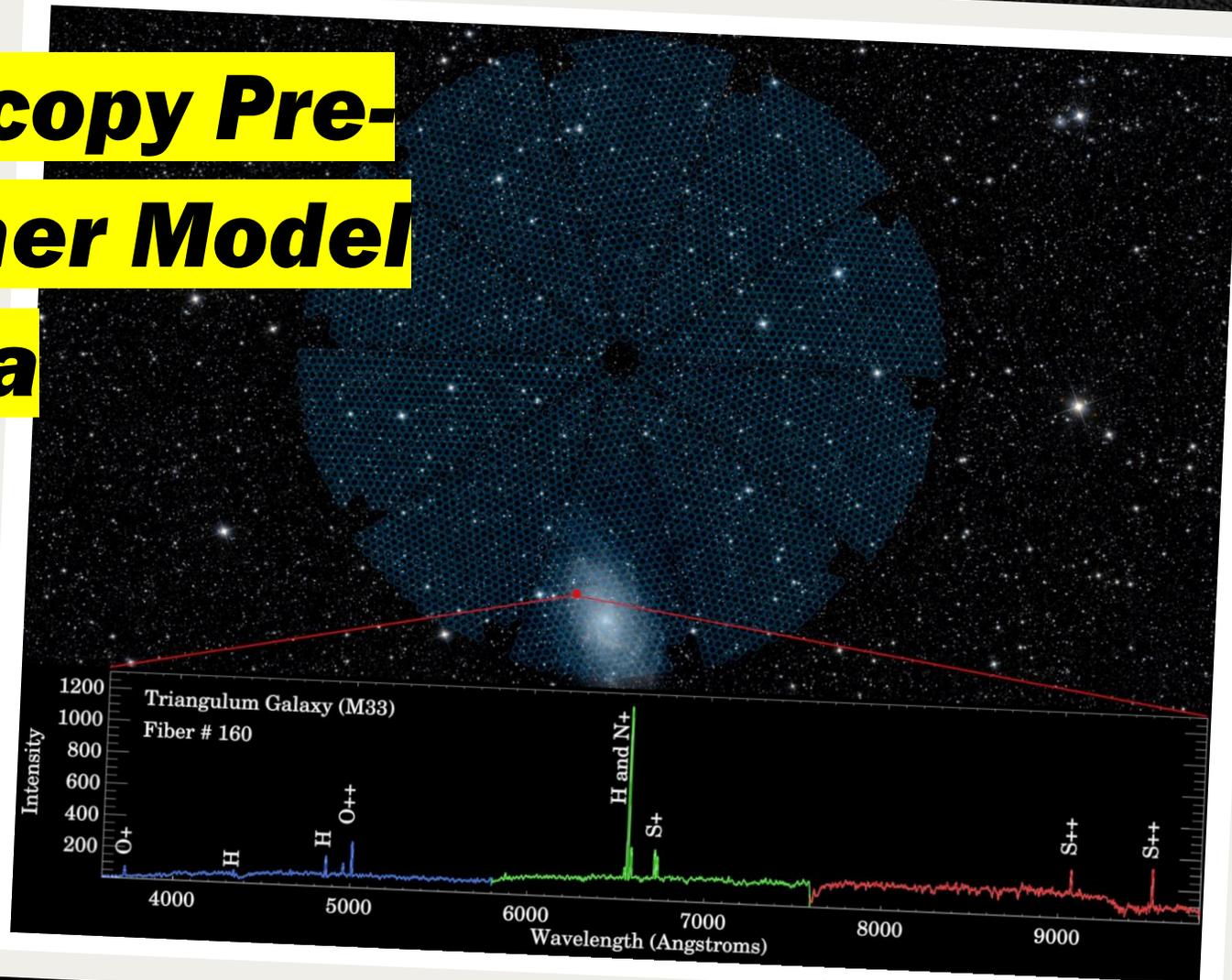


# SpecPT: Spectroscopy Pre-trained Transformer Model for Galaxy Spectra

Rohan Pattnaik  
PhD Student @ RIT  
Advisor: Dr. Jeyhan Kartaltepe

AI/ML in Astronomy and  
Astrophysics Workshop, IUCAA

January 7<sup>th</sup>, 2025



Credit: DESI Collaboration



# Background: Galaxy

## Evolution and

## Redshifts

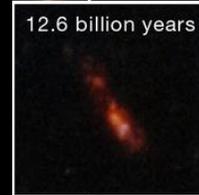
- Understanding galaxy evolution and large-scale structure of the universe requires studying the **effects of local environments on ISM properties.**
- Accurate redshift measurements are essential to place galaxies in their proper cosmic context.
- Redshift can be measured using **Spectroscopy or Photometry**

Photometric Redshifts	Spectroscopic Redshifts
Obtained using brute force fitting of an SED to photometric data.	Identify spectral features and calculate shifts using visual inspection/Cross-correlation codes.
<i>Pros:</i> Easier to collect data, High SNR	<i>Pros:</i> Highly accurate
<i>Cons:</i> Can be order of mag. less accurate	<i>Cons:</i> Time and labor-intensive

# Spectroscopic Redshift Measurement

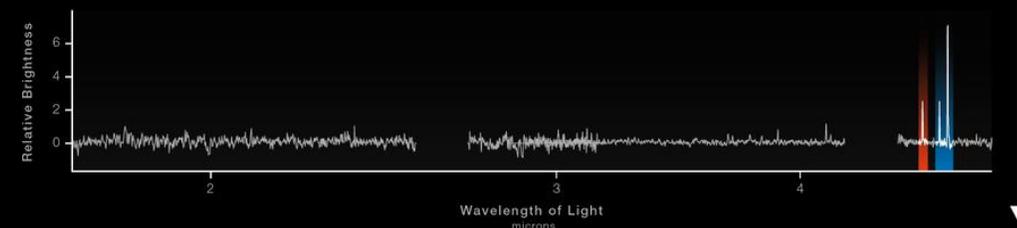
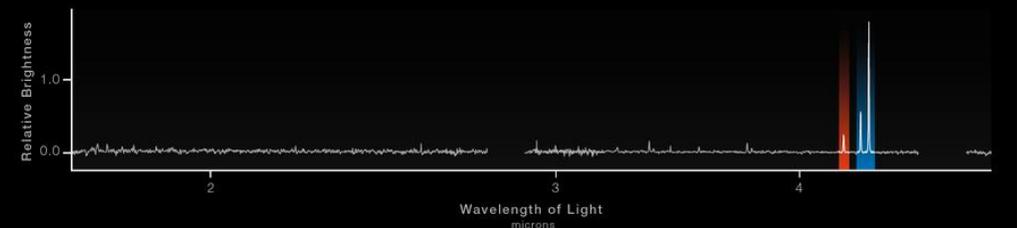
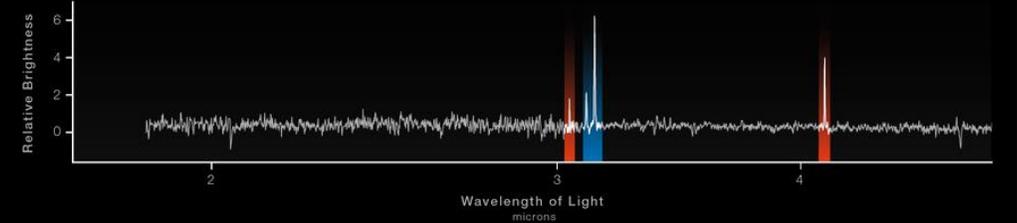
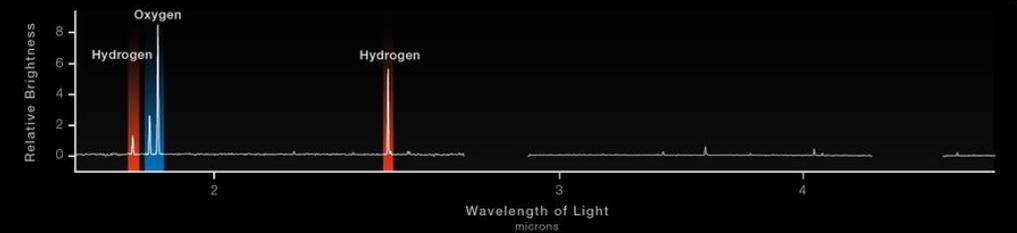
- Spectroscopic redshifts are derived by identifying key spectral features:
  - Emission lines (e.g.,  $H\alpha$ , [O II]).
  - Absorption lines (e.g., CaII, H & K)
  - Continuum features (e.g., 4000Å break).
- Traditional methods of specz measurement:
  - Visual Inspection
  - Cross-Correlation Codes

