Ariel Machine Learning Data Challenges

ESA/ESAC
September 2022

Ingo Waldmann

The Alan Turing Institute

UK SPACE AGENCY

Horizon 2020

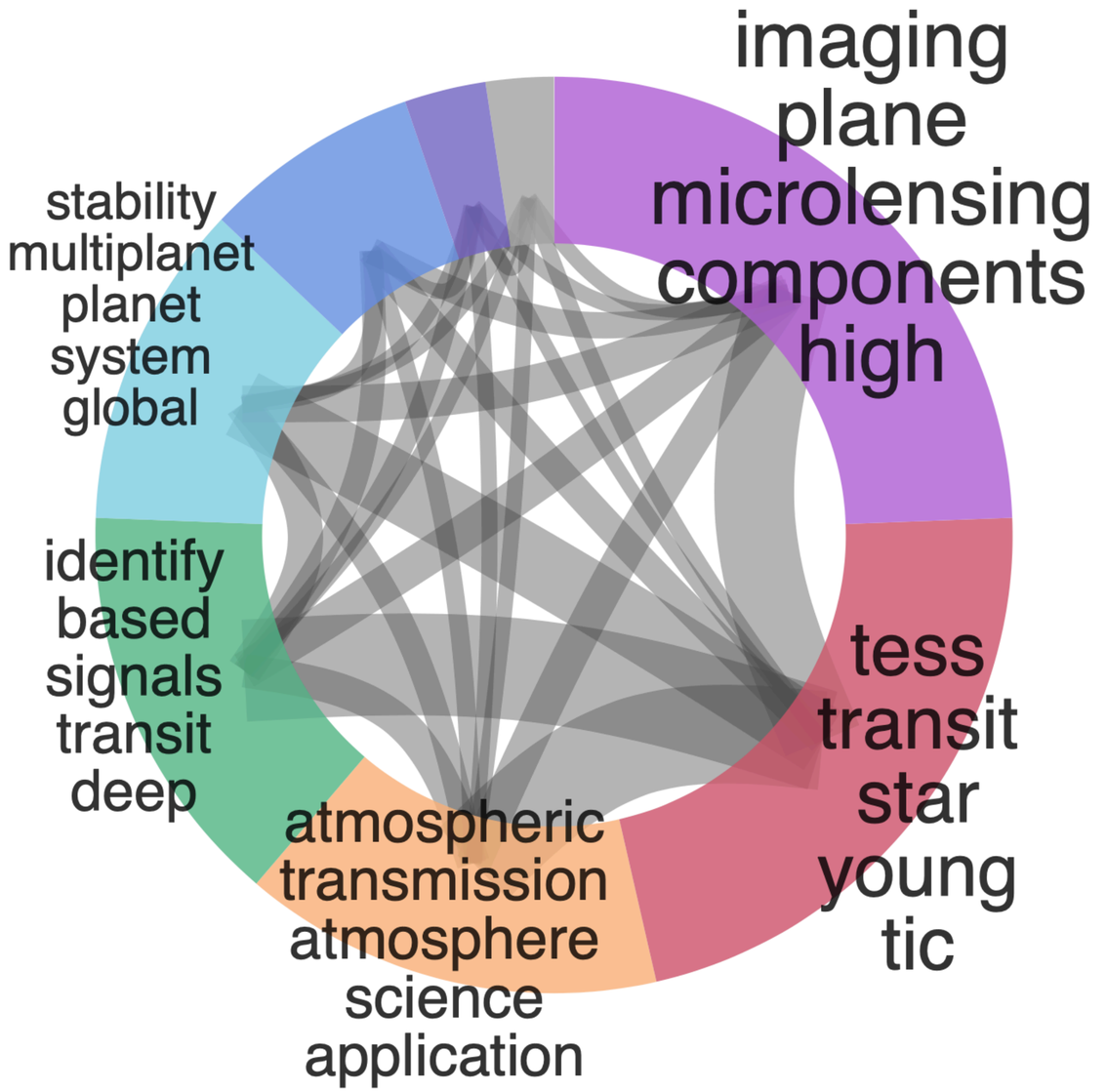
erc
European Research Council
Established by the European Commission

Science & Technology Facilities Council

UCL

State of ML in Exoplanets

- Still very new to exoplanets
- Mainly by three sub-fields: Transit detection, direct imaging, atmospheric modelling
- Number of serious ML/AI papers increased dramatically since 2018/19
- Many concurrent issues with other fields but little cross-talk yet

