# Ariel Data Challenge 2024: Extracting exoplanetary signals from the Ariel Space Telescope

**Kai Hou Yip**[*] Lorenzo V. Mugnai Andrea Bocchieri Andreas Papageorgiou
Orphée Faucoz Tara Tahseen Virginie Batista Angèle Syty Enzo Pascale
Quentin Changeat Billy Edwards Paul Eccleston Clare Jenner Ryan King
Theresa Lueftinger Nikolaos Nikolaou Pascale Danto Sudeshna Boro Saikia
Luís F. Simões Giovanna Tinetti Ingo P. Waldmann `kai.hou.yip@ucl.ac.uk`

## Abstract

The Ariel Data Challenge 2024 tackles one of astronomy's hardest data analysis problems - extracting faint exoplanetary signals from noisy space telescope observations like the upcoming Ariel Mission. A major obstacle are systematic noise sources, such as "jitter noise" arising from spacecraft vibrations, which corrupts spectroscopic data used to study exoplanet atmospheres. This complex spatio-temporal noise challenges conventional parametric denoising techniques. In this challenge, the jitter time series is simulated based on Ariel's payload design and other noise effects are taken from in-flight data from JWST, in order to provide a realistic representation of the effect.

To recover minute signals from the planet's atmosphere, participants must push boundaries of current approaches to denoise this multimodality data across image, time, and spectral domains. This requires novel solutions for non-Gaussian noise, data drifts, uncertainty quantification, and limited ground truth. Success will directly improve the Ariel pipeline design and enable new frontiers in characterising exoplanet atmospheres - a key science priority in the coming decades for understanding planetary formation, evolution, and habitability.

**Keywords**
Astronomy, Data drift, Time Series Denoising, Exoplanets, Instrument systematics

## 1 Competition description

### 1.1 Background and impact

The discovery of exoplanets - planets orbiting stars other than our Sun - has transformed our cosmic perspective, challenging conventional notions about Earth's uniqueness and the potential for life elsewhere. As of today, we are aware of over 5,600 exoplanets. Detecting these worlds is the initial step; we must also comprehend and characterise their nature by studying their atmospheres. With James Webb Space Telescope (JWST [1]) and the forthcoming ESA Ariel mission [2], we are entering an era of high-quality, high-precision data on an unprecedented number of exoplanets. In 2029, Ariel will conduct the first comprehensive study of 1000 extrasolar planets in our galactic neighbourhood.

We are the first generation that can realistically make progress towards answering the age-old questions of life's provenance in our universe. Understanding the formation and evolution of extrasolar planets, as well as the conditions for their habitability, has recently been highlighted as one of the key themes of scientific development and investment by the US National Academies' Decadal Survey [3] and ESA's Voyage2050 report [4] for the 2020's, 2030's and beyond.

However, observing these atmospheres is one of the hardest data-analysis problems in contemporary astronomy. When an exoplanet transits its host star in our line of sight, a tiny fraction of starlight

[*]kai.hou.yip@ucl.ac.uk

Preprint. Under review.

(20 - 100 photons per million) passes through the planet's atmospheric annulus and interacts with its chemistry, clouds, and winds. These faint signals typically range from 50ppm (for Super-Earth like planets) to 200ppm (for Jupiter like planets) in magnitude and are regularly corrupted by the noise of the instrument. A major component of this noise is due to the inevitable vibration of the spacecraft in space, known as 'jitter noise'. This noise arises from the difficulties of maintaining precise pointing in low-gravity environments, as the spacecraft relies on spinning momentum wheels for stability. Akin to taking long-exposure images with a shaky hand, this noise poses a far greater challenge than the motion blur encountered in commercial photography applications. The photometric variation ($\sim$200 ppm) caused by jitter noise alone is comparable to the variation exhibited by the planetary signal we aim to detect, undermining signals from small planets like Earths and super-Earths. Coupled with other sources of correlated and uncorrelated noises, it is proving difficult for parametric model-based pipelines to achieve the strict technical requirement of the Ariel Payload design to maintain flux stability at 20-100 ppm over 1 hour.

ML approaches have been studied in the literature to denoise and extract signals directly from observation data [5, 6, 7, 8, 9]. However, most of these have so far been worked on light curve data to avoid the problem of high dimensionality in the raw spectro-photometric data. This neglects both the spatial and spectral information about the noise characteristics. By failing to leverage the full richness of the data, crucial details about the jitter noise are lost, rendering the denoising efforts so far inadequate to bring the noise down to the specified requirements for Ariel.

The jitter noise presents an intriguing challenge for the wider AI/ML communities interested in high-fidelity image denoising, image stabilisation, or motion blur in videos. This noise is sub-pixel in amplitude but exhibits correlations in both spatial and time domains over long time-series of images, spanning 3 to 10 hours with exposures on the order of seconds. It is characterised by non-white power spectral densities across a wide frequency range, see Figure 1 for an illustration of the jitter noise.

The NeurIPS community is well-acquainted with denoising problems in computer vision. However, this particular denoising challenge presents several complex issues that push the boundaries of current state-of-the-art methods. These challenges include:

- Detecting signals in an ultra low signal-to-noise environment.
- Detrending both spatial and time-dependent non-Gaussian noise (jitter) in the presence of substantial Gaussian noise.
- Handling shifts in data distribution (i.e. data and concept drifts) caused by unforeseen variations in instrument characteristics and variations in signal characteristics of the star and exoplanet.
- Uncertainty quantification in the presence of data drifts and non-Gaussian noise.
- Limited access to or a general absence of ground truth data.

