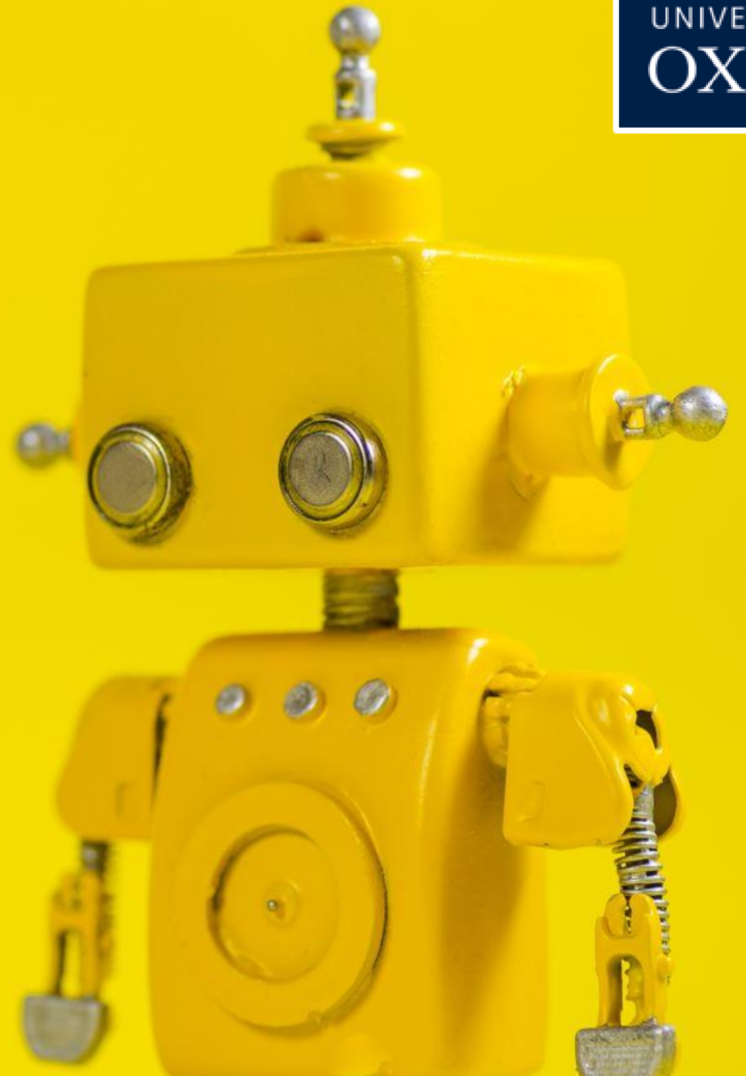
A.I.

Just because you can, doesn't mean you should...



Dr. Heloise F. Stevance
Eric & Wendy Schmit A.I. in Science Fellow
University of Oxford, Reuben College

hfstevance@gmail.com | ML for Transients 12/12/23 | JBYCDMYS

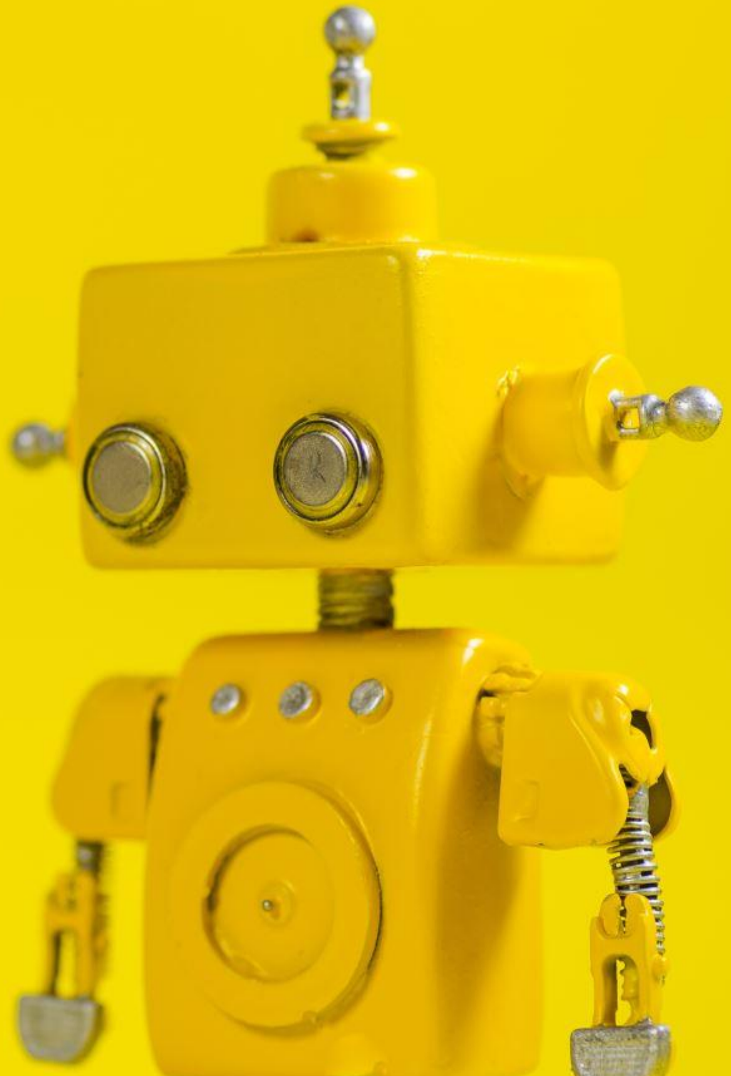
01) My project

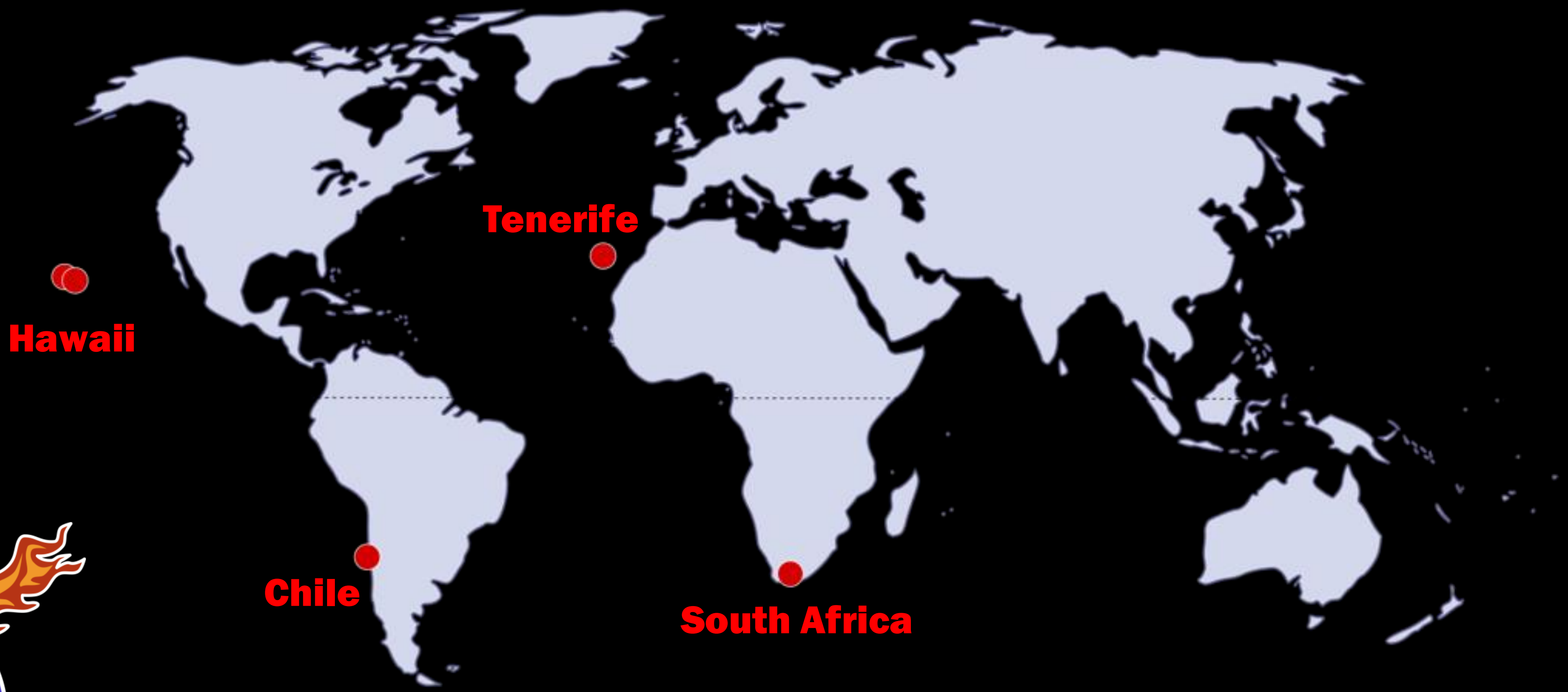
ATLAS and the Virtual Research Assistant

10) Lessons and Reflections

So many proof of concepts, no solutions

Constructing Impactful Machine Learning Research for Astronomy: Best Practices for Researchers and Reviewers





Asteroid Terrestrial-impact Last Alert System

[Tonry et al. 2018]

See Josh's talk

Current Procedure

Smith et al. 2021

thespacedoctor/sherlock

ML

Decision Tree

Alert → Real Vs Bogus → Sherlock → Eye-balling

Real Vs Bogus

Josh Weston's talk

Sherlock

AGN
Variable Star
Nuclear Transient
Supernova
Cataclysmic Variable
Orphan (no host gal)
UNCLEAR

Eye-balling

Identify follow-up candidates (remove bogus, spot contaminants)

The mission

Virtual Research Assistant

Make a tool to **assist human** eye-ballers in that ATLAS sky survey by automating some of the **early decision** jobs and flagging interesting targets for follow-up

Current Procedure

Smith et al. 2021

Alert → **ML**: Real Vs Bogus → thespacedoctor/sherlock **Decision Tree**: Sherlock → Eye-balling

V.
R.
A.

Real Vs Bogus

Josh Weston's talk

Sherlock

AGN
Variable Star
Nuclear Transient
Supernova
Cataclysmic Variable
Orphan (no host gal)
UNCLEAR

Eye-balling

Identify follow-up candidates (remove bogus, spot contaminants)

The original plan...