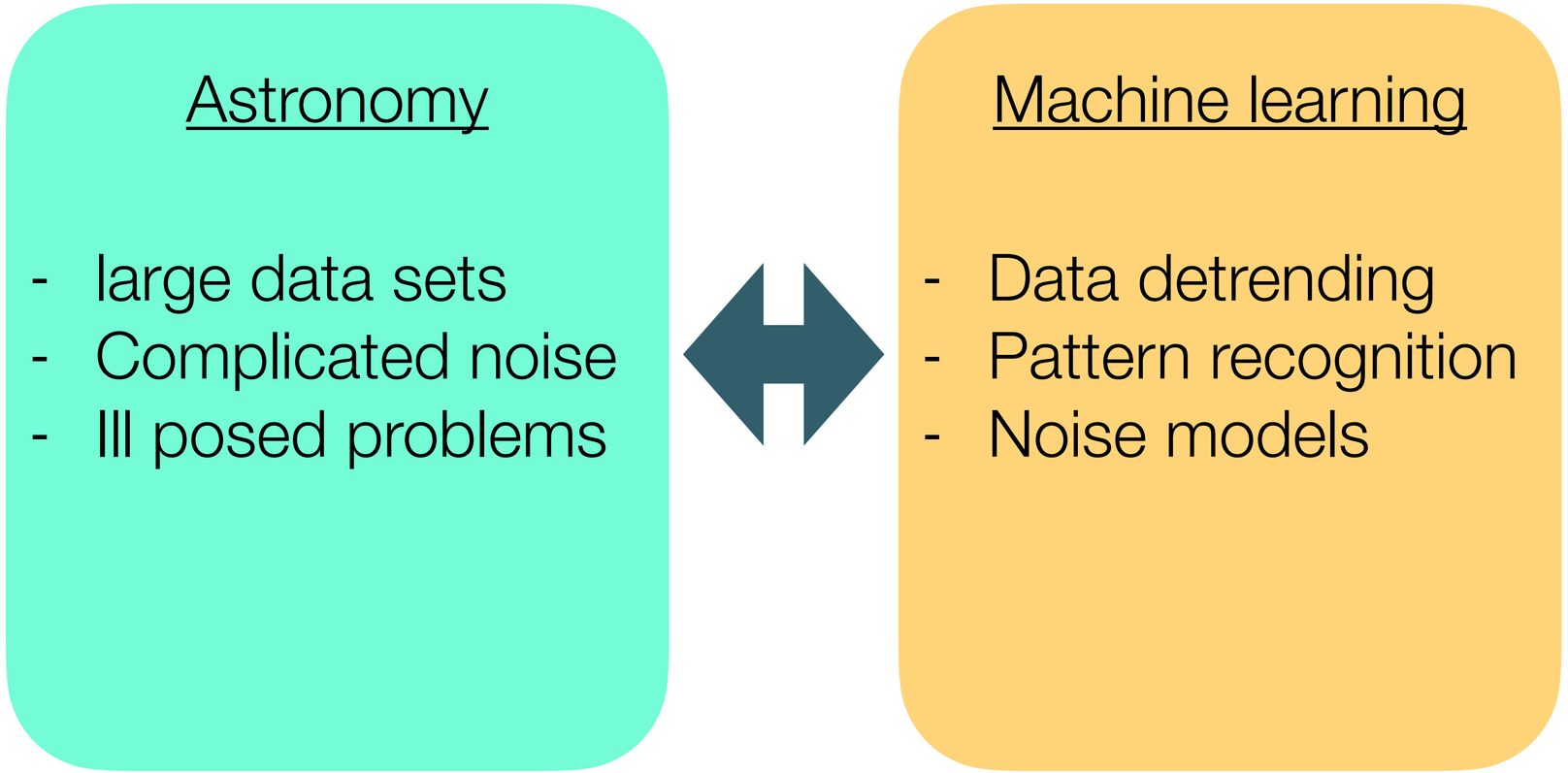


Search term

(full:"extrasolar planet" or full:"exoplanet") and (full:"machine learning" or full:"artificial intelligence" or full:"neural network")

