

Delegating decision making to the computer in Science

Dr. Heloise Stevance

Schmidt AI in Science Fellow, Department of Physics

Green Templeton Associate Fellow

Caroline Herschel Lectureship Prize 2024



Changes in the night sky





Stellar Explosion s





**1 Billion
visible sources**

A night sky photograph featuring the Milky Way galaxy stretching diagonally across the frame. The galaxy's core is visible as a bright, dense region of stars and dust. The foreground shows the dark silhouette of a mountain range, with a small structure visible on a peak. The text "10 Million Differences per night" is overlaid in a large, white, serif font. A small, stylized star icon is positioned to the left of the text.

10 Million Differences per night

Legacy Survey of Space and Time

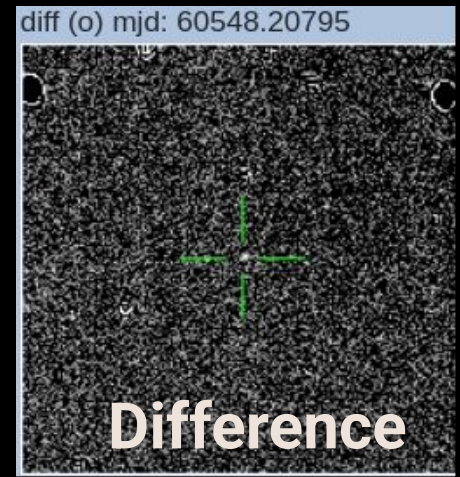
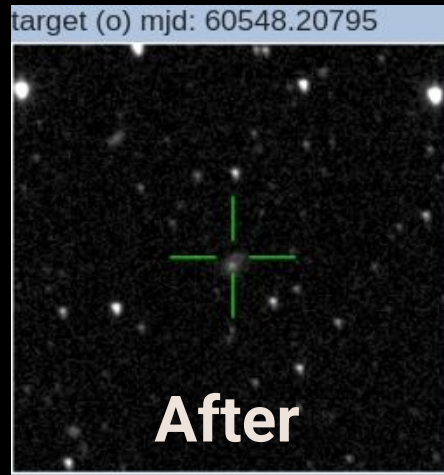
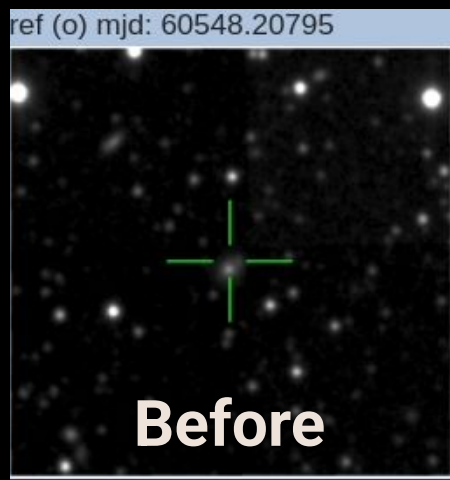
Data
x100



Why we delegate

1. **Data Volume** (inc. dimensionality)
2. **Time Pressure**
- ...
3. **Both!**

How we delegate...



How we delegate:

