Exploring the connection between optical and X-ray emissions in Active Galactic Nuclei

Rebeka L. Bottger[*1,2,3] bottger.rebeka@gmail.com
Rocco Di Tella[*1,2] rocco.di_tella@cfa.harvard.edu
Carolina Cuesta-Lazaro[4,5]
Jason Reeves[6]
Nico Cappelluti[3]
Rafael Martínez-Galarza[1,2]

[*]Equal contribution, author order was determined by a coin toss.
[1]Center for Astrophysics | Harvard & Smithsonian [2]AstroAI
[3]Department of Physics, University of Miami
[4]Flatiron Institute, Simons Foundation [5]Institute for Advanced Studies
[6]Department of Astronomy, University of Illinois Urbana-Champaign

Abstract

Supermassive black hole accretion is most directly visible in X-rays, while host galaxies are richly observed in optical and infrared surveys. Quantifying how much X-ray information is shared with these non-X-ray modalities constrains how accretion, gas, dust, and galaxy structure are observationally connected. We attach a probabilistic normalizing-flow head to frozen AION multimodal embeddings and predict eROSITA broad-band X-ray flux from DESI spectra, Legacy Survey imaging, WISE mid-infrared photometry, and redshift. The full multimodal model reaches $R^2 = 0.549$ and $\exp(\mathrm{IG}) = 1.405$, outperforming a classical optical emission-line baseline ($R^2 = 0.392$). Information gain depends on the available modality combination, source class, and redshift: spectra and imaging dominate at low redshift, WISE and redshift become more informative at high redshift, and redshift adds little once spectra are present.

1 Introduction

Understanding the co-evolution of galaxies and their central supermassive black holes (SMBH) remains one of the central problems in extragalactic astrophysics (Hickox and Alexander [2018], Netzer [2015]). Observational findings suggest that processes occurring in the immediate vicinity of the black hole are linked to galaxy-wide evolution (Madau and Dickinson [2014]). One of the most prominent examples is the well-established correlation between black hole mass and the stellar velocity dispersion of the host galaxy bulge (the $M_{\mathrm{BH}} - \sigma$ relation; (Ferrarese and Merritt 2000, Gebhardt et al. 2000) suggesting a coupled evolutionary history. However, the physical mechanisms that underpin this connection remain unclear, particularly given the large separation in spatial scales between black hole accretion-driven phenomena and global galaxy properties.

Optical observations (in the visible wavelengths) reflect large-scale galaxy properties such as stellar kinematics, star formation, and morphology, while X-ray emission probes high-energy processes in the nuclear region of the galaxy, such as black hole accretion. Despite this scale separation, empirical correlations between X-ray luminosity and optical emission-line diagnostics (e.g., [O III]) have been reported (Heckman et al. [2005], Panessa et al. [2006]), indicating a link between the central engine and gas on larger scales. While optical and infrared emission are tightly related through dust reprocessing, the optical–X-ray connection is more elusive due to obscuration and anisotropy.

Recent work has shown that pre-trained machine learning models can use optical data to accurately predict infrared observables much more effectively than classical methods [Jespersen et al., 2026].

Conference on Physics and AI at Stanford University (PAI 2026).

Foundation models represent a new way of doing empirical science, especially apt to relate large dimensional structures that are still poorly understood. Astronomical data is sparse, plentiful and spread across multiple scales, wavelengths and modalities. AION-1 was pre-trained in a self-supervised regime across multiple astronomical modalities, including images, spectra and photometry, using modality-specific tokenizers followed by a shared transformer encoder [Parker et al., 2025]. We use this pre-trained multimodal encoder with an attached probabilistic head to ask how much information do combinations of input modalities carry about observed X-ray flux.