Generate using Astrophysics Data System

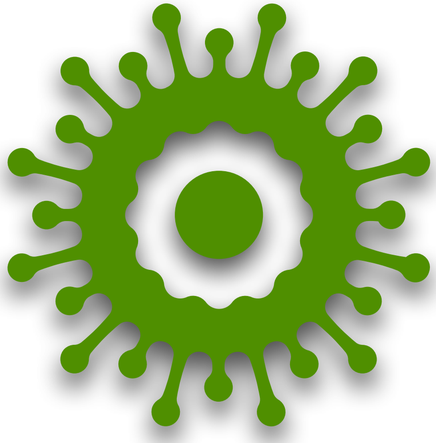
A natural synergy with the AI community



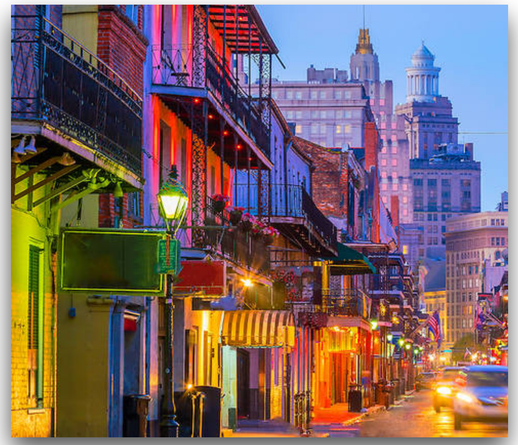
Three Machine Learning Data Challenges

ECML 2019
Würzburg

ECML 2021
Online

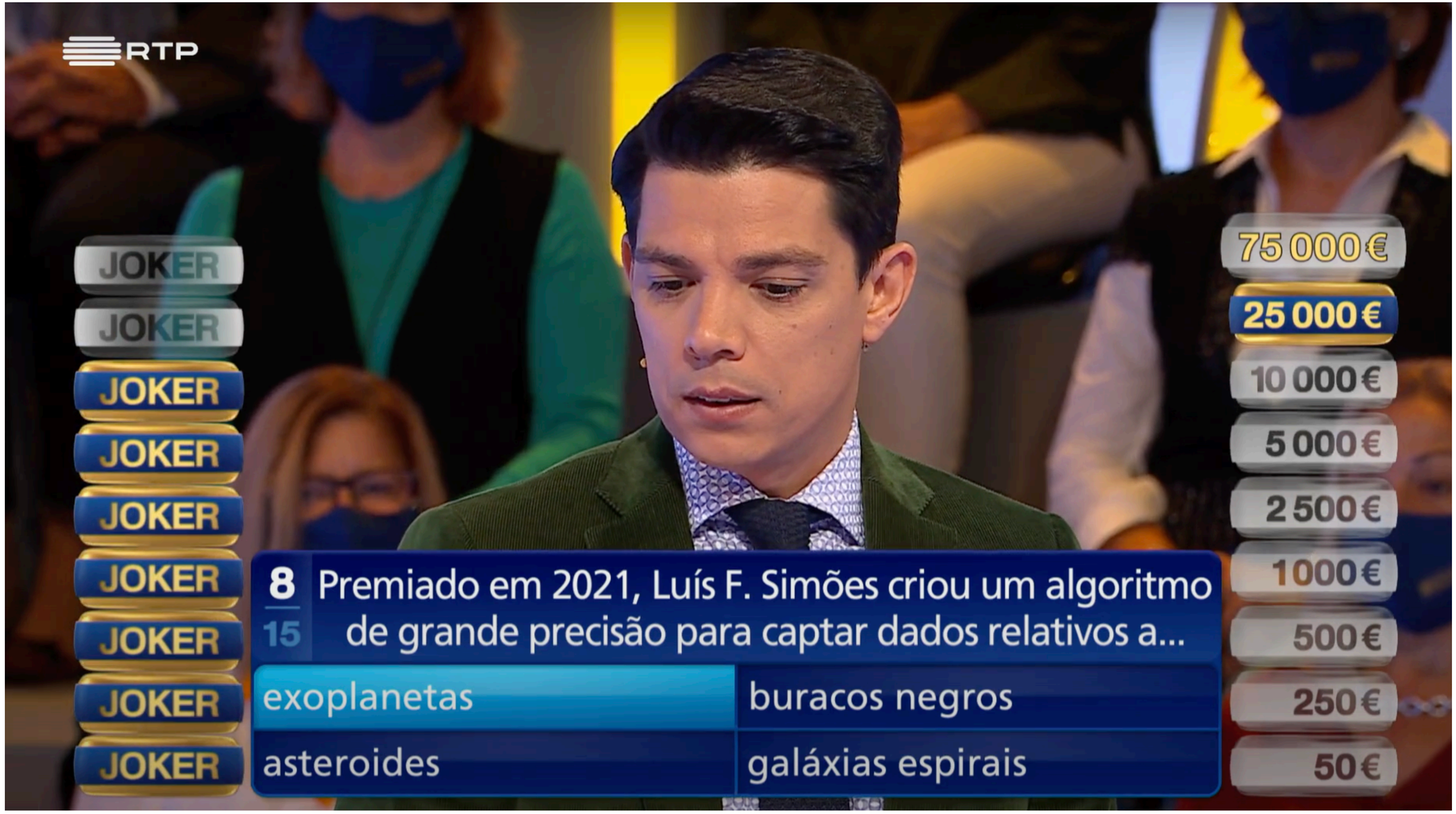


NeurIPS 2022
New Orleans



Ariel MLDC

Accidentally entered pop culture in Portugal…



Ariel MLDC

Can we model exoplanet atmospheres with AI?

- What if we can train an AI to quickly and reliably classify and measure planet atmospheres?
- One AI that works for varying instruments, noise levels and systematics?