The problem presents a major obstacle in conducting high precision observations for many legacy, current and upcoming space telescopes (such as Ariel). This challenge extends beyond astronomy and is pervasive in various satellite applications requiring high-precision remote sensing data for monitoring crucial environmental parameters and climate changes. In particular, high-resolution optical Earth observations taken from low to high-altitude commercial orbits have highly similar noise characteristics and problems. Results from this challenge will directly feed into the post processing of Ariel data and will find wide application to a large number of current and future NASA/ESA space-missions, satellites, as well as balloon-borne and ground-based experiments.

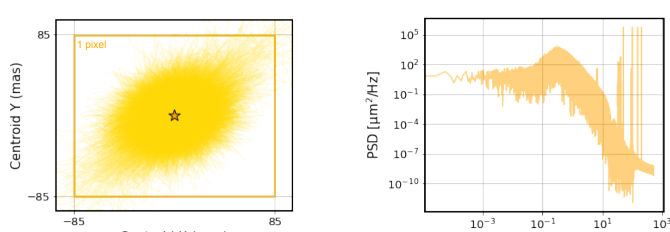


Figure 1: Left: illustration of the jitter effect on the focal plane. Jitter noise acts on the pointing (star) by displacing the star both in the x-axis and y-axis. Amplitudes of this shift are sub-pixel. Right: Power spectral density (PSD) of the jitter noise over time. Jitter noise has a complex frequency structure over a large frequency range.

### 1.2 Previous Experience, Success and Future Outlook

Our team has ample experience in running data challenges and our platform, data generation and validation pipelines have been tried and tested. We have run successful data challenges from 2019, 2021 - 2023 as part of ECML-PKDD and NeurIPS (2022) with between 200 - 300 participating teams ranging from academia and industry in over 51 countries. Each Ariel Data Challenge is designed to address a different aspect of the complex data analysis and modelling of the Ariel Space Mission. By setting self-contained challenges for the AI/ML and Astronomy communities, we invite out-of-the-box thinking on difficult, yet unsolved problems. This has already resulted in several new approaches to data analysis and modelling [9, 10, 11] that were included into the design of this upcoming space mission. A testament to the success of this approach is the backing of the European Space Agency (ESA), the Centre National D'Etudes Spatiales (CNES) and the UK Space Agency (UKSA). We detail some of the key results of last-year's challenge below:

- The press release of the 2023 edition is reported in 32 different articles, spanning 15 different countries, and into 12 different languages
- Three leading academic institutes used our data challenge as Masters thesis projects.
- The European Planetary Science Congress (the largest planetary science conference worldwide) now issues the Ariel Data Challenge Award to foster interdisciplinary collaborations between the Planetary Sciences and the AI/ML community.

As is now customary, winners of this challenge will be invited to give talks at the Ariel Consortium Conferences hosted at the European Space Agency (ESA) and partnering institutions. We will also issue a joint ESA, CNES and UKSA press release promoting the winners.

### 1.3 Novelty

Our Ariel Data Challenge (ADC) 2024 is an entirely new challenge built from the ground up. It is the first challenge of its kind aiming to address noise detrending problems on image time sequences for space-based telescopes. There have been several challenges such as Pose Estimation 2019, 2021 Challenges, spotGEO challenge that are using noisy data. However, all of them are using data either from ground based telescopes or earth bound satellites, while Ariel will be situated at the L2 point, far away from the Earth. There are also major differences in the type of data (time series of spectro-spatial data VS image) and the prediction task is different (regression VS detection).

As far as we know, there is no existing public dataset where a mission approved simulator is used to generate the data nor are we aware of any data challenges addressing this problem. In 2019 and 2021 we ran two ADC challenges on detrending spectro-photometric time series data at ECML-PKDD. We note that the challenge here is substantially different and significantly more realistic (and harder) due to the following aspects:

1. Previous challenges lacked realistic noise models derived from an actual flying telescope. For the first time, genuine detector and noise profiles from an operational mission, such as JWST at the L2 point —where Ariel will be positioned post launch— are employed in simulations to ensure realistic outcomes.
2. Impacts from pointing jitter has never been accounted for in previous challenges. This dataset will be using realistic jitter noise time series provided by Airbus and based on a complete physical representation of the satellite (Ariel) payload.
3. Previous challenges focused on the impact from the host star (e.g. star spots), while this challenge focuses on the impact from the full realistic noise profile, and even goes beyond this attempting to identify models robust to changes in instrument behaviour.

### 1.4 Data

We utilised `ExoSim 2` [12], an end-to-end time-domain simulator for astronomical telescopes developed by Ariel scientists. We adapted `ExoSim 2` with Ariel's payload configuration to generate the datasets for this challenge proposal. As a first-of-its-kind dataset dedicated to machine learning challenges and benchmarks, we will release it as a public dataset. Here we outline the data generation steps:

**Step 1: Prepare the Inputs:** `ExoSim 2` requires the following inputs for a single simulation: properties of the planetary system, atmospheric properties (high resolution theoretical spectrum) and the jitter time series. For the planetary systems, we have selected several representative systems from

