



Seldonian Toolkit Competition



DALL-E's interpretation of "An abstract art painting with hints of technology and safety."

Description. The first Seldonian Toolkit Competition took place in March and April 2023. The contest was designed as a way for undergraduate and graduate students in the US and Canada to learn about and develop safe and fair machine learning algorithms using the recently launched Seldonian Toolkit. This toolkit was created to make it easier for data scientists to apply machine learning responsibly: with high-confidence safety and fairness constraints on the machine learning system's behavior.

This competition was intentionally open ended -- any application of the toolkit that includes a high-confidence constraint was acceptable. Students picked topics that they found interesting.

Talks. To help give participants a more complete view of AI safety and fairness topics, we invited speakers who are not working with Seldonian algorithms, but who are still studying issues related to safety and fairness.

Important Dates.

- **[Early Registration] March 2, 2023:** Participating teams should register by midnight (anywhere on Earth) on March 2, 2023. Registration is free and non-binding. Teams can register right now by clicking the button below and filling out the provided form.
- **[Kick-Off Event] March 6, 2023:** This event will consist of:
 - Prof. Philip Thomas providing a high-level introduction to AI safety and fairness topics using Seldonian algorithms.
 - Dr. Austin Hoag providing coding examples of how the Seldonian Toolkit can be used.

This event will be held from 5:30pm - 7:00pm Eastern Time on Zoom ([link](#)), though participants are welcome to attend in-person at UMass Amherst (CS Building Room 151).

- **[Late Registration] March 10, 2023:** Participating teams **must** register by midnight (anywhere on Earth) on March 10, 2023 in order to be eligible for any awards. Registration is free and non-binding. We strongly encourage teams to register before the March 2 deadline so that they can receive information about the timing and location of the Kick-Off Event. However, registration by March 10 is sufficient to participate in the contest.
- **[Early-Contest Q&A] March 17, 2023, 3pm Eastern:** Prof. Philip Thomas and Dr. Austin Hoag will host an open Q&A session for participating teams. This event will take place at 3pm Eastern at [this Zoom link](#).
- **[Invited Talk: [Prof. Yair Zick](#)] March 22, 2023, 1pm Eastern:** Prof. Zick will present a talk titled "A Simple, General Framework for Fair Allocation Under Matroid Rank Valuations." The talk will be hosted on [Zoom](#), as well as in-person in room 203 of the CS building at UMass. There is limited seating, so people interested in attending in-person, please confirm with Prof. Thomas that a seat is available.
- **[Mid-Contest Q&A] March 27, 2023, 4pm Eastern:** Prof. Philip Thomas and Dr. Austin Hoag will host an open Q&A session for participating teams. This event will take place at 4pm Eastern at [this Zoom link](#).
- **[Invited Talk: [Dr. Vael Gates](#)] March 28, 2023, 5pm Eastern:** Dr. Gates will present at [this Zoom link](#).
- **[Late-Contest Q&A] April 13, 2023, 4pm Eastern:** Prof. Philip Thomas and Dr. Austin Hoag will host an open Q&A session for participating teams. This event will take place at 4pm Eastern at [this Zoom link](#).
- **[Invited Talk: [Dr. Thomas Gilbert](#)] April 19, 2023, 4pm Eastern:** Dr. Thomas Gilbert will present a talk titled "Reward Reports for Reinforcement Learning" at [this Zoom link](#)
- **[Final Submissions] April 21, 2023:** Final submissions are due at midnight (anywhere on Earth) on April 21, 2023. See the "Submissions" section below for more information about what should be submitted.
- **[Award Announcement] April 28, 2023:** We aim to announce the winners of the competition by April 28, 2023. The exact date will depend on the volume of submissions.

Participation and Submission

Participating teams should select an application of the Seldonian Toolkit. While there are no restrictions on the allowed applications, we recommend that you select an application for which you have access to training data and which your team members are familiar with. This could range from predicting how far landslides will travel based on features of the slope, with safety guarantees related to the chance of under-predictions, to predicting whether a tumor is benign or malignant with safety guarantees with respect to the false negative rate, to predicting whether someone will commit a crime in the future while enforcing fairness constraints with respect to race, gender, or age. Some teams might already have applications in mind, while others might begin by brainstorming possible applications. For teams still trying to select an application, we recommend searching for datasets that relate to machine learning problems where safety or fairness guarantees would be beneficial.

After selecting an application, teams should apply the Seldonian Toolkit. In almost all cases, teams should use the Experiments component of the toolkit to show how effective the Seldonian Toolkit is for their application. The Experiments component is described in [this](#) tutorial. It provides plots that show how accurate the learned models are, how much data was required before the system could reliably return solutions, and how often the system violated the desired safety or fairness constraints.

Next, teams should put together a report describing their application, its importance, how the Seldonian Toolkit was applied, and the resulting performance of their system. This report should be provided as a markdown (.md) file in a GitHub repository that contains the source code for the project. The markdown file should clearly indicate the team name somewhere near the top. Each team should then fill out the submission form linked below, which asks for a link to the GitHub repository. [This](#) is an example of what a team might submit.

Awards and Evaluation Criteria

The panel of judges, consisting of AI faculty from UMass Amherst (Professors Philip S. Thomas, Bruno Castro da Silva, and Scott Niekum), Stanford University (Professor Emma Brunskill), and Brown University (Professor George Konidaris), selected the following two submissions as a tie for the "Best Overall Student Project" award (each will receive \$600).

- **Fairness for Breast Cancer Recurrence Prediction** by Derek Lacy: [Link](#).
- **Fairness in Student Course Completion Based on Student Data** by Sahil Yerawar, Pranay Reddy, and Varad Pimpalkhute: [Link](#).

These projects can now be found on the [examples](#) page.

Support

Participants are encouraged to post questions on the GitHub issues pages [links: [Engine](#), [Experiments](#), and [GUI](#)]. We will do our best to answer these questions in a timely manner. For questions related to this competition but not directly related to the use of the Seldonian Toolkit, we encourage teams to ask during the kick-off event or the various Q&A sessions held throughout the contest. Teams can also email Seldonian@cs.umass.edu. However, responses to these emails may be slow depending on the volume of participants and questions.

The UMass Data Science club has also created a Discord server where participants can interact with each other.

Discord Server

Sponsors

This contest is a collaboration between the Autonomous Learning Laboratory (ALL) at the University of Massachusetts and the Berkeley Existential Risk Initiative (BERI). The awards will be provided by BERI.