Brokers, recommendation systems and adaptable machine learning

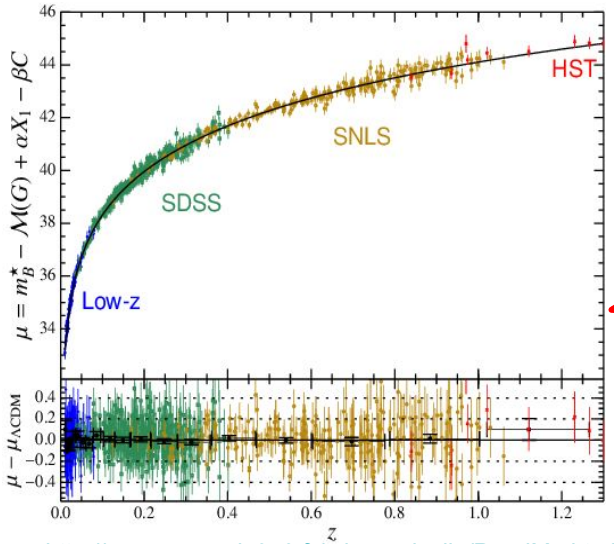
7 June 2024 - Exploring the Dark Side of the Universe Tools
Île de Noirmoutier, France

Emille E. O. Ishida
Laboratoire de Physique de Clermont, CNRS - Université Clermont-Auvergne
Clermont Ferrand, France

Science case: SNIa cosmology

Standard candles used to measure cosmological distances

https://supernova.eso.org/exhibition/images/1111_E_549779main_pia14095_full/



http://supernovae.in2p3.fr/sdss_snls_jla/ReadMe.html

Ω_Λ

The paradigm shift:

Big data in astronomy

SDSS

120 TB

2000 - now

Primary mirror: 2.5 m

ZTF

1.4 TB/night

Primary mirror: 1.2 m

2018 - now

VERA C. RUBIN OBSERVATORY

15 TB/night

From 2025

Primary mirror: 8.4 m

The paradigm shift:

Big data in astronomy

SDSS

2000 - now

120 TB

Primary mirror: 2.5 m

PS: this is all photometry …

ZTF

2018 - now

1.4 TB/night

Primary mirror: 1.2 m

VERA C. RUBIN OBSERVATORY

From 2025

15 TB/night

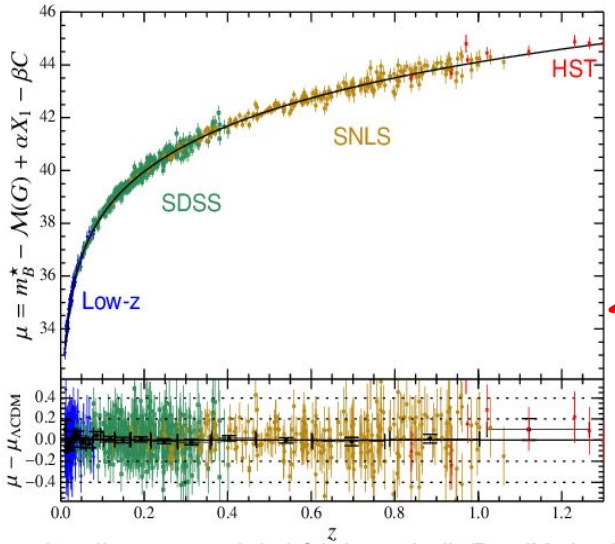
Primary mirror: 8.4 m

Photometric SNIa cosmology

Standard candles used to measure cosmological distances

https://supernova.eso.org/exhibition/images/1111_E_549779main_pia14095_full/

Without SN spectra!



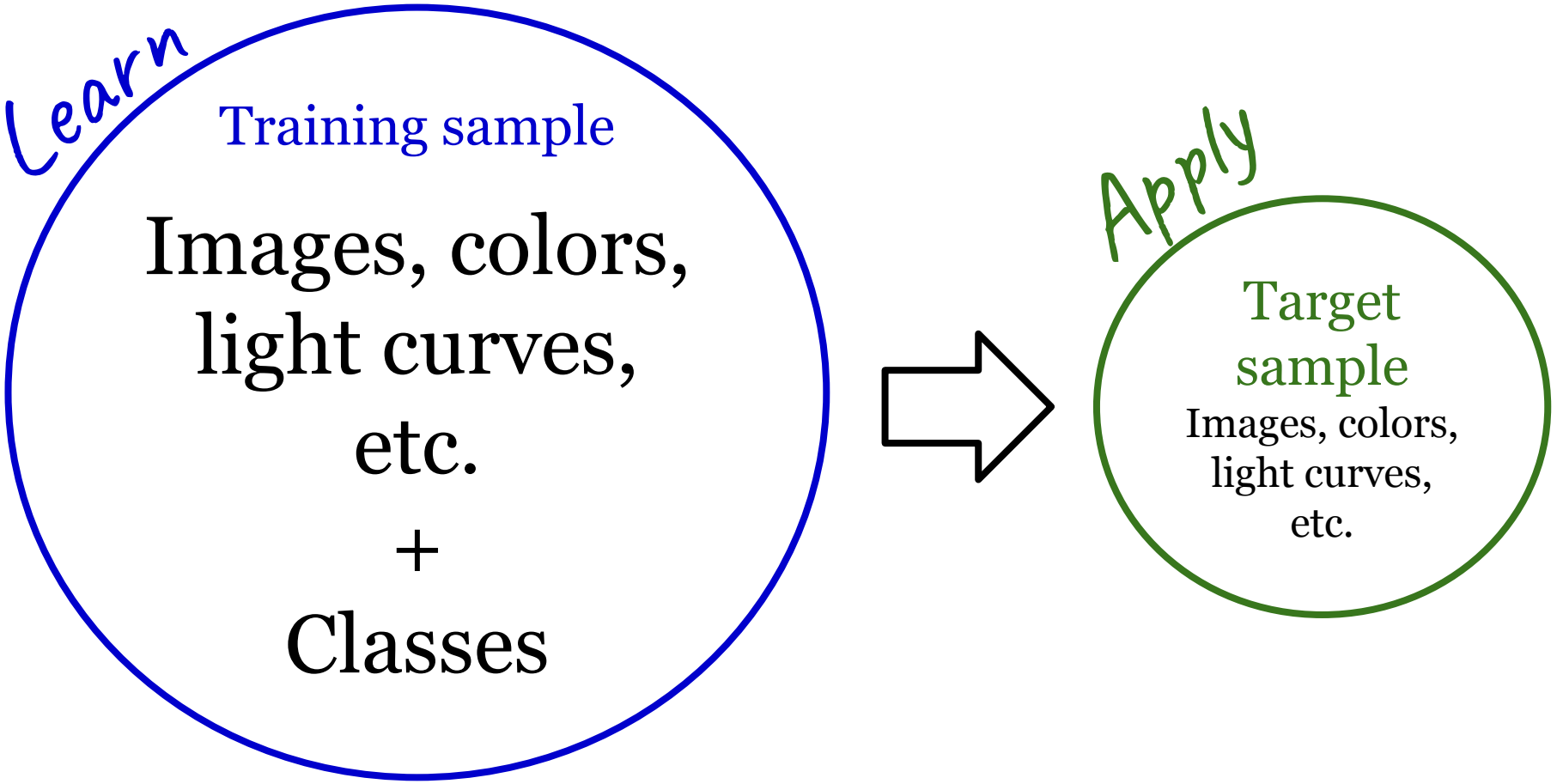
http://supernovae.in2p3.fr/sdss_snls_jla/ReadMe.html