RAPID: Early Classification of Explosive Transients Using Deep Learning

Daniel Muthukrishna[1], Gautham Narayan[2,7], Kaisey S. Mandel[1,3,4], Rahul Biswas[5], and Renée Hložek[6]
[1] Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK; daniel.muthukrishna@ast.cam.ac.uk
[2] Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA
[3] Statistical Laboratory, DPMMS, University of Cambridge, Wilberforce Road, Cambridge, CB3 0WB, UK
[4] Kavli Institute for Cosmology, Madingley Road, Cambridge, CB3 0HA, UK
[5] The Oskar Klein Centre for CosmoParticle Physics, Department of Physics, Stockholm University, AlbaNova, Stockholm SE-10691, Sweden
[6] Department of Astronomy and Astrophysics & Dunlap Institute, University of Toronto, 50 St. George Street, Toronto, ON M5S 3H4, Canada
Received 2019 January 18; accepted 2019 March 26; published 2019 September 30

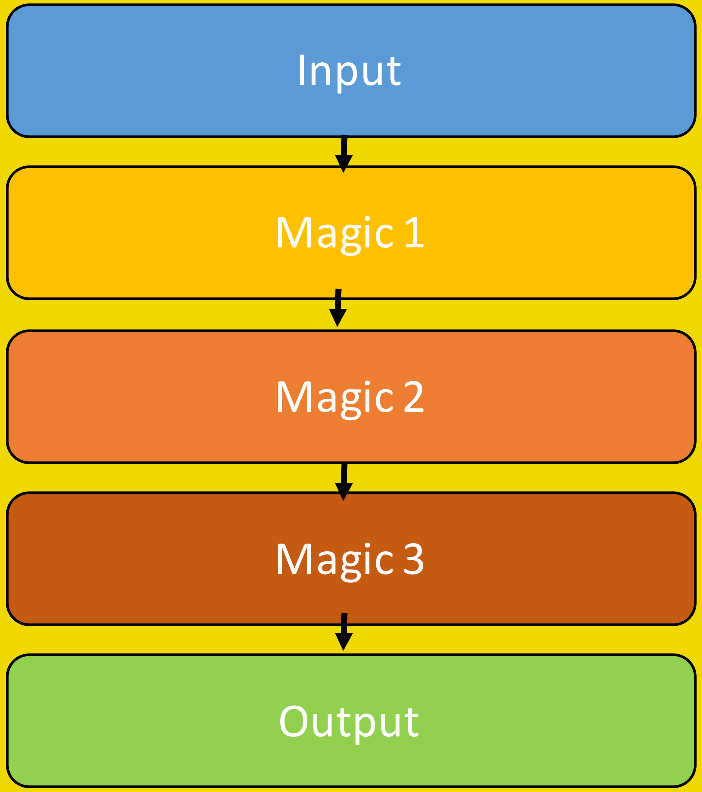
Abstract

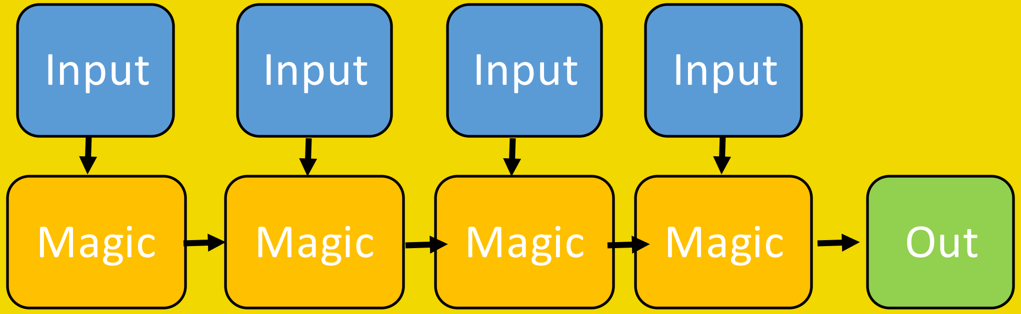
We present Real-time Automated Photometric IDentification (RAPID), a novel time series classification tool

[Muthukrishna et al. 2019]

with some adjustments...

Recurrent NN in 1 slide?





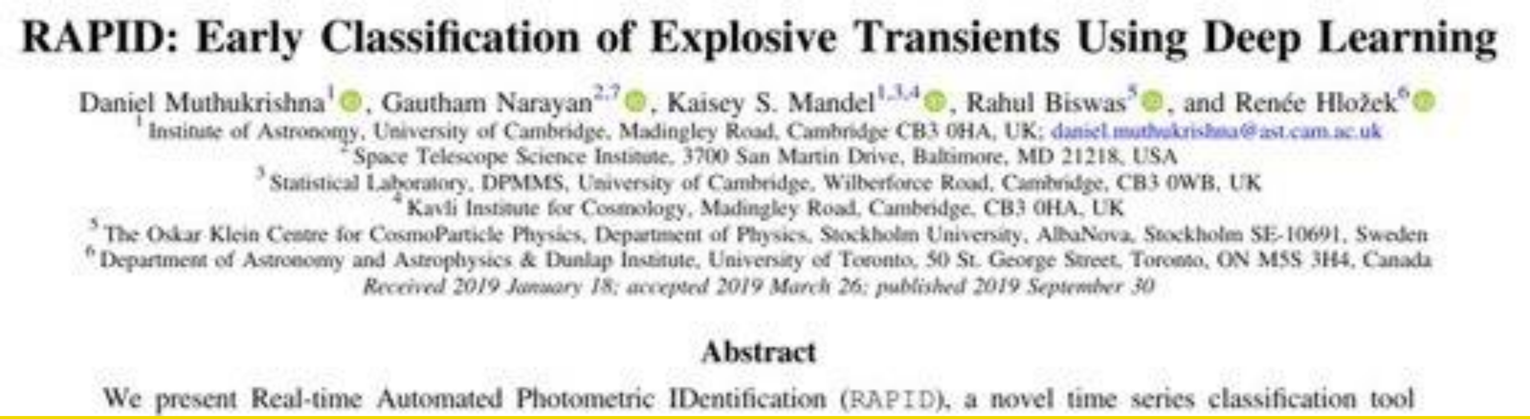
RNNs > variable length

(so it's great for time series or language analysis)

CNNs, NNs > fixed length input

Good place to start to gain intuition:
[Youtube – StatQuest RNN]

The original plan...



RAPID: Early Classification of Explosive Transients Using Deep Learning

Daniel Muthukrishna[1], Gautham Narayan[2,7], Kaisey S. Mandel[1,3,4], Rahul Biswas[5], and Renée Hložek[6]
[1] Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK; daniel.muthukrishna@ast.cam.ac.uk
[2] Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA
[3] Statistical Laboratory, DPMMS, University of Cambridge, Wilberforce Road, Cambridge, CB3 0WB, UK
[4] Kavli Institute for Cosmology, Madingley Road, Cambridge, CB3 0HA, UK
[5] The Oskar Klein Centre for CosmoParticle Physics, Department of Physics, Stockholm University, AlbaNova, Stockholm SE-10691, Sweden
[6] Department of Astronomy and Astrophysics & Dunlap Institute, University of Toronto, 50 St. George Street, Toronto, ON M5S 3H4, Canada
Received 2019 January 18; accepted 2019 March 26; published 2019 September 30

Abstract

We present Real-time Automated Photometric IDentification (RAPID), a novel time series classification tool

Muthukrishna et al. 2019

with some adjustments...