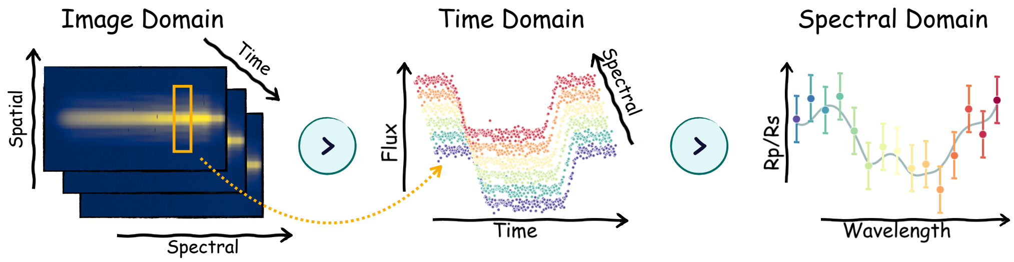

Figure 2: Schematic of the data provided. Left: Time consecutive stack of detector images containing the spectrally dispersed light. Middle: Time-series representation of a pixel/spectral-bin (orange rectangle) in the detector image stack. The dip (so-called 'lightcurve') is the exoplanet obscuring some of the stellar light. By measuring the exact depth of this lightcurve, we can identify the amount of light absorbed by the exoplanet atmosphere as a function of wavelength. Right: Atmospheric spectrum of exoplanet. Each measurement is a lightcurve depth extracted from the time series measurement. These spectra encode the chemical compositions, temperatures and cloud coverages of exoplanet atmospheres.

the Ariel Mission Candidate Sample [13], a carefully curated list of planetary candidates that align with Ariel's observation capabilities. As for the planet's atmosphere, we used a radiative transfer model, `TauREx` 3[14], to simulate a realistic primary atmosphere with trace gases, and included other atmospheric effects. The jitter data is based on simulations performed by the Ariel payload contractor, Airbus; we augmented it by phase-shifting the original jitter time-series to increase its diversity.

**Step 2: Setting up `ExoSim 2`:** We simulated observations using one of Ariel's three photometers (FGS1) and one of its three spectrometer channels (AIRS-CH0) [2]. For this challenge, we considered a detector model with a diverse range of noise sources, quantum effects (such as pixel-to-pixel quantum efficiency variation, readout and dark current), and realistic detector defects (such as non-linearity effects, bad pixels and gain drifts). The detector models (including pixel-to-pixel quantum efficiency variation, readout, dark current, bad pixel and pixel non-linearity) are based on actual in-flight calibration data from JWST. We note that the resulting detector model represents a conservative estimate of the Ariel telescope's capability. For simplicity, we assumed that all observations are conducted with the same integration time and last for 7.5 hours, a typical duration of exoplanet observations using the transit method.

**Step 3: Data Generation and Output:** Based on the above settings and inputs, `ExoSim` 2 generates a total of 250 time sequences of spectro-photometric data, which is $\sim$60 GB in total. This is a sizeable dataset without placing a significant burden on participants' hardware. Each simulation produces a three-dimensional data cube comprising a time series of exposures from Ariel's photometer (FGS-1) and spectrometer (AIRS-CH0), in the following format: *Time dimension* $\times$ *spatial dimension* $\times$ *spatial/spectral dimension*. The number of frames obtainable from a 7.5-hour observation varies between the two instruments. A single FGS-1 observation produces 135000 frames (or time stamps) of 32 $\times$ 32 images, whereas a single AIRS-CH0 observation produces 11250 frames (or time stamps) of 32 $\times$ 356 images. Depending on the chosen approach, participants can build a substantial dataset from these frames (further descriptions in Section 1.5).

**Step 4: Target Generation:** In parallel, the high resolution theoretical spectra from `TauREx` 3, free from any noise effect, are convolved with Ariel's instrument response function to generate uncontaminated instrument level spectra $\mathcal{S}$, i.e. a measurement of changes in the radius ratio between the planet ($R_p$) and the star ($R_s$) squared over wavelength ($\lambda$), i.e. $\mathcal{S} = (x_\lambda)_{\lambda \in \lambda_1 \ldots \lambda_{283}}$, where $x = (\frac{R_p}{R_s})^2$. The output from this process will serve as the ground truth for this challenge and will be kept confidential until the challenge is concluded.

### 1.4.1 Training data

Participants will receive a total of $N_{examples}$ = 150 observations for training. To ensure that the lessons learned from the data challenge apply to real-life scenarios, participants will work with the same amount of information available to an astronomer before the data extraction step; these include:

1. `all_observations.hdf5`: contains all the raw observations obtained via simulated FGS1 ($N_{examples}$ $\times$ 135,000 frames $\times$ 32 $\times$ 32) and AIRS-CH0 observations ($N_{examples}$ $\times$

11,250 frames $\times$ 32 $\times$ 356), as output from `ExoSim` 2. Calibration files for each observation are attached to each example.
2. `all_targets.hdf5`: contains the corresponding pixel level target spectra ($N_{examples}$ $\times$ 283 wavelengths) for each observation in `all_observation.hdf5`.

#### 1.4.2 Test data

The test data aims to mimic the real-world conditions the telescope will encounter in space, unexpected changes in detector performance whilst in-orbit as well as unanticipated atmospheric effects from the exoplanets themselves. The exploratory nature of astronomy implies that scientists cannot predict the type of data they will receive in advance. Consequently, any algorithms used for removing data trends must be resilient to out-of-distribution inputs (i.e. not contained in the training data), while still maintaining high performance on the expected in-distribution data.

To thoroughly test the participants' performance, the test data have different out-of-distribution test cases. The out-of-distribution test cases are created by changing each aspect one at a time, while keeping the rest the same as the training data. To ensure we can extract statistically significant results, each case will have 20 examples, making a total of 100 examples in the test case.

1. Detector specification: The algorithm's performance is evaluated under different detector models (i.e. differences in pixel-to-pixel quantum efficiency variation, dark current, read noise, pixel non-linearity and bad pixels).
2. Planetary system: The algorithm is tested for different planet-star systems. We simulated additional, unseen stars and planetary systems.
3. Atmospheric assumptions: Any algorithm should faithfully extract the spectrum without prior assumptions about the planet's atmosphere. We simulated different atmospheres to test the performance of the test set.