Ω

Λ

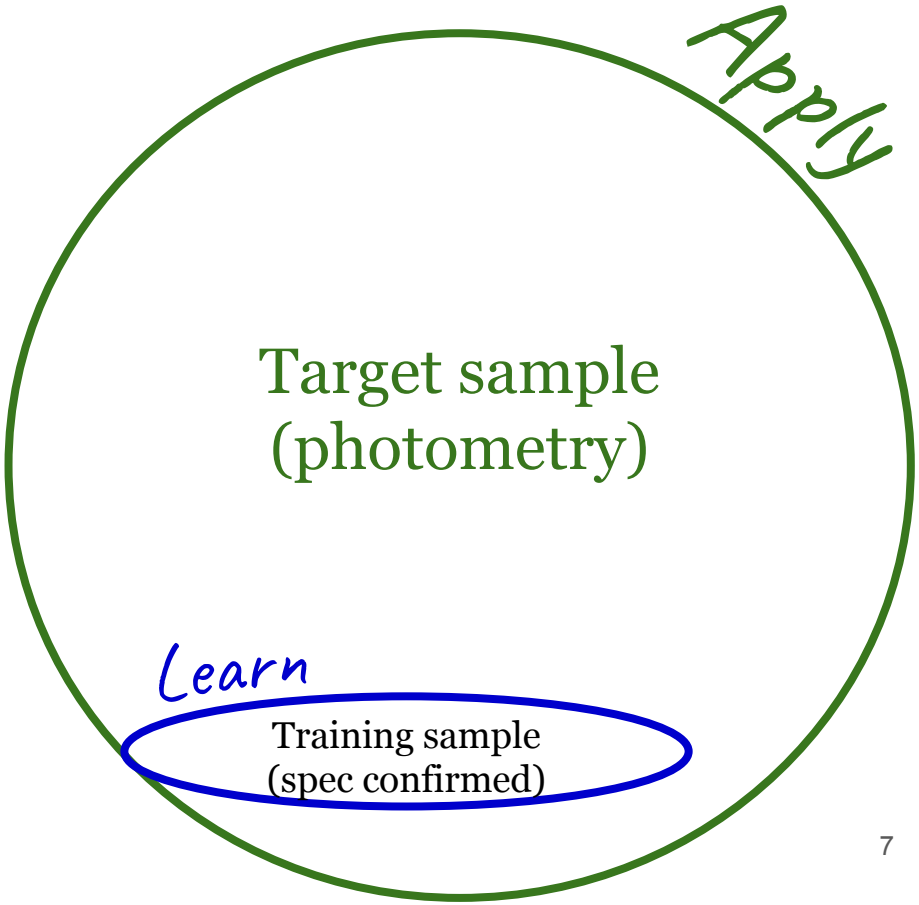
If only it were that simple …

Ideal Supervised learning situation



In astro, training means spectra

Real astro-learning situation



Proposal solution …

Let the algorithm choose the training

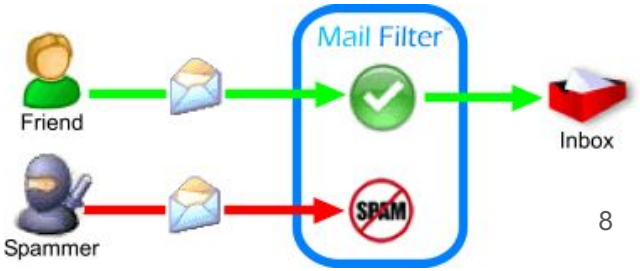
Given limited resources, we need **recommendation systems**!

amazon

35% OF AMAZON'S REVENUE ARE GENERATED BY IT'S RECOMMENDATION ENGINE.