1. **Use real data instead of synthetic**
2. **Reduce number of supernova classes to 4 (IIP, Ia, Ibc, IIb)**
3. **Include CVs**
4. **Use *all* the data in the stream not just LC**

The plan evolved a bit…

To train an RNN (or transformer or else) you need to do data augmentation…

… Gaussian Processes (GP)…

…GP (default) no ideal for SN (Stevance et al. 2023) - need to fix this before we try to train…

⇩

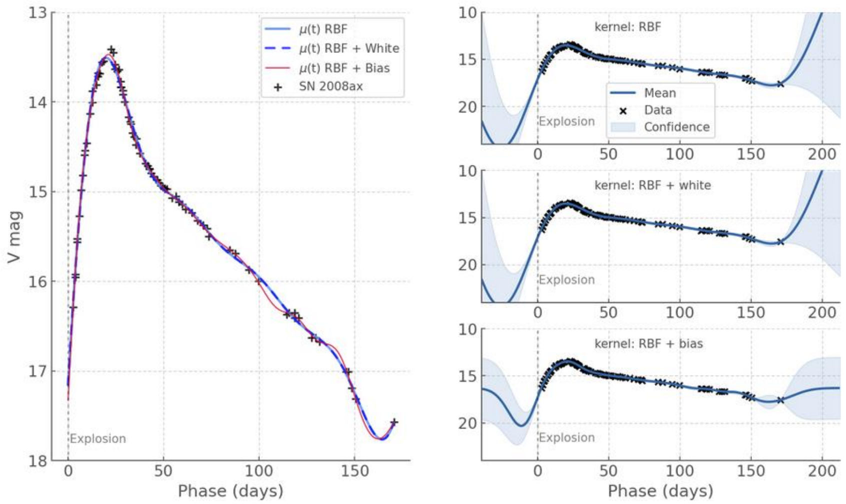
⚠ **Basic GPs assume STATIONARY time series – this is broken by transient physics!** (leads to overfitting)

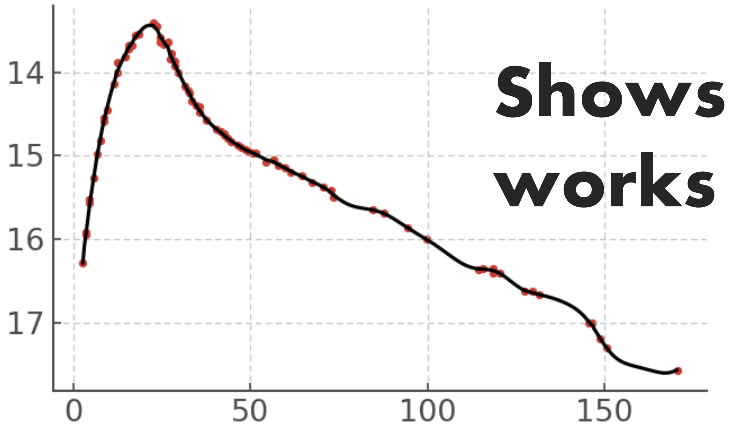
Jupyter Notebooks on GitHub

Reproduce figures of the paper

```
ax[1].text(1, 22, 'Explosion', color='k', alpha=0.5)

ax[2].text(40,12, 'kernel: RBF + white')
ax[2].text(1, 22, 'Explosion', color='k', alpha=0.5)

ax[3].text(40,12, 'kernel: RBF + bias')
ax[3].text(1, 22, 'Explosion', color='k', alpha=0.5)
ax[3].set_xlabel('Phase (days)')

#plt.savefig('sn2008ax.png', dpi=180, bbox_inches='tight')
```

Out[22]: Text(0.5, 0, 'Phase (days)')





Shows you what works well

HUZZAH! It's still over fitting a bit the latter part of the LC but it's not insane. And you'll notice that if you add a k_white kernel above the kit doesn't change, whereas perviously it went from squiggly to basically a smoothly decaying curve. Why is that? That's because earlier on the data kernel (RBF or MAtern) was doing such a terrible job that the optimizer just fitted the data with noise (if you plotted the uncertainties they'd cover the data).

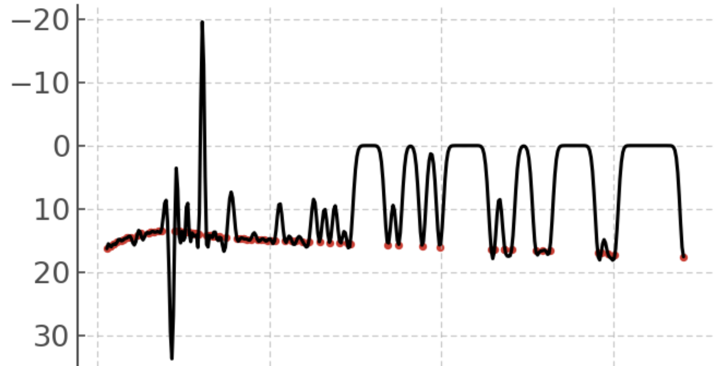
Jupyter Notebooks on GitHub

4.1 RBF and Rational Quadratic

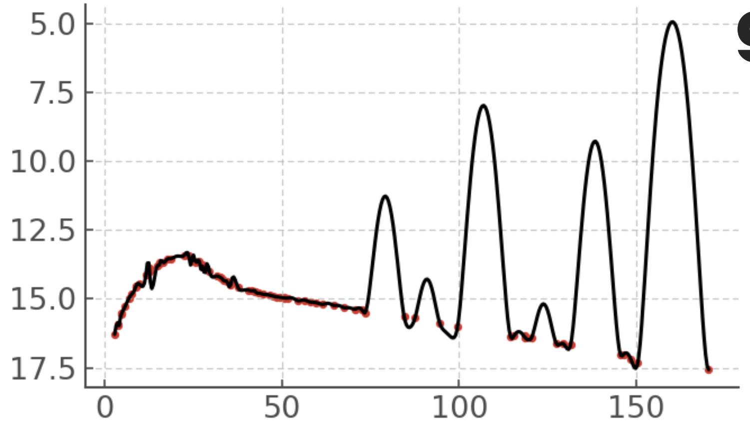
As we'll see not all kernels are created equal. The RBF in sklearn falls over its face in a dramatic fashion

In [55]:
```
gp = sklgp.GaussianProcessRegressor(kernel=k_rbf, n_restarts_optimizer=100, normalize_y=False)
gp.fit(X, y)

x=np.atleast_2d(np.linspace(t_norm.min(), t_norm.max(),400)).T
_mag_interp, sigma = gp.predict(x, return_std=True)
_time_interp = x.T[0]

plt.plot(x, _mag_interp, c='k', label='GP')
plt.scatter(t_norm, y, label='data')
plt.gca().invert_yaxis()
```

```
/home/fste075/.local/lib/python3.8/site-packages/sklearn/gaussian_process/_gpr.py:370: UserWarning:Predicted varia
nces smaller than 0. Setting those variances to 0.
```



https://scikit-learn.org/stable/modules/preprocessing.html



And what goes wrong!

It's terrible! and this is a very good demonstration of what it means for our kernel to be stationary and to overfit the data because it *expects* large deviations on short timescales. Now obviously those large deviations we see in the plot are only one part of the solution - I've not plotted the **ucnertainties** we get from the predicted covariance in these regions of the LC, adn they would be massive! But they needen't be. We know a SN LC will not behave like that - our kernel is **misinformed**.

… and then

Using Post follow-up labels

Goal is to help humans determine WETHER to follow-up

× **Using Post follow-up labels**

Goal is to help humans determine WETHER to follow-up

× **Still focusing on transients**

(and forgetting the contaminants, e.g. CVs)

- ✗ **Using Post follow-up labels**
  Goal is to help humans determine WETHER to follow-up
- ✗ **Still focusing on transients**
  (and forgetting the contaminants, e.g. CVs)
- ✗ **Still obsessing about light-curves**
  (when there is so much more info in the data stream)

Let's start again

The mission

> Make a tool to **assist human** eye-ballers in that ATLAS sky survey by automating some of the **early decision** jobs and flagging **interesting targets for follow-up**

Start with the humans

Not the technology

1) What do they do?

2) What do they need?