### 1.5 Tasks and application scenarios

The task of this competition is to extract the atmospheric spectra for each observations with an estimate of their level of uncertainty. In order to obtain such a spectrum, we require the participant to detrend a large number of sequential 2D images of the spectral focal plane taken over several hours of observing the exoplanet as it eclipses its host star. Performing this detrending process to extract atmospheric spectra and their associated errorbars from raw observational data is a crucial and common prerequisite step for any modern astronomical instrument before the data can undergo scientific analysis

The jitter noise is correlated both in the image domain (i.e. x-y shifts on the detector image, see Figure 1) as well as time domain of the image stack (i.e. time-correlated noise), coupled with other correlated noises. Here, each pixel (in the image domain of Figure 2) constitutes a time series at a given wavelength of light which encodes the atmospheric information as a time-dependent signal. We ask the participants to consider all available data, i.e. image, time and spectral domains as this follows the real-life data processing steps of exoplanet data reduction. By simulating a realistic data workflow, we not only provide the user with the highest possible amount of flexibility in their analysis approach but also ensure the direct applicability of methods/strategies developed here to real-life astrophysical data processing and the development of the Ariel Space Mission data pipeline.

This is a multimodal supervised learning task. The participants can choose to detrend this jitter noise in either modality (i.e. the image, time or spectral domains) or combinations thereof. Each modality bears different advantages. Here we outline two common training strategies.

1. Approach 1: Train directly on the full 3D data cube and extract the corresponding spectra $\mathcal{S}$. This approach leverages the rich information content but is limited by the availability of training examples (Image –> Spectral Domain in Figure 2).
2. Approach 2: Extract a time series for each wavelength and train a model to infer a single $x_{\lambda=\lambda_n}$ at wavelength $\lambda_n$ (Time –> Spectral Domain in Figure 2). This approach allows one to build a substantial dataset for model training (i.e. 283 wavelengths $\times$150 training observations = 42,450 examples), but it discards crucial information from the spatial and spectral domains by focusing solely on a single wavelength.

The second approach aligns with the conventional data reduction method in astronomy, where image domain data are reduced and analysed individually. However, neither approach is optimal for denoising jitter time series and we anticipate the winning solutions to include information from all three domains.

### 1.6 Metrics

We will evaluate the quality of the participants' submitted predicted spectra, $(\tilde{x}_\lambda)$, and their corresponding uncertainties at different wavelengths, $(\tilde{\sigma}_\lambda)$, $\tilde{\mathcal{S}} = (\tilde{x}_\lambda, \tilde{\sigma}_\lambda)$ against the ground truth pixel level spectrum from the test set $\mathcal{S}_{test} = (x_\lambda)$ (See Step 4 in Section 1.4 for its generation process).

For a given test sample $\mathcal{S}_{test,i}$, we will calculate the Gaussian Log-likelihood (GLL) value for each pair of $x$ in $\mathcal{S}_{test,i}$ and $\tilde{\mathcal{S}}_i$, i.e.

$$GLL\Big(x_\lambda, \tilde{x}_\lambda, \tilde{\sigma}_\lambda\Big) = -\frac{1}{2}\log(2\pi\tilde{\sigma}_\lambda^2) - \left(\frac{\tilde{x}_\lambda - x_\lambda}{2\tilde{\sigma}_\lambda}\right)^2 \tag{1}$$

the GLL values from each point will be summed across all wavelengths and across the entire test set via $\mathcal{L} = \sum_i \sum_\lambda GLL\Big(x_{i,\lambda}, \tilde{x}_{i,\lambda}, \sigma_{i,\lambda}\Big)$. The $\mathcal{L}$ value will be transformed into a score using the following conversion function:

$$score = 10000 \times \frac{\mathcal{L} - \mathcal{L}_{ref}}{\mathcal{L}_{ideal} - \mathcal{L}_{ref}} \tag{2}$$

Here we denote the $\mathcal{L}_{ideal}$ as the ideal case scenario, which is when the submission has $(\tilde{x}_\lambda)$ = $(x_\lambda)$, i.e. predictions from the model are exactly the same as the ground truth values, with $\tilde{\sigma}$ = 1 ppm uncertainty in transit depth space ($x^2$), which is better than Ariel's Stability Requirement. $\mathcal{L}_{ref}$ is the reference case where the prediction of the model is the overall mean of the training set and the uncertainty is the overall standard deviation of the training set. The score will be returned as an integer in the interval $[0, 10000]$, with higher scores corresponding to better performing models, any score that is lower than 0 will be treated as 0.

### 1.7 Baselines, code, and material provided

Based on current literature in the field, we developed a 2D CNN with Monte Carlo Dropout model based on the first approach (outlined in Section 1.5) as a baseline and entry point for the participants. It has achieved a score at 5145 based on our preliminary test set.

A detailed description of the network, pre-processing steps, training procedure, and evaluation metric will be made available to the participants as part of the starter kit released in May on Github during the beta test period. The main tools used in data generation tool, `TauREx3` and `ExoSim 2`, are publicly available on Github[2]. We will also have a limited number of slots of HPC resources sponsored by DiRAC HPC facility (see resource section below).