NIRCam Imaging



**WEBB**  
SPACE TELESCOPE

NIRSpec Microshutter Array Spectroscopy



Credit: NASA, ESA, CSA, and STScI

# Current Challenges in Specz Measurement

- **Traditional Methods Face Scalability Issues:**
  - **Visual Inspection:**
    - Requires experts to manually analyze spectra to identify key features.
    - Highly time and labor intensive – takes multiple people months to measure redshifts for just a few thousand spectra.
  - **Cross-Correlation Codes:**
    - Automates redshift determination by comparing spectra to predefined templates.
    - Struggles with **complex or noisy spectra**, often leading to inaccuracies.
    - Limited by the quality and diversity of templates used, making it less effective for spectra with unusual features.
- With the exponential growth of datasets (e.g., DESI, Euclid, JWST), these methods are increasingly impractical for processing millions of spectra efficiently.

# The Promise of Deep Learning in Astronomy

- ML/DL methods revolutionize large dataset analysis in astronomy.
- Examples: Classifying transient events, X-ray binaries, detecting exoplanets, and photo-z measurements.

## Photometric redshifts from SDSS images using a Convolutional Neural Network

Johanna Pasquet, Emmanuel Bertin, Marie Treyer, Stéphane Arnouts, Dominique Fouchez

## The Zwicky Transient Facility Bright Transient Survey. III. BTSbot: Automated Identification and Follow-up of Bright Transients with Deep Learning

Nabeel Rehemtulla, Adam A. Miller, Theophile Jegou Du Laz, Michael W. Coughlin, Christoffer Fremling, Daniel A. Perley, Yu-Jing Qin, Jesper Sollerman, Ashish A. Mahabal, Russ R. Laher, Reed Riddle, Ben Rusholme, Shrinivas R. Kulkarni

## Combining statistical learning with deep learning for improved exoplanet detection and characterization

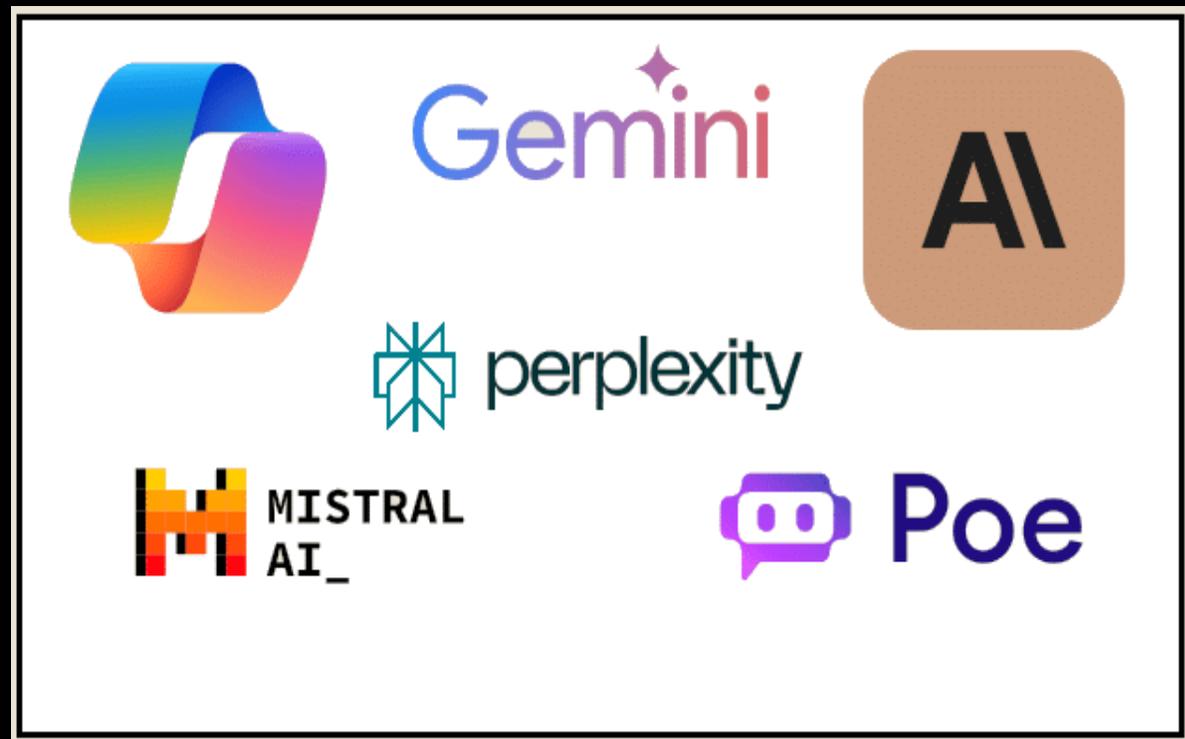
Olivier Flasseur, Théo Bodrito, Julien Mairal, Jean Ponce, Maud Langlois, Anne-Marie Lagrange