Start with the humans

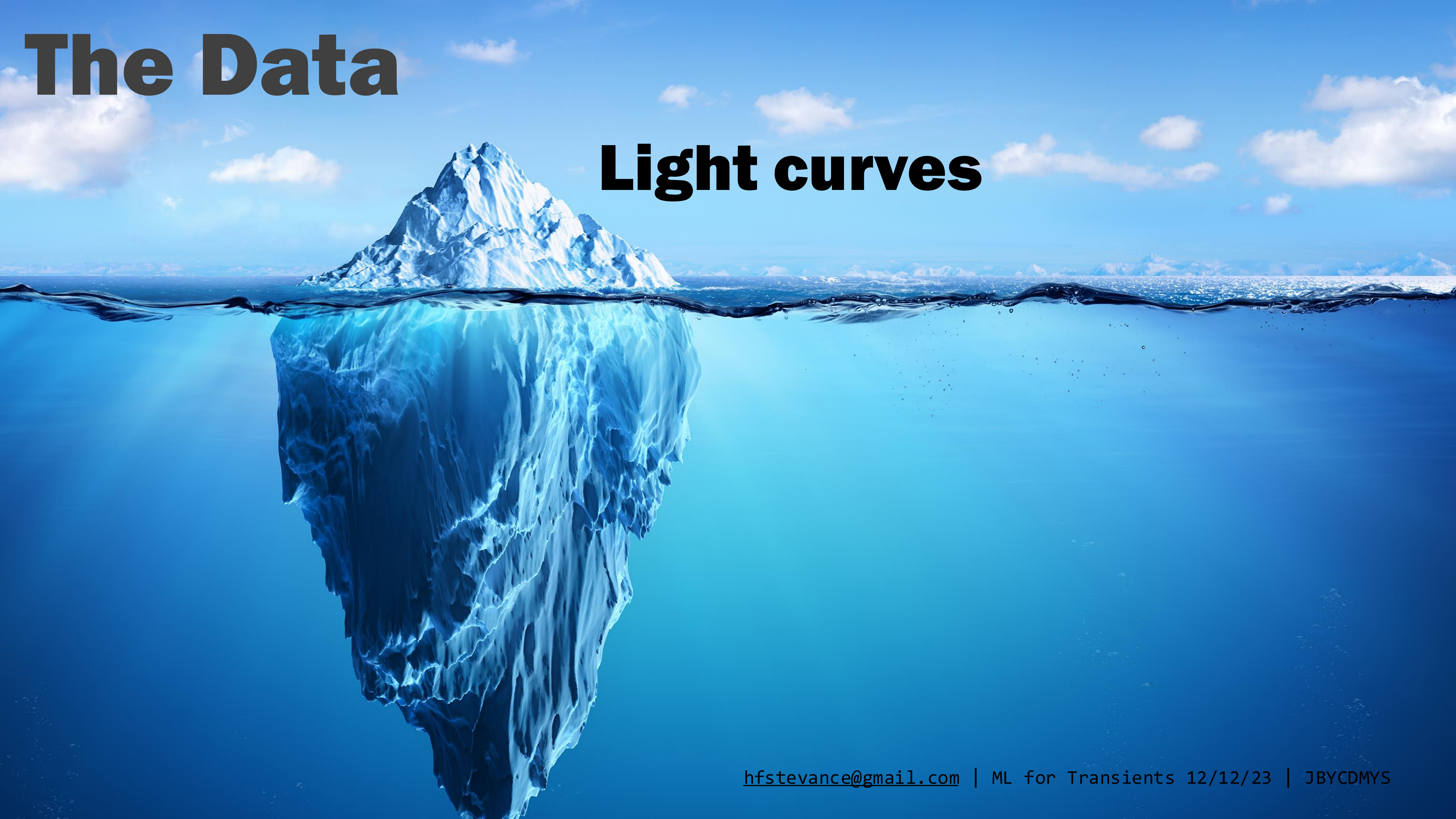
Not the technology

1) What do they do?

2) What do they need?

Conducted Needs Assessment interviews with experienced eyeballers

The Data



Light curves

1024904050142814300

Coords:

02:49:04.05 -14:28:13.0

(42.26690 -14.47028)

Galactic (l,b)

(194.09003,-59.84375)

Realbogus Factors

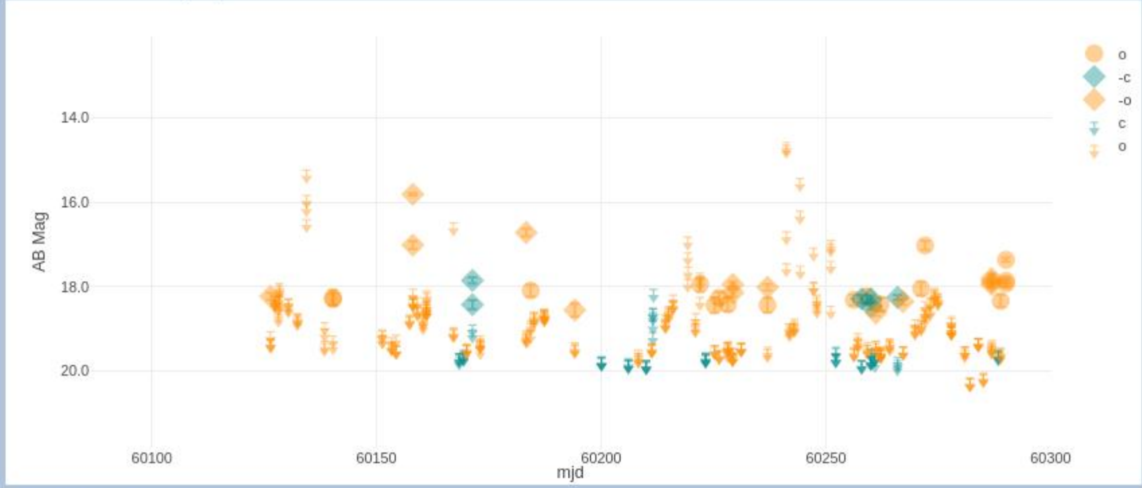
0.09 (DEW) 0.46 (TF)

Flag Date:

12 Dec 2023, midnight

Generate AstroNote

Unforced Photometry Lightcurve



Raw Unforced Data

Current MJD (vertical line): 60290.33725

Number of Detections:

37

Object List:

pending

Processing Flags:

moons stamps eph

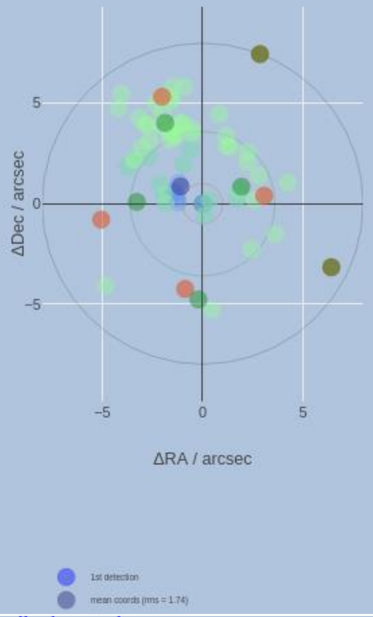
Context:

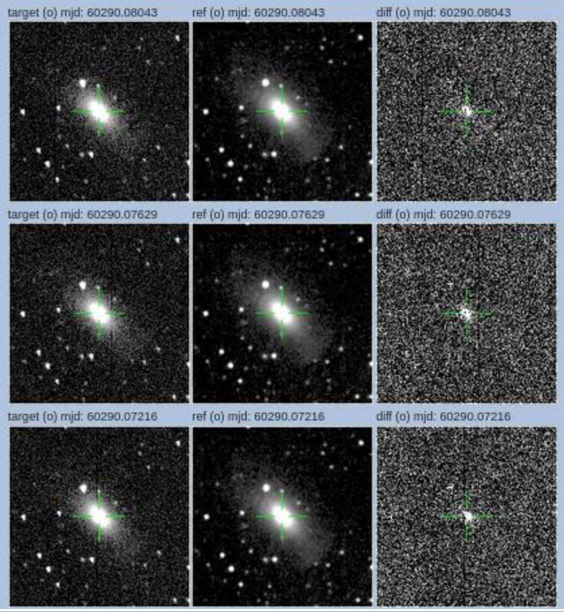
NT

The transient is synonymous with *NGC1120*; a 14.83 mag galaxy found in the NED_D catalogue. It's located 1.1" (0.7 Kpc) from the galaxy core. host distance of 131.0 Mpc(z=0.028) implies a *m* - *M* = 35.59.

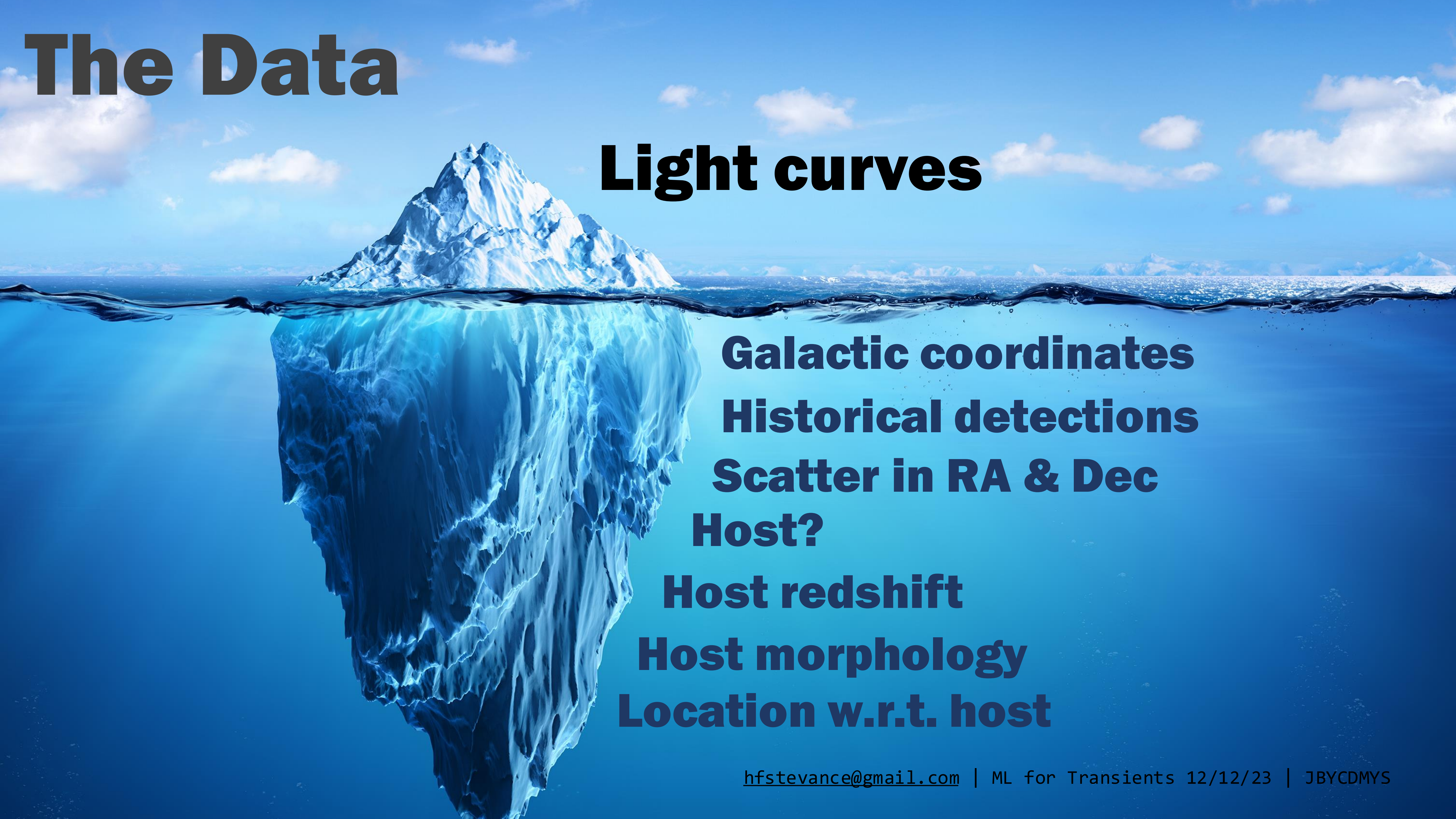
Possible GW Events Association:

S231129ac qub map (MJD = 60277.34567). Time since GW trigger 11.53 days. Within 90% contour (87.29%). Position specific distance 3964 (+/- 1513) Mpc. BBH (99%) Terr (1%) Alert time: 29 Nov 2023, 8:26 p.m.. Alert type: UPDATE.