### 1.8 Website, tutorial and documentation

The competition will be hosted on `https://www.ariel-datachallenge.space/`. The format will be similar to previous challenges (see website). We aim to make the website and its resources easily accessible to the participants. The website will be organised into the following areas: (1) Home (Front Page) (2) Getting Started (3) Leaderboard (4) Tutorials & Documentation (5) Contact Us (6) FAQ. The backend is complete, tried and tested from the 2021 - 2023 challenge and the new frontend design will be finished by mid April.

The data challenge platform will be running on a dedicated webserver (128 cores, 150 Tb space). Our hardware setup allows for real-time calculations of participant performance metrics and leaderboard updates even in the event of large number of simultaneous submissions.

To equip participants with sufficient background knowledge. We will design a jupyter-notebook tutorial to illustrate the data generation process, discuss any assumptions and demonstrate the atmospheric model. We will also provide a blog-style explanation of the problem and physics and a list of more in-depth background literature for participants.

[2] Documentation: `https://taurex3-public.readthedocs.io/en/latest/` and `https://exosim2-public.readthedocs.io/en/latest`

# 2 Organizational aspects

## 2.1 Protocol

In order to enter the competition, participants will need to create an account for themselves/their team. The data is publicly available from the website. To evaluate their results, the participants will only be submitting results (i.e. their models' predictions and error estimates on the test set examples) to our online website. We will automatically evaluate the uploaded results based on our metric and report the submissions' score on a leaderboard maintained on our webpage. We also provide a personal timeline of scores and statistics to each user after upload. The competition will not involve multiple phases and participants are free to make as many submissions as they wish. The only restriction is that consecutive submissions cannot be made within 24 hours of one another. This is a measure to (i) prevent the participants from producing a trivially/randomly different model just to climb a bit higher on the leaderboard and (ii) to limit the extend to which the provided solutions overfit to the test set contained on the server. These are both dangers due to the presence of the leaderboard which is effectively causing some information leakage from the test set. To further mitigate this danger, the leaderboard score will only be calculated on 50% of the available test data (chosen uniformly at random from the full test set), while the final score for each solution (which will determine the final ranking of solutions at the end of the competition) will be calculated on the other 50% of the test set.

To prevent cheating, shortly after the end of the competition, the authors of the ten top-ranked entries will be requested to provide a brief description of their solution (data preprocessing steps, model and training details) to the organisers (1–2 pages) and a copy of their code[3]. Provided they do so and we do not identify any of the following issues:

- The authors had access to ground truth.
- The model relies heavily (as judged by the organisers) on hard-coded elements that are solely deemed to be due to test-set leakage.
- The authors are participating in the competition under multiple aliases.

Then the top three participants will be deemed the winners of the competition and receive their prizes. If any of the above issues are identified, the corresponding entrant will be disqualified and removed from the leaderboard, in which case the same procedure will be repeated on the remaining entries. The online platform will be operational by mid April (see timeline below). From this time until mid May, members of the organising team and selected postgraduate students will beta test the platform.

## 2.2 Rules and Engagement

**Draft of contest rules**

1. Each participant must have a unique alias.
2. No participant can submit more than 1 entry in a 24 hour interval. If they do, their second entry will be automatically ignored.
3. After the end of the competition and before announcing the winners, the authors are required to provide a brief description of their solution to the organisers (1–2 pages) and a copy of their code. Failure to do so in the agreed upon time frame will lead to their disqualification.
4. The organisers will use the provided code, models and solutions only for the purposes of checking whether the contest rules (5) & (6) are not broken. The organisers will not use any result without the authors' explicit permission.
5. Participants must not have access to the ground truth before the competition's closing date. If they do, they will be disqualified.
6. For an entry to count as a winning entry, it must not rely heavily (as judged by the organisers) on hard-coded elements that are solely deemed to be due to test-set leakage.
7. Participants are allowed to form teams. In this case, they must notify the organisers to deactivate their individual aliases and they must create a joint (group) alias. The group alias shall then be treated as a single participant for all purposes, including the restriction on the number of daily submissions and the prize received. All members of a winning team will be listed as winners of the competition and invited for further collaboration (e.g. for publishing the results with the organisers). However, a single prize will be awarded to the entire team. The team can -of course- decide how to receive the prize (e.g. nominate a single member to receive it, or split a monetary equivalent among them).

[3]We will offer participants the option of signing a Non-Disclosure Agreement before submitting their code and algorithmic description

The purpose of Rules (1) & (2) is to prevent the participants from generating multiple trivial random variations of their models which would lead to overfitting to the test set score signal. To enforce Rule (1) we will request each participant to provide their full name, email address and affiliation (if applicable). We plan for Rule (2) to be automatically implemented in our online platform. They also allow for limiting the amount of traffic on the competition's platform.

The aim of Rule (3) is to check whether the contest Rules (5) & (6) are not broken. Rule (4) guarantees that the organisers will not use any result without the authors' permission. If a submitted solution is novel and useful to the field, we plan to collaborate with its author (see [15] as an example). Rules (5) & (6) are measures for preventing cheating and overfitting to the test data. Rule (7) allows for and handles team formation.

To allow for the broadest possible participation, the set of rules is the minimal possible. There is no restriction on the models, algorithms or data pre-processing techniques, neither on the programming languages, environments or tools used for their implementation. The participants are also free to use data augmentation techniques, pre-trained models or any prior domain knowledge not included in the provided dataset. Finally, they are free to choose their own way of splitting the training data between training and validation sets.