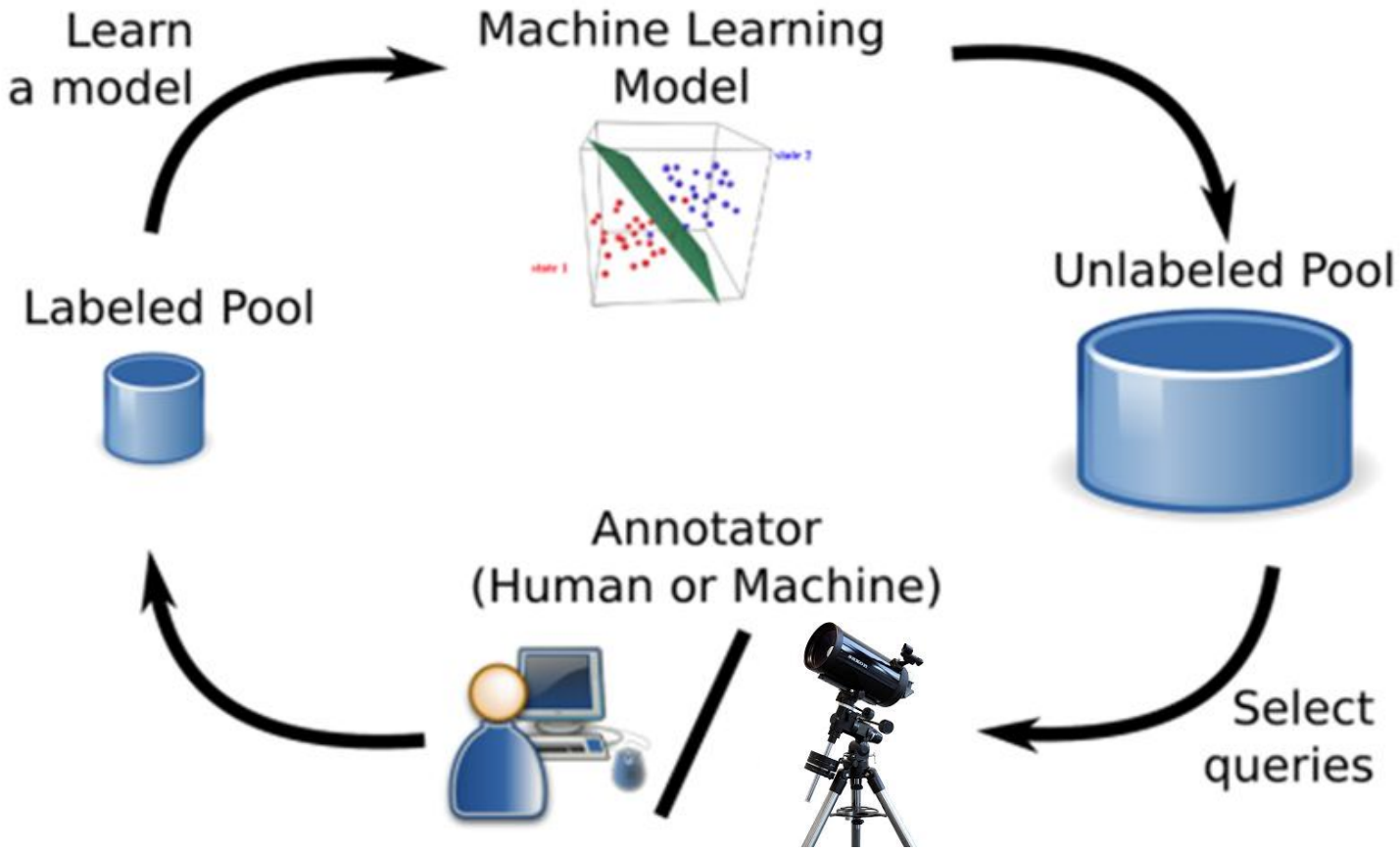
NETFLIX

75% OF USERS SELECT MOVIES BASED ON NETFLIX'S RECOMMENDATIONS.



Active Learning

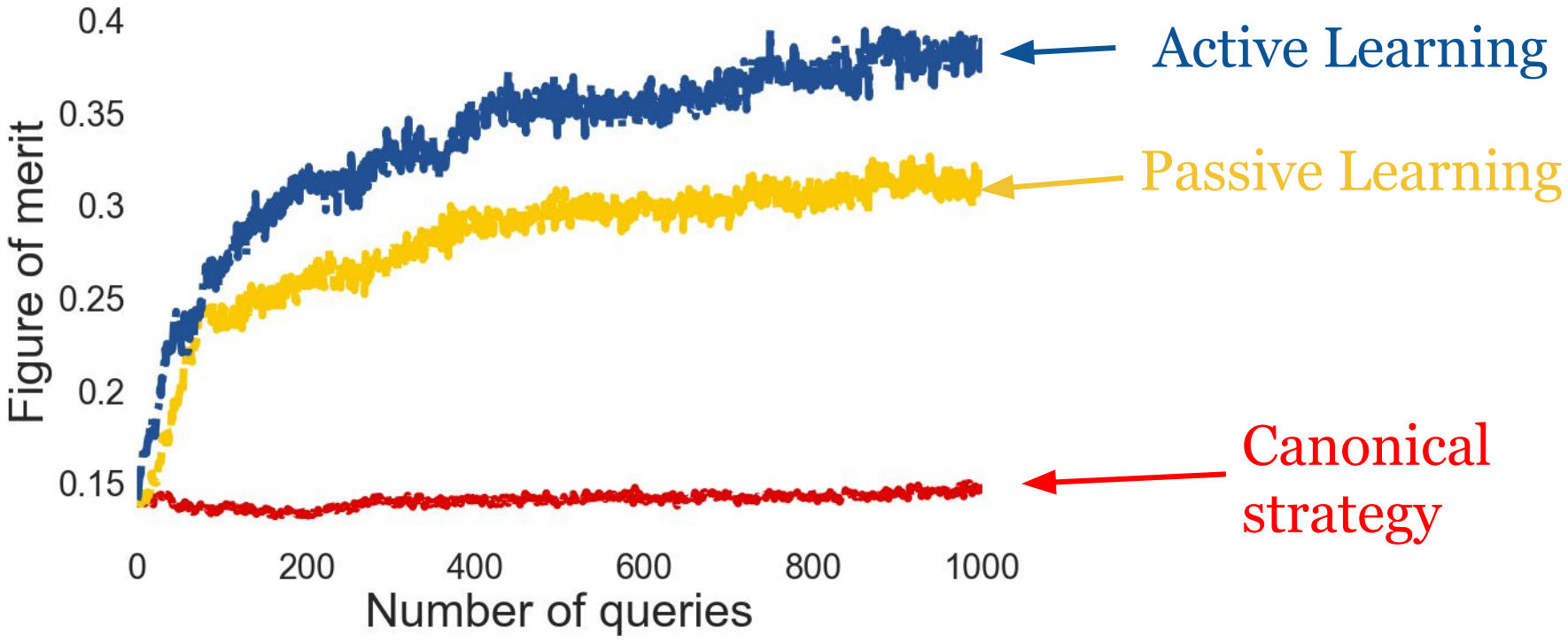
Optimal classification, minimum training



Plot modified from [Chowdhury et al., 2021, SPIE Medical Imaging]