Det id	Ra	Dec	Mag	Dmag	X	Y	Major	Minor	Phi	Det	Chin	Pvr	Ptr	Pmv	Pkn	Pno	Pbn	Pcr	Pxt	Psc	Dup	Wpflx	Dflx	Image group id	Mjd	Obs	Mag5sig	Atlas metadata id	Texp	Filt	Obj	fpRA	fpDec
4952	42.2669	-14.47064	-18.232	0.081	3125.39	10114.0	2.14	1.79	104.3	5	5.27	0	0	0	0	0	0	0	0	0	0	30936.7	16.4	—	60126.41072825	04a60126o1018o (04a)	19.26	—	30.0	o	SN041S17	41.11854	-16.96247
3054	42.267	-14.47036	18.262	0.184	5145.21	10050.6	2.77	2.16	101.7	0	1.45	34	0	0	0	0	0	0	0	965	0	31795.7	15.1	—	60140.30465715	04a60140o0783o (04a)	18.97	—	30.0	o	SU042S17	42.18987	-16.93055
2736	42.26606	-14.47033	18.298	0.145	5171.08	10094.9	2.35	2.09	108.6	0	6.37	0	0	0	0	0	0	0	0	999	0	18175.1	20.9	—	60140.33084325	04a60140o0839o (04a)	19.22	—	30.0	o	SU042S17	42.20386	-16.95295
3767	42.2673	-14.47088	-15.809	0.03	9000.48	41.13	6.11	2.52	179.0	5	2.49	0	0	0	0	0	0	0	0	0	0	25770.6	12.6	—	60158.10550145	03a60158o0882o (03a)	18.0	—	30.0	o	SE044S12	44.17885	-11.73932
3666	42.26728	-14.4707	-17.015	0.09	9017.52	81.65	5.89	2.47	179.8	5	1.41	0	0	0	0	0	0	0	0	0	0	28073.4	12.4	—	60158.10271705	03a60158o0876o (03a)	18.05	—	30.0	o	SE044S12	44.18875	-11.75976
3234	42.26748	-14.47059	-17.852	0.075	8664.69	7835.87	2.66	2.37	121.7	5	2.65	0	0	0	0	0	0	0	0	0	0	15429.7	6.4	—	60171.34355125	04a60171o0835c (04a)	19.22	—	30.0	c	SK044S16	44.071	-15.79781
3083	42.26736	-14.47128	-18.43	0.106	8644.44	7950.74	2.7	2.37	128.6	5	2.04	0	0	0	0	0	0	0	0	0	0	10573.9	3.9	—	60171.33810335	04a60171o0823c (04a)	18.97	—	30.0	c	SK044S16	44.05977	-15.85766


The Data
Light curves
Galactic coordinates
Historical detections
Scatter in RA & Dec
Host?
Host redshift
Host morphology
Location w.r.t. host
hfstevance@gmail.com | ML for Transients 12/12/23 | JBYCDMYS

The Data

Light curves

At early days only 2 to 3 points!

Galactic coordinates

Historical detections

Scatter in RA & Dec

Host?

Host redshift

Host morphology

Location w.r.t. host

Needs Assessment: Lessons learned

> these require much more than the time series

High level labels

- **Real** (vs bogus)
- **Extra-galactic** (vs galactic)
- **Fast Follow-up Candidate** (vs can wait for more data)
- **Follow-up Candidate** (vs don’t care)

Needs Assessment: Lessons learned

> the time series classes are very broad

Important LC labels

- SPEED: Fast | Normal SN | Slow
- LUMINOSITY: Faint | Normal SN | Bright
- STRUCTURE: Early peak + Main peak

MORE DETAILED CLASSES WOULD NOT BE TRUSTED

Forget the SN classification

Especially at early days!

P(Real)	P(ExtraGal)	P(Fast)
P(Bogus)	P(Galactic)	P(Normal or Slow)

Forget the SN classification

Especially at early days!

P(Real) | **P(ExtraGal)** | **P(Fast)**

P(Bogus) | P(Galactic) | P(Normal or Slow)

NOT MUTUALLY EXCLUSIVE

Real

CV, Variable Stars

AGN, SNe, KNe, TDE etc..

Galactic

XtrGal

Edge artifacts (bright stars)

Gal. Nucleus bad subtraction

Bogus

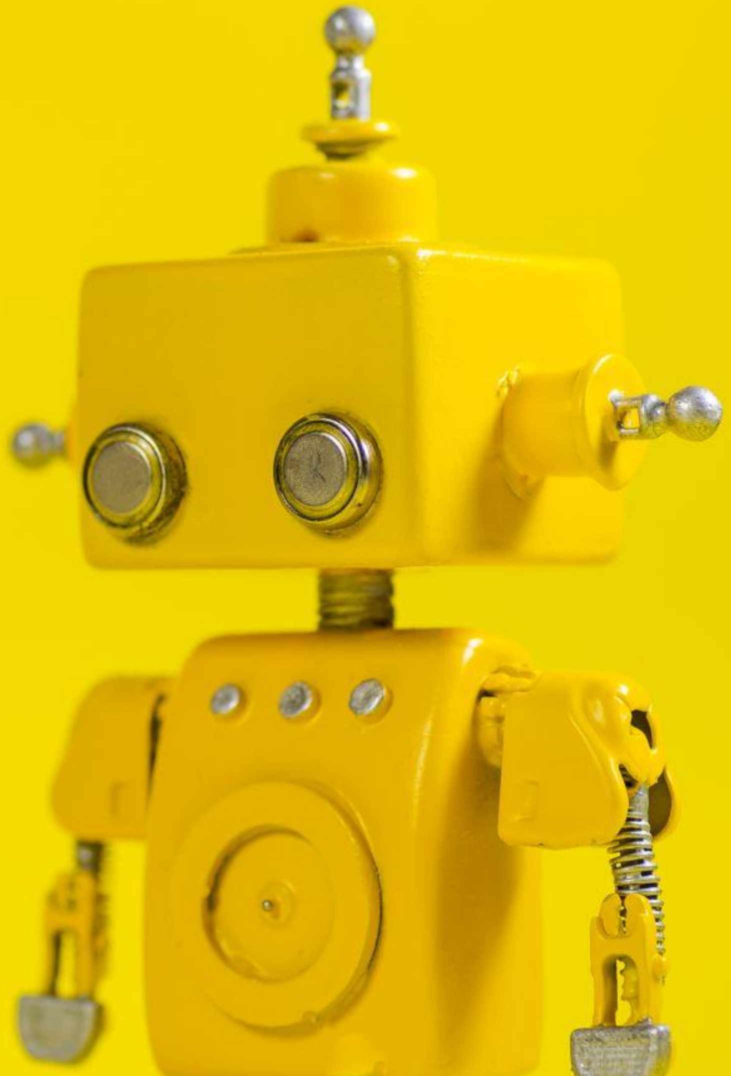
Forget the SN classification

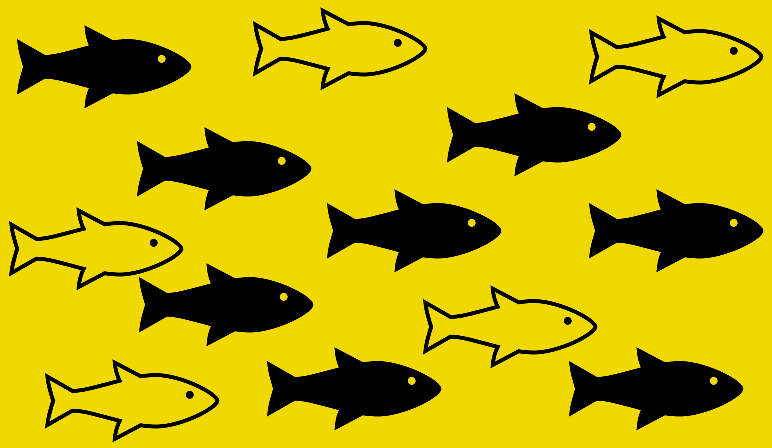
Especially at early days!

