University of Colorado Denver
Department of Physics
1200 Larimer Street, Suite 3123
Campus Box 157, PO Box 173364
Denver, CO 80217-3364

Using Generalized Adversarial Networks to ensure trusted science results and maximize science reach within the dark matter community

PI Dr. Amy Roberts, PhD (Assistant Professor), amy.roberts@ucdenver.edu

Co-PI Dr. Farnoush Banaei-Kashani (Associate Professor), farnoush.banaei-kashani@ucdenver.edu

Artificial Intelligence Research For High Energy Physics (DE-FOA-0002705)

Using Generalized Adversarial Networks to ensure trusted science results and maximize science reach within the dark matter community

PI Amy Roberts (Physics), Co-PI Farnoush Banaei-Kashani (Computer Science)

Abstract

SuperCDMS is a direct-detection dark matter search that aims to answer a top-priority question of the DOE[1], "What is dark matter?" It is one of a suite of experiments that will begin data taking in the next two years. All these dark matter experiments, including SuperCDMS, are facing unprecedented challenges in interpreting their data. Our work looks for a way to test and improve the trustworthiness of these results and addresses two DOE Primary Research Directives: massive scientific data analysis and intelligent decision-support for complex systems[2].

Massive scientific data analysis (AI for HEP)

The upcoming dark matter experiments will be the cleanest, quietest, rare-event detectors ever built. But the dark matter community doesn't yet know how to use this data to effectively search for dark matter. One challenge faced by the dark matter community is how to perform a "blinded" analysis on incoming data. Scientists use blinded analyses to reduce the effect of human bias on the final science result. Blinded analysis is the gold standard within the physics community, but it remains unclear how to effectively perform blinded analysis on the upcoming generation of dark matter data.

Generalized Adversarial Networks (GANs) could make a promising blinding technique – "salting" – feasible for use with dark matter data. "Salting" adds simulated signal to a data set. This allows analyzers to look at the full data set, making it ideal for dark matter analysis. However, creating realistic simulated signal currently takes years. GANs specialize in generating new data that is indistinguishable from the training set and have the potential to create realistic simulated signal within weeks. Using GANs for data salting could provide a much-needed analysis tool to the dark matter community.

Massive scientific data analysis (HEP for AI)

Initial efforts to create "salt" are promising, but it may be possible to improve on initial results by incorporating physics constraints into the GAN model. This is an under-developed area of AI and could result in improved methods for GANs.

Supporting decisions in complex systems (AI for HEP)

The upcoming dark matter experiments together represent millions of dollars of investment from the scientific community. Previous dark matter experiments consistently lost half of all science data because humans cannot analyze that data in real-time.

The work on GANs has the potential to identify changes in incoming data fast enough for scientists to identify and fix issues with the experiment before weeks or months of costly data has been lost. Like the issue of data blinding, improving the time scale of data-quality monitoring would impact the entire dark matter field.

[1] DOE, "Basic Research Needs for Dark Matter Small Project New Initiatives," 2018.
[2] N. Baker *et al.*, "Workshop Report on Basic Research Needs for Scientific Machine Learning: Core Technologies for Artificial Intelligence," Feb. 2019.