Nikita Rawat 21156058

COMP717: Artificial Intelligence

Assignment 2: AI Exploration and Reflection

Part 1: The Future of AI, Me, and Astronomy and Space Science

1. Motivation

Astronomy and space science have long fascinated humanity, offering insights into the universe's beginnings and mechanisms. One of the most dynamic and rapidly evolving areas within this domain is the detection and classification of celestial bodies, including exoplanets and galaxies. The labour-intensive and time-consuming nature of typical astronomical data analysis techniques limits the rate of discovery (Wright et al., 2015). Astronomers are faced with the task of analysing enormous volumes of data in order to find significant patterns, especially with the introduction of large public datasets from space telescopes like Kepler, TESS, and the Sloan Digital Sky Survey (SDSS). Promising approaches to automate and speed up these processes are offered by AI and machine learning, which could revolutionise how we explore the universe (Shallue & Vanderburg, 2018).

The night sky has always held a special place in my heart. As a child growing up in a small village in the Northern Hemisphere, far from the pollution and noise of cities, I was fortunate to witness a crystal-clear view of the stars. My extended family and I used to spend warm summer evenings sleeping outside on our two-story home's rooftop. Our elders have woven science and mythology into the fabric of the cosmos as we lie beneath the wide expanse of stars, telling us stories about the constellations. I vividly recall learning about constellations like the Three Sisters, the Pole Star (Dhruv Tara), and the Big Dipper (Sapt Rishi Tara Mandal). My lifelong obsession with space began in those peaceful times, not as an abstract scientific concept but rather as something solid, cultural, and intensely personal.

That fascination deepened as I grew older. Through images released by NASA and visual documentaries on the solar system and galaxies, I discovered new ways to explore the universe. These visualisations made distant planets, nebulae, and black holes feel more real and reachable. I sought every opportunity to learn about space science—an interest that eventually led me to study at AUT, where I chose **Astronomy and Space Science** as my minor. Simultaneously, I pursued **Data Science** as my major, recognising its transformative potential across many fields, especially astronomy, where traditional methods struggle with the sheer scale of data produced today (Borucki et al., 2010; Abazajian et al., 2009).

We now live in an era of *big space data*. Projects like **NASA's Kepler and TESS missions** and massive astronomical surveys such as the **Sloan Digital Sky Survey (SDSS)** have produced petabytes of observational data—light curves, time-series spectra, and terabytes of high-resolution galaxy images (Borucki et al., 2010; Abazajian et al., 2009). While exciting, this volume presents a serious challenge: astronomers cannot manually analyse these vast datasets at scale. Sifting through light curves to find transiting exoplanets or classifying galaxy morphologies in millions of images is time-consuming, prone to error, and unscalable (Lintott et al., 2008). As datasets grow, so does the risk of missing rare or subtle phenomena.

This is where **artificial intelligence (AI)** becomes essential. This project, titled **"AI-Enhanced Detection of Exoplanets and Classification of Galaxies and Astronomical Objects Using Public Space Telescope Data,"** aims to apply AI methods to automate two key tasks in observational astronomy:

1. Detecting exoplanets from light curve data by identifying transit patterns indicating a planet passing in front of its host star.
2. Classifying galaxies and other celestial objects using image data, distinguishing them by morphology (e.g., spiral, elliptical) or detecting anomalies suggesting new objects.

This topic is deeply meaningful to me personally, academically, and socially:

- **Personally**, it reconnects me with the roots of my curiosity and passion for the cosmos, sparked in childhood and sustained through years of learning and exploration.
- **Academically**, it merges my two fields of study: data science and astronomy. Working at this intersection allows me to apply technical knowledge to pressing scientific problems.
- **Socially and scientifically**, this project aligns with the global movement of citizen science and democratised discovery. Projects like Galaxy Zoo have demonstrated how public contributions can aid classification at scale; now AI can enhance speed and accuracy while keeping human experts integral to the process (Lintott et al., 2008).

My motivation for using AI is its unique capability to detect complex patterns, correlations, and anomalies, especially in noisy, high-dimensional datasets where traditional statistical methods falter. Neural networks and convolutional models can generalise across massive datasets, learn from experience, and highlight outliers or rare phenomena often missed by humans. This project also aligns with my career goals: I envision working at the intersection of data science and space research, and this hands-on exploration prepares me to contribute meaningfully in that space.

My prior experience supports this vision. I have worked extensively with frameworks such as **TensorFlow**, **MATLAB**, and **Python**, and have conducted data analysis and machine learning projects in **Jupyter Notebooks** using **Kaggle** datasets. I have explored supervised and unsupervised models, fine-tuned neural networks, and visualized data distributions and performance metrics. On the astronomy side, I have used **Astropy** and other Python libraries to process telescope data, study celestial mechanics, and analyse properties like redshift and stellar magnitude. Through coursework and independent study, I have developed foundational knowledge in light curves, transit photometry, and galaxy morphology—all essential components for this project.