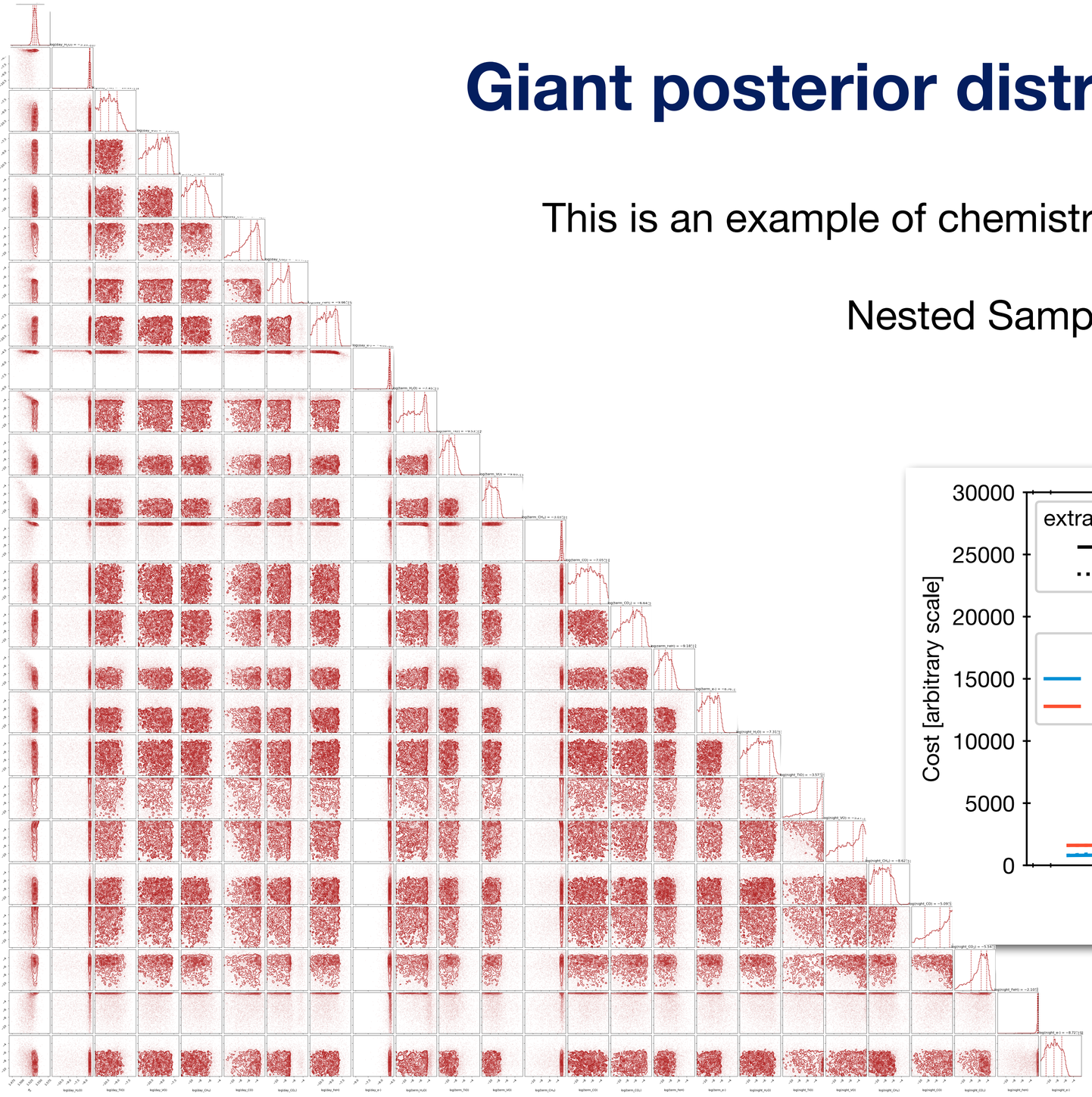
Yip et al. 2022

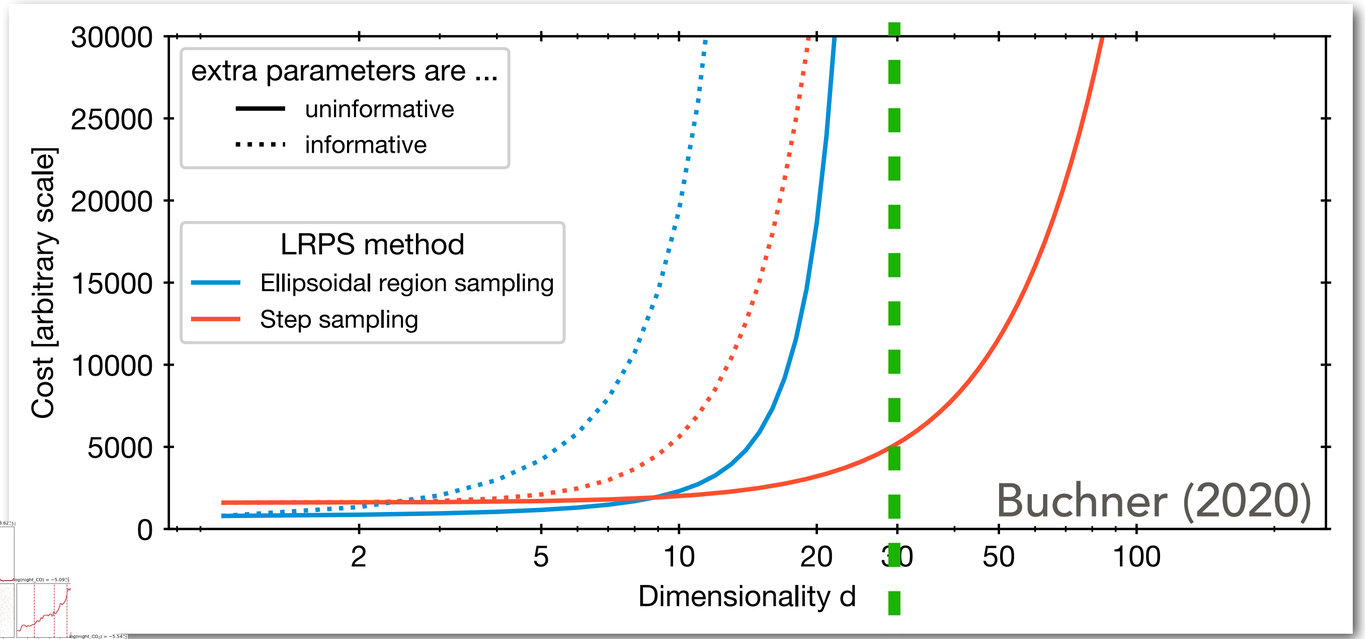
Giant posterior distributions are here to stay

This is an example of chemistry parameters only for phase curve retrievals

Nested Sampling efficiency maxes out at ~30 parameters

The real parameter space is ~2x as big





We are here

Current ML retrieval progress in the field

- Radom Forrests (e.g. Marquez-Neila et al. 2018)
- Feed forward models (e.g. Waldmann 2016)