## A Machine Learning Approach For Classifying Low-mass X-ray Binaries Based On Their Compact Object Nature

R. Pattnaik, K. Sharma, K. Alabarta, D. Altamirano, M. Chakraborty, A. Kembhavi, M. Mendez, J.K. Orwat-Kapola

# The Era of Transformers and LLMS

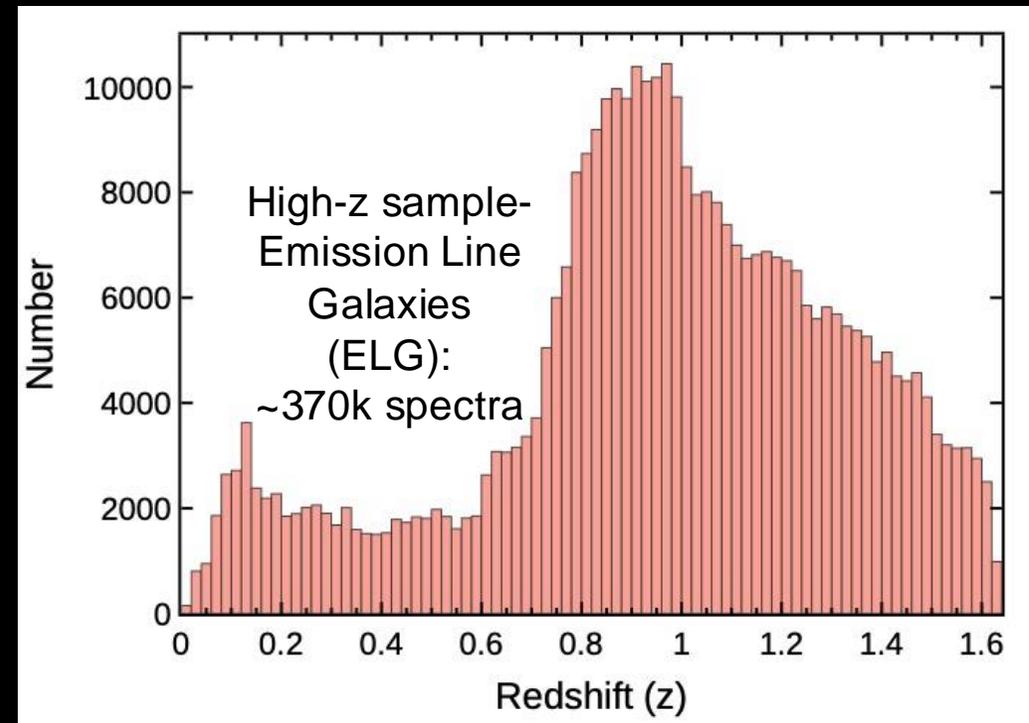
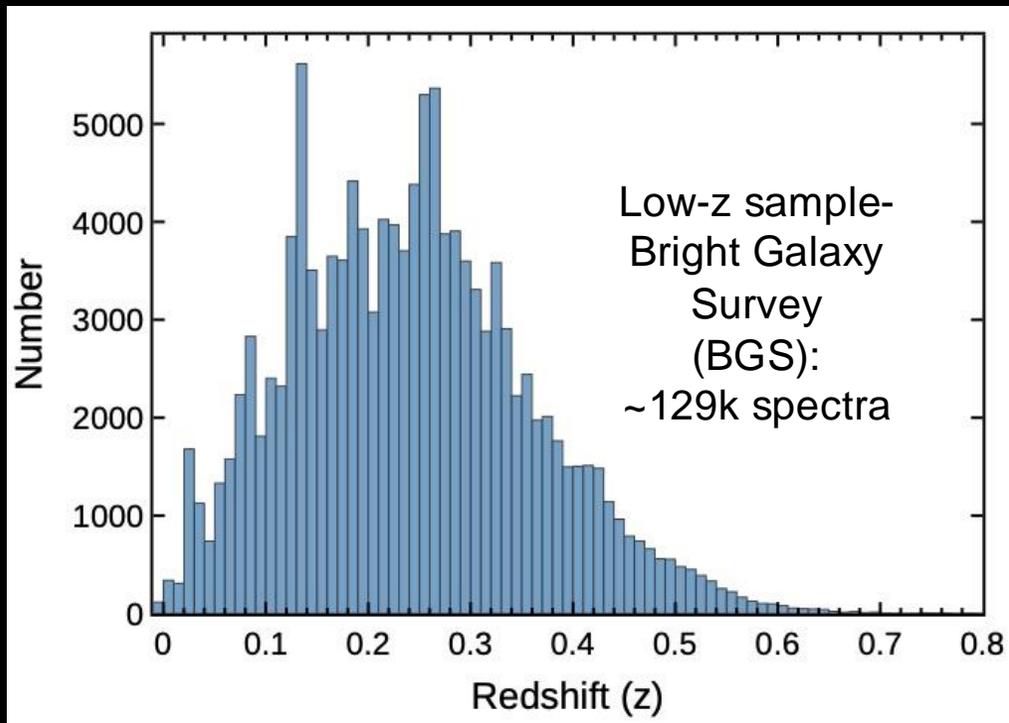
- Recently, Transformers, the backbone of LLMs like ChatGPT, have transformed AI.
- *Why Transformers for Spectroscopy?*
  - Spectra are sequential like text, with patterns in emission/absorption lines (e.g., [OIII] follows H $\beta$ ).
  - Transformer models excel at capturing patterns and contextual relationships (via attention mechanisms).
  - Beyond redshifts: A transformer model could also generalize to tasks like measuring ISM properties and anomaly detection.



Credit: datascientest.com

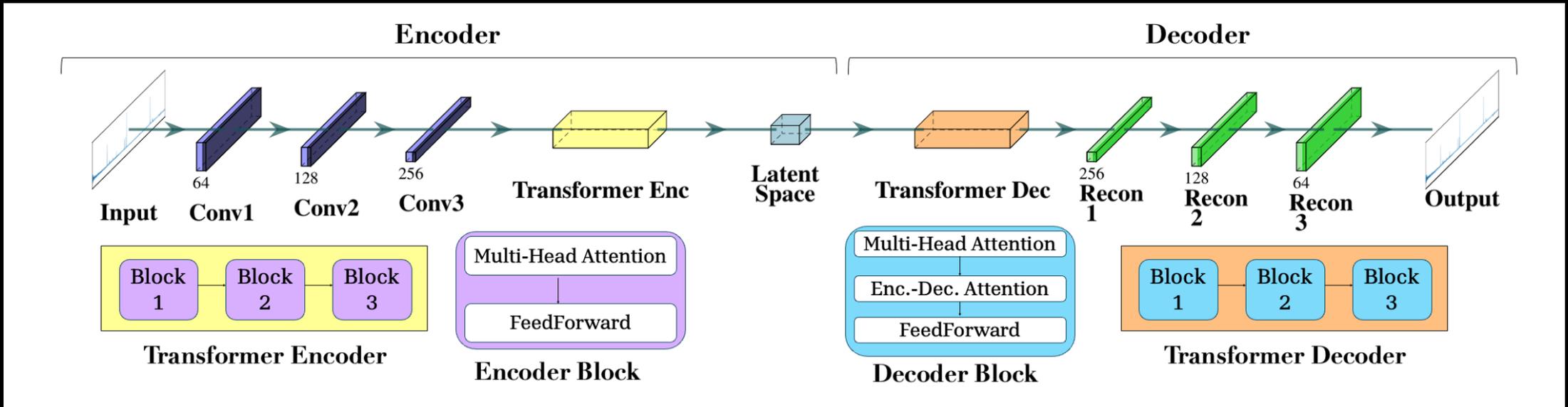
# Training data: DESI Early Data Release (EDR)

- Training transformer models requires a lot of data.
- DESI EDR is the ideal dataset for this purpose.