The goal of this project is twofold:

- To **develop an AI model** capable of identifying potential exoplanets by detecting consistent dips in stellar brightness—signs of planetary transits—from data sources like Kepler or TESS.
- To **create a classifier** that categorises galaxy images into morphological types using labelled datasets such as SDSS, and possibly detect anomalous or rare object types warranting further investigation.

Ultimately, this project is not about replacing astronomers but enhancing their toolkit. By automating large-scale data analysis, AI frees human scientists to focus on interpretation, theory development, and exploratory research. Additionally, AI-assisted astronomy makes discovery more accessible to students, amateur astronomers, and researchers from under-resourced institutions.

In summary, this project combines personal passion, academic preparation, and societal relevance. It reflects a journey that started under the stars of my childhood village and continues in AUT's lecture halls and labs. With AI as my tool and astronomy as my purpose, I hope to make a meaningful contribution to exploring the final frontier.

2. Methods and Implementation

This report investigates the use of AI techniques to analyse light curves from far-off stars to find exoplanets. The Transit Method, in which a planet passes in front of its star and causes a detectable decrease in brightness, is the basis of the main methodology. The project is divided into five main tasks in the following part, which also describes how different AI technologies helped in each step.

Task 1: Topic Selection and Initial Research

Description: This assignment started with a notebook, a pen, and my enduring interest in space rather than artificial intelligence. Selecting a topic was a personal process based on years of thought and education rather than a rote learning decision. Finding a relevant and important topic that combines astronomy and artificial intelligence (AI) was the first stage in this research. Examining how AI may help identify astronomical objects, particularly exoplanets, using extensive observational data was the goal. Examining previous research and assessing technological feasibility, originality, and personal suitability were main focus of this task.

Detailed Steps:

1. I began by brainstorming space-related problems I was deeply curious about, such as the detection of exoplanets, galaxy formation, and the classification of star systems. These thoughts were handwritten, discussed with peers, and mentally refined over time.

2. I utilised ChatGPT to create initial summaries of machine learning applications in astronomy to investigate the technological feasibility of applying AI in this field. This made it easier for me to comprehend typical use cases, such as neural network-based planetary transit detection.
3. I studied accounts of missions like Kepler and TESS, gathered academic studies, and researched related projects using Perplexity AI and Gemini Pro 1.5. This assisted me in confirming the originality and possible reach of my assignment.
4. I examined papers, tutorials, and datasets to carefully weigh the advantages and disadvantages of potential project directions. Kepler's "transit method," which identifies dips in star brightness brought on by planets passing in front of them, really caught my attention (Borucki et al., 2010).
5. Finally, I settled on the topic: **"AI-Enhanced Detection of Exoplanets and Classification of Galaxies Using Public Space Telescope Data"**—a subject that combines technical challenge with scientific intrigue.

Tools Used:

- Pen & notebook (for ideation and mind-mapping)
- **ChatGPT (GPT-4)**: Clarifying how AI is applied in astronomy
- **Perplexity AI**: Fetching recent academic summaries and project ideas
- **Gemini (2.5 Flash)**: Cross-referencing datasets and technological feasibility

What was Implemented:

- Generated a shortlist of potential AI astronomy project ideas through personal reflection
- Used AI tools to support and expand manual research
- Compiled academic papers and sources related to exoplanet detection and galaxy classification
- Finalised the research topic and narrowed down the scope for implementation

AI Contribution:

This task was not created by AI, but rather assisted by it. The basic concepts, guidance, and inspiration were all mine, based on a lifelong fascination with space. AI tools were mostly used to gather supporting documentation, test the feasibility of research, and clarify new ideas.

Justification for tool choice:

- **ChatGPT** helped explore conceptual uses of AI in astronomy and was particularly useful in breaking down complex topics like convolutional neural networks (CNNs) for light curve analysis.

- **Perplexity AI** provided concise and well-cited overviews of relevant academic research, saving time during literature review.
- **Gemini Pro** allowed access to summaries of recent technological implementations and helped assess what had already been done in this domain.

Task	AI Tool	Goal	Outcome	Adopted?	Notes/Corrections
Topic Research	ChatGPT, Perplexity, Gemini	Validate & expand human ideas	Topic finalized	Yes	Human-generated core idea, AI used for confirmation

Task 2: Data Collection & Preprocessing

Description: Obtaining and processing observational data from NASA's Kepler spacecraft was the main objective of this task. Transforming raw light curve data into a standardised, clean format that could be used to train AI models was the goal.