- Convolutional models (Ardevol Martinez et al. 2022)
- Generative models (Zingales & Waldmann 2018)
- Ensemble forward networks (e.g. Cobb et al 2019)
- Variational Inference (Yip et al. 2022
- Surrogate models (e.g. Himes et al. 2021)
- Transverse Vector Decomposition (Matchev et al. 2022)

Why a challenge on retrievals?

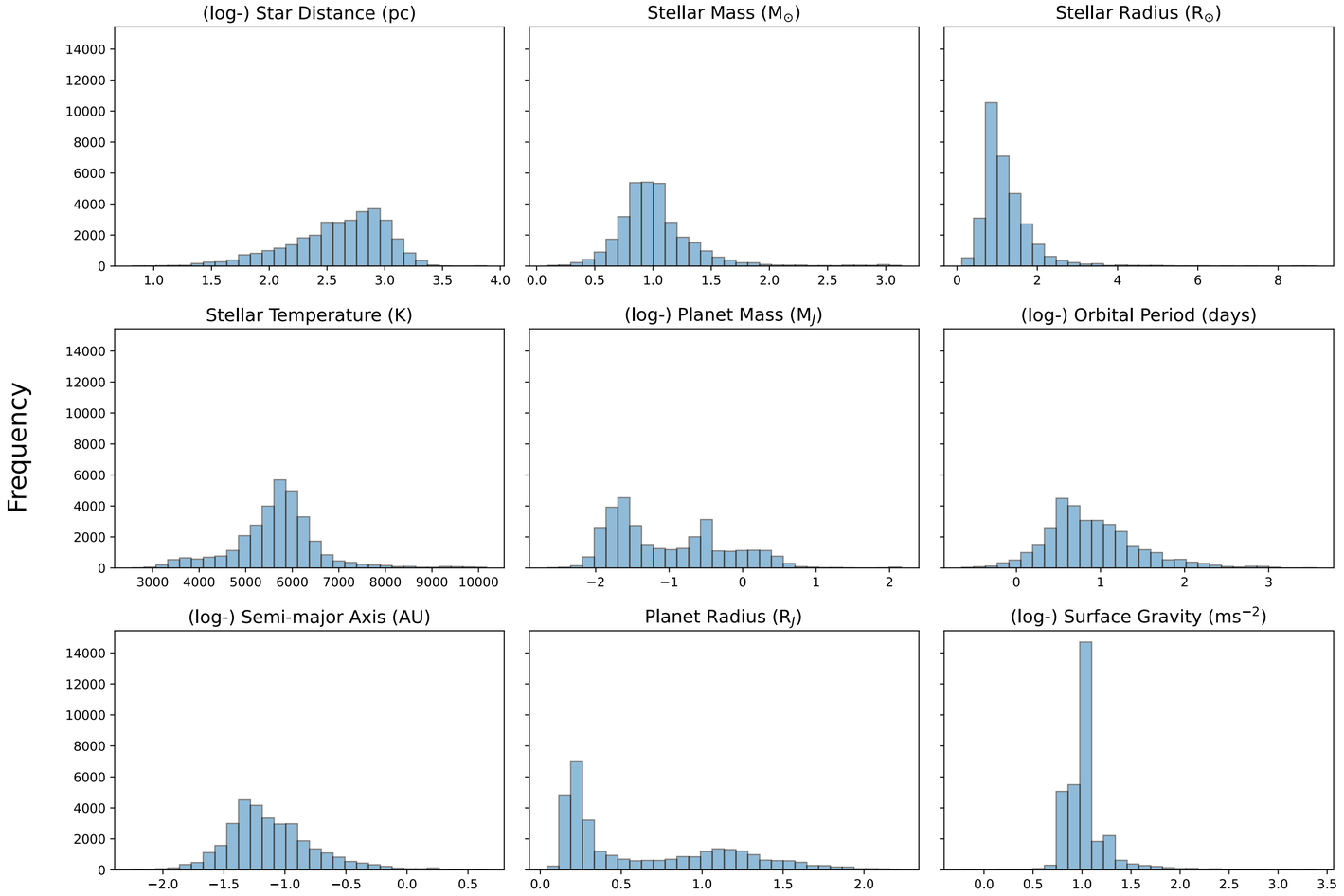
Generating approximate posterior distributions using machine learning over a wide parameter range is **one of the hardest tasks in contemporary ML**