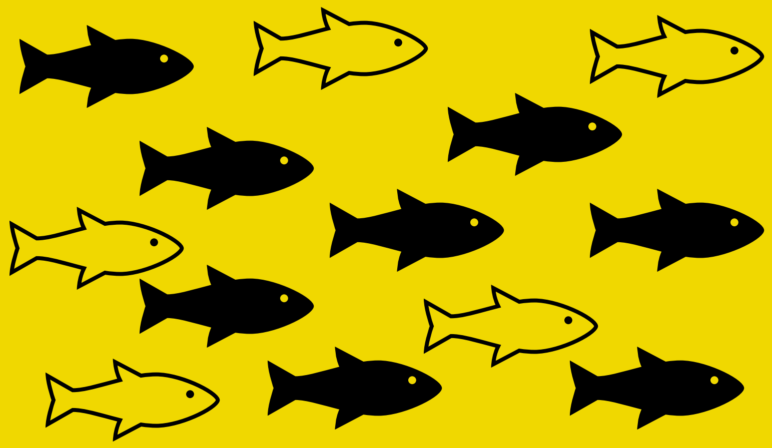
P(Real)	P(ExtraGal)	P(Fast)
P(Bogus)	P(Galactic)	P(Normal or Slow)

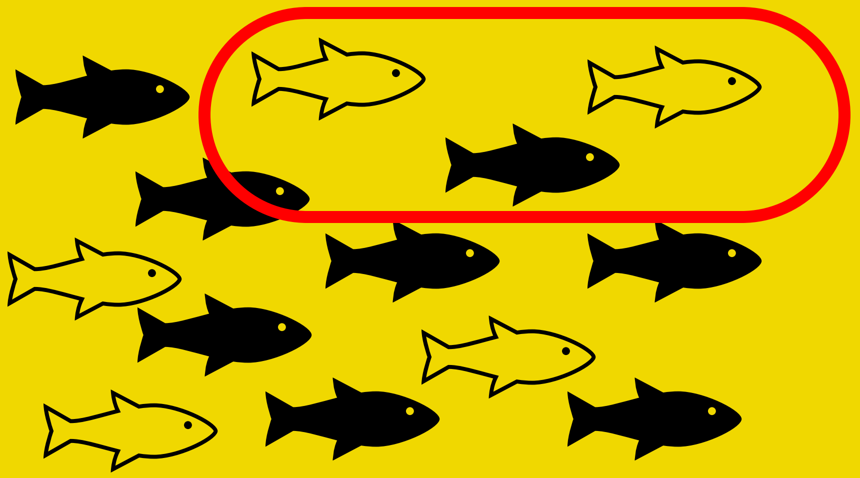
Haven't we already done this?

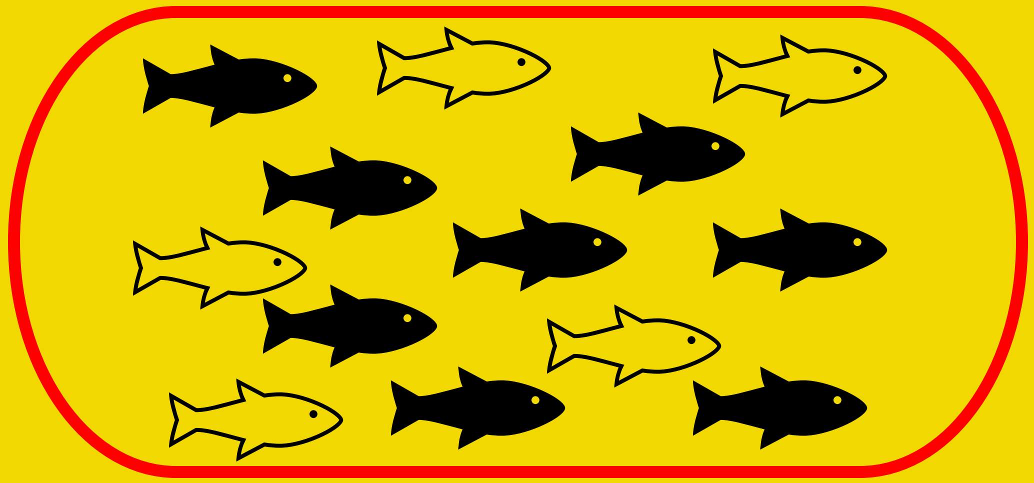
Contaminants are a feature, not a bug











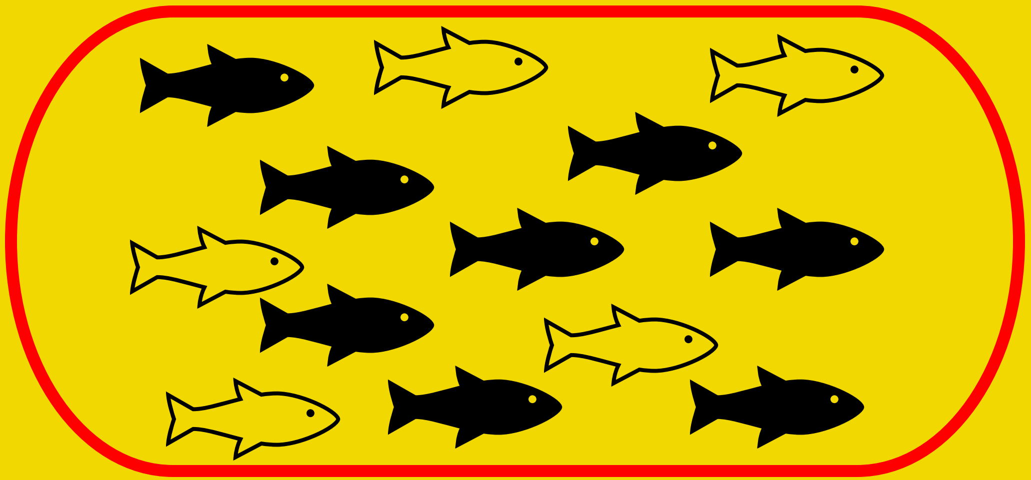
Purity

Completeness (recall)

For our science case we want to maximise RECALL

(you don't want to bin the next cool transient)

<u>Even in theory</u>, the "ideal" Real/Bogus classifiers will let through a lot of bad alerts



<u>Completeness</u>

(recall)

P(Real | Data)

P(ExtraGal | Data)

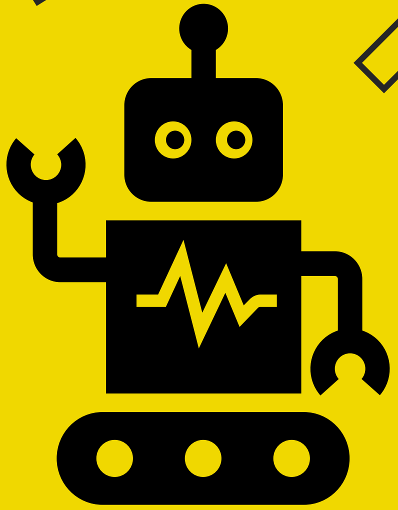
P(Fast | Data)

XGBoost or similar (<u>not NN</u>)

Sort alerts in eyeball list

See the most important + urgent at the top

V.R.A



Update everyday

Bogus will weed itself out without humans having to look at it!

1

First alerts (1-2 days)

First estimates of P(real), P(XtrGal), P(Fast)
Immediate attention required? Can wait for more data?

2

Updating (3-21 days)

Updating and settling of P(real), P(XtrGal), P(Fast)
Garbage discarded, Light Curve labels start being applied

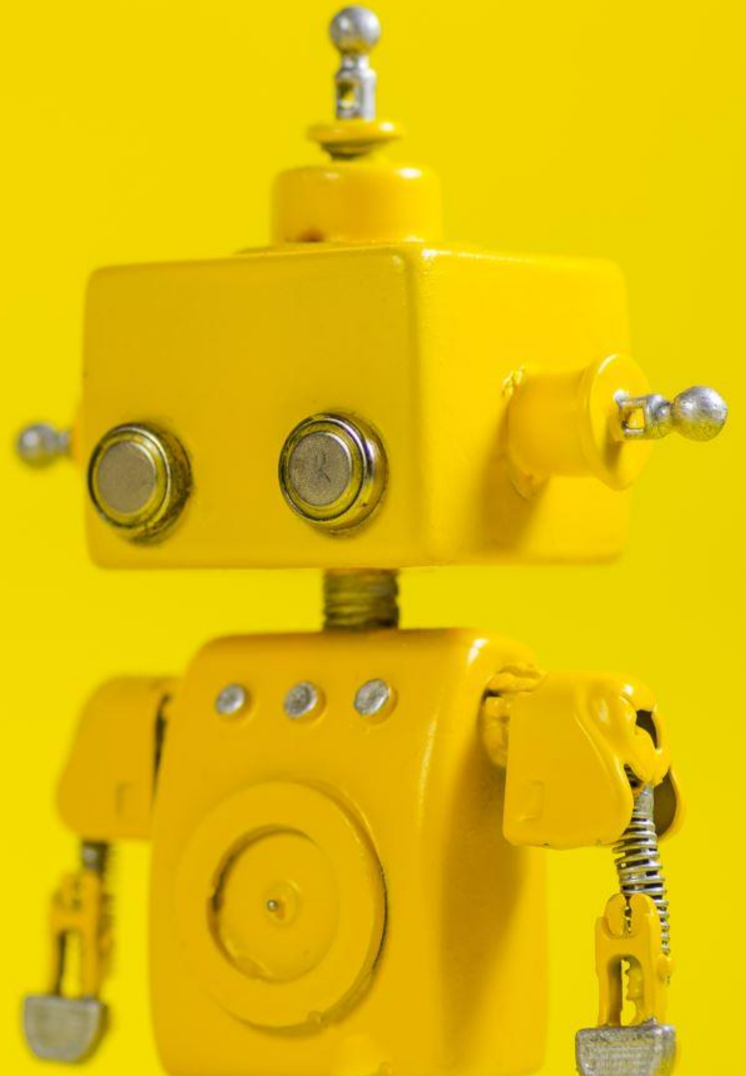
3

Long Term monitoring (>21 days)

follow-up YES: monitor for anomalies
follow-up NO: alert when observing window is closing