AL for SN classification

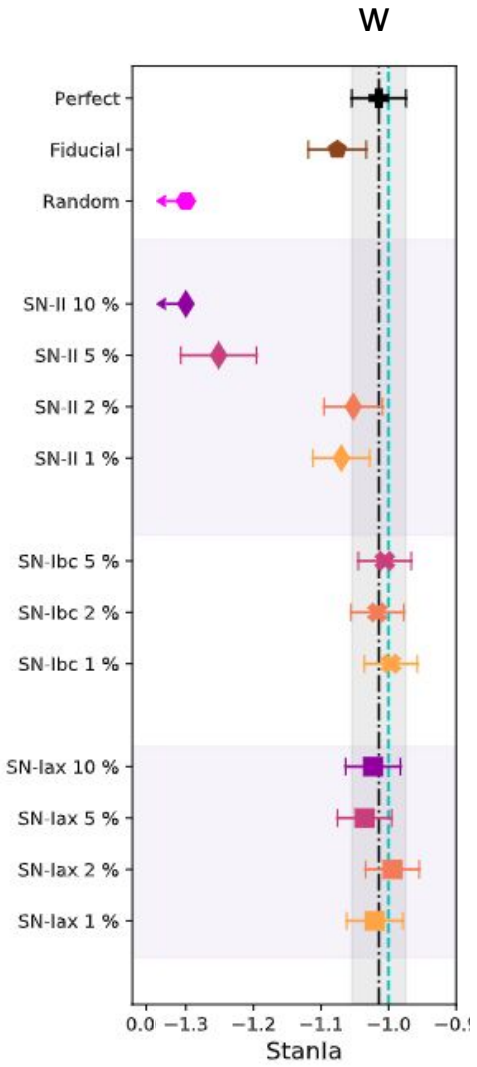
Static results from simulations



From COIN Residence Program #4, ***Ishida*** *et al., 2019, MNRAS, 483 (1), 2–18*

Classification ⇒ Cosmology ?

5% contamination can be catastrophic … **depending on the contaminant class**



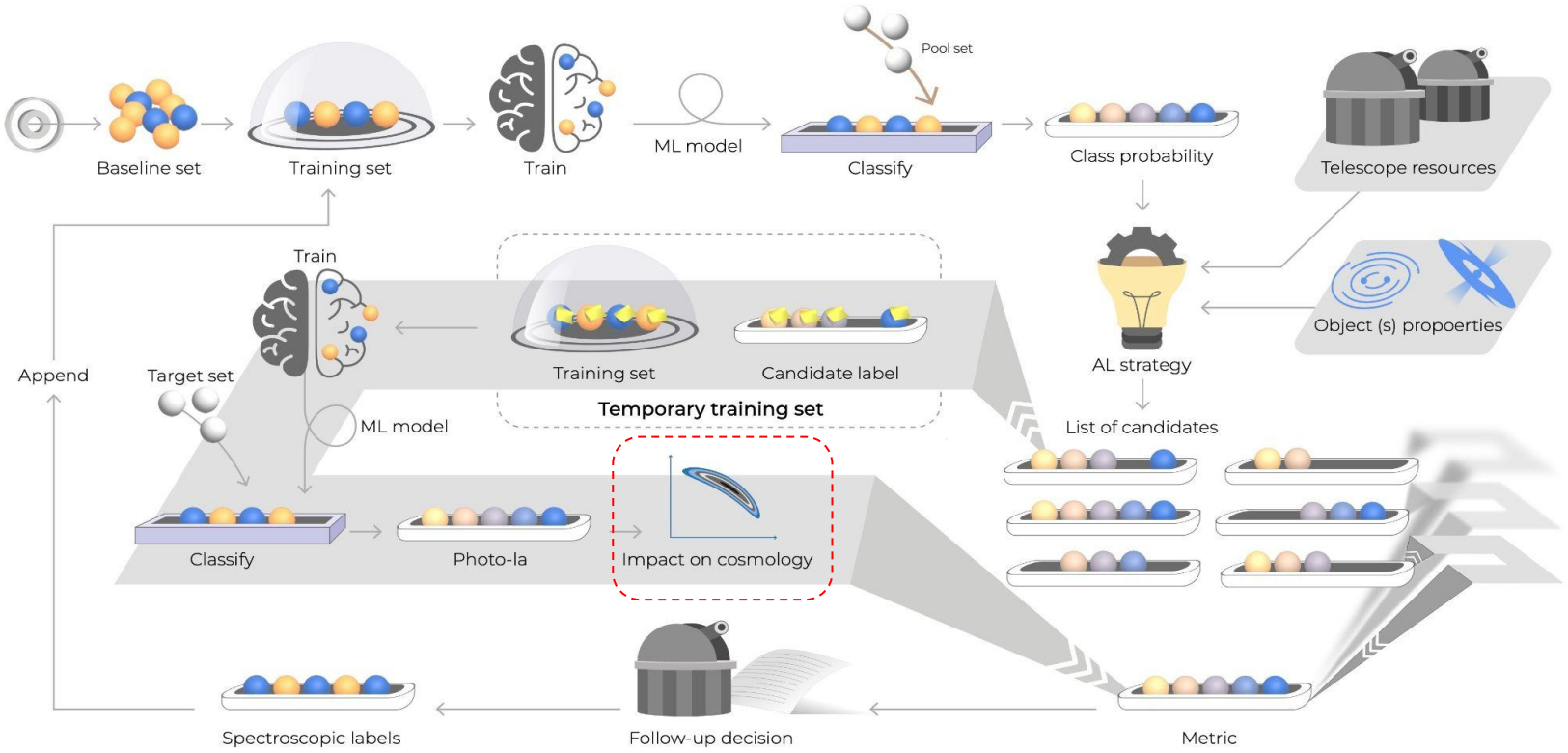
LSST DESC & COIN
RESSPECT
Recommendation System for Spectroscopic Follow-up

Malz et al. (the RESSPECT team), 2023 - arXiv: astro-ph/2305.14421

LSST DESC & COIN
RESSPECT
Recommendation System for Spectroscopic Follow-up

RESSPECT:

Recommendation System for Spectroscopic Follow-up



Ishida et al. (the RESSPECT team) - in prep

https://github.com/COINtoolbox/RESSPECT

Take home message:

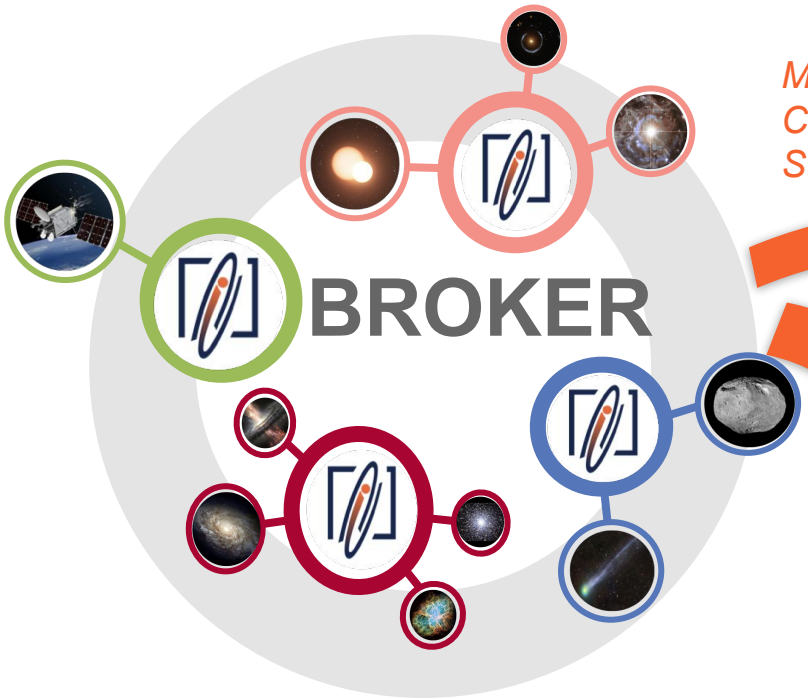
- “Big data” in astronomy means more **photometry**
- Be suspicious of off-the-shelf ML applications to fundamental sciences
- Optimal results from machine learning will depend on **optimized recommendations systems** with well defined science goals

Take home message:

- “Big data” in astronomy means more **photometry**
- Be suspicious of off-the-shelf ML applications to fundamental sciences
- Optimal results from machine learning will depend on **optimized recommendations systems** with well defined science goals

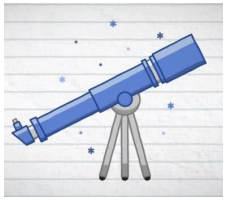
How will you (or anyone) get the data in the era of Rubin?

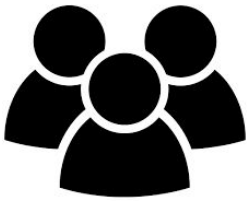
The data path

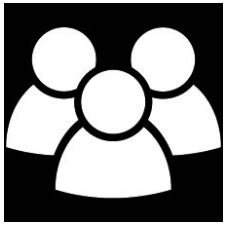


Machine learning
Catalog association
Streams join

ZTF: 300k alerts per night
LSST: 10 million alerts per night…







We would like the interesting ones ...

Ongoing science projects

Satellite glints – arxiv:2202.05719, 2310.17322

https://fink-broker.org/papers

Solar System Objects – linkage problem + spins, arxiv:2305.01123, 2403.20179

Anomaly detection - in prep

Young Stellar objects - partnership with COIN - 2303.09409

Multi-class classifiers - arxiv: 2404.08798, 2303.08951

Kilonova – ML-based partnership with GRANDMA – arxiv:2202.09766,2210.17433

AGN – ML-based + catalogs – arxiv:2211.10987

***Supernovae & Core-collapse** – ML & DL-based – arxiv:2111.11438,2207.04578,2303.08951*

Pair-instability Supernovae – ML-based

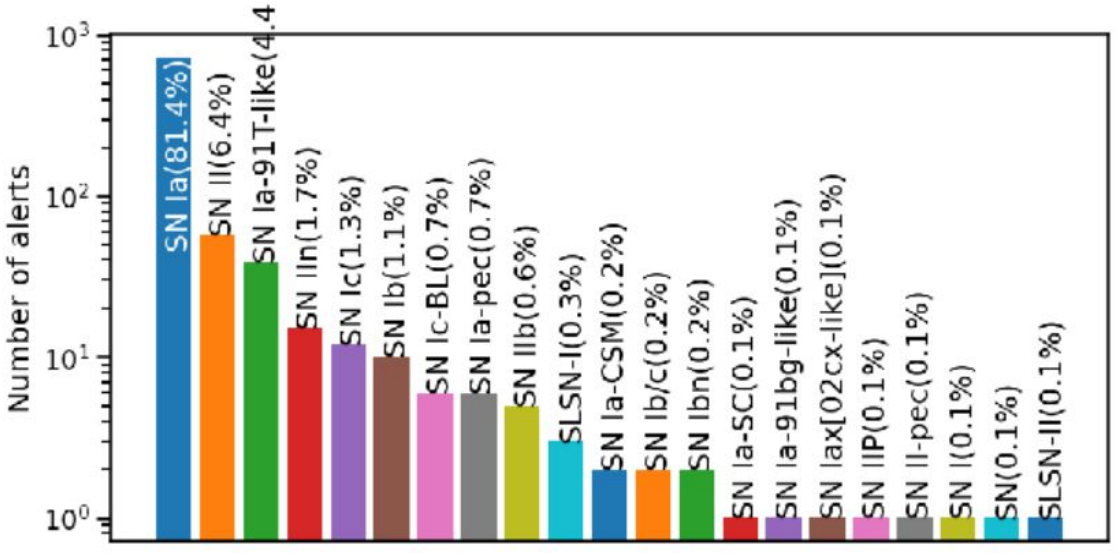
GRB follow-up + Orphan GRB – Fermi, Swift + partnership with SVOM

\+ *Multi-messenger analysis and others!*

Fink EarlySNIa candidates

Fink Early SNIa module:

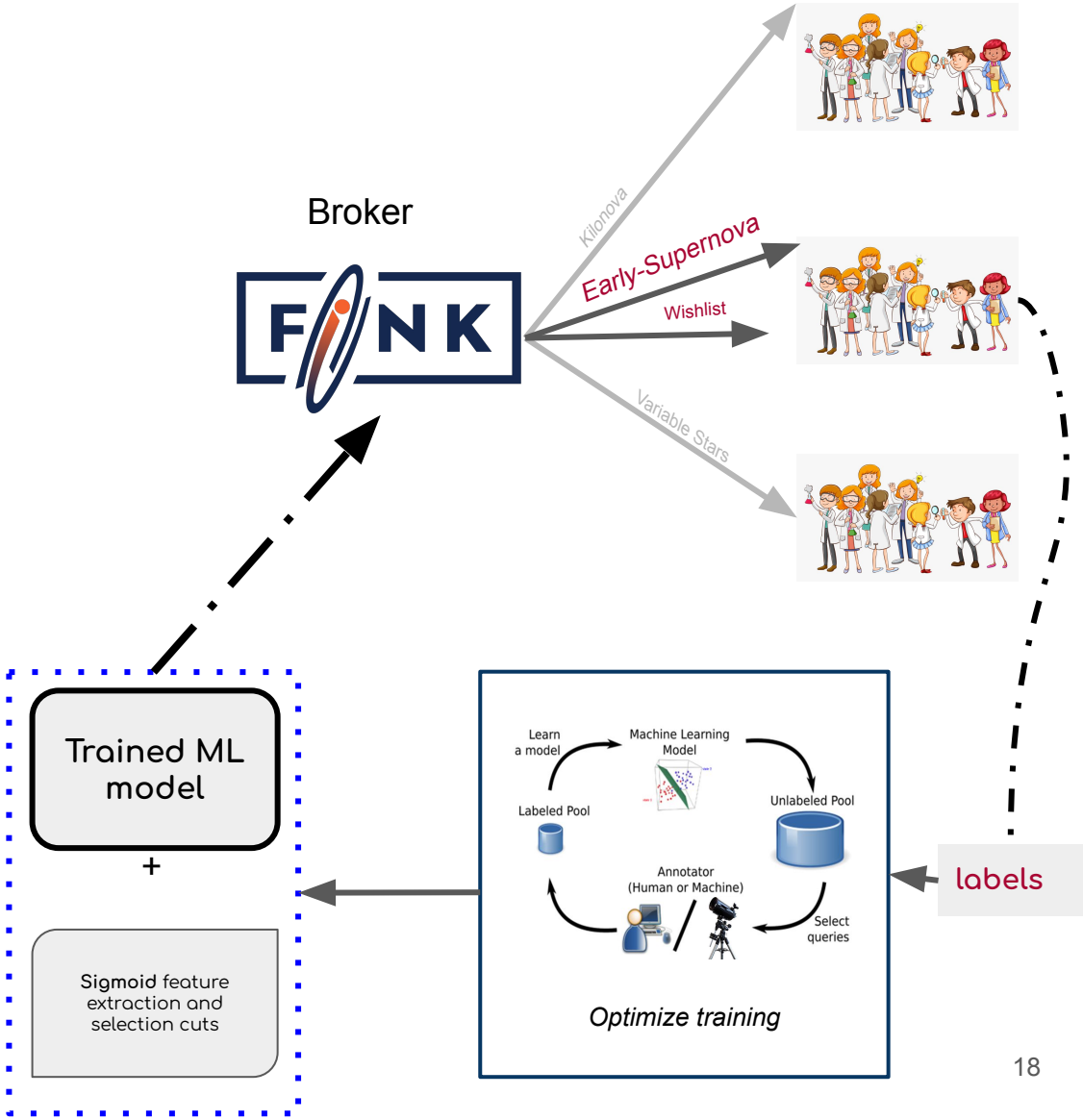
- Agreement with SuperNNova
- Only rising candidates
- from November/2020 to November/2022 Fink communicated 1,533 early SNIa candidates to TNS



- 908 (59%) of which where followed-up & spectroscopically by facilities around the world
- 788 (86%) were spectroscopically confirmed as SNIa

For Rubin …

Automatic re-training
once informative new labels are made available



Take home message:

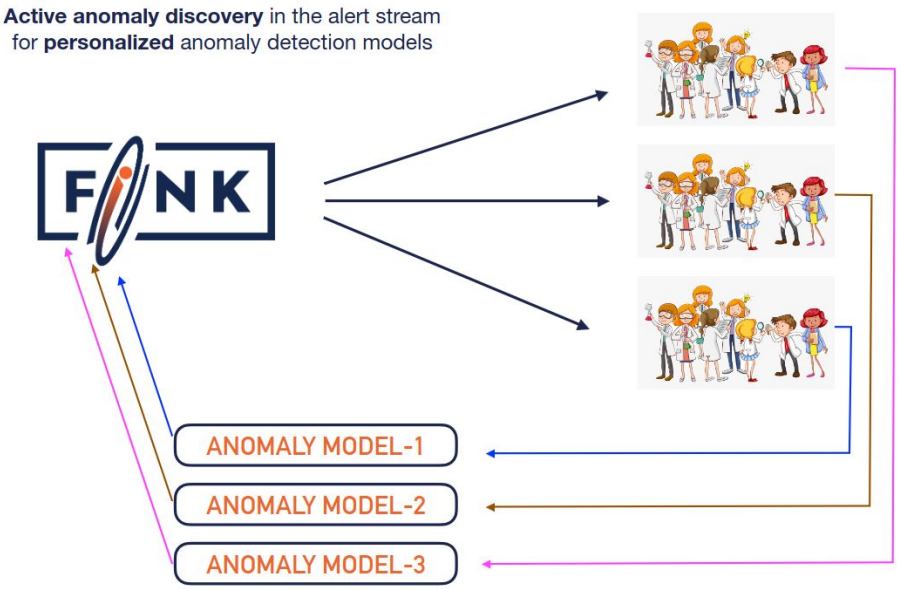
- Recommendation system can (and should) be used in a variety of science cases
- Fink infrastructure was built to allow smooth integration of adaptive learning
- Automatization of the machine learning cycle is essential
- Coordination will be a challenge .. but different players are preparing for it