In this work, we predict eROSITA X-ray flux from DESI optical spectra, Legacy Survey imaging, WISE mid-infrared photometry, and redshift. These modalities trace complementary physics: spectra capture emission lines produced by ionized gas, including high-ionization lines such as OIII and NeV and Balmer recombination lines such as Hα and Hβ [Kewley et al., 2019]; imaging captures spatial structure, morphology, and central concentration; and WISE photometry measures mid-infrared dust emission heated by AGN activity or star formation [WISE Team, 2020, Wright et al., 2010]. By testing different modality combinations, we quantify how much information about accretion-linked X-ray emission is shared across these survey views.[1]

2 Methods

2.1 Data

We predict X-ray flux using four different modalities as predictors: DESI optical spectra (S), scalar redshift (Z), WISE mid-infrared photometry (W), and Legacy Survey imaging (I). X-ray source properties are taken from the first eROSITA all-sky survey catalogue (eRASS1; Merloni et al. 2024). Optical spectra and redshifts come from DESI DR1 [DESI Collaboration et al., 2026]; mid-infrared photometry uses WISE W1–W3 fluxes [Wright et al., 2010, WISE Team, 2020]; and optical images are Legacy Survey DR10 *griz* cutouts [Dey et al., 2019]. Cross-matching these sources by sky position within 3 arcsec yields a sample of 30,762 sources. We adopt an 80/10/10 train/validation/test split, stratified by redshift and grouped so repeated measurements of the same source stay in the same fold.

Sources are also labeled as either `QSO` (Quasi Stellar Object) or `GALAXY`. Since these are X-ray visible galaxies, we know they have an Active Galactic Nuclei (AGN).

As a baseline, we train a separate model on four catalogue emission-line fluxes: [O III] $\lambda5007$, [Ne V] $\lambda3426$, Hα, and Hβ. These lines are common optical diagnostics of AGN activity and ionized gas [Kewley et al., 2019]; previous work has also shown empirical correlations between [O III] and X-ray emission [Heckman et al., 2005, Panessa et al., 2006]. The baseline tests whether learned multimodal representations add information beyond these classical features.

2.2 Model

We use the frozen AION-base encoder (300M params), a transformer-based multimodal foundation model [Parker et al., 2025]. The input consists of 576 image tokens, 273 spectral tokens, 3 WISE tokens, and 1 redshift token, each of dimension 768. Missing modalities are omitted, so each input combination returns a $T \times 768$ token sequence.

Each token first receives a learned per-modality affine transformation. We then summarize the variable-length sequence into a fixed-size representation with an attention-pooling head (Fig. 1): four learned query vectors, conditioned on which modalities are present, attend to the AION tokens over two 8-head pooling blocks with residual connections, yielding a 4×768 summary. An MLP ($3072 \to 512 \to 512 \to 256$) maps this to a 256-dimensional context vector C that conditions a one-dimensional neural spline flow (NSF; Durkan et al. 2019) implemented with Zuko [Rozet et al., 2022], yielding a full conditional density rather than a point estimate. We train one joint model across all 15 non-empty combinations of S, Z, W, and I.

We also train two supporting probabilistic models: an unconditional one-dimensional KDE prior over the training targets with Scott bandwidth [Scott, 1992], to use as a prior and a NSF with the four spectral-line conditioning as a baseline.

[1] Code is available at `https://github.com/RoccoDiTella/erosita-desi-aion-flow`.

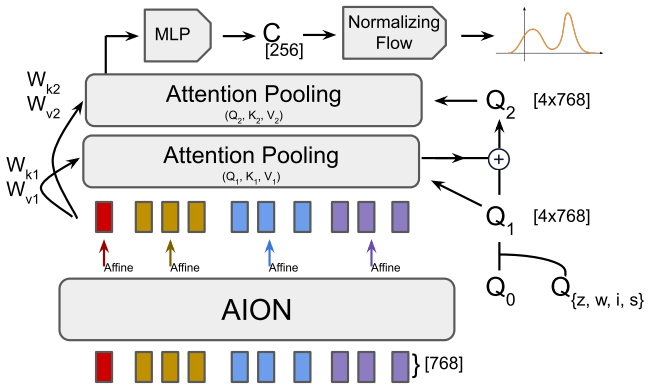

Figure 1: Model architecture. A presence-conditioned query pools the affine-calibrated AION tokens through two attention pooling layers into a 4×768 summary, which an MLP maps to a 256-dimensional context C conditioning the neural spline flow.

2.3 Metrics

For each modality group g and test object i, the flow gives the conditional log density of the true standardized target, $\log p_\theta(y_i \mid g)$. We compare it to an unconditional KDE prior density $\log p_{\mathrm{KDE}}(y_i)$ fit only on the training targets, defining $\mathrm{IG}_g = N^{-1} \sum_{i=1}^{N} [\log p_\theta(y_i \mid g) - \log p_{\mathrm{KDE}}(y_i)]$. IG is reported in nats, and positive values mean that the conditional model assigns higher likelihood to the true X-ray flux than the target-only prior. We report $\exp(\mathrm{IG})$ as the geometric mean likelihood-gain factor; for example, $\exp(\mathrm{IG}) = 1.4$ means the observed flux is 1.4 times more likely than under the KDE prior. Finally, we report R^2 as a familiar point-estimate diagnostic, using the inverse-standardized posterior sample mean as the prediction.

3 Results

Table 1 reports R^2, information gain, and $\exp(\mathrm{IG})$ for each input combination and the spectral-line baseline. Multimodal learning improves over single-modality inputs, but the value of a modality depends on which other information is already available.

Using the full spectra outperforms the classical spectral-line baseline ($R^2 = 0.480$ vs. 0.392), supporting that learned representations carry information beyond traditional features. The best performance comes from combining all input modalities ($R^2 = 0.549$, $\mathrm{IG} = 0.340$, $\exp(\mathrm{IG}) = 1.405$). Not all modalities are complementary: for example, adding redshift when spectra are already present does not detectably change performance, as expected since the spectra already carry redshift information.

Figure 2 shows the same complementarity as a function of redshift: spectra and images perform better for lower-redshift objects, which are closer and easier to observe, while redshift and WISE contribute more at higher redshift, where AGN-heated dust contributes more strongly to the WISE W1 band [Landt, 2023].

When breaking down results by object type, we observe that AGNs benefit more from combining multiple modalities than QSOs do (an information-gain improvement of $\sim$0.08 vs $\sim$0.04 nats). QSOs are typically more luminous and less host-dominated [Netzer, 2015, Padovani et al., 2017], so their spectra already trace accretion well. AGN objects can have stronger host-galaxy emission, dust obscuration, and star-formation contamination, making the optical or mid-infrared link to X-ray flux less direct. This is consistent with imaging and photometry helping isolate accretion-linked emission from host-galaxy light [Kewley et al., 2019, Comerford et al., 2020]. The information gained from a spectrum is therefore not a fixed value: it depends on the source population and on whether the X-ray-relevant optical emission can be isolated from other physical processes.

z	Spectra	WISE	Images	R^2	Information Gain	exp(IG)
Z				**0.140** ± 0.013	**0.045** ± 0.009	**1.046** ± 0.010
		W		**0.302** ± 0.014	**0.166** ± 0.011	**1.180** ± 0.013
Z		W		**0.314** ± 0.015	**0.175** ± 0.011	**1.191** ± 0.013
			I	**0.418** ± 0.021	**0.248** ± 0.014	**1.281** ± 0.018
Z			I	**0.455** ± 0.020	**0.290** ± 0.014	**1.336** ± 0.019
Z	S			**0.475** ± 0.021	**0.275** ± 0.016	**1.317** ± 0.020
		W	I	**0.477** ± 0.020	**0.291** ± 0.015	**1.337** ± 0.020
	S			**0.480** ± 0.020	**0.278** ± 0.016	**1.320** ± 0.021
Z		W	I	**0.498** ± 0.019	**0.317** ± 0.015	**1.373** ± 0.020
Z	S	W		**0.507** ± 0.019	**0.301** ± 0.015	**1.351** ± 0.021
	S	W		**0.515** ± 0.019	**0.306** ± 0.016	**1.358** ± 0.021
Z	S		I	**0.525** ± 0.020	**0.317** ± 0.017	**1.373** ± 0.023
	S		I	**0.530** ± 0.020	**0.315** ± 0.017	**1.370** ± 0.023
Z	S	W	I	**0.549** ± 0.019	**0.340** ± 0.017	**1.405** ± 0.023
	S	W	I	**0.554** ± 0.019	**0.344** ± 0.016	**1.411** ± 0.023

Baseline Spectral Lines	R^2	Information Gain	exp(IG)
[O III] 5007 + [Ne V] 3426 + Hα + Hβ	**0.392** ± 0.021	**0.218** ± 0.012	**1.243** ± 0.015

Table 1: Predicting eROSITA broad-band X-ray flux from redshift (Z), optical spectra (S), WISE mid-infrared photometry (W), and Legacy Survey imaging (I). Rows indicate available input modalities. Metrics are R^2, information gain (IG) relative to a KDE prior, and $\exp(\mathrm{IG})$, the expected likelihood-gain factor; along with 1σ bootstrap intervals. The bottom row shows the classical optical emission-line baseline.

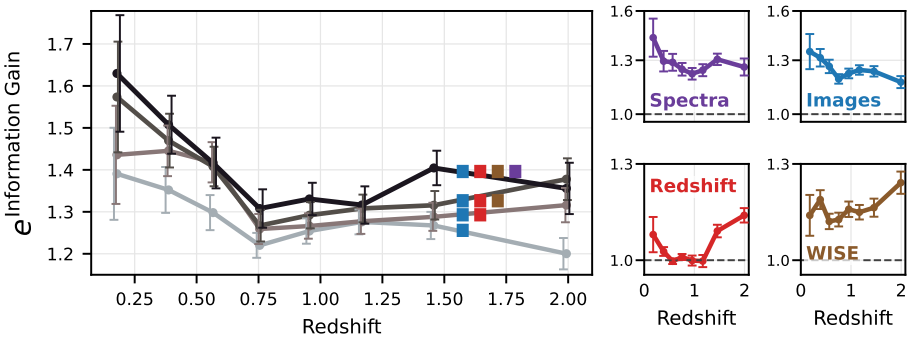

Figure 2: **Model performance as a function of redshift.** Right panels show single-modality performance for spectra (S), imaging (I), redshift (Z), and WISE photometry (W) as a function of redshift. The main panel shows likelihood-gain factor as a function of redshift for I, I+Z, I+Z+W, and I+Z+W+S; colored squares identify each modality combination. Error bars are 1σ bootstrap intervals.

4 Conclusions

Optical and X-ray emission carry information about different physical scales. Our approach places lower bounds on the shared information between these modalities, achieves higher predictive performance than traditional emission-line baselines, and supports the use of astronomical foundation models as scientific instruments for cross-modal inference. Such cross-modal models can map where the optical, mid-infrared, and X-ray signatures of AGN are connected and which observations are most informative for studying the co-evolution of supermassive black holes and their host galaxies.

References

Julia M. Comerford, James Negus, Francisco Müller-Sánchez, Michael Eracleous, Dominika Wylezalek, Thaisa Storchi-Bergmann, Jenny E. Greene, R. Scott Barrows, Rebecca Nevin, Namrata Roy, and Aaron Stemo. A Catalog of 406 AGNs in MaNGA: A Connection between Radio-mode AGNs and Star Formation Quenching. *ApJ*, 901(2):159, 2020. doi: 10.3847/1538-4357/abb2ae.

DESI Collaboration, M. Abdul Karim, A. G. Adame, D. Aguado, et al. Data Release 1 of the Dark Energy Spectroscopic Instrument. *The Astronomical Journal*, 171(5):285, 2026. doi: 10.3847/1538-3881/ae4c43.

Arjun Dey, David J. Schlegel, Dustin Lang, Robert Blum, Kaylan Burleigh, Xiaohui Fan, et al. Overview of the DESI Legacy Imaging Surveys. *Astronomical Journal*, 157(5):168, 2019. doi: 10.3847/1538-3881/ab089d.

Conor Durkan, Artur Bekasov, Iain Murray, and George Papamakarios. Neural Spline Flows. In *Advances in Neural Information Processing Systems*, volume 32, 2019. URL `https://arxiv.org/abs/1906.04032`.

Laura Ferrarese and David Merritt. A fundamental relation between supermassive black holes and their host galaxies. *ApJ*, 539:L9, 2000.

Karl Gebhardt et al. A relationship between nuclear black hole mass and galaxy velocity dispersion. *ApJ*, 539:L13–L16, 2000. doi: 10.1086/312840.

Timothy M. Heckman, A. Ptak, A. Hornschemeier, and G. Kauffmann. The relationship of hard x-ray and optical line emission in low-redshift active galactic nuclei. *ApJ*, 634:161–168, 2005. doi: 10.1086/491665.

Ryan C. Hickox and David M. Alexander. Obscured active galactic nuclei. *ARA&A*, 56:625, 2018.

Christian K. Jespersen, Peter Melchior, David N. Spergel, Andy D. Goulding, ChangHoon Hahn, and Kartheik G. Iyer. The Optical and Infrared Are Connected. *The Astrophysical Journal*, 1002(2): 132, 2026. doi: 10.3847/1538-4357/ae5c9a.

Lisa J. Kewley, David C. Nicholls, and Ralph S. Sutherland. Understanding Galaxy Evolution Through Emission Lines. *Annual Review of Astronomy and Astrophysics*, 57:511–570, August 2019. doi: 10.1146/annurev-astro-081817-051832.

Hermine Landt. The outer dusty edge of accretion disks in active galactic nuclei. *Frontiers in Astronomy and Space Sciences*, 10:1256088, 2023. doi: 10.3389/fspas.2023.1256088.

Piero Madau and Mark Dickinson. Cosmic Star-Formation History. *Annual Review of Astronomy and Astrophysics*, 52:415–486, August 2014. doi: 10.1146/annurev-astro-081811-125615.

A. Merloni, G. Lamer, T. Liu, M. E. Ramos-Ceja, H. Brunner, E. Bulbul, et al. The SRG/eROSITA all-sky survey: First X-ray catalogues and data release of the Western Galactic hemisphere. *Astronomy & Astrophysics*, 682:A34, 2024. doi: 10.1051/0004-6361/202347165.

Hagai Netzer. Revisiting the unified model of active galactic nuclei. *ARA&A*, 53:365, 2015.

Paolo Padovani et al. Active galactic nuclei: what's in a name? *Astronomy & Astrophysics Review*, 25:2, 2017. doi: 10.1007/s00159-017-0102-9.

F. Panessa, L. Bassani, M. Cappi, M. Dadina, X. Barcons, F. J. Carrera, L. C. Ho, and K. Iwasawa. On the x-ray, optical emission line and black hole mass properties of local seyfert galaxies. *A&A*, 455:173–185, 2006. doi: 10.1051/0004-6361:20064894.

Liam Parker, Francois Lanusse, Jeff Shen, Ollie Liu, Tom Hehir, Leopoldo Sarra, et al. AION-1: Omnimodal Foundation Model for Astronomical Sciences, 2025. URL `https://arxiv.org/abs/2510.17960`.

François Rozet et al. Zuko: Normalizing flows in PyTorch, 2022. URL `https://pypi.org/project/zuko/`.

David W. Scott. *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley & Sons, 1992.

WISE Team. WISE All-Sky Source Catalog. NASA IPAC DataSet, IRSA142, January 2020.

Edward L. Wright, Peter R. M. Eisenhardt, Amy K. Mainzer, Michael E. Ressler, Roc M. Cutri, Thomas Jarrett, J. Davy Kirkpatrick, Deborah Padgett, Robert S. McMillan, Michael Skrutskie, S. A. Stanford, Martin Cohen, Russell G. Walker, John C. Mather, David Leisawitz, Thomas N. Gautier, III, Ian McLean, Dominic Benford, Carol J. Lonsdale, Andrew Blain, Bryan Mendez, William R. Irace, Valerie Duval, Fengchuan Liu, Don Royer, Ingolf Heinrichsen, Joan Howard, Mark Shannon, Martha Kendall, Amy L. Walsh, Mark Larsen, Joel G. Cardon, Scott Schick, Mark Schwalm, Mohamed Abid, Beth Fabinsky, Larry Naes, and Chao-Wei Tsai. The Wide-field Infrared Survey Explorer (WISE): Mission Description and Initial On-orbit Performance. *Astronomical Journal*, 140(6):1868–1881, December 2010. doi: 10.1088/0004-6256/140/6/1868.