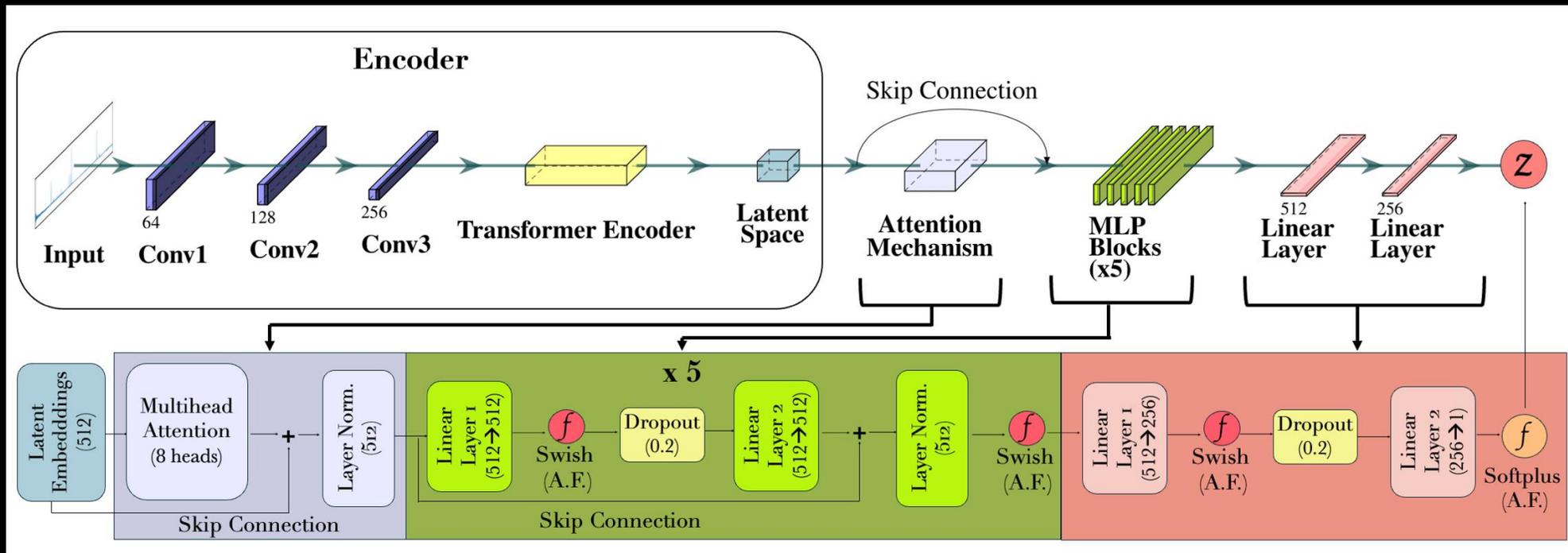
# SpecPT Autoencoder Architecture

- Input spectrum processed through convolutional and pooling layers for feature extraction.
- Transformer encoder generates compact embeddings.
- Transformer decoder and linear layers reconstruct the original spectrum



# SpecPT Redshift Measurement Architecture

- Integrates trained SpecPT encoder with residual MLP blocks.
- Includes attention mechanism for refined feature processing.
- Linear layers transform embeddings into redshift predictions.

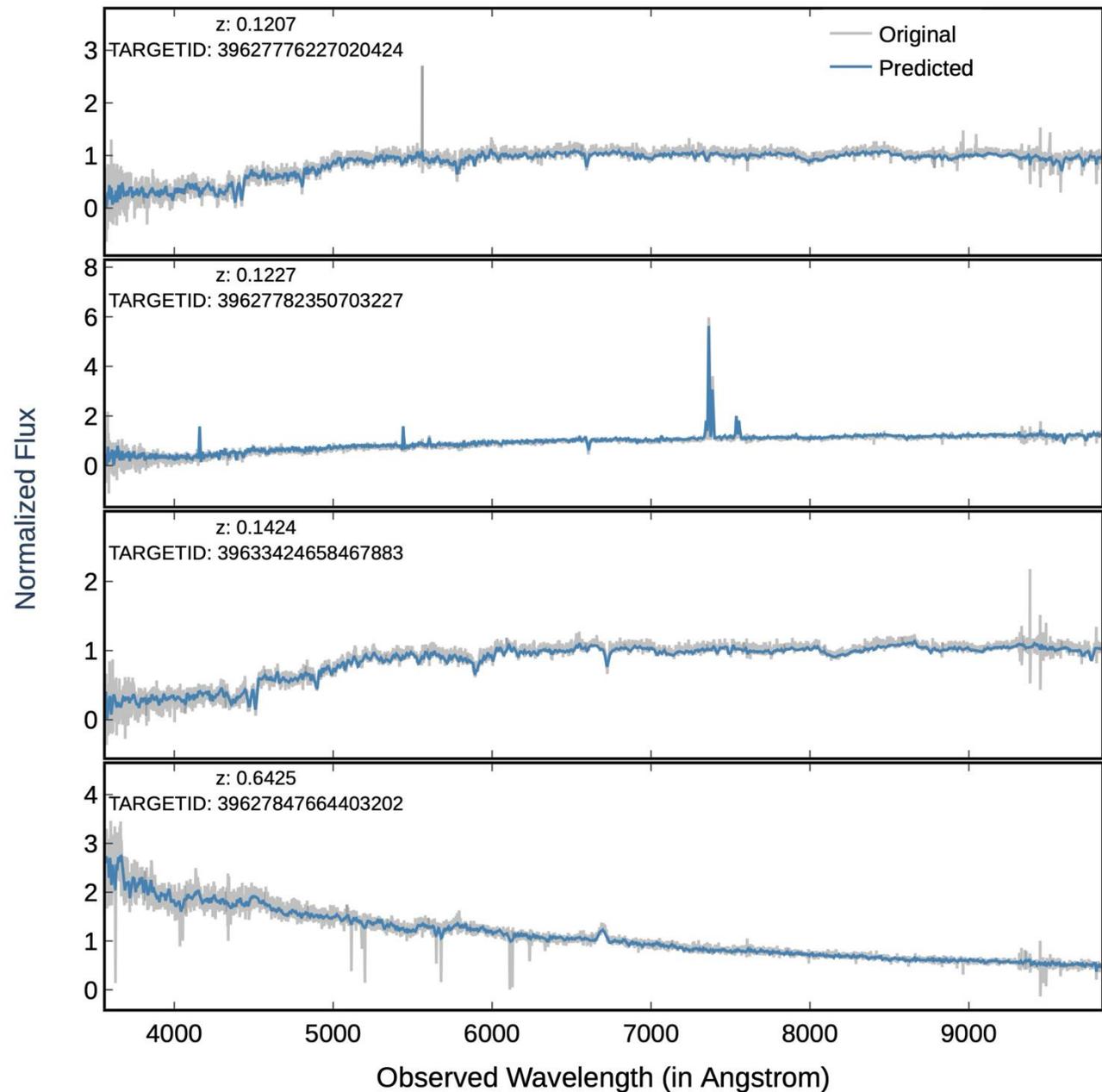


**Results:**

**Spectrum**

**Reconstruction**

**(BGS/low-z)**



Does not reconstruct  
sky lines (noise)!