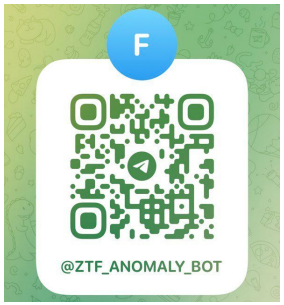
The beauty of an observational science

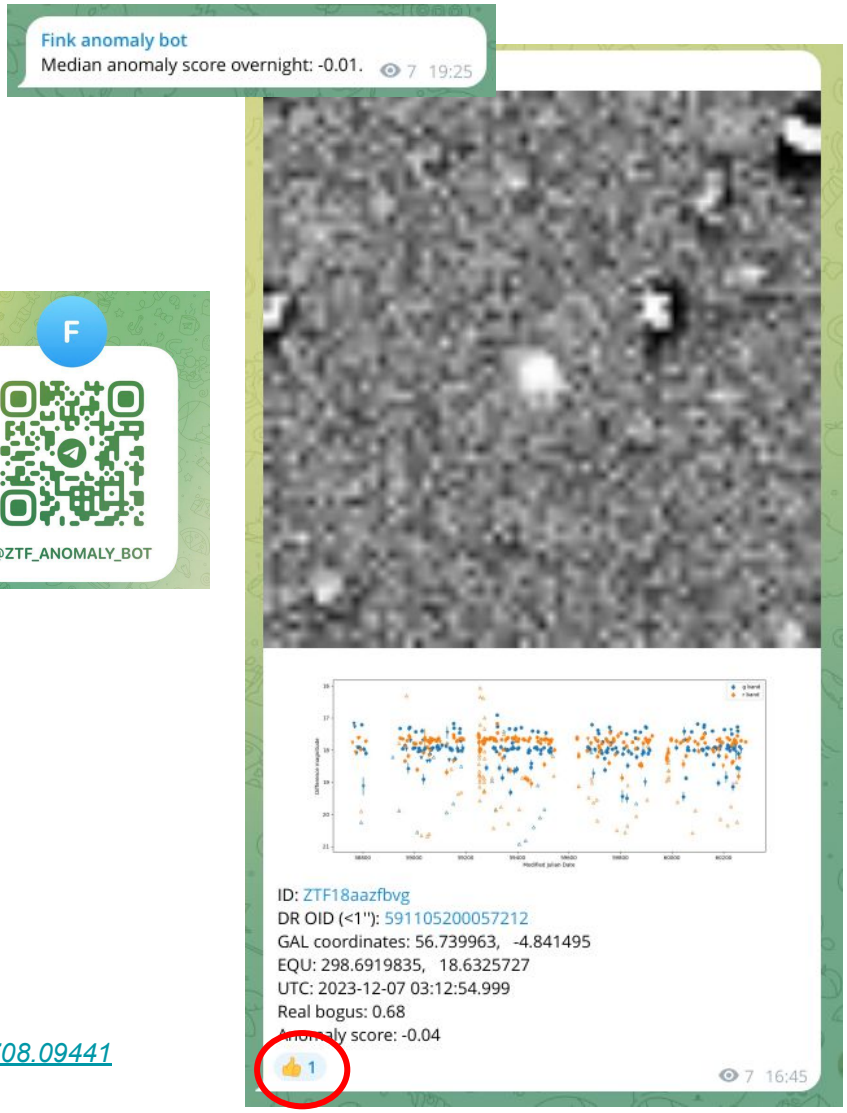
"… telescopes that merely achieve their stated science goals have probably failed to capture the most important scientific discoveries available to them."

Norris, R. (2017). Discovering the Unexpected in Astronomical Survey Data. Publications of the Astronomical Society of Australia, 34, E007. doi:10.1017/pasa.2016.63