### 2.3 Schedule and readiness

| Preparation Timeline | Launch Timeline |
|---|---|
| Early Feb: Data generation is initiated [✓] | mid Jul: Competition launch |
| Late Feb: Beta version dataset ready for inspection [✓] | mid Jul - late July: Application for HPC resources |
| Early Mar: Competition data generation initiated | mid Jul - Mid Oct: Competition promotion campaign |
| Mid Mar: 1st batch of data ready [ ✓] | Late Oct: Competition ends |
| Early Apr: Data generation completed [50% ✓] | Early - mid Nov: Evaluation stage |
| Early Apr - Mid Apr: Documentation on dataset, start kit and metric [50% ✓] | mid Nov: Winners are announced |
| Mid Apr: Website operational [50% ✓] | Late Nov: Winners are invited to present at NeurIPS 2024 |
| Late Apr - Late May: Beta test period | |
| late May - late Jun: Beta launch & buffer period | |

### 2.4 Competition promotion and incentives

Our organising team has a diverse background spanning research institutes across the UK, Europe, North America, and Japan. Our promotion campaign will leverage these global networks through invited talks, social media platforms, international conferences, and outreach at respective institutes to ensure broad reach to our potential audience worldwide. Below is a list of selected channels where we will promote our challenge.

| Astrophysics | Machine Learning |
|---|---|
| European Space Agency (ESA) | DiRAC HPC Facility |
| United Kingdom Space Agency (UKSA) | The Alan Turing Institute |
| Centre National D'Etudes Spatiales (CNES) | SciML Research Group, RAL Space |
| The Ariel Space Mission Consortium | ESA Datalabs |
| Instituto Nazionale di Astrofisica | UCL Centre for Data Intensive Science and Industry |
| Centre à l'Energie Atomique (CEA) | UCL Computer Science |
| UCL Physics and Astronomy | Data Science network @ UniVie |
| UK Exoplanet mailing list | ALLSTATS mailing list |
| NASA Exoplanet mailing list | Machine learning for Planetary Science (ML4PS) |
| JWST exoplanet mailing list | Royal Statistical Society |
| European Planetary Science Congress mailing list | Machine Learning News group |
| Europlanet Society magazine and mailing list | |

Incentives: Monetary awards and invitation to NeurIPS (subject to availiability) for top 3 teams, private tour to ESA facilities (subject to tour availability) and invited talks to an Ariel Consortium meeting. The top 3 teams will be invited to join our result paper, they will be listed after the main organising team. UK STFC DiRAC is sponsoring $\sim$100 GPU hours on NVIDIA A100 for 20 participants from underprivileged backgrounds without sufficient HPC access, offered on a first-come, first-served basis.

### 2.5 Resources provided by organizers

The organisers received support from UKSA, University College London (UCL) and STFC DiRAC to handle day-to-day operations in organising the data challenge, generating the dataset and evaluating solutions. Our sponsors are providing the following resources: Monetary awards (UCL and CNES) and private tours (ESA). Awards will be provided by our sponsors this year. DiRAC is providing access to HPC facilities.

## 3 Resources

### 3.1 Organizing team

Dr. Kai Hou (Gordon) Yip, *University College London* (coordinator, evaluator, beta-tester), is a Research Fellow at UCL's ExoAI group focusing on designing ML & Explainable AI solutions for the modelling of exoplanet atmospheres. He is the lead organiser for both Ariel Data Challenge 2022 and 2023, and is responsible for coordinating Machine Learning efforts within the ESA Ariel Space Mission. During his PhD he has employed statistical methods to characterise planetary atmospheres. He also works on generative adversarial models for detecting and simulating ultra-low signal-to-noise observations of extrasolar planets.

Dr. Lorenzo V. Mugnai, *Cardiff University* (data provider), serves as a Postdoctoral Research Associate specialising in the intersection of astronomical instrument development and the scientific exploitation of data. His work focuses on developing and improving data analysis tools for precise data interpretation, which is crucial for conducting reliable comparative studies within the field of Exoplanetary Science. He is the creator of advanced simulation tools such as ExoRad and ExoSim 2. These tools are instrumental in optimising the design and data reduction strategies for the scientific exploitation of astronomical instruments. His contributions significantly benefit missions like the ESA's Ariel Space mission, NASA's balloon-borne EXCITE mission, and private space endeavours such as Twinkle and Mauve by BSSL.

Dr. Andrea Bocchieri, *Sapienza Università di Roma* (data provider), is a postdoctoral researcher in exoplanet science at Sapienza University of Rome, where he works with Prof. Enzo Pascale, the Ariel mission scientist. His main research topic is the atmospheric characterisation of exoplanets with Ariel. He has expertise in atmospheric retrieval methods, instrument simulation pipelines, calibration procedures, detrending techniques, and Ariel mission performance analysis. His work bridges the gap between instrumentation, observation, and interpretation of astronomical observations.

Dr. Andreas Papageorgiou, *Cardiff University* (data provider), has over 15 years of extensive experience in all phases of space telescope missions since 2008. Throughout his career, he has played a key role in the definition, implementation, and operation of science ground segments for several high-profile projects. Currently involved with the ESA Ariel mission planned for launch in 2029, his expertise spans managing ground systems for major observatories like Herschel, as well as proposed missions like EChO and BETTII.

Mr. Orphée Faucoz, *Centre National d'Etudes Spatiales* (baseline method provider, evaluator, beta tester), is a Data Scientist within the CNES AI Team working in different subjects related to the space field using Machine Learning, notably on autonomous ultrasonography, anomaly detection and quantum machine learning. He is involved in the ARIEL ML consortium whose goal is to use machine learning to better understand exoplanets.