1. DATA PROCESSING

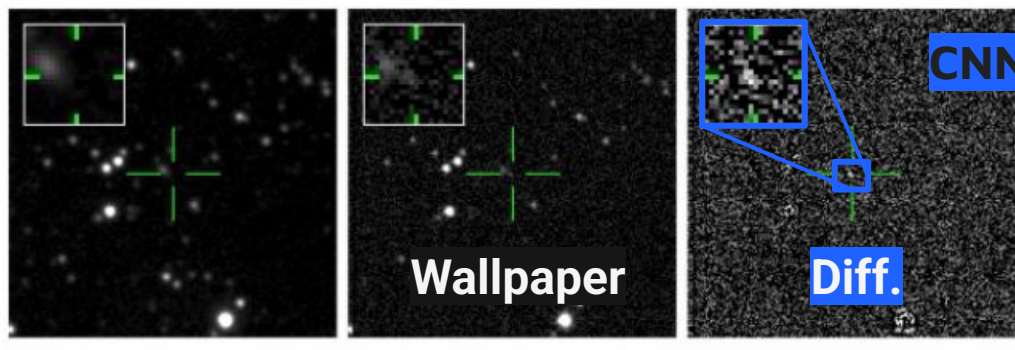


AGN

The transient is synonymous with [SDSS J054724.7+003734.9](#); a J=16.31 mag AGN found in the SDSS/GAIA/PS1/2MASS catalogues. It's located 0.5" from the AGN core.

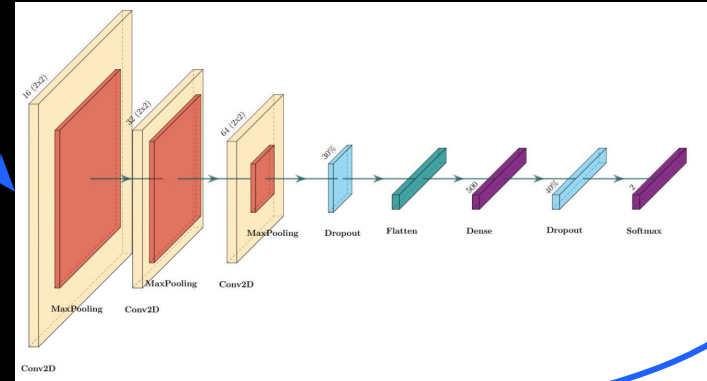
How we delegate:

2. BASIC STATS & CONDITIONS



CNN input

Diff.

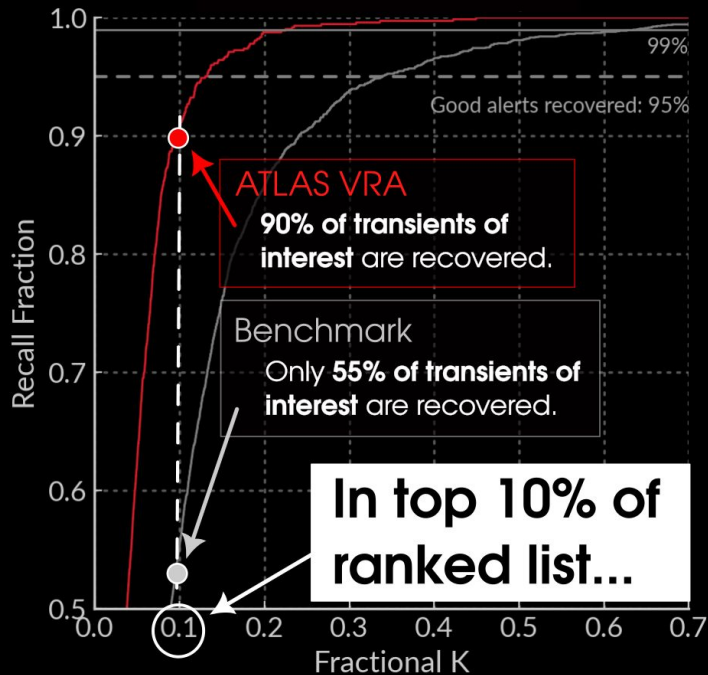



Score between 0 (bogus) and 1 (real)

$< 0.2 \rightarrow$ Garbage



How we delegate:

3. COMPUTER VISION



 **st3ph3n** APP 11:13 PM

Ingest Complete

-  1 objects with rank > 7. (2 with low ranks): [Fast Track List](#)
-  17 objects with rank > 7. [Extragalactic Candidates](#)
-  3 objects. [Galactic Candidates](#)

How we delegate:

4. RANKING ALGORITHM

Writing Code

ChatGPT 5.2 ▾

Correct solution: single source → generate the other

Step 1: Decide the source of truth

Pick one:

- ☒ `lvra_env.yml` (common in science)
- ☐ both (guaranteed pain)

Assume: `lvra_env.yml` is the source of truth.