A.I. in Science…

Going beyond the proof of concept



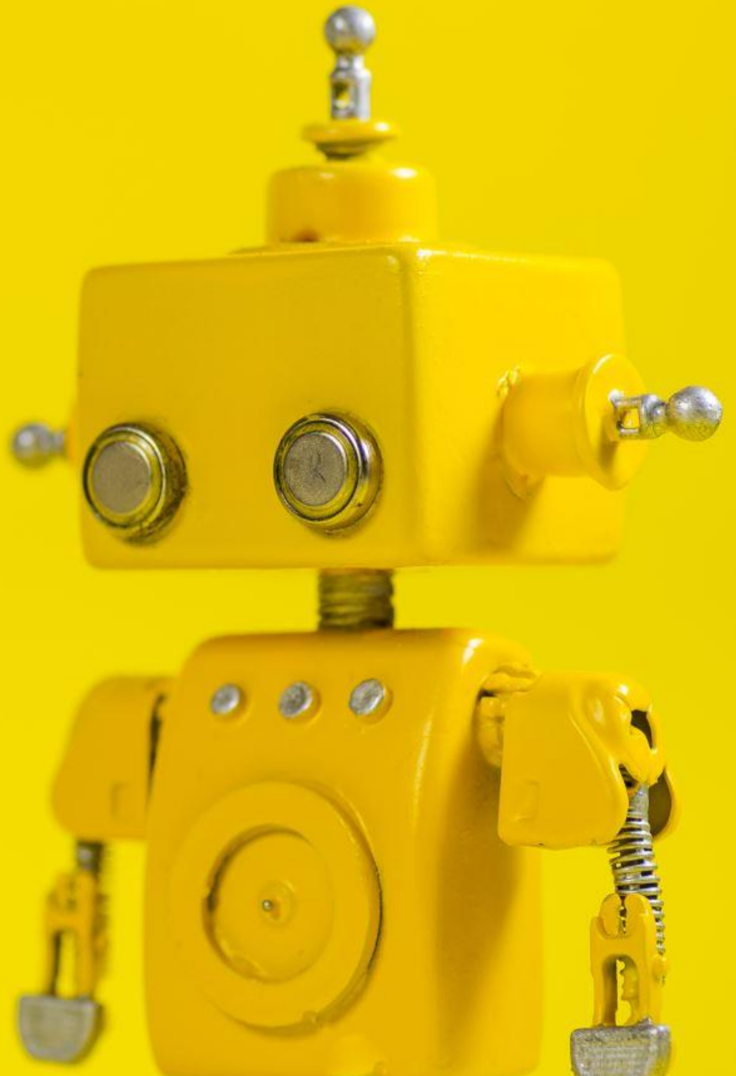
Asking the right questions

Data Drift

see also Tom's talk

Metrics

Is the cost justified?



1 Asking the right questions: ML & transients

What is your classifier for?

1 Asking the right questions: ML & transients

What is your classifier for?

LC only classification w/o spectroscopy?

- OR -

Preliminary Classification before follow-up?

1 Asking the right questions: ML & transients

What is your classifier for?

LC only classification w/o spectroscopy?

- OR -

Preliminary Classification before follow-up?

Work alongside humans?

- OR -

Fully automated decisions?

1 Asking the right questions: ML & transients

What is your classifier for?

LC only classification w/o spectroscopy?

- OR -

Preliminary Classification before follow-up?

Work alongside humans?

Where do they come into the process?

- AND -

What do they need?

- OR -

Fully automated decisions?

1 Asking the right questions: ML & transients

What is your classifier for?

You need to truly understand the use case of your tool to pick the right metrics and evaluate its usefulness over each step of development.

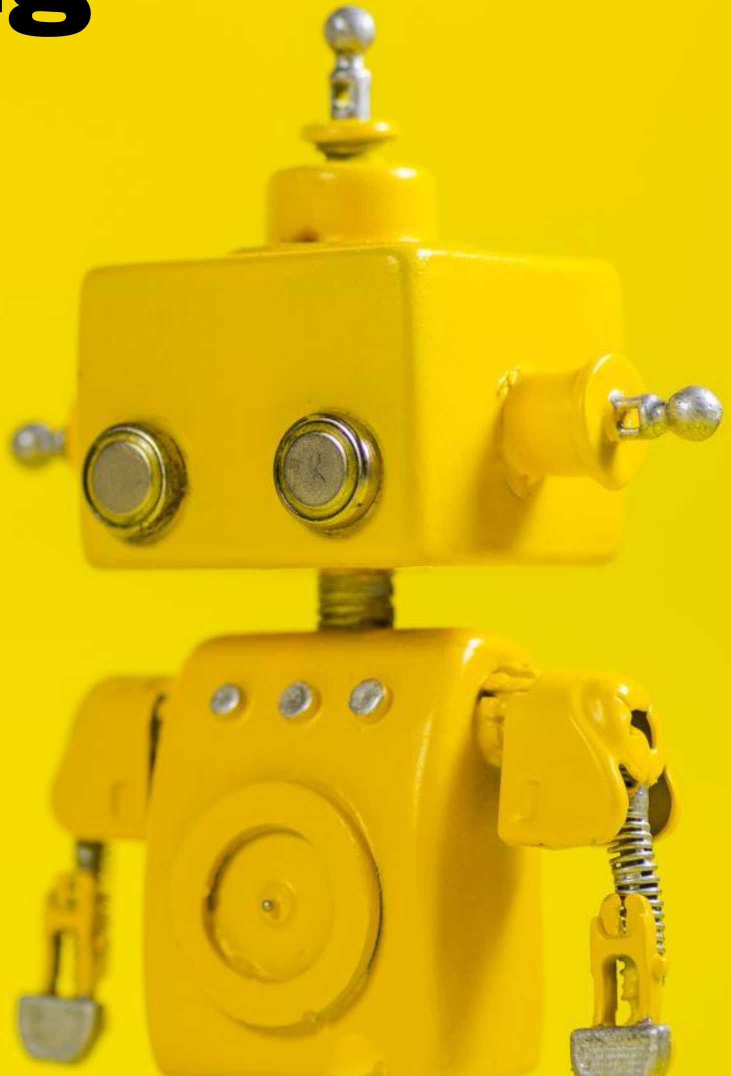
2 Data Drift

The properties of your training sample are <u>not drawn from the same distributions</u> as those in the real-life use data

⇨ Why training on only **synthetic data** has been found to lead **poor results** IRL.

⇨ Why you **SHOULDN'T IGNORE CONTAMINANTS** (CVs, M-dwarfs, etc...)

If a telescope is going to see it, test on it...



2 Data Drift

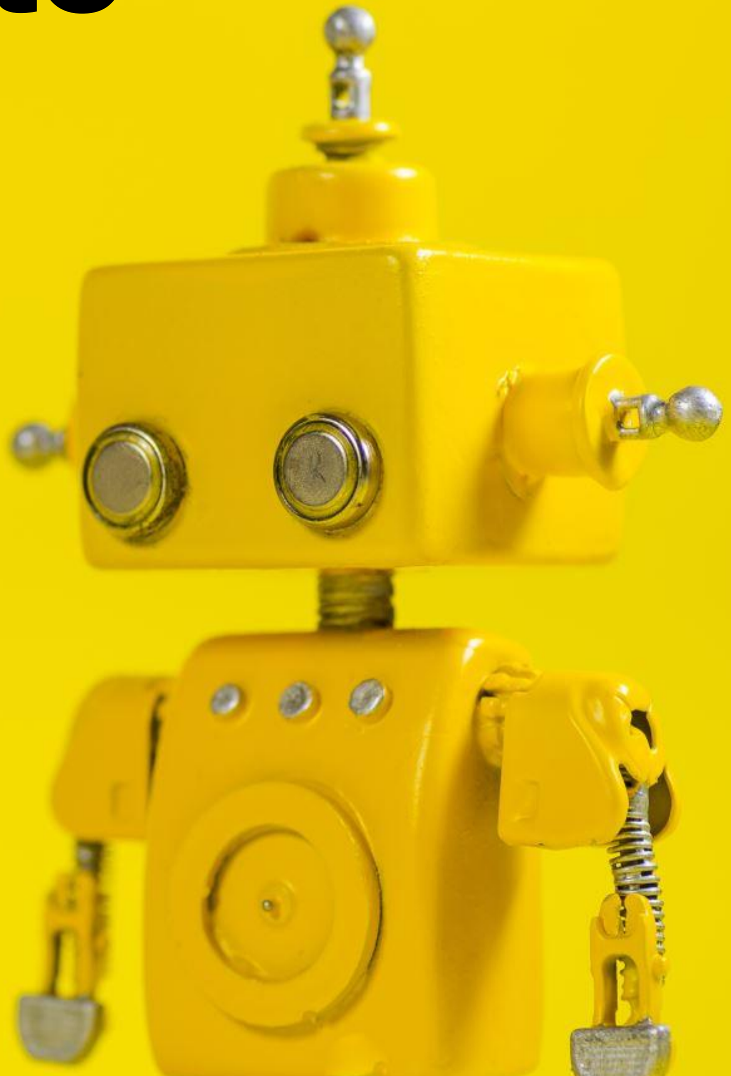
Outside-of-scope predictions are A BAD TIME....