Ms Tara Tahseen, *University College London* (baseline method providers, beta testers), a PhD student at UCL's ExoAI group, she works on applying machine learning to enhance atmospheric modelling of exoplanets. Formerly a Data Scientist at a specialist insurance provider, she has collaborated on machine learning projects ranging from an NLP project on coreference resolution at The Guardian, to characterising information flow through the brain at the Consciousness and Cognition Lab, Department of Psychology, University of Cambridge.

Prof. Enzo Pascale, *Sapienza Università di Roma* (data provider), is an associate professor at Sapienza University of Rome. He has accrued extensive experience in designing, implementing, operating, and exploiting astronomical instrumentation across the electromagnetic spectrum from microwaves to visible light. Since 2010, his research has centred on developing instruments for spectroscopic characterisation of exoplanetary atmospheres using transit spectroscopy, spanning the visible to mid-infrared range. Notably, he serves as the instrument scientist for NASA's Excite suborbital mission and the mission scientist for ESA's Ariel Space Mission. Under his leadership, his group has successfully developed cutting-edge radiometric and time domain instrument simulators (ExoSim) that optimise instrument designs, refine data reduction techniques, enhance instrument performance, and improve data quality.

Dr. Quentin Changeat, *European Space Agency* (atmospheric model provider), is a ESA research fellow at STScI and an expert in modelling and data interpretation of exoplanet atmospheres. He is a leader in the analysis of data from the Hubble Space Telescope and has exploited this instrument to understand the chemistry and thermal structure of many exoplanets. He is also the coordinator of the Spectral Retrieval working group for the future ESA telescope Ariel.

Dr. Billy Edwards, *SRON, Netherlands Institute for Space Research* (data provider), is staff scientist at SRON. His work has a general focus on exoplanets, but with particular emphasis on spacecraft capability and operations studies. He works on modelling the performance of the JWST and Ariel space missions, exploring their expected capabilities and optimising target selection. He is providing Ariel's Mission Candidate Samples.

Dr. Paul Eccleston, *STFC RAL Space* (consultant), is the Chief Engineer for RAL Space since 2016 and also manages the Ariel Mission Consortium, a collaboration of over 60 institutes across 19 countries in Europe and North America. He has played leading roles in several major science instruments, including serving as the engineering manager for the Spectral Investigation of the Coronal Environment on Solar Orbiter from 2012-2015. Eccleston worked in various capacities on the Mid-Infrared Instrument for the James Webb Space Telescope, coordinating the consortium of 40+ institutes to build, test, and calibrate the instrument. He is the organising team's liaison with UK STFC RAL Space.

Dr. Clare Jenner, *Distributed Research utilising Advanced Computing (DiRAC)* (consultant), is Deputy Director of the STFC DiRAC High Performance Computing Facility. She directs the facility's training programmes supporting the prestigious DiRAC user-community's educational, research and innovation activities. She will be providing advice on HPC resource management. She is the organising team's liaison with STFC DiRAC.

Mr. Ryan King, *UK Space Agency* (consultant), is a Lead Science Programme Manager at UKSA. With over a decade of experience at the UK Space Agency, he has held pivotal leadership roles in managing the agency's major space science programs and missions, including SMILE mission and Ariel mission. He is the organising team's liaison with UKSA.

Dr. Theresa Lueftinger, *European Space Agency* (consultant), is a Project Scientist for the ESA Ariel space mission. Prior her appointment to ESA, she is a Staff Astronomer at the University of Vienna. Her research focuses on how stellar activity, triggered by the magnetic field, shapes exoplanets and their atmospheres. She is the organising team's liaison with ESA.

Dr. Nikolaos Nikolaou, *University College London* (consultant, evaluator, beta-tester), is an Assistant Professor in Data Intensive Science. His research interests focus on machine learning –in particular uncertainty quantification, interpretability, causal inference & resource-efficiency– with applications in astrophysics, clean energy and pharmaceutics & medicine. He was the main organiser of our 2019 ECML-PKDD Data Challenge and a co-organiser of our 2021– 2023 ECML-PKDD Data Challenges.

Ms Pascale Danto, *Centre National d'Etudes Spatiales* (consultant), she has over a decade experience working at Centre National d'Etudes Spatiales (CNES) in various project management and technical leadership roles for major space missions and instrumentation projects. Currently, she is working as the Project Manager for the French contributions to the ESA ARIEL exoplanet mission and the NASA Roman Space Telescope by NASA. She is the organising team's liaison with CNES.

Dr. Sudeshna Boro Saikia, *University of Vienna* (consultant, evaluator), is an Assistant Professor at the University of Vienna. Her primary research interest is to understand the factors influencing planetary habitability. Specifically, she focuses on the influence of host stars on planetary atmospheric photochemistry, evolution and loss, as well as the characterisation of atmospheres surrounding terrestrial exoplanets. She serves as a science team coordinator for Ariel mission consortium's star planet interaction subWG and a science team member for PLATO mission.

Dr. Luís F. Simões, *ML Analytics* (consultant, evaluator), Lead Data Scientist and co-founder at ML Analytics, where he leads the development and operations of Machine Learning models for various industries. Former member of the European Space Agency's Advanced Concepts Team and Frontier Development Lab. Researcher on applications of AI to various space science problems, including spacecraft landing, trajectory optimisation, distributed control and process automation. Winner of the 2021 Ariel Machine Learning Data Challenge, having then joined the consortium of ESA's Ariel mission. Collaborating with the mission on matters of model verification & validation and MLOps.

Prof. Giovanna Tinetti, *University College London* (consultant), is a Professor of Astrophysics at UCL and the Principal Investigator (PI) and science lead of the Ariel space mission and scientific lead of the Twinkle Space Mission. Additionally, she is the PI of the ExoLights group at UCL working on

Bayesian inverse retrievals of exoplanet atmospheric data and atmospheric modelling. Prof. Tinetti is the director of UCL's Centre for Space Exoplanet Data (UCL-CSED).

