The nature of dark energy is one of the biggest mysteries in physics and astronomy today. To quote Michael Turner, dark energy is "a problem for the 22nd century discovered by accident in the 20th century." The Dark Energy Spectroscopic Instrument (DESI) is a massive next-generation survey that will attempt to constrain the dark energy equation of state. DESI will measure the spectra of over 30 million galaxies and determine their redshift. These redshifts, combined with measurements of Type Ia supernovae (SNe Ia), provide the strongest measurements of cosmological distances, our way of knowing the expansion rate of the universe.

DESI offers the potential to spectroscopically observe ~$10^5$ supernovae (SNe) including many SNe Ia [1]. Interestingly, while SNe Ia are used as standard candles, **their origins are not fully understood.** Evidence exists for two types of progenitors: a degenerate white dwarf accreting matter from a giant companion star, or two coalescing white dwarfs [2]. Identifying a large population of SNe Ia will help answer the progenitor question and provide constraints on dark energy. My background in handling large astrophysical datasets (see: Personal Statement) has prepared me make valuable contributions in this area. **I propose to develop an efficient computational procedure to identify SNe Ia in the DESI survey.**

In order to obtain accurate, unbiased distance measurements from SNe Ia, spectroscopic measurements are crucial in calibrating corresponding photometric observations. In particular, DESI will be able to spectroscopically complement future large-sky surveys such as the Large Synoptic Sky Telescope (LSST), which will begin its science observing in 2021 but collect immensely more data (expected ~1 million transient alerts per night with ~1 million SNIa observed over a decade [2]), and the Zwicky Transient Factory. In doing so, we can understand just how "standard" these standard candles are and provide complementary redshift coverage. Further, it is to our advantage to identify SNe Ia in real-time and send out alerts for followup observations. **Moving forward, generalized data-analysis pipelines that are able to handle enormous quantities of data will be fundamental to the success of future experiments.**

I propose the following analysis and timeline to develop such a computational procedure: **1.** (year 1) Identify SNe Ia in galactic spectra. **2.** (years 2-4) Extend the identification algorithm to detect and classify galactic spectral anomalies (outliers) *in general.* **3.** (final year) Develop an automated pipeline that will run in real time to identify transients.

In step **1.** I will focus on identifying SNe Ia in the DESI catalog. This will require construction of spectroscopic models of galaxies and applying a statistical test to a large sample of galaxy spectra to look for deviations from the model expectations. Spectra with significant deviations (anomalies) will be tested further by fitting a SN Ia spectral template to see if the anomaly is in fact a supernova.

**I propose to develop several complementary tests to identify SNe Ia spectra.** Previous studies have searched for SNe Ia in the Sloan Digital Sky Survey (SDSS) catalog using singular-value-decomposition of a large sample of galaxy spectra to construct a basis of eigenspectra. The eigenbasis is used to fit galaxy spectra, and the residual spectrum is searched for features corresponding to a SNe Ia (note that the model in this case, i.e. the individual eigenspectra, does not represent a physical object - only the linear combination has a physical interpretation). Using this method, Graur & Maoz [3] report 90 SNe Ia in SDSS data release 7.

A first alternative approach involves constructing a $\chi^2$ statistic (or likelihood test) of SNe Ia templates with the observed spectra, defining anomalies based on the $\chi^2$ goodness of fit. The advantage is that no unphysical basis is being used. As a sanity check, I will cross-correlate my sample of SNe Ia with those found in previous studies [3]. With this sample, I will calculate an estimate of the SNe Ia rate to help answer the progenitor question and make first-order distance estimates using SNe Ia redshifts to build a framework for future photometric calibrations.

While useful for cosmology, this method should be capable of identifying more than just SNe. In step **2.** I will generalize this procedure to allow the classification to include any number of other astrophysically interesting phenomena. As a first step, this will involve adding templates for each source of interest, for example a two-galaxy-spectrum model to represent sources of strong gravitational lensing, which offer excellent tests of the general theory of relativity. Once the algorithm is efficient at detecting several different classes of objects, I will drop all *a priori* assumptions. In this case, the proposed algorithm will be general and unassuming of any particular class of object, allowing **the potential of discovering new phenomena.** A successful algorithm will also identify bad spectra either due to instrumental issues or other errors, learn what those patterns look like, and avoid or even correct for them in the future. At this stage, **a plethora of astrophysical phenomena beyond just SNe may be observed and studied.**

Previous studies have tried general approaches to anomaly detection using predictive learning algorithms such as a random forest (decision tree), and have even found larger yields than targeted identification algorithms [4, 5]. I will study these approaches and develop a machine-learning algorithm to classify spectroscopic anomalies with DESI. I will work to boost the efficiency as well as the yield of outlier spectra while minimizing false-positives. In step **3.** I will maximize the efficiency of the algorithm so that it can be used in real-time in a pipeline.

These approaches will each be tested and trained on existing spectroscopic data obtained by SDSS. I will also test simulated spectra generated for DESI. Simulated data in particular will give an estimate of the classification purity of the pipeline for the various phenomena we are aiming to observe. Once DESI is online, the learning algorithm will be applied to the real data, being continuously trained as the dataset grows.

An integral part of my work on this project will be engaging the public in active participation. The citizen-science project *Galaxy Zoo Supernovae* demonstrated that members of the general public are remarkably good transient-spotters: 93% of supernovae were correctly identified by the public with no false-positives [6]. **I will make the results of my pipeline, including sample data, available to existing citizen-science platforms** which have a proven track record of engagement and popularity with the public. Participants will enjoy hands-on involvement in the data analysis alongside lessons detailing the qualitative astrophysics of supernovae and how their signature can be seen in galaxy spectra through the heavy elements produced in the explosion. The data collected by citizen scientists can support actual results by providing confirmation or rejection of suspected outlier spectra. My findings, combined with the citizen-science results, will lead beyond DESI and into future LSST analysis and other data-intensive astronomy projects.

**References**

[1] DESI Collaboration, 2016, Final Design Report I
[2] Maoz & Mannucci, 2012, PASA, 29, 447
[3] Graur & Maoz, 2013, MNRAS, 430, 1746
[4] Baron & Poznanski, 2017, MNRAS, 465, 4530
[5] Buisson et al., 2015, MNRAS, 454, 2026
[6] Smith et al., 2011, MNRAS, 412, 1309