Even the smart-looking {ChatGPT, DallE}-like generative models – fail catastrophically outside of scope [e.g. Yadlowsky et a. 2023]. They seem smart BECAUSE THEIR CORPUS OF DATA IS VAST.

c.f. "On the dangers of Stochastic Parrots" (Bender et al. 2021)

"Do I have the data I need to get the results I want from a given algorithm?"

You can't "outcompute" inadequate data



3 Metrics

[Metrics Reloaded: Recommendations for image analysis validation]

Accuracy (N_true / N_totalSample)
Purity (N_truePositive / N_predictedPositive)
Completeness (N_truePositive / N_positiveSample)
et al.

What are your priorities? *e.g discussion in Alex A.'s talk*
What are your benchmarks for success?

- For the <u>science goal</u> to be achievable/workable IRL.
- For your tool to be an improvement on existing methods

4 Why this method? Is it worth it?

Is there a simpler* way?

***Less compute needed / more interpretable method**

What problems could you be creating?

e.g. Transient classifiers using Black Box algos will have hard to track biases which could trickle into our transient rate calculations or be v. tricky to account for.

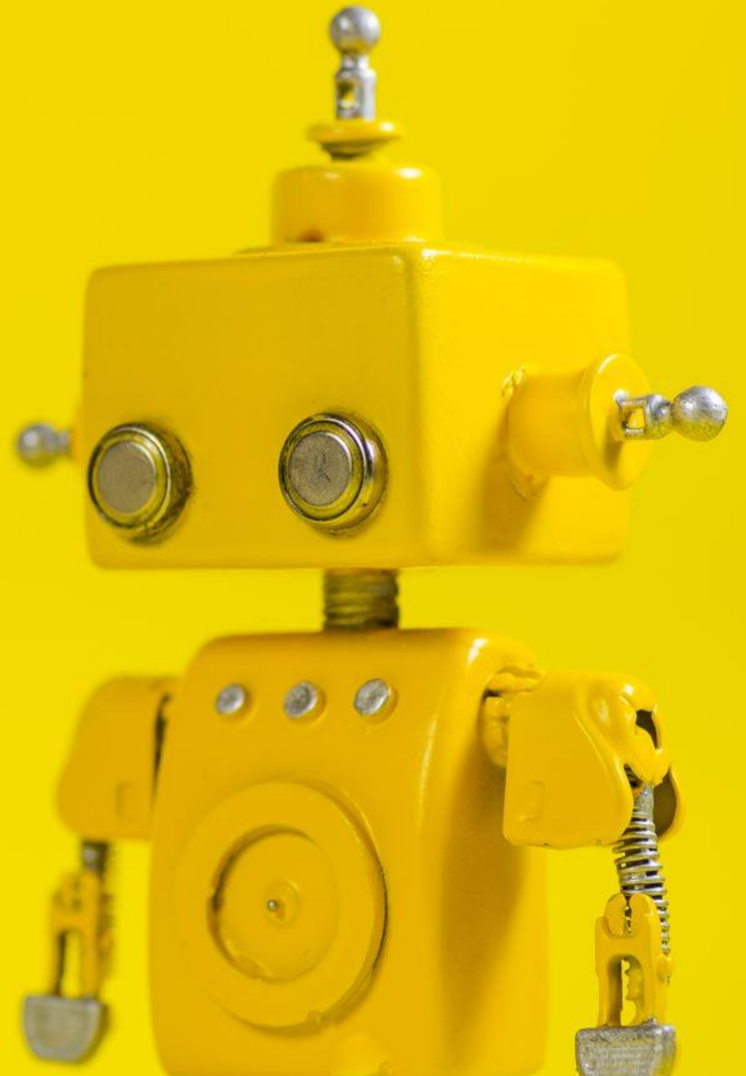
Go beyond the proof-of-concept

DO Carefully scope your use-case

DO Pick metrics that reflect it

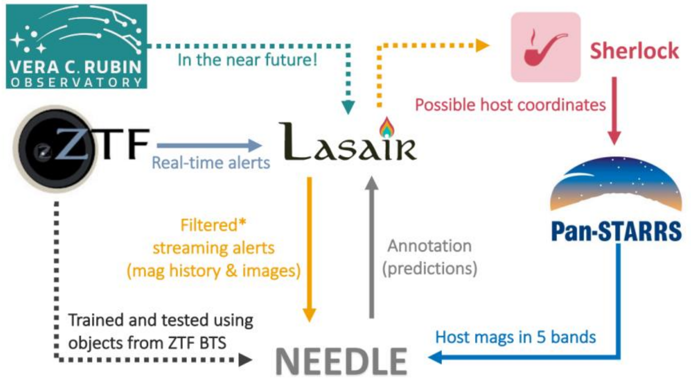
DON'T forget the humans

DON'T forget the contaminants



Xinyue Sheng

PhD Student, ARC, QUB

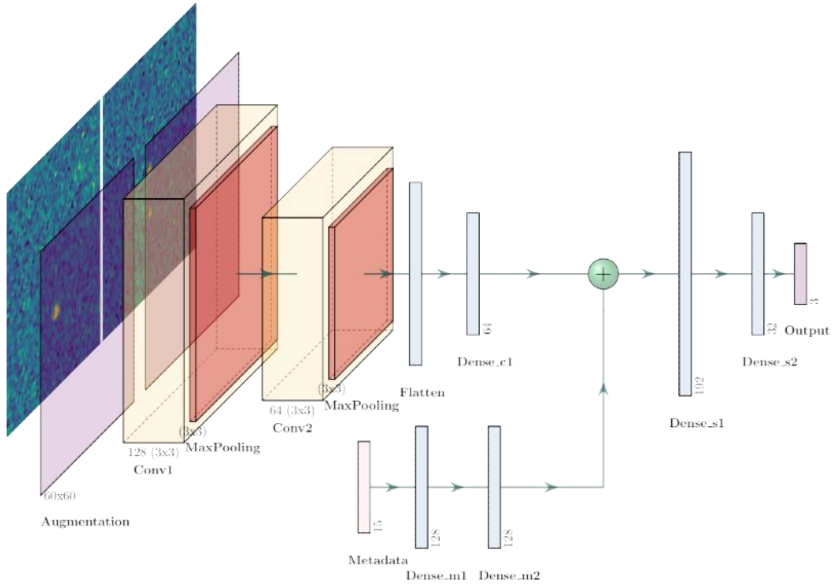


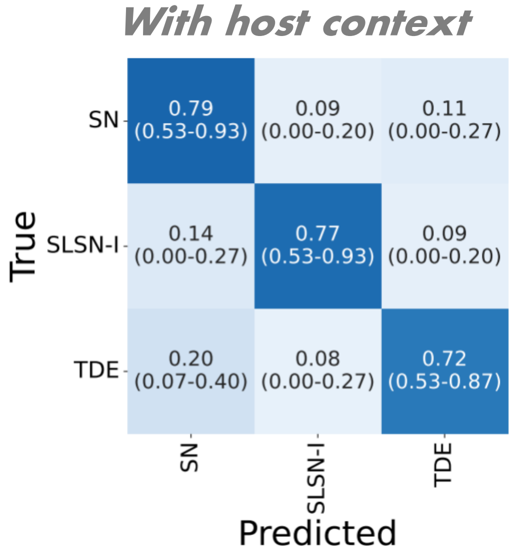
Sheng, Nicholl, et al. (2023) Submitted to MNRAS

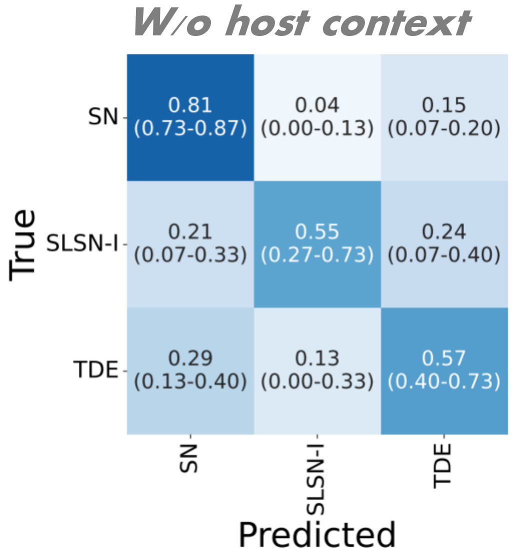
NEural Engine for Discovering Luminous Events (NEEDLE):

identifying rare transient candidates in real time from host

- **Early-stage probability provided**
- **Without redshift information**
- **One-stamp imaging**
- **Using host photometric information**
- **Deal with real-time alerts**
- **Annotation on Lasair**







Go beyond the proof-of-concept

hfstevance@gmail.com

DO Carefully scope your use-case

DO Pick metrics that reflect it

DON'T forget the humans

DON'T forget the contaminants