Prof. Ingo P. Waldmann, *University College London / The Alan Turing Institute* (coordinator, evaluator, platform administrators), is an Professor of Astrophysics at UCL, deputy director of the Centre for Space Exochemistry Data (CSED), a Turing Fellow at the Alan Turing Institute. The research group focuses on designing and implementing machine learning approaches in the data analysis and modelling of exoplanet and solar-system data. He is leading the Machine Learning work of the European Space Agency Ariel mission consortium. He is furthermore the UCL representaive on AI to the European Space Agency (ESA-labs) and UCL AI steering board member. He is a director and co-founder of Spaceflux Ltd focusing on AI applications in global robotic telescope space situational surveillance.

### 3.2 Support requested

We do not require any support from the conference.

## References


[1] Jonathan P. Gardner et al. "The James Webb Space Telescope". In: 123.4 (Apr. 2006), pp. 485–606. DOI: 10.1007/s11214-006-8315-7. arXiv: astro-ph/0606175 [astro-ph].

[2] Giovanna Tinetti et al. "Ariel: Enabling planetary science across light-years". In: arXiv e-prints, arXiv:2104.04824 (Apr. 2021), arXiv:2104.04824. arXiv: 2104.04824 [astro-ph.IM].

[3] Engineering National Academies of Sciences and Medicine. Origins, Worlds, and Life: A Decadal Strategy for Planetary Science and Astrobiology 2023-2032. Washington, DC: The National Academies Press, 2023. ISBN: 978-0-309-47578-5. DOI: 10.17226/26522. URL: https://nap.nationalacademies.org/catalog/26522/origins-worlds-and-life-a-decadal-strategy-for-planetary-science.

[4] Pedro Russo et al. "ESA's Voyage 2050 Long-term Plan for Education and Public Engagement: White Paper". In: arXiv e-prints, arXiv:1908.01546 (Aug. 2019), arXiv:1908.01546. DOI: 10.48550/arXiv.1908.01546. arXiv: 1908.01546 [astro-ph.IM].

[5] James G. Ingalls et al. "Repeatability and Accuracy of Exoplanet Eclipse Depths Measured with Post-cryogenic Spitzer". In: The Astronomical Journal 152.2, 44 (Aug. 2016), p. 44. DOI: 10.3847/0004-6256/152/2/44. arXiv: 1601.05101 [astro-ph.IM].

[6] Jessica E. Krick et al. "Random Forests Applied to High-precision Photometry Analysis with Spitzer IRAC". In: The Astronomical Journal 160.3, 99 (Sept. 2020), p. 99. DOI: 10.3847/1538-3881/aba11f. arXiv: 2006.14095 [astro-ph.IM].

[7] Mario Morvan et al. "Detrending Exoplanetary Transit Light Curves with Long Short-term Memory Networks". In: The Astronomical Journal 159.3, 109 (Mar. 2020), p. 109. DOI: 10.3847/1538-3881/ab6aa7. arXiv: 2001.03370 [astro-ph.EP].

[8] Mario Morvan et al. "Don't Pay Attention to the Noise: Learning Self-supervised Representations of Light Curves with a Denoising Time Series Transformer". In: Machine Learning for Astrophysics. July 2022, 11, p. 11. DOI: 10.48550/arXiv.2207.02777. arXiv: 2207.02777 [astro-ph.IM].

[9] Nikolaos Nikolaou et al. "Lessons learned from the 1st Ariel Machine Learning Challenge: Correcting transiting exoplanet light curves for stellar spots". In: RAS Techniques and Instruments 2.1 (2023), pp. 695–709.

[10] Kai Hou Yip et al. "Lessons Learned from Ariel Data Challenge 2022 Inferring Physical Properties of Exoplanets From Next-Generation Telescopes". In: 2023.

[11] Mayeul Aubin et al. "Simulation-based Inference for Exoplanet Atmospheric Retrieval: Insights from winning the Ariel Data Challenge 2023 using Normalizing Flows". In: arXiv e-prints, arXiv:2309.09337 (Sept. 2023), arXiv:2309.09337. DOI: 10.48550/arXiv.2309.09337. arXiv: 2309.09337 [astro-ph.EP].

[12] Lorenzo V. Mugnai et al. "ExoSim 2. The new time-domain simulator applied to the Ariel space mission". In: European Planetary Science Congress. Sept. 2022, EPSC2022-370, EPSC2022–370. DOI: 10.5194/epsc2022-370.

[13] Billy Edwards and Giovanna Tinetti. "The Ariel Target List: The Impact of TESS and the Potential for Characterizing Multiple Planets within a System". In: 164.1, 15 (July 2022), p. 15. DOI: 10.3847/1538-3881/ac6bf9. arXiv: 2205.05073 [astro-ph.EP].

[14] A. F. Al-Refaie et al. "TauREx 3: A Fast, Dynamic, and Extendable Framework for Retrievals". In: 917.1, 37 (Aug. 2021), p. 37. DOI: 10.3847/1538-4357/ac0252. arXiv: 1912.07759 [astro-ph.IM].

[15] Nikolaos Nikolaou et al. "Lessons Learned from the 1st ARIEL Machine Learning Challenge: Correcting Transiting Exoplanet Light Curves for Stellar Spots". In: arXiv e-prints, arXiv:2010.15996 (Oct. 2020), arXiv:2010.15996. arXiv: 2010.15996 [astro-ph.IM].