargons (ngpus vs gpus-per-node, many technical terms, other abbreviations)

Considerations

- AI has biases
- AI is good at being confidently incorrect
- It's a tool, not a replacement for qualified staff
- The ethics of using AI shouldn't be ignored -
  - "The Exploited Labor Behind Artificial Intelligence" by Adrienne Williams, Milagros Miceli, and Timnit Gebru from the Distributed AI Research Institute
  - AIethicist.org – Repo of material for AI ethics

Questions?

- Brandon.Biggs@inl.gov

Idaho National Laboratory

Battelle Energy Alliance manages INL for the U.S. Department of Energy's Office of Nuclear Energy. INL is the nation's center for nuclear energy research and development, and also performs research in each of DOE's strategic goal areas: energy, national security, science and the environment.