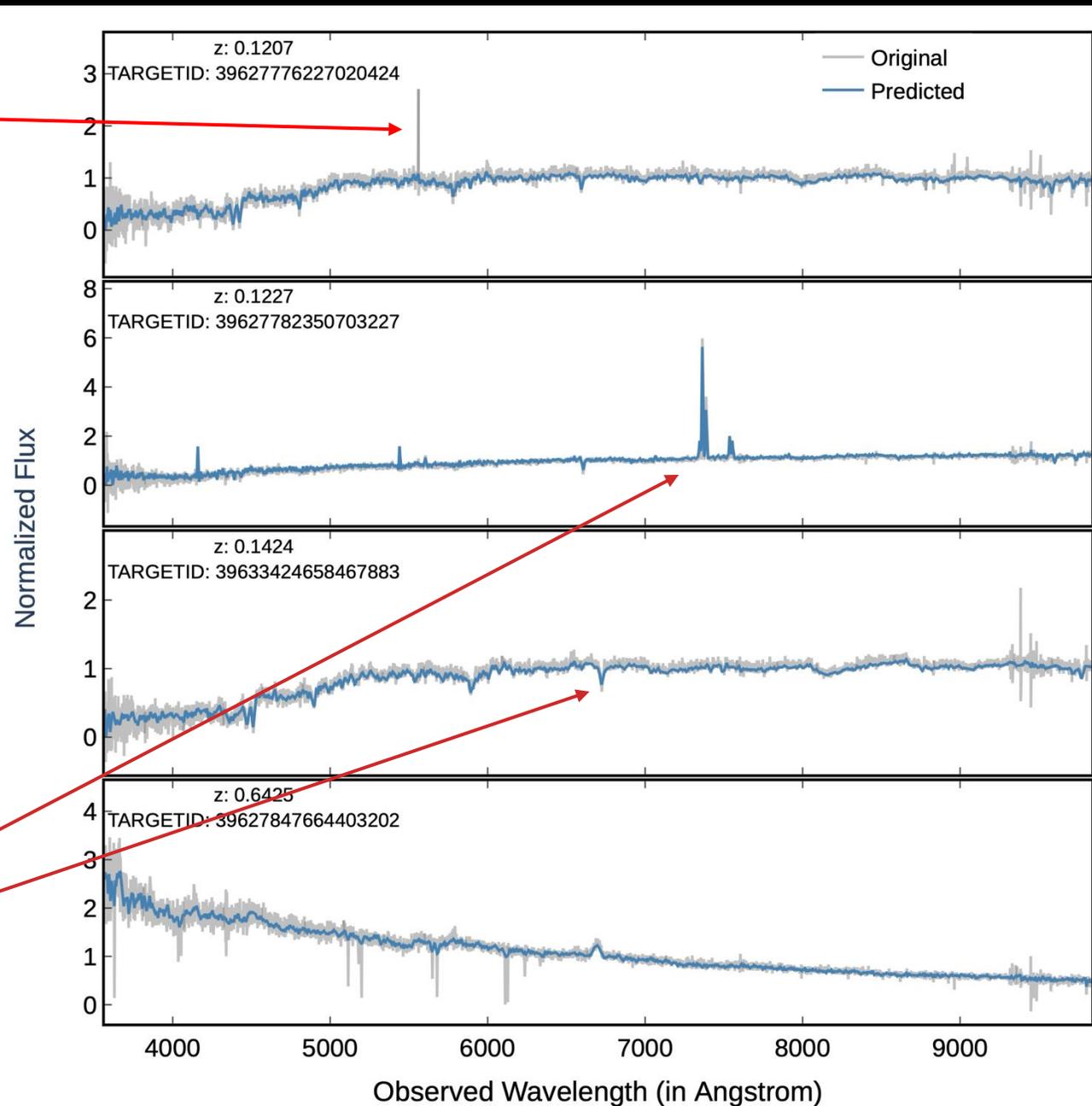
**Results:**

**Spectrum**

**Reconstruction**

**(BGS/low-z)**

Accurately reconstructs  
emission & absorption  
lines, along with the  
continuum shape