Understanding decision making and biases is fundamentally important to an adoption of these methods in science

Last year's NeurIPS Bayesian inference challenge featured 2D parameter spaces -> we're asking for 6D

Putting retrievals to the ML community

Largest simulation set to date:
100,000 forward models - 26,000 retrievals



SciML AI Benchmark Dataset and continued development platform

- The dataset will be made into a standard AI benchmarking set
- We hope to maintain this data set / platform as development platform and for future data workshops

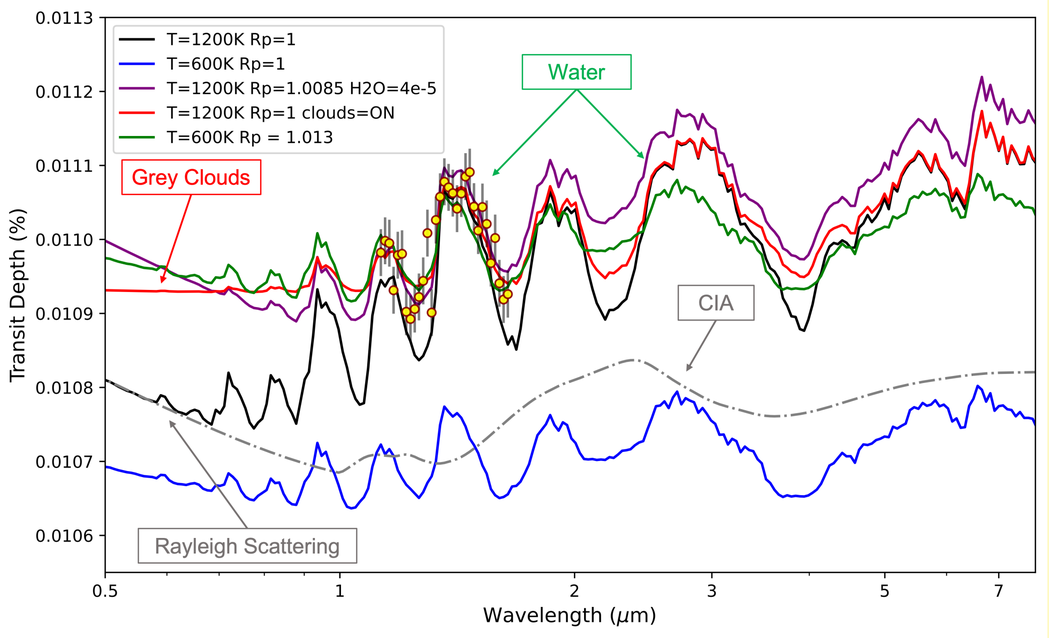
DiRAC

SciML

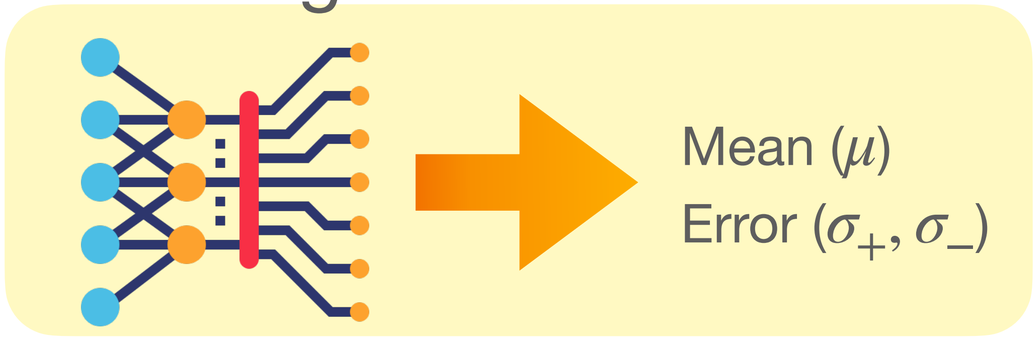
Open Source Software for Scientific Machine Learning

Science & Technology Facilities Council
Rutherford Appleton Laboratory

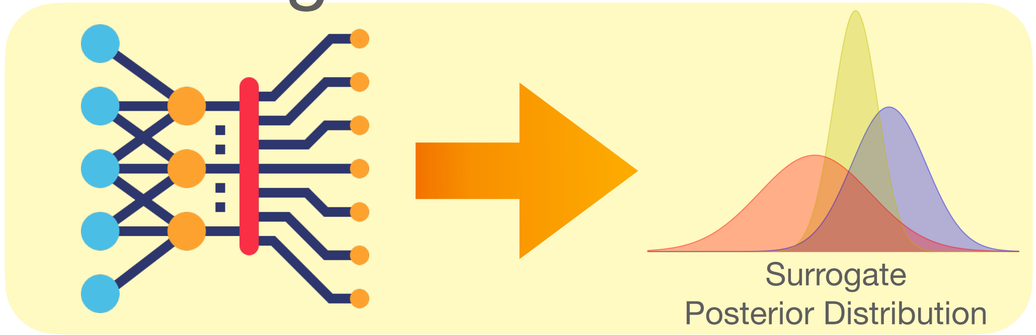
Ariel Competition Tracks



Light Track

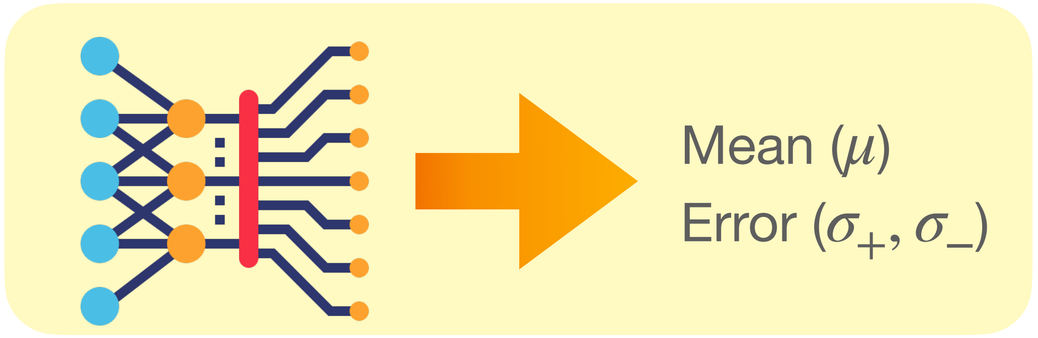


Regular Track



Ariel Competition Tracks

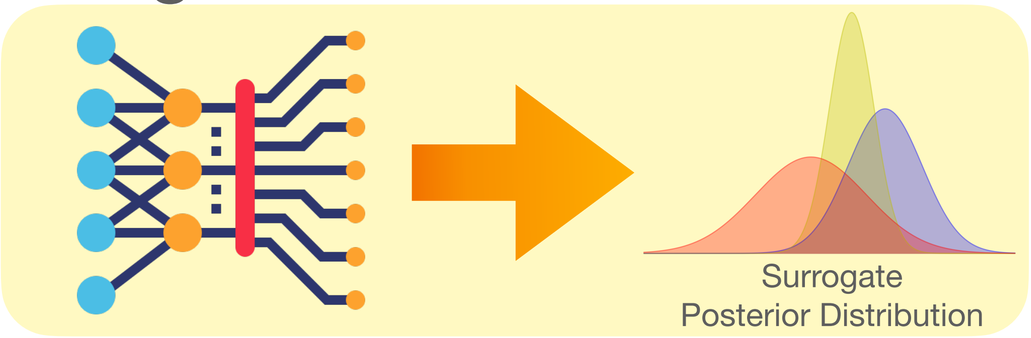
Light Track



- Provide summary statistics
- 500 planets for leader board
- 3000 planets for final solution