Step 2: Auto-sync versions into `pyproject.toml`

Add a small script, e.g. `scripts/sync_versions.py`:

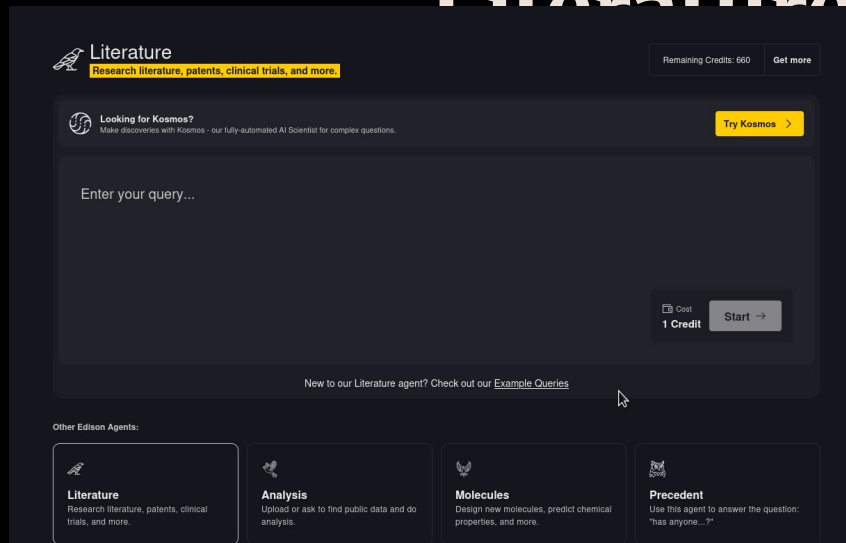
```
python
```

```
from pathlib import Path
import re

env_file = Path("lvra_env.yml")
toml_file = Path("pyproject.toml")
```

Copy code

Searching Literature



How we delegate:

5. Large Language Models

How we delegate

1. Data Processing
2. Basic Stats and Conditions
3. Computer Vision
4. Ranking Algorithm
5. Large Language Models

How we delegate

1. Data Processing
2. Basic Stats and Conditions
3. Computer Vision
4. Ranking Algorithm
5. Large Language Models



**Increasing
Complexit**

y

The Problems

Tech Debt

How hard will it be to maintain a tool/pipeline/method?

Intellectual Debt

How much skill and information do you forsake?

Sovereignty

How much data and skill is controlled by someone else?

Science is

Reproducible

Falsifiable

Aware of biases

Science is

Reproducible

Falsifiable

Aware of biases

Our shared responsibility

My research ethics

1. Software is only open if the data is open

Especially when training algorithms, data is everything, including the way that data has been cleaned and handled.