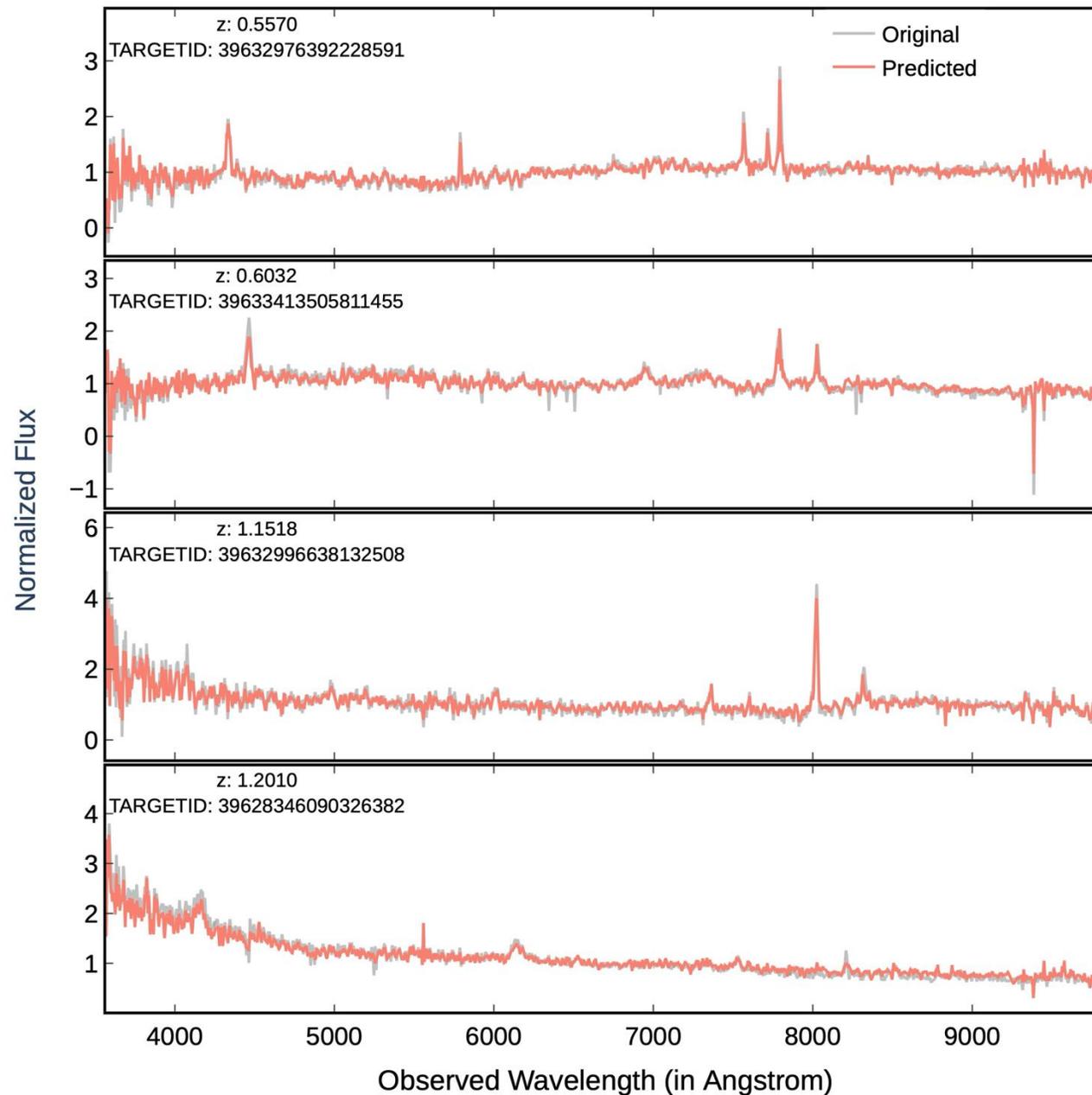


**Results:**

**Spectrum**

**Reconstruction**

**(ELG/high-z)**



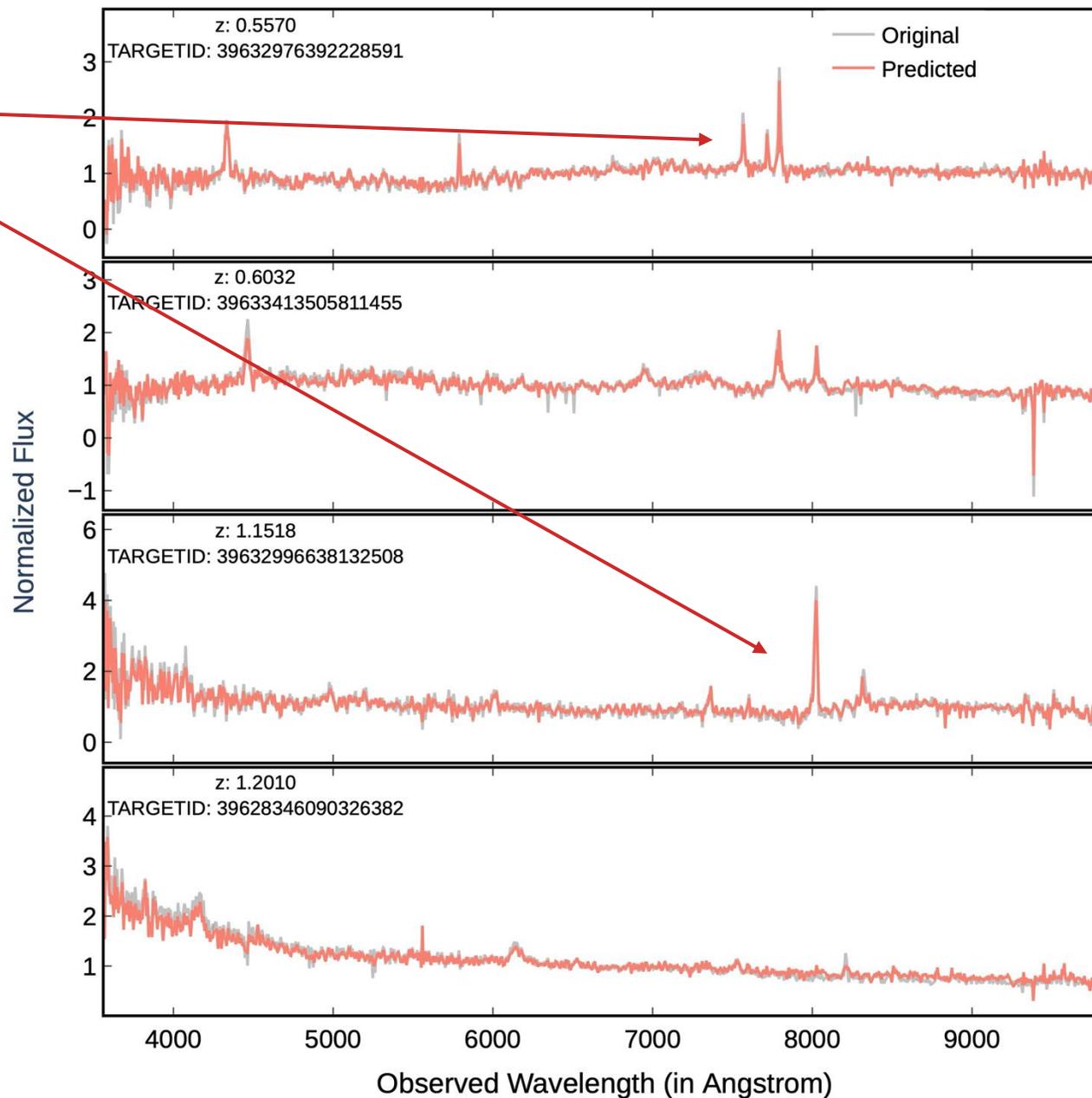
Reconstructs emission & absorption lines, and the continuum shape accurately like the low-z sample!