- 1st Prize: $1000
- 2nd Prize: $500

Regular Track



- Provide full posteriors
- 200 planets for leader board
- 3000 planets for final solution

- 1st Prize: $2000
- 2nd Prize: $500

Participating and Sponsoring partners

DiRAC

esa

SPACEFLUX

INAF ISTITUTO NAZIONALE DI ASTROFISICA

cea

UK SPACE AGENCY

European Research Council
Established by the European Commission

SciML
Open Source Software for Scientific Machine Learning

UCL
CENTRE FOR SPACE EXOPLANET DATA

ARIEL

Science & Technology Facilities Council
Rutherford Appleton Laboratory

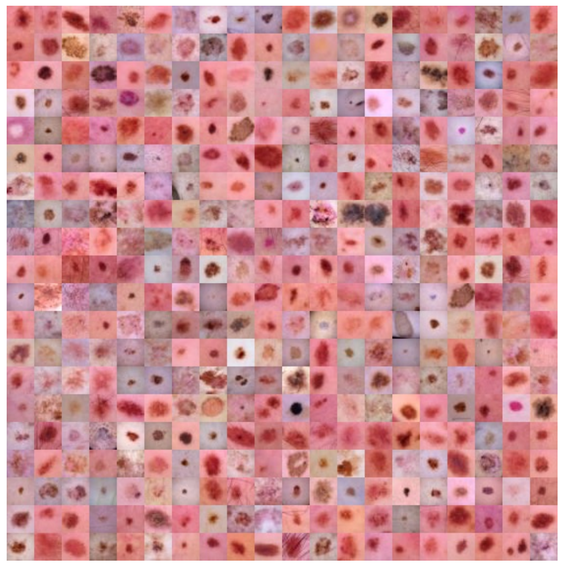
The Alan Turing Institute

FLATIRON INSTITUTE

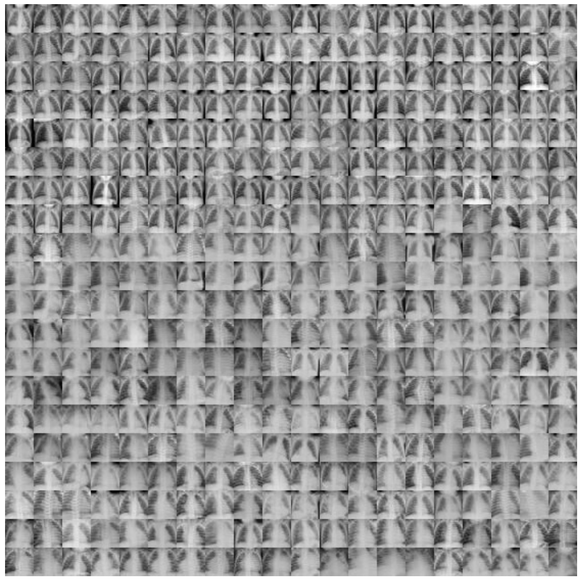
Important beyond planets

- Bayesian Sampling of high parameter spaces is a general problem
- Applicable in all complex inverse processes in physical science
- Big problem in medical physics for example

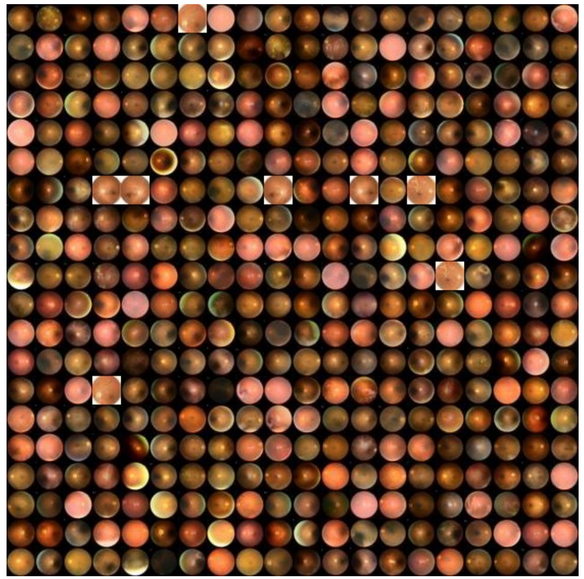
DermaMNIST



PneumoniaMNIST



RetinaMNIST



Yang et al. 2021

Ariel ML Data Challenge

About The Science Leader Board Personal Score Download Information Logout

Start: 30th June
End: 17th October
NeurIPS 2022, 26 November - New Orleans

https://www.ariel-datachallenge.space


NeurIPS 2022
It's all just an excuse to go to New Orleans
Thank you.
HARRAH'S
CANAL
2023
Please Have Exact Fare
Or RTA Pass Ready
Magazine
Decatur
the INSECTARIUM
FLY
HOP