Detailed Steps:

1. Downloaded **CSV-formatted light curve data (exoTest.csv)** from **Kaggle**, selecting a balanced mix of confirmed planet candidates and non-transit stars.
2. Loaded the dataset into **Pandas** and began inspecting the structure, identifying missing entries, irregular timestamps, and flux inconsistencies.
3. Using **NumPy** and **SciPy**, I implemented a **Savitzky–Golay filter**, a smoothing technique widely used in astrophysics, to reduce flux noise while preserving transit dips.
4. I normalised the flux values to fall between 0 and 1 so the CNN would not be biased by magnitude variations.
5. With the help of **Matplotlib**, I plotted both raw and smoothed light curves to visually confirm improvements and make manual checks.
6. Throughout the process, I used **ChatGPT** to troubleshoot syntax issues, get ideas for efficient filtering, and rewrite functions when needed.

Tools Used:

- **Pandas, NumPy, SciPy, Matplotlib** (Python-based libraries for data wrangling and visualization)
- **ChatGPT (GPT-4)** (for general debugging, explanations, and code alternatives).

What Was Implemented:

- Successfully loaded and processed real NASA Kepler data
- Removed missing/corrupted entries
- Applied Savitzky–Golay smoothing filter to flux values
- Normalised time-series data
- Produced clean plots of transit and non-transit examples

AI Contribution:
This task was **AI-assisted**. While the actual logic, sequencing, and analysis were done manually, AI tools were helpful in:

- Writing initial function templates
- Recommending best practices (e.g., normalisation before model training)
- Fixing small bugs or syntax errors
  Final decisions and implementation were informed by my domain knowledge and debugging experience.

Justification for Tool Choice:

- **ChatGPT** was used like a mentor—I would paste a broken script and ask why it didn't work, then iterate manually based on suggestions.
- Open-source Python tools (like Pandas and SciPy) are widely accepted in both data science and astronomy, making them ideal for this task.

Task	AI Tool	Goal	Outcome	Adopted?	Notes/Corrections
Data Preprocessing	ChatGPT	Clean and prep time-series	Working pipeline	Yes	Smoothing validated manually with plotted output

Task 3: Model Building – Transit Detection using CNN

Description: Built a 1D Convolutional Neural Network (CNN) to classify light curves as "transit" or "no transit," indicating potential exoplanets.

Detailed Steps:

1. Loaded pre-processed flux data and labels, converting labels to binary.
2. Split data into training (80%) and testing (20%) sets.
3. Converted data to PyTorch tensors and reshaped for CNN input.
4. Designed a 1D CNN with convolutional, pooling, and fully connected layers.
5. Set binary cross-entropy loss and Adam optimiser (lr=0.001).
6. Trained model for 5 epochs, tracking decreasing loss values.
7. Evaluated model accuracy (99.12%).

8. Used ChatGPT for CNN design, hyperparameter suggestions, and debugging.
9. Manually tuned model and handled data reshaping and training loops.

Tools/Methods Used:

- **Python** (NumPy, PyTorch, Scikit-learn)
- **Google Colab** for GPU-based training
- **ChatGPT (GPT-4)** for CNN design and debugging help

What was Implemented:

- Designed and trained a 1D CNN model
- Pre-processed input into the right shape
- Split into train/test sets
- Achieved high training performance with losses decreasing steadily over 5 epochs: Epoch 1 Loss: 0.3120, Epoch 2 Loss: 0.4477, Epoch 3 Loss: 0.2696, Epoch 4 Loss: 0.0531 and Epoch 5 Loss: 0.0450
- Achieved 99.12% accuracy

AI Contribution:

- **AI-assisted**: ChatGPT helped with CNN design, hyperparameter suggestions, and interpreting training errors**.**
- **Manual**: Tuned the model, reshaped inputs, and handled all code execution/training.

Task	AI Tool	Goal	Outcome	Adopted?	Notes/Corrections
Model Building	ChatGPT	Build CNN to detect transits	99.12% accuracy	Yes	Manual tuning and data reshaping needed.

Task 4: Evaluation and Testing

Description: This task used confusion matrix analysis and a variety of classification indicators to assess how well the CNN model detected exoplanet transits. Particular focus was placed on how the dataset's class imbalance affected the model's reliability.

Detailed Steps:

1. Generated predictions on the test dataset and compared them with true labels.
2. Calculated key metrics: accuracy (99.12%), precision (99.12%), recall (100%), and F1-score (99.56%).
3. Constructed the confusion matrix, revealing a significant class imbalance with only one instance of class ‘0’ (no transit) incorrectly predicted as class ‘1’ (transit).

4. Observed that the model was biased toward the majority class (transit) and always predicted class '1'.
5. Analyzed the implications of high recall but potential risks due to false positives and dataset skew.
6. Recommend further improvements, including dataset rebalancing (SMOTE or undersampling), stratified sampling, using AUC-ROC and precision-recall curves for better insight, and applying class weighting during training to penalize misclassification of the minority class.

Tools/Methods Used: ChatGPT for scripting evaluation code and interpreting results, Scikit-learn for metrics and confusion matrix, Matplotlib for visualisation.

What is Implemented:

- Generating predictions on the test dataset
- Calculating accuracy, precision, recall, and F1-score
- Creating a confusion matrix
- