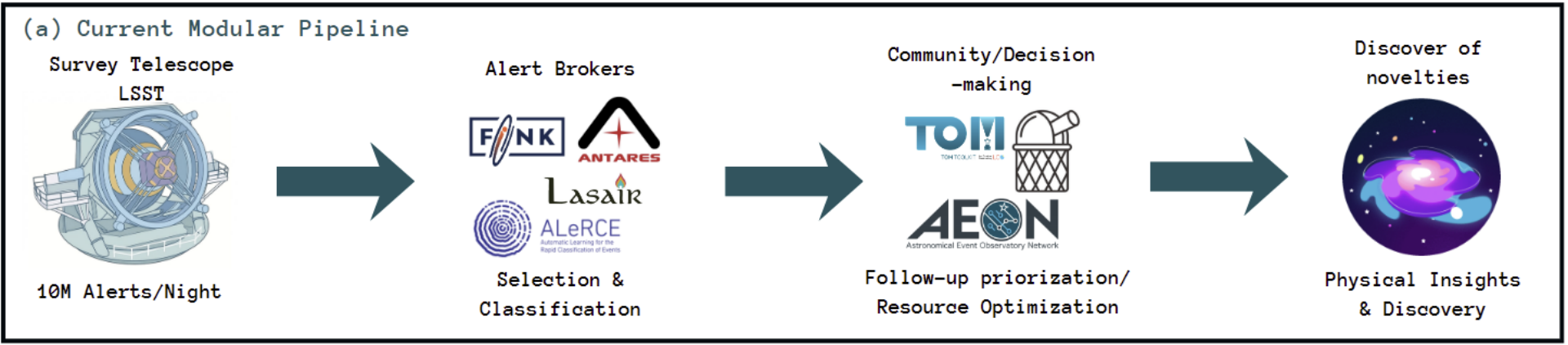
Centro Brasileiro de Pesquisas Físicas

TOWARD DECISION-AWARE AI FOR LSST-SCALE TIME-DOMAIN ASTRONOMY

Clecio R. Bom,
for TVS Science Collaboration
debom@cbpf.br



The General Time-Domain paradigm



- "Classification Centric"
- Great if you can look and make a complete assessment of the full stream, but How about milions of alerts?

§1 · THE RUBIN ERA

THE FOLLOW-UP BOTTLENECK

Q1

Which transients should we prioritize for follow-up?

Q2

How unique or interesting is each transient?

Q3

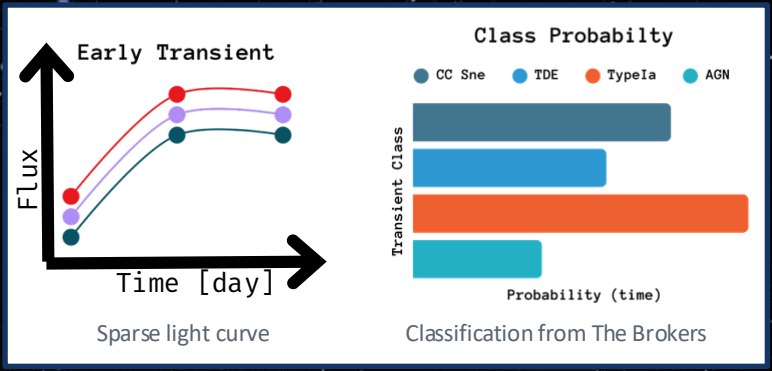
What is the best observing strategy — IR? Spectroscopy?

THE QUESTIONS

How do we optimally allocate scarce telescope resources in real time — balancing novelty, rarity, and population completeness?

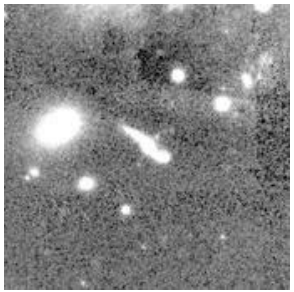
What if we could probe the most interesting explosive transients automatically?

AN AMBIGUOUS EARLY TRANSIENT



Sparse light curve

Classification from The Brokers



Host galaxy (g-band) — nuclear or offset? Star-forming or quiescent?

Example Detected in two bands over 48 hours — consistent with four different physical events:

- **Young Type Ia supernova**
- **Core-collapse supernova**
- **Tidal disruption event**
- **Unusual AGN flare**

> Its scientific value depends less on the *most probable class* than on whether timely follow-up can **distinguish these interpretations** before diagnostic information is lost.

At LSST scale, such cases are routine.

§1 · THE RUBIN ERA

WHY PER-ALERT ACCURACY IS NOT ENOUGH

IRREVERSIBLE

Missing early colour evolution of a rare core-collapse supernova

The diagnostic window closes within hours — the information is gone forever.

LOW PENALTY

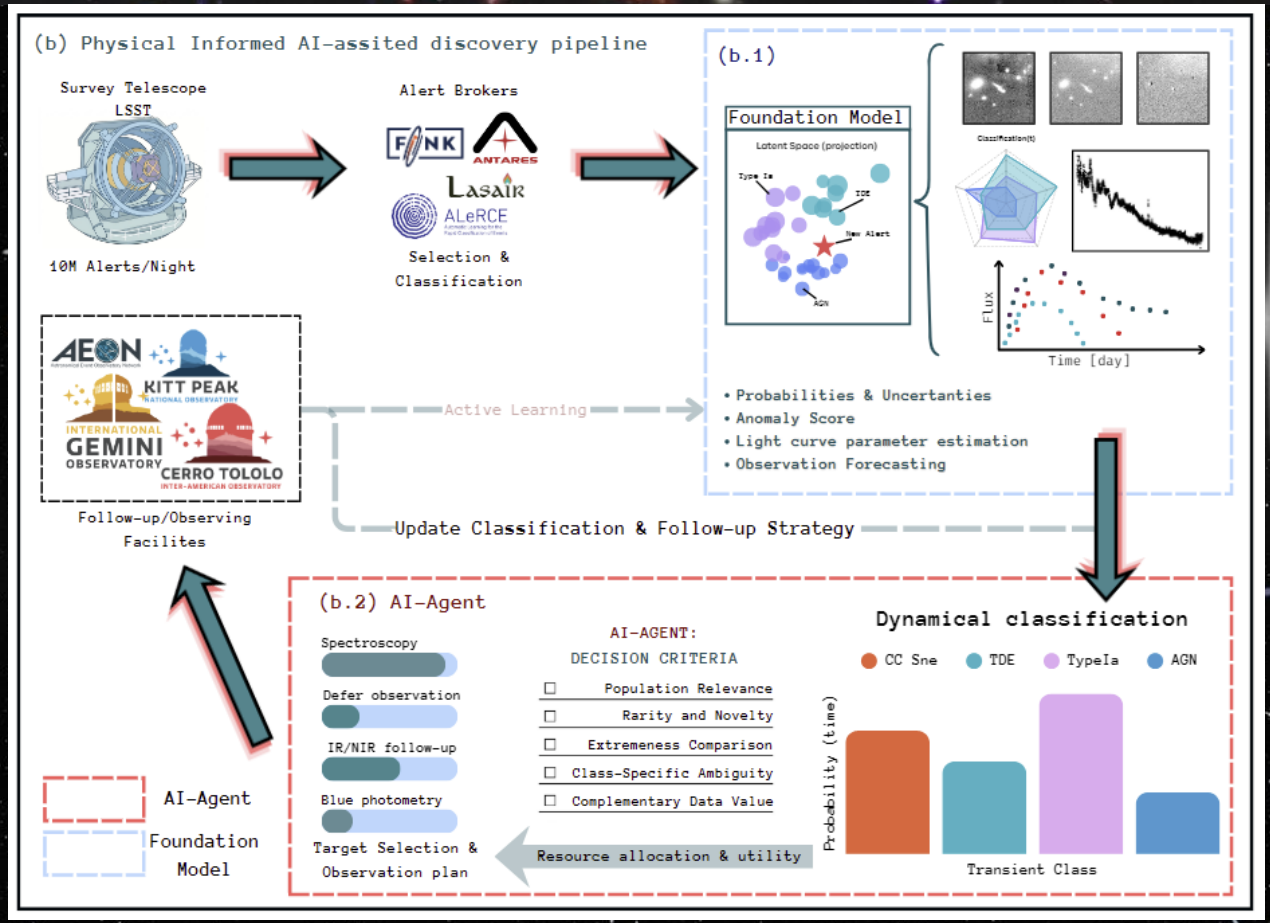
Delaying spectroscopy of a well-understood Type Ia supernova

Little scientific cost — the population is already well characterized.

The cost of a wrong action is asymmetric — per-alert accuracy encodes none of this structure.

The AI-Assisted Time-Domain

- Science Question driven
- Reasoning and personalized recomendation, plans of efective follow-up and Science relevance at scale.



Foundation Models as Infrastructure AI

Foundation Models

- A single **large model** pretrained on **massive, diverse data**
- Learns general representations
- Reused for many downstream tasks with little or no retraining

Why?

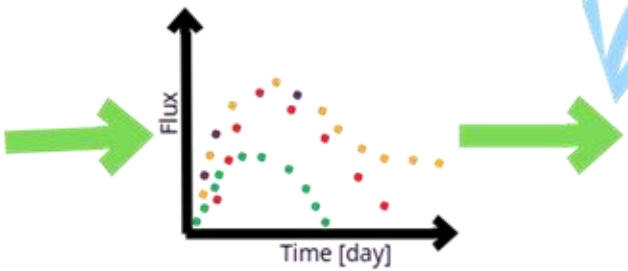
- Replace many task-specific models
- Exploit **unlabeled data** (self-supervised learning)
- Enable **transfer learning** across domains
- Better **scaling behavior** with more data & parameters

WHISPER



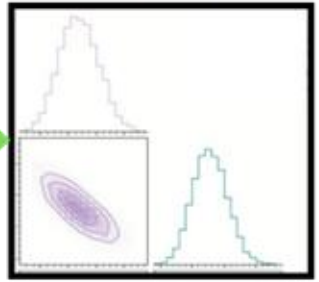
- Kilonova Models
- GRB afterglow models
- Tidal disruption models
- Supernova Models
- Magnetar driven ejecta models
- Combined models (KNe + afterglow)
- Binary Black Hole optical flare Models

- Bayesian Model Selection
- Fast Parameter Inference (SBI – hundreds per night)
- Physics informed Neural Networks
- Nested Sampling/MCMC (only for a handful)



Analysis Model

Physical Constraints from Transient Models (Candidate Data)



If I observe one more epoch with this Telescope would I be able to resolve the correct model? On which band, on what Limiging Magnitude? How many follow-up do I need? The system can recomend spectroscopy.

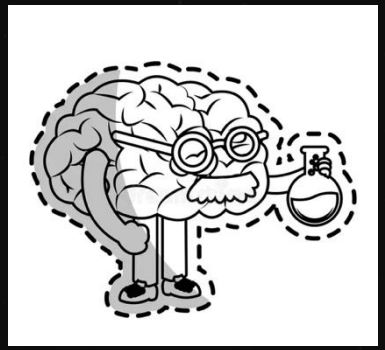


Redback (nikhil-sarin)

- Kilonova Models
- GRB afterglow models
- Tidal disruption models
- Supernova Models
- Magnetar driven ejecta models
- Combined models (KNe + afterglow)
- Binary Black Hole optical flare Models

MOSFiT

WHISPER



AUTOMATED ANALYSIS FOR TIME-DOMAIN ASTRONOMY

Approximate Bayesian Computation (ABC)

Speed: ★★★★½

Constraining Power: ★½☆☆☆

Simulation Based Inference (SBI - NPE/NLE)

Speed: ★★½☆☆

Constraining Power: ★★½☆☆

MCMC/Nested Sampling

Speed: ★½☆☆☆

Constraining Power: ★★★★½

Light Curves
Multi-band photometry

Host Images
Galaxy morphology & environment

Metadata
Coordinates, redshift, alerts

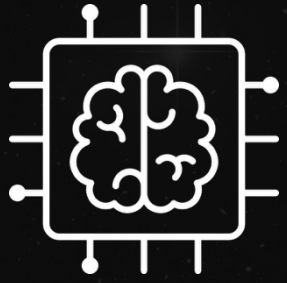
WHISPER
Posterior distribution + Metrics

Trained on Personalized Dataset

arXiv

- **Graph knowleage representation over the whole history of arxiv**

AI-AGENT



AI-ASSITED DISCOVERY PIPELINE

LLM Integration Layer

Natural Language Interface: Scientists query the pipeline in plain English: "Show me all TDE candidates above z=0.3 observed in the last 24h." LLM translates to structured API calls.

Reasoning & Justification: LLM generates human-readable explanations for every Agent decision: why spectroscopy was chosen over deferral, which decision criteria dominated.

Adaptive Science Goals: Community encodes observing objectives as natural-language GOALS (e.g., "prioritize rare SNe-I over AGN this season").

Anomaly Narrative: When the LLM Model flags a “rare” event, the LLM synthesizes context from literature and internal catalogs to generate a discovery Report.

Agentic Tool Use: LLM directly calls observatory APIs, broker endpoints, and scheduling systems — acting as the orchestration layer that translates model outputs into telescope commands. **(Under development and testing in Small Telesope)**

WHISPER – Real Case

ZTF20achncvv

ZTF PHOTO-Z 0.128 ± 0.056

Replayed alert by alert: at each point below WHISPER only knew the detections up to that moment. Its class is revealed at the end of the page.

SURVEY	REDSHIFT	FIRST DETECTION	DATA USED	SURVEY LIMITS (5Σ)
ZTF (alerts via ALeRCE)	0.128	MJD 59130.473	first 30 days: 23 detections (g 11, r 12)	g 19.53, r 19.81, i 19.80

LAST NON-DETECTION

r > 19.92, 1.96 d before

Decision timeline

+5.0 d	+7.0 d	+9.0 d	+15.0 d	+19.0 d	+27.0 d
ABC (CPU) · 3 DETECTIONS	ABC (CPU) · 6 DETECTIONS	EMCEE-JAX (GPU) · DATA TO +10 D · 8 DET.	EMCEE-JAX (GPU) · DATA TO +15 D · 13 DET.	EMCEE-JAX (GPU) · DATA TO +20 D · 16 DET.	EMCEE-JAX (GPU) · DATA TO +30 D · 23 DET.
Wait for more data	Wait for more data	Spectroscopy within 48 h	Spectroscopy within 48 h	Wait for more data	Wait for more data
best: **Arnett (Ni-56 decay)**	best: **Arnett (Ni-56 decay)**	best: **Arnett (Ni-56 decay)**	best: **Arnett (Ni-56 decay)**	best: **Arnett (Ni-56 decay)**	best: **Arnett (Ni-56 decay)**
interest: medium	interest: medium	interest: medium	interest: medium	interest: medium	interest: medium

▾ **+5.0 d** ABC (CPU) · 3 detections Wait for more data **best: Arnett (Ni-56 decay)** D_{max} = 1.4 at i +3 d

Wait for more data INTEREST: MEDIUM confidence: medium

Blue early alert could be either Arnett or TDE; survey should decide

Three detections give BIC 2.3 for Arnett (Ni-56 decay) against 3.4 for the TDE Gaussian rise, delta_BIC 1.1, strength inconclusive, with the magnetar close behind at delta_BIC 3.3; nothing is ranked decisively. The g-r colour of -0.01 is blue and the brightest g detection of 18.71 implies an approximate absolute magnitude of -20.3, but that number rests on a fitted redshift: the modulus moves by -1.34 to +0.87 mag at one sigma, and the Arnett posterior pulls z to 0.0182, which would make the object much less luminous. The luminosity claim is therefore weak. All three top models forecast a flat, slowly brightening light curve near 18.6-18.8 in g, with D_max only 1.35 at i +3 d, so one photometric point cannot separate them; that alone does not justify the ToO. A moderate-cost spectrum now would capture nothing that a spectrum at the next decision point would miss: the object is not fading, so its cost tier should not rise quickly. Nothing here is time-critical, and most alerts are common events, so waiting for the survey to break the degeneracy is the right call; dropping would be premature given the competing tidal-disruption interpretation and the blue colour.

Why it is medium interest: Bright, blue, possibly luminous early alert, but ranking is inconclusive between Arnett and the TDE model with only three detections.

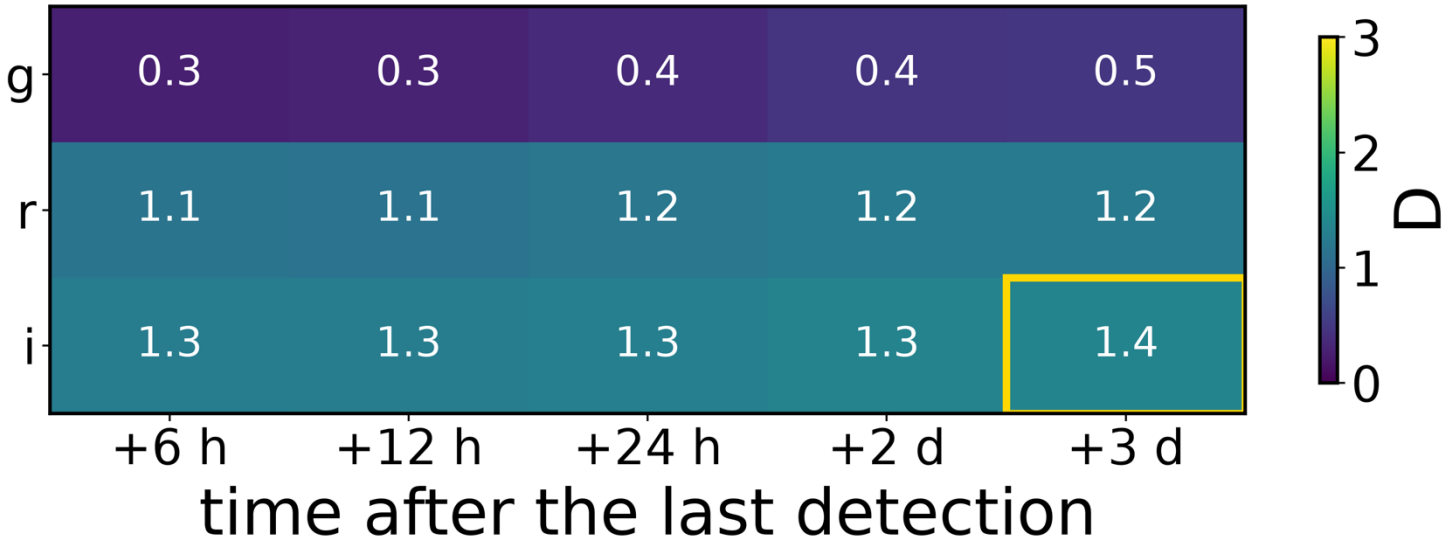
Written by an AI astrophysicist (claude-opus-5) from WHISPER's fact sheet for this moment only: no name, no class, no later data. Automatic check: passed.

Model ranking (BIC; lower is better)

#	MODEL	K	BIC	ΔBIC	WEIGHT	Z FITTED	RMSE G	RMSE R	COV. 68/95	SAMPLER
1	**Arnett (Ni-56 decay)**	**8**	**2.3**	**0.0**	**0.51**	**0.018** [0.006-0.045]	**0.000**	**0.000**	**1.00 / 1.00**	**ABC: 777 accepted**
2	TDE Gaussian rise + cooling envelope	9	3.4	1.1	0.30	0.106 [0.067-0.154]	0.000	0.000	0.67 / 1.00	ABC: 46,618 accepted
3	Magnetar-powered	11	5.6	3.3	0.10	0.109 [0.057-0.166]	0.000	0.000	1.00 / 1.00	ABC: 9,047 accepted
4	CSM shock breakout + Arnett	12	6.7	4.4	0.06	0.120 [0.066-0.175]	0.000	0.000	1.00 / 1.00	ABC: 49,662 accepted
5	Shock cooling + Arnett	13	7.8	5.5	0.03	0.030 [0.008-0.086]	0.000	0.000	1.00 / 1.00	ABC: 1,644 accepted

WHISPER – Real Case

Where Arnett (Ni-56 decay) and TDE Gaussian rise + cooling envelope disagree most
(D in combined sigmas; hatched: fainter than the survey limit; gold box: max D)



#1 Arnett (Ni-56 decay) ΔBIC 0.0, weight 0.97

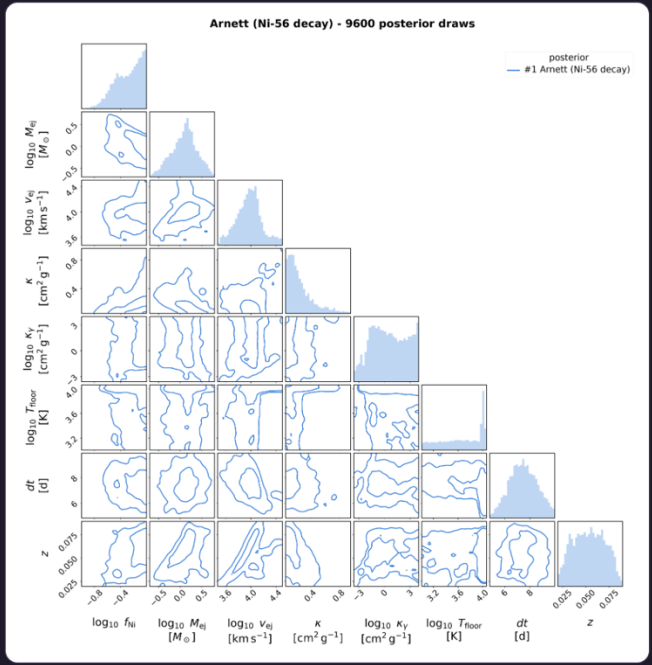


Figure 25. Posterior of Arnett (Ni-56 decay) at this decision point, from the emcee-JAX posterior samples, each with its own time origin dt. Each diagonal panel is the histogram of one parameter; each panel below the diagonal shows two parameters against each other, with contour lines enclosing about 39 % and 86 % of the draws (the two-dimensional equivalents of 1σ and 2σ). Parameters with a log-uniform prior are drawn on $\log_{10}$ axes (labelled $\log_{10}$), the others on linear axes; each axis spans the central 99 % of the draws, not the whole prior. A narrow peak means the data pin that parameter down; a histogram

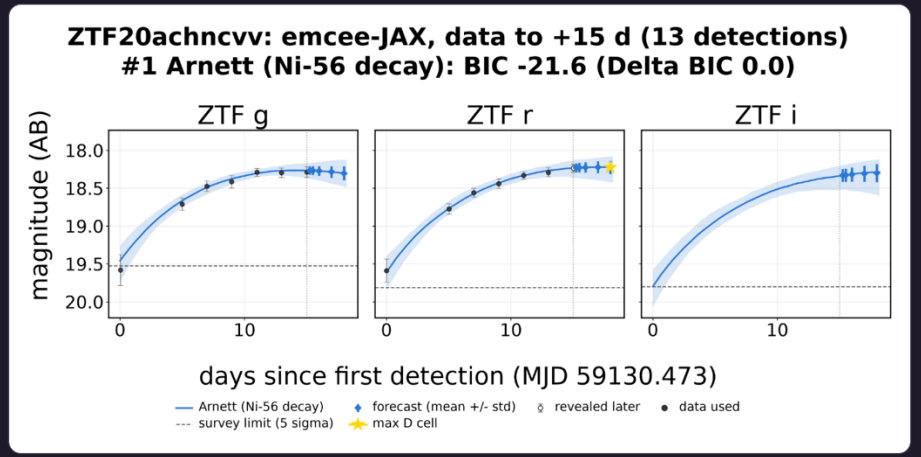


Figure 26. Posterior predictive check of Arnett (Ni-56 decay) (rank 1). One panel per band, apparent magnitude against days since first detection, brighter upwards. The solid line is the median and the shaded band the central 95 % of the model light curves computed from 200 of the emcee-JAX posterior samples, each with its own time origin dt. Filled points with error bars are the detections used in this fit; hollow points are detections that arrived later (up to 3 days after the last one used) and were not used; the dotted vertical line marks the last detection used. Diamonds with error bars are the forecast at +6 h, 12 h, 24 h, 2 d and 3 d (mean ± standard deviation over up to 500 draws), the dashed horizontal line is the survey's 5σ limit, and the gold star is the forecast cell where the two best models differ most; forecasts and limits are drawn for g, r and i only. Read it as: filled points inside the band mean the model reproduces the data it was given; hollow points inside the band mean its forecast held up.

#1 Arnett (Ni-56 decay) ΔBIC 0.0, weight 0.97

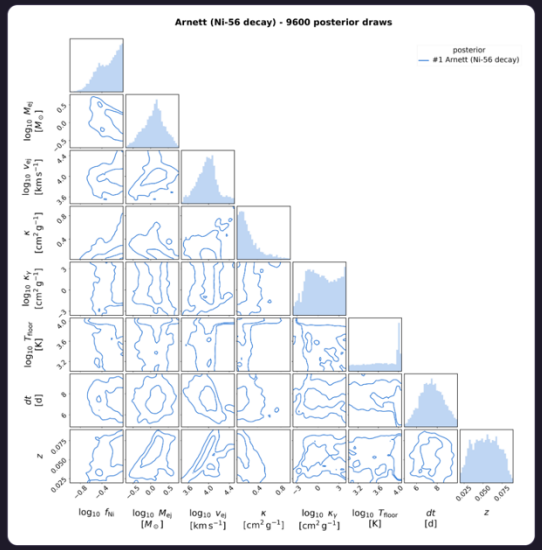


Figure 25. Posterior of Arnett (Ni-56 decay) at this decision point, from the emcee-JAX posterior samples, each with its own time origin dt. Each diagonal panel is the histogram of one parameter; each panel below the diagonal shows two parameters against each other, with contour lines enclosing about 39 % and 86 % of the draws (the two-dimensional equivalents of 1σ and 2σ). Parameters with a log-uniform prior are drawn on $\log_{10}$ axes (labelled $\log_{10}$), the others on linear axes; each axis spans the central 99 % of the draws, not the whole prior. A narrow peak means the data pin that parameter down; a histogram

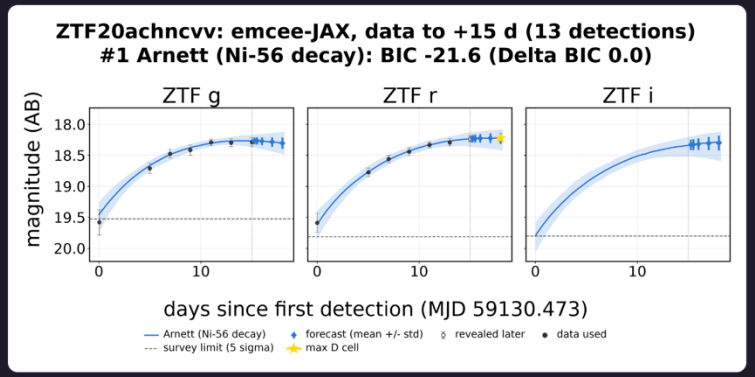


Figure 26. Posterior predictive check of Arnett (Ni-56 decay) (rank 1). One panel per band, apparent magnitude against days since first detection, brighter upwards. The solid line is the median and the shaded band the central 95 % of the model light curves computed from 200 of the emcee-JAX posterior samples, each with its own time origin dt. Filled points with error bars are the detections used in this fit; hollow points are detections that arrived later (up to 3 days after the last one used) and were not used; the dotted vertical line marks the last detection used. Diamonds with error bars are the forecast at +6 h, 12 h, 24 h, 2 d and 3 d (mean ± standard deviation over up to 500 draws), the dashed horizontal line is the survey's 5σ limit, and the gold star is the forecast cell where the two best models differ most; forecasts and limits are drawn for g, r and i only. Read it as: filled points inside the band mean the model reproduces the data it was given; hollow points inside the band mean its forecast held up.

Disclaimer: ZTF20achncvv was classified as **SLSN-II** (MVP_TEST targets.py (TNS classification). Without the recomended actions and no more than ZTF data WHISPER could not resolve.

AI-SCOPE

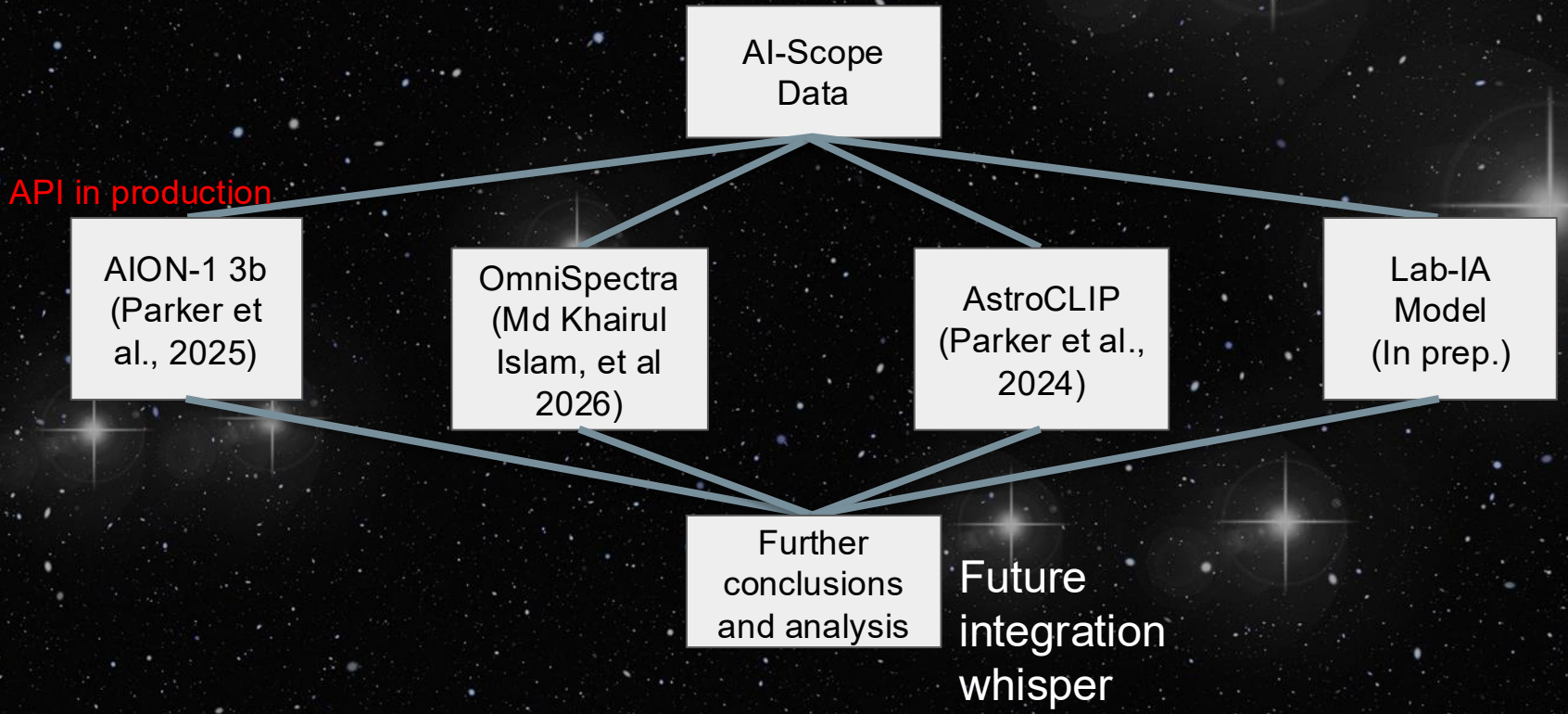
- Alert Streams, GCN notices & circulars, brokers (FINK) powered by LLMs.
- Probability maps (LIGO Skymaps) or transient positions + uncertainties (GRB, FXT).
- Catalogues (DES, DELVE, DESI, SPLUS, open LSST, GAIA, ZTF).
- Processing and analysis of images and spectra

A restrict version is open oline. If you are interested to unleash the full (fun) funcionalities let me know

https://ai-scope.cbpf.br/



Foundational Models as API service with AI-Scope



AGENTIC ASTRONOMER

TVS LLMS and Foundational Models

Coordinators: Clecio R. Bom

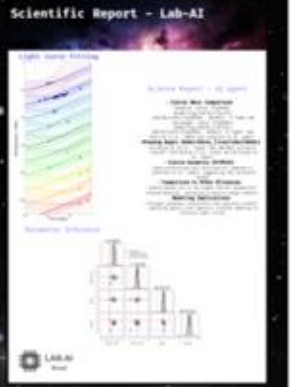
Slack channel: #tvs-llms-foundational-models

Telecons: TBD

Goals

- Develop TVS related Agents and LLMs to assist TVS community
- Integration between informed AI agents, LLMS and recommendation systems with Brokers
- Built community efforts in AI Agents and Generative AI in TVS





Current Status

Portal for development of AI agents integrating public alerts + public avalible datasets already-up

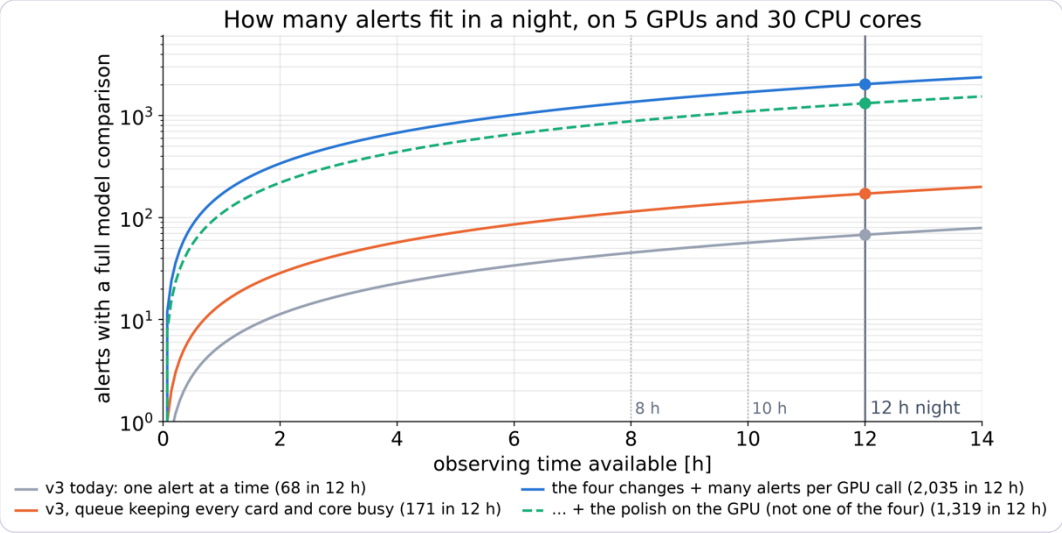
HOW FAST – performance measurement

2,035 alerts in a 12-hour night on 5 GPUs

2,035

alerts with a full 5-model comparison, 12 h

GPU s per alert	106
CPU core s per alert	159
runs out first	the GPUs
10 GPUs (extrapolated)	~4,069



WHISPER

How WHISPER will modelling these peculiar sources?

Kilonova Spectral Inverse Modelling with Simulation-based Inference: An Amortized Neural Posterior Estimation Analysis

P. Darc[1], C. R. Bom[1], B. Fraga[1], and C. D. Kilpatrick[2]
[1] Centro Brasileiro de Pesquisas Físicas, Rua Dr. Xavier Sigaud 150, CEP 22290-180, Rio de Janeiro, RJ, Brazil; phelipedarc@cbpf.br
[2] Center for Interdisciplinary Exploration and Research in Astrophysics (CIERA), Northwestern University, Evanston, IL 60201, USA
Received 2024 March 13; revised 2024 May 18; accepted 2024 May 29; published 2024 August 7

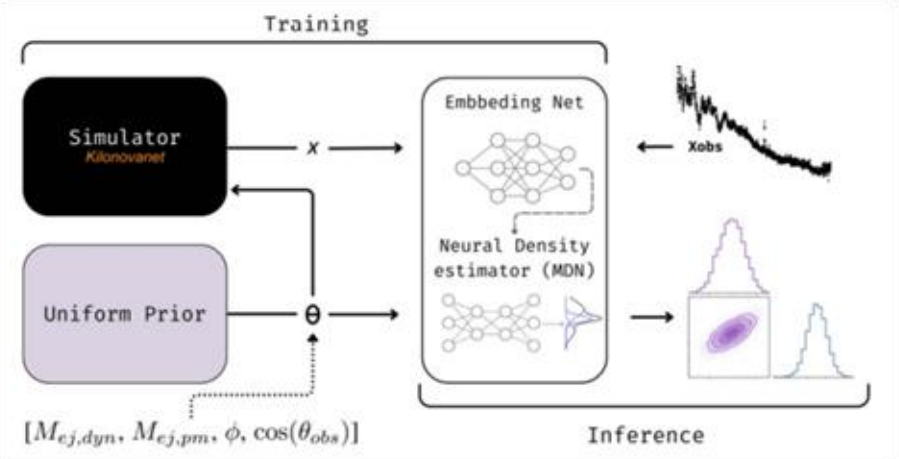


Figure 2. Inference pipeline using ANPE—amortized inference enables rapid inference for every new spectrum observed. In the training phase, a neural-network-based density estimator (MDN) learns the probabilistic relationship between the model's parameters and simulated spectra, at high computational cost. Then, in the inference phase, the trained density estimator takes the observed real data as input and infers the posterior distribution of the parameters, at low computational cost: less than a second.

WHISPER

How WHISPER will modelling these peculiar sources?

Multi-Wavelength Analysis of Kilonova Associated with GRB 230307A: Accelerated Parameter Estimation and Model Selection Through Likelihood-Free Inference

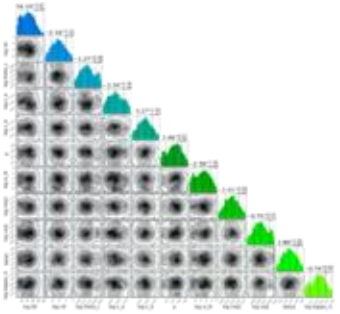
Phelipe Darc*
Centro Brasileiro de Pesquisas Físicas (CBPF)
Rua Xavier Sigaud, 150, Urca, Rio de Janeiro, Brazil.

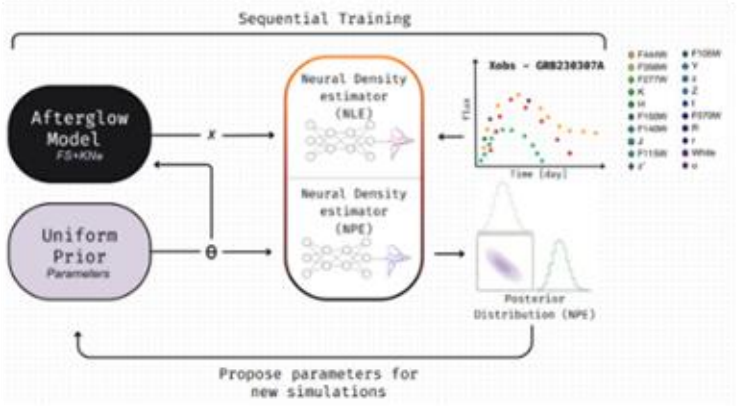
Clecio R. Bom
Centro Brasileiro de Pesquisas Físicas (CBPF)
Rua Xavier Sigaud, 150, Urca, Rio de Janeiro, Brazil.
Centro Federal de Educação Tecnológica Celso Suckow da Fonseca (CEFET-RJ),
Av. Maracanã, 229,Maracanã, Rio de Janeiro, Brazil.

Gabriel S. M. Teixeira
Centro Brasileiro de Pesquisas Físicas (CBPF)
Rua Xavier Sigaud, 150, Urca, Rio de Janeiro, Brazil.

Charles Kilpatrick
Center for Interdisciplinary Exploration and Research in Astro- physics (CIERA)
and Department of Physics and Astronomy
North-western University, Evanston, IL 60208, USA.







WHISPER

How WHISPER will modelling these peculiar sources?

The progenitor of the GRB 230307A: a decision through a Bayesian analysis

Viviane Alfradique, Rodrigo da Mata, Clécio R. Bom
Centro Brasileiro de Pesquisas Físicas,
Rua Dr. Xavier Sigaud 150,
22290-180 Rio de Janeiro, RJ, Brazil

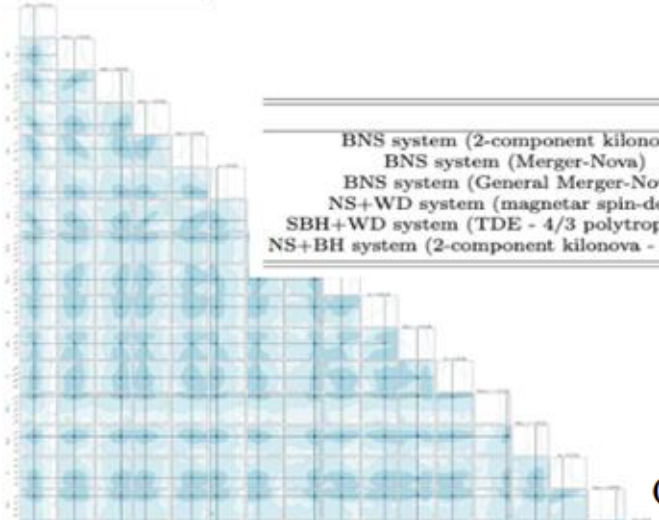


TABLE I. Summary statistical results.

	$\ln(\mathcal{Z}^1/\mathcal{Z}^{\rm ref})$	Maximum $\ln(\mathcal{L})$	BIC	Dimensions	KN model	Reference
BNS system (2-component kilonova)	-23.07	-108.12	382.86	17		[29]
BNS system (Merger-Nova)	-98.45	-182.50	525.12	15		[30]
BNS system (General Merger-Nova)	-98.05	-181.06	515.82	13		[31]
NS+WD system (magnetar spin-down)	-	-87.19	323.97	14		[16]
SBH+WD system (TDE - 4/3 polytropes stars)	-8398.99	-8458.49	17126.20	15		[32]
NS+BH system (2-component kilonova - Ejecta rel)	-3849.79	-3932.98	8027.80	15		[29]

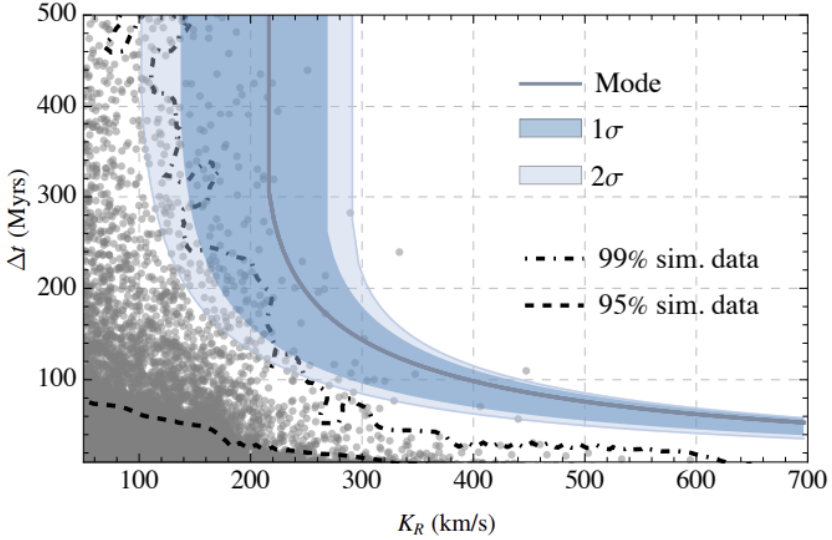
Our results show a statistical preference for a BNS or NS–WD progenitor producing a kilonova powered by a magnetar and ^{56}Ni decay, characterized by a ^{56}Ni mass of $\sim 4\times10^{-4}\,M_\odot$ and an ejecta mass of $0.06\,M_\odot$. Furthermore, under the assumption of a BNS origin within this model, we infer binary component masses of $m_1 = 1.81^{+0.46}_{-0.61}\,M_\odot$ and $m_2 = 1.61^{+0.65}_{-0.41}\,M_\odot$,

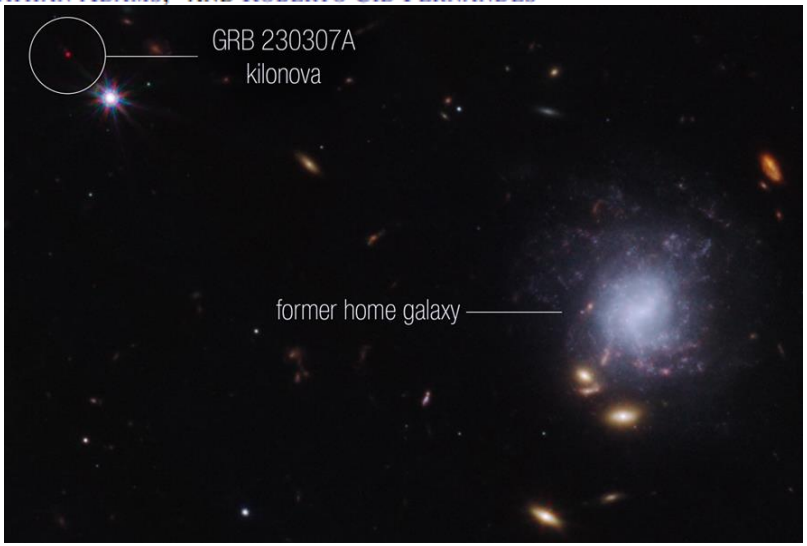
How AI-Scope can accelerate Transient Research?

- The LLMs report evaluates population level features and call attentions to interesting and outliers

Probing the origin of the kilonova candidate GRB 230307A: analysis of host galaxy and offset

Clecio R. Bom,[1] Davi C. Rodrigues,[2,1] Arianna Cortesi,[3,4] Amanda E. Araujo-Carvalho,[4] Daniel Ruschel-Dutra,[5] Giuliano Iorio,[6] Luidhy Santana-Silva,[1] Charles D. Kilpatrick,[7] Fabricio Ferrari,[8] Luis Lomelí-Núñez,[4] Thomas Harvey,[9] Duncan Austin,[9] Christopher J. Conselice,[9] Nathan Adams,[9] and Roberto Cid Fernandes[5]



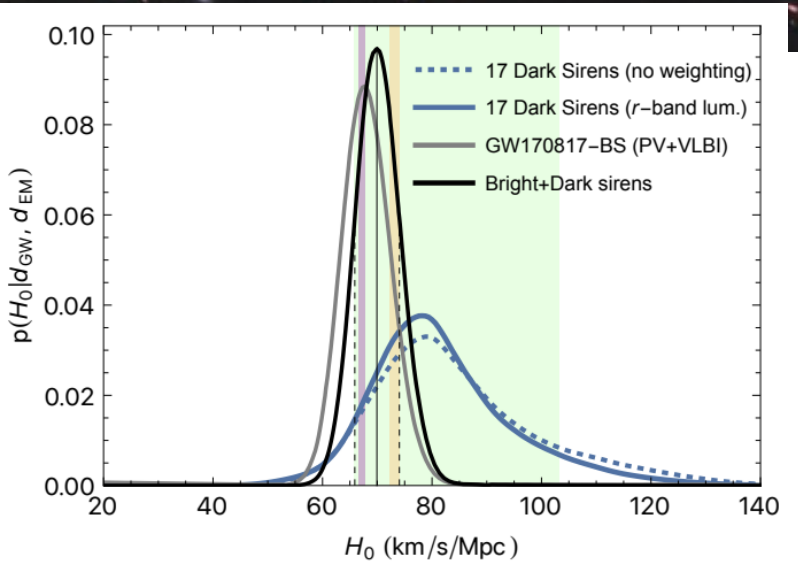


The time to arrive at the merger position departing from the disk compared to the total kick of binary system, given a galaxy mass model and BNS population simulation overlap only 0.1% within the 2σ region of the galaxy mass model. Accepted **to ApJ Letters 2026**.

How AI-Scope can accelerate Research?

Improved constraint on the Hubble constant from dark sirens with LIGO/Virgo/KAGRA O4a

Viviane Alfradique,* Clécio R. Bom, Gabriel Teixeira, and André Santos
Centro Brasileiro de Pesquisas Físicas,
Rua Dr. Xavier Sigaud 150,
22290-180 Rio de Janeiro, RJ, Brazil



Event(s)	Method(s)	H_0 (km/s/Mpc)	σ_{H_0} (km/s/Mpc)
O1-O3 – 47 dark sirens [1]	Catalog (K-band luminosity-weighted) + BNS/NSBH/BBH population (PLP model with fixed parameters)	67^{+13}_{-12}	12.5 (18%)
O1-O4a – 142 dark sirens [2]	Catalog (K-band luminosity-weighted) + BNS/NSBH/BBH population (FullPop-4.0 model)	$81.6^{+21.5}_{-15.9}$	18.7 (23%)
O1-O4a – 142 dark sirens [3]	Catalog (K-band luminosity-weighted) + BNS/NSBH/BBH population (BPL+3peak model)	$82.5^{+16.8}_{-14.3}$	15.6 (19%)
O1-O4a – 142 dark sirens [2]	BNS/NSBH/BBH population (FullPop-4.0 model)	$76.4^{+23.0}_{-18.1}$	20.6 (27%)
O1-O4a – 137 dark sirens [4]	BBH population (semiparametric B-spline model)	$57.8^{+21.9}_{-20.6}$	21.3 (37%)
O1-O4a - 15 dark sirens [5]	Catalog (no luminosity weighting)	$70.4^{+13.6}_{-11.7}$	12.6 (18%)
GW170817 [6]	Bright (v_p corrected)	$68.8^{+17.3}_{-7.6}$	12.5(18%)
GW170817 [7]	Bright (v_p+VLBI)	$68.3^{+4.6}_{-4.5}$	4.6(7%)
O1-O4a – 17 dark sirens [8]	Catalog (no luminosity weighting; GW likelihood Gaussian approximation, eq. 5)	$72.6^{+11.6}_{-11.3}$	11.4 (16%)
O1-O4a – 17 dark sirens [8]	Catalog (r-band luminosity-weighted; GW likelihood Gaussian approximation, eq. 5)	$72.8^{+12.6}_{-14.3}$	13.5 (18%)
O1-O4a – 17 dark sirens [8]	Catalog (no luminosity weighting; full GW likelihood, eq. 6)	$78.8^{+14.6}_{-12.2}$	13.4 (17%)
O1-O4a – 17 dark sirens [8]	Catalog (r-band luminosity-weighted; full GW likelihood, eq. 6)	$78.2^{+12.0}_{-11.0}$	11.5 (15%)
O1-O4a – 17 dark + 1 bright siren [8]	Bright (host) + Catalog (r-band luminosity-weighted; full GW likelihood, eq. 6)	$75.2^{+8.0}_{-7.9}$	8.0(11%)
O1-O4a – 17 dark + 1 bright siren (EM) [8]	Bright (v_p+VLBI) + Catalog (r-band luminosity-weighted; full GW likelihood, eq. 6)	$69.9^{+4.1}_{-4.0}$	4.1 (6%)

Summary

- AI can do more than classification.
- Agentic AI can encapsulate contextual information efficiently
- AI can help people get informed decisions on what is the best transient for follow-up based on the possible scientific outcome, and optimize spectro and photometric telescope time.
- Local tunned Infrastructure for astronomy, Local Run and development of LLMs, fine-tunning, Rags.
- Article https://arxiv.org/abs/2606.05285 - accepted for publication in Nature Astronomy

Demonstration vídeo:
https://www.youtube.com/watch?v=WojdkXPqhPw - ABC and LLM report (LLMs detects a fail fit due to a weak choice of modelling right away)
https://youtu.be/vlzQ3tLftxY - Foundation Model API and LLM coding for Astro

The Rubin Static Strong Lensing Data Challenge

A COMMUNITY BENCHMARK AT LSST YEAR-1 QUALITY



21
classification teams

5
regression teams

>1,500
submissions

>60
participants

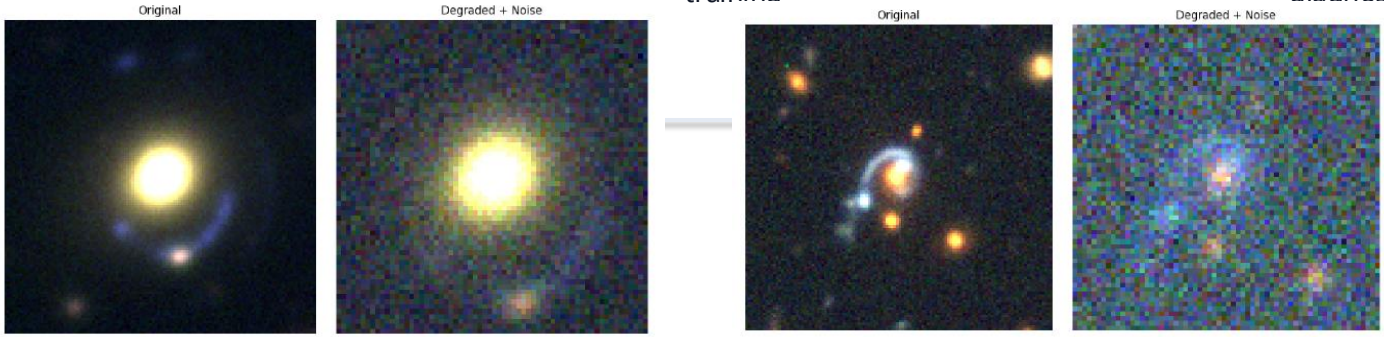
DESIGN
Physically motivated populations + survey metrics

DATA
SLSim + SIMCT + real HSC matched to LSST Y1

BLIND TEST
Real confirmed lenses held out from training

RUN
1 Aug – 15 Oct 2025 public, non-competitive

RESULTS
Discovery + regression + ensemble diagnostics



Two tasks: classification (find lenses) and **regression** (estimate physical parameters).

•Grespan et al. 2026, RSLDC ApJ draft

Centro Brasileiro de Pesquisas Físicas

TOWARD DECISION-AWARE AI FOR LSST-SCALE TIME-DOMAIN ASTRONOMY

Clecio R. Bom,
for TVS Science Collaboration
debom@cbpf.br