**Results:**

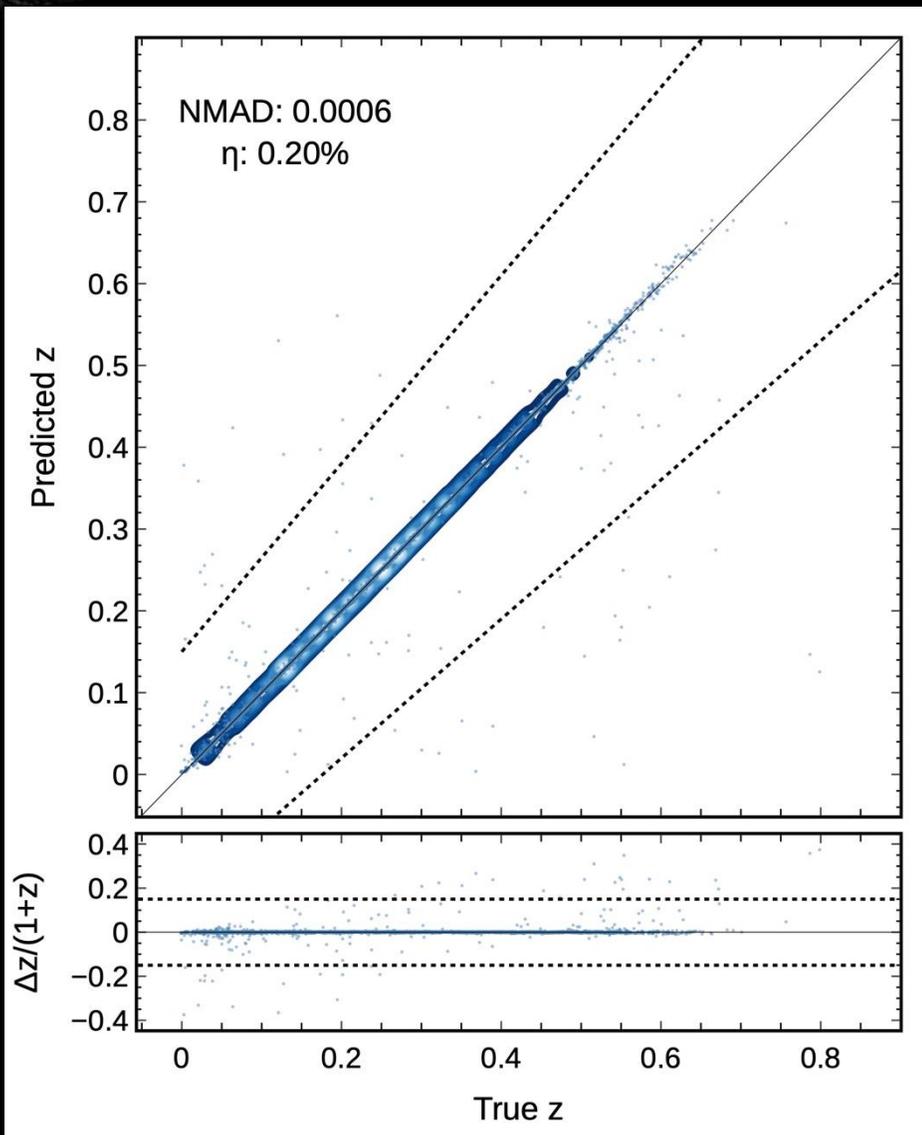
**Spectrum**

**Reconstruction**

**(ELG/high-z)**

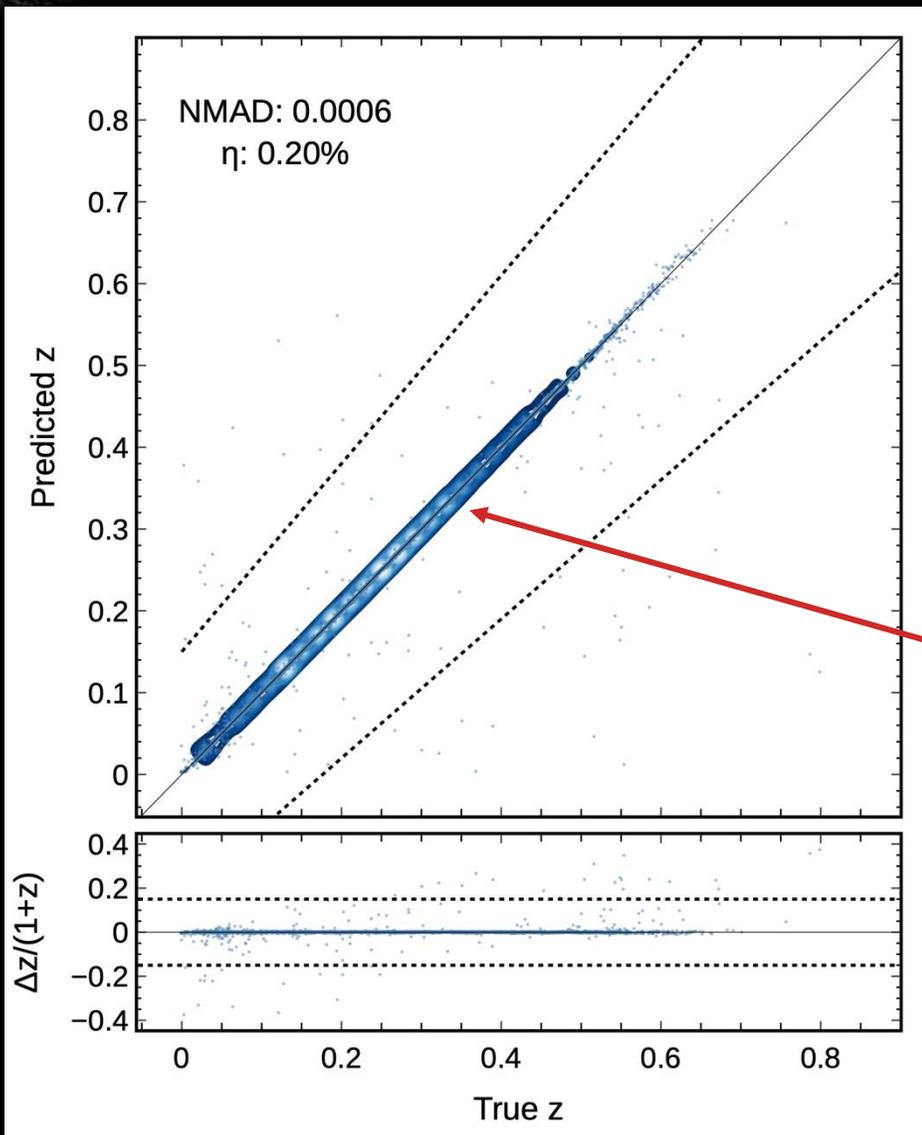


# Results: Redshift Measurement (BGS/low-z)



$$\text{NMAD} = 1.48 \times \text{median} \left( \left| \frac{\Delta z}{1 + z_{\text{true}}} \right| \right)$$

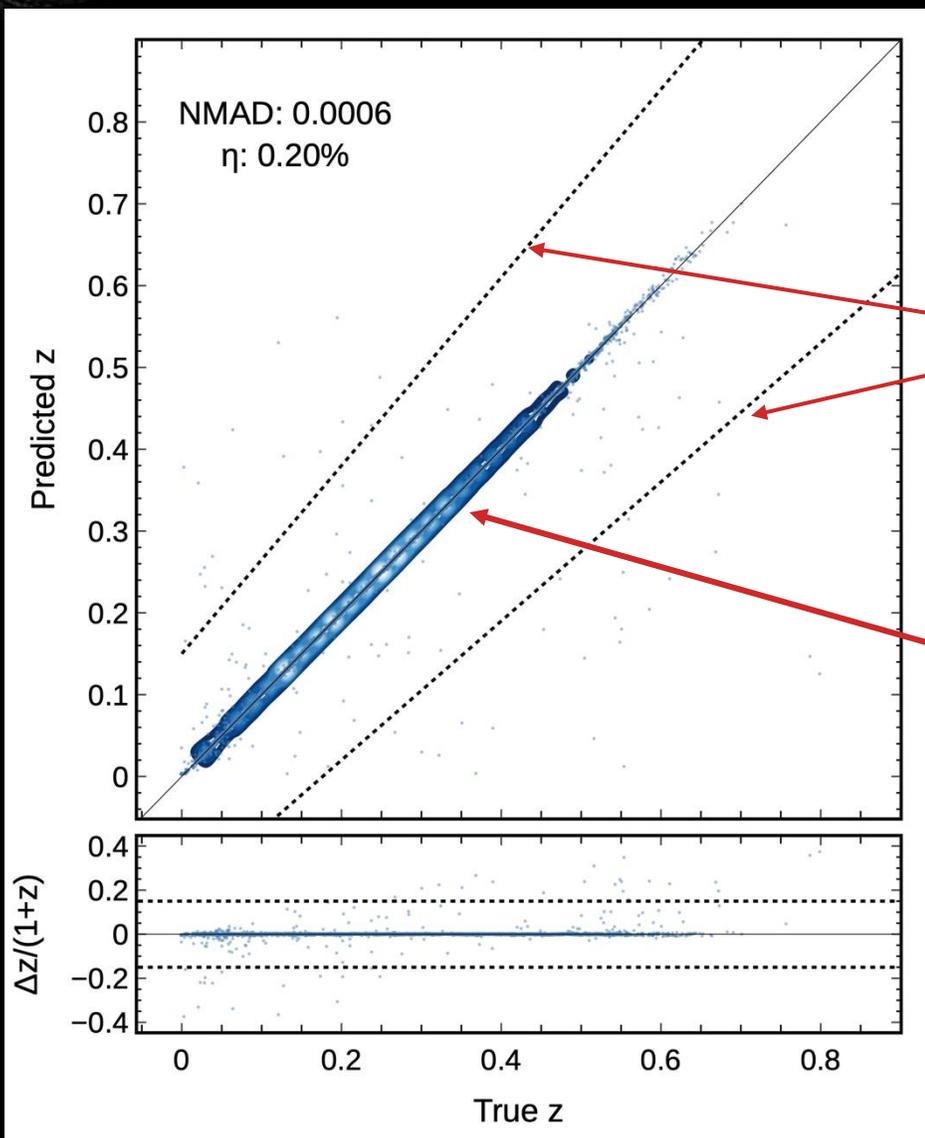
# Results: Redshift Measurement (BGS/low-z)



$$\text{NMAD} = 1.48 \times \text{median} \left( \left| \frac{\Delta z}{1 + z_{\text{true}}} \right| \right)$$

Majority of the points lie on y=x line!