Fink Anomaly Detection module



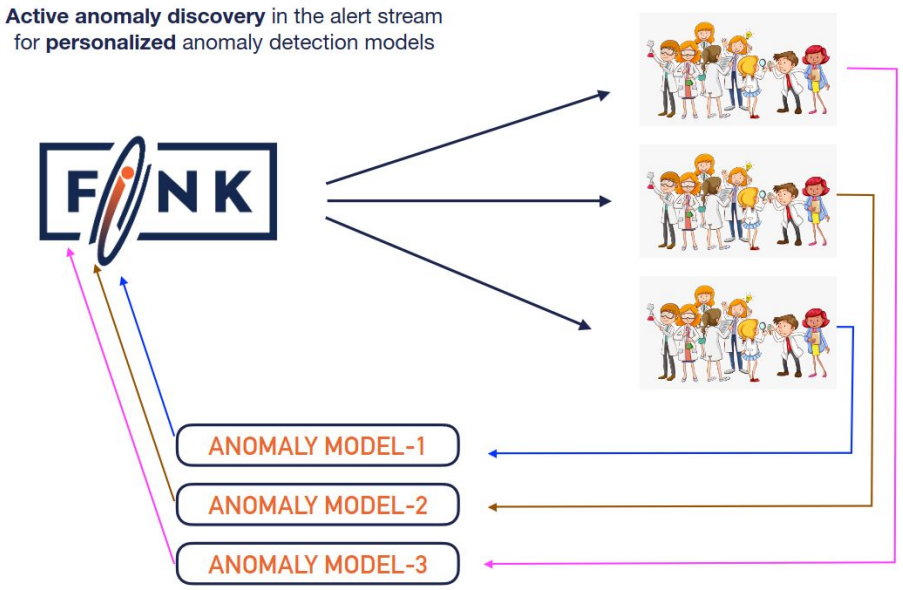


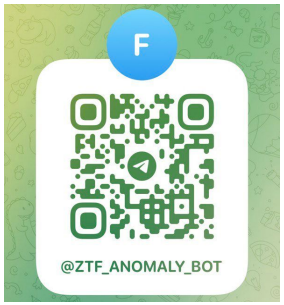


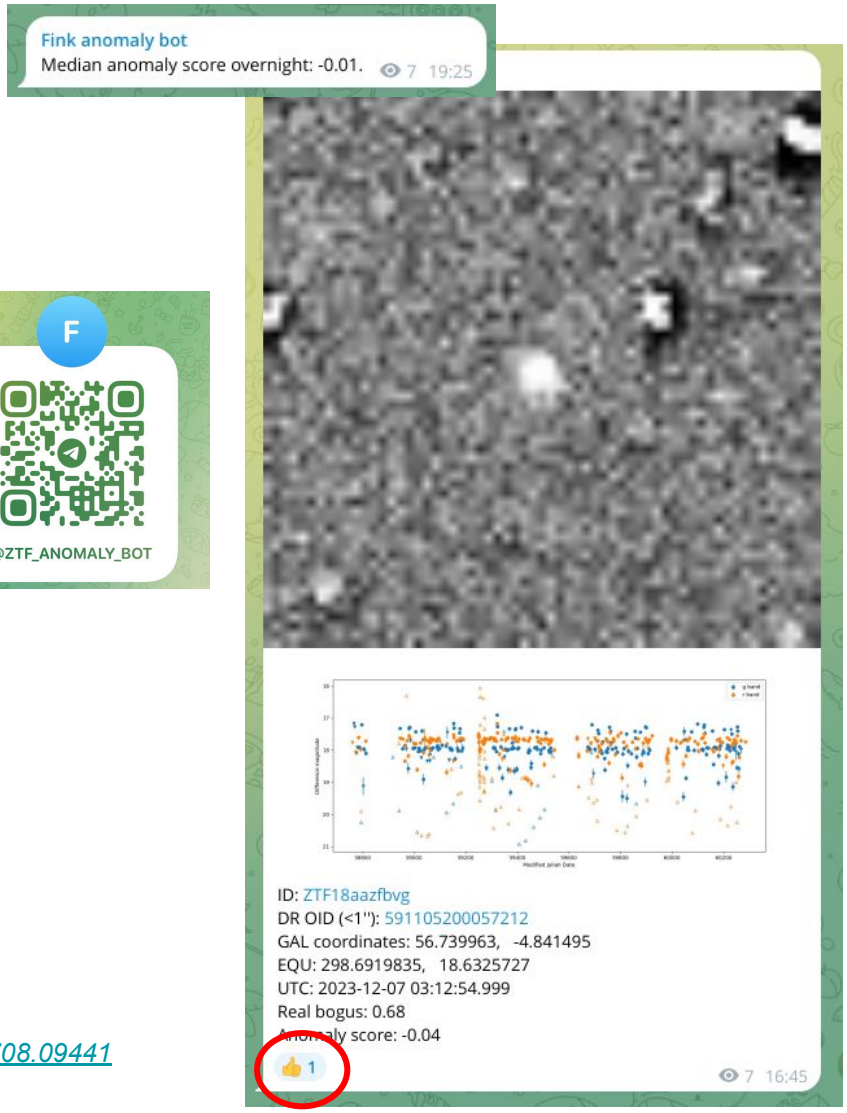
Work lead by Maria Pruzhinskaya (LPCA)

Algorithm from Das, S., et al., 2017, in DEA'17, KDD workshop, arXiv:cs.LG/1708.09441
Implementation by the SNAD team, via coniferest package

Fink Anomaly Detection module







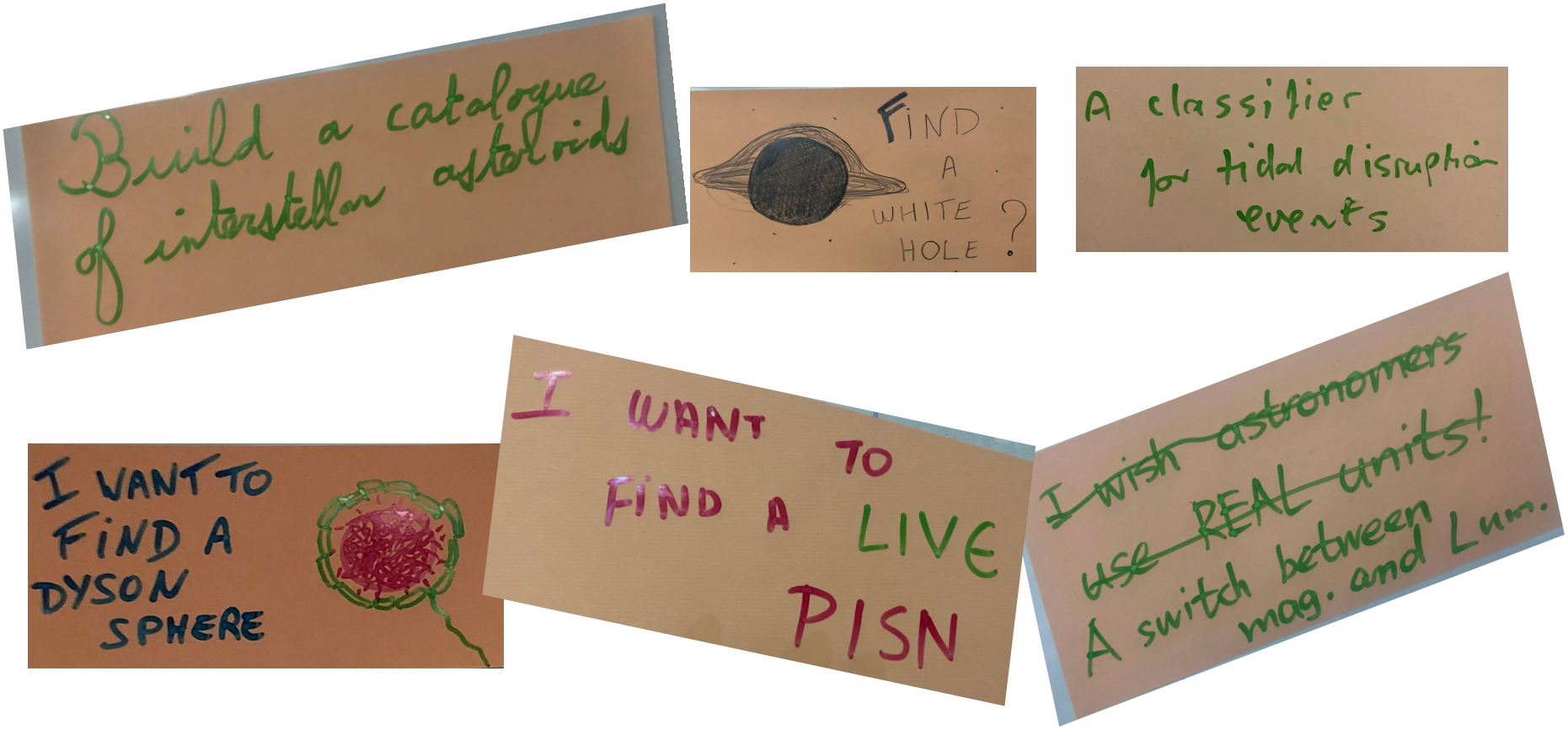
COMING SOON

Personalized channels for a few users

Work lead by Maria Pruzhinskaya (LPCA)

Algorithm from Das, S., et al., 2017, in DEA'17, KDD workshop, arXiv:cs.LG/1708.09441
Implementation by the SNAD team, via coniferest package

#FinkDreamShots



From OzFink 2023 - Melbourne, Australia - https://www.ozgrav.org/ozfink-workshop-2023.html

What do you want to see?