My research ethics

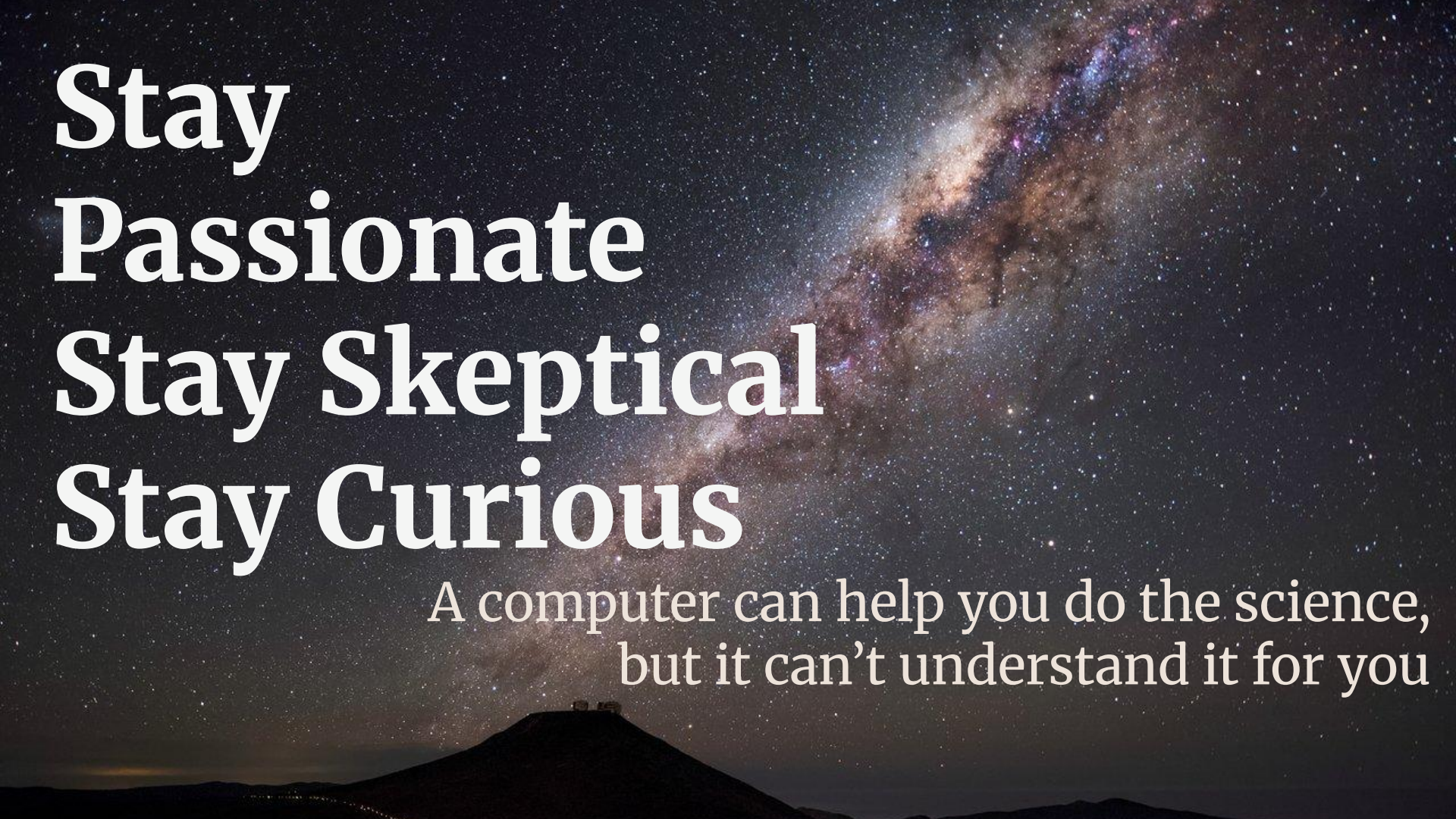
2. Use the simplest solution that works

Whenever we create a new tool and add it to our scientific process, that tool has to be maintained and updated (tech debt) but it also has to be understood so that the data that it generates or currates is understood with all its caveats (intellectual debt). The more complex the algorithms, the fewer people understand it, making it harder to ensure scientific rigour. To minimise this issue, I will always use the simplest tool/algorithm that produces sufficient performance for success.

My research ethics

3. If it works but I don't understand it, *it does not work*

Complex methods have a lower barrier to entry for non-specialists now because they are more user friendly and LLMs like Chat gpt or Claude can help get code that runs without it being understood. It can be tempting to apply a new complex algorithm especially if it looks like it works, but as scientists we have to test our tools thoroughly and we cannot do that if we do not understand them or their caveats.

The background of the image is a deep night sky filled with stars. A prominent feature is the Milky Way galaxy, which appears as a dense, colorful band of light stretching diagonally from the upper right towards the center. The colors range from warm oranges and reds to cooler blues and purples. In the foreground, at the bottom of the frame, is a dark silhouette of a mountain range. The peaks are jagged, and the overall scene conveys a sense of vastness and cosmic wonder.

**Stay
Passionate
Stay Skeptical
Stay Curious**

A computer can help you do the science,
but it can't understand it for you