# Results: Redshift Measurement (BGS/low-z)

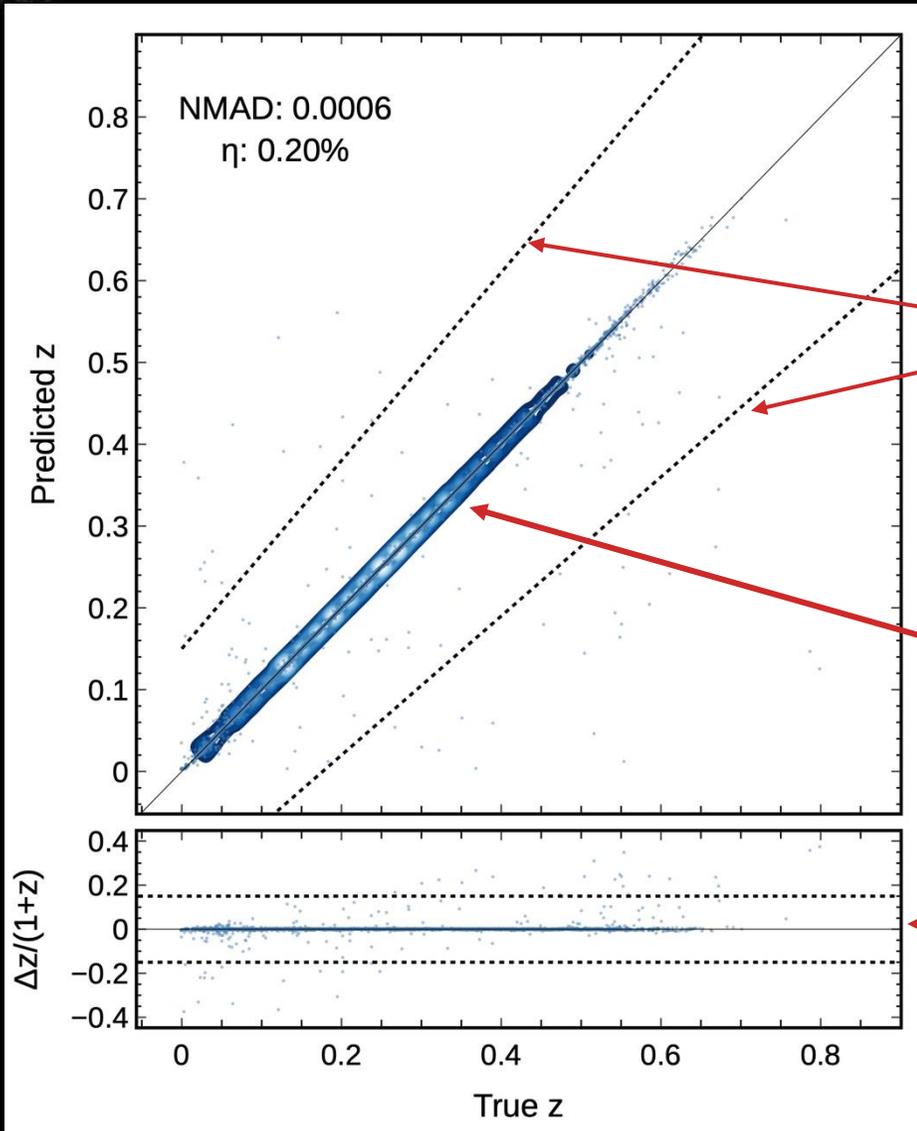


$$\text{NMAD} = 1.48 \times \text{median} \left( \left| \frac{\Delta z}{1 + z_{\text{true}}} \right| \right)$$

- $\eta$  (catastrophic outliers) =  $|\Delta z| / (1 + z) > 0.15$
- Only 0.2% of 19,345 low-z test spectra were catastrophic outliers!

Majority of the points lie on  $y=x$  line!

# Results: Redshift Measurement (BGS/low-z)



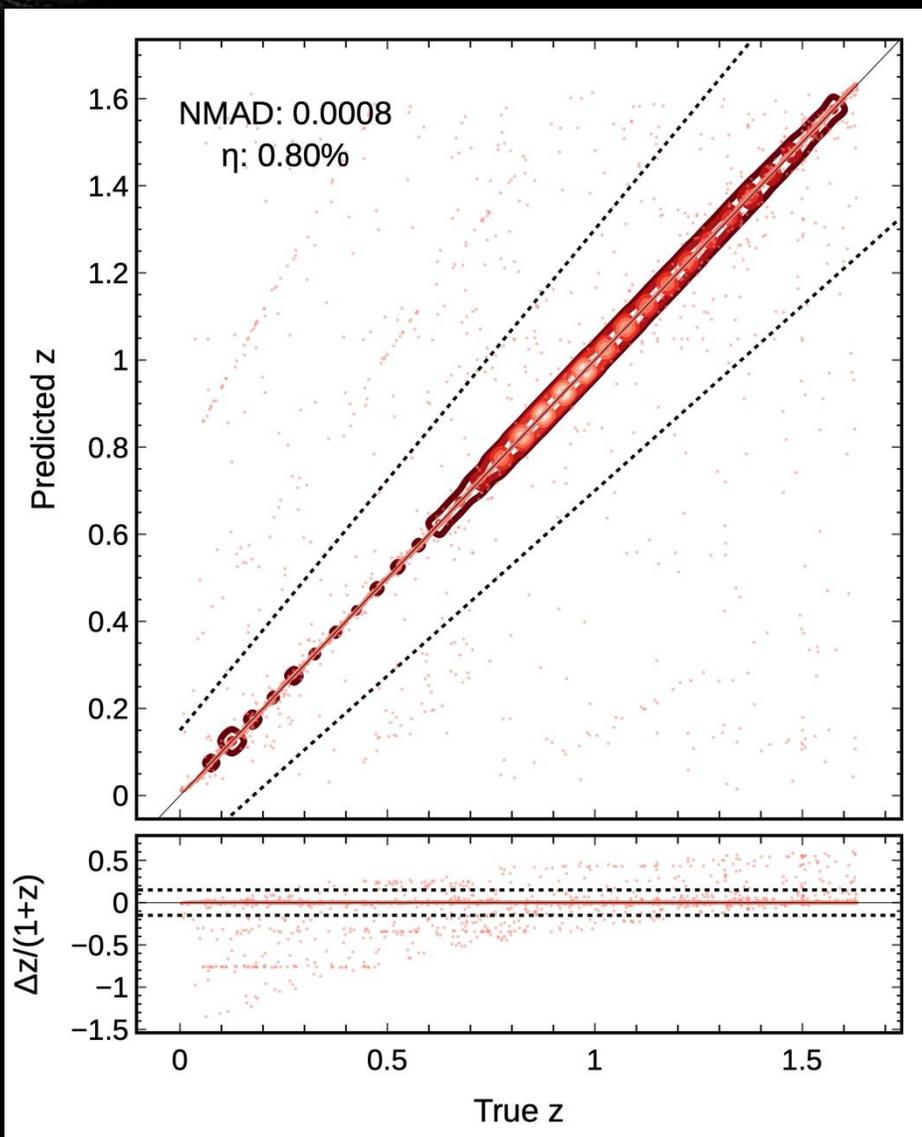
$$\text{NMAD} = 1.48 \times \text{median} \left( \left| \frac{\Delta z}{1 + z_{\text{true}}} \right| \right)$$

- $\eta$  (catastrophic outliers) =  $|\Delta z| / (1 + z) > 0.15$
- Only 0.2% of 19,345 low-z test spectra were catastrophic outliers!

Majority of the points lie on  $y=x$  line!

Same performance is also reflected in the distribution of normalized redshift residuals ( $\Delta z / (1 + z)$ ) plot below

# Results: Redshift Measurement (ELG/high-z)



Achieves comparable performance for the high-redshift sample as well!

# Summary

- **The Problem:** Traditional spectroscopic redshift measurement methods are time-intensive and struggle with large datasets and noisy spectra.

# Summary

- **The Problem:** Traditional spectroscopic redshift measurement methods are time-intensive and struggle with large datasets and noisy spectra.
- **The Solution:** Introduced SpecPT, a novel transformer-based architecture for spectroscopic analysis, focusing on spectral reconstruction and accurate redshift prediction.

# Summary

- **The Problem:** Traditional spectroscopic redshift measurement methods are time-intensive and struggle with large datasets and noisy spectra.
- **The Solution:** Introduced SpecPT, a novel transformer-based architecture for spectroscopic analysis, focusing on spectral reconstruction and accurate redshift prediction.
- **Key Results and Insights:**
  - **Spectral Reconstruction:**
    - Autoencoder captured key spectral features, reduced noise, and avoided artifacts like skylines.
    - Demonstrated strong generalization across diverse datasets.
  - **Redshift Measurement:**
    - Achieved NMAD values of  $0.0006$  (BGS) and  $0.0008$  (ELG).
    - Outlier fractions:  $0.20\%$  (BGS) and  $0.80\%$  (ELG).
    - Delivered robust performance across varied redshift ranges.

## Future Work

- Combine BGS, ELG, LRG, and QSO datasets for balanced redshift distribution.
- Apply SpecPT to other datasets (e.g., HST, DEIMOS).
- Extend applications to outlier detection and ISM property estimation. Scale SpecPT to full DESI dataset for next-generation spectroscopic surveys.

FIRST PAPER OUT ON ARXIV NOW!  
<https://arxiv.org/abs/2501.01070>



**Thank You for  
listening!  
Questions?**

Contact information-

Email: [rp2503@rit.edu](mailto:rp2503@rit.edu)

Linkedin:

<https://www.linkedin.com/in/rohan-pattnaik/>

Paper link-

<https://arxiv.org/abs/2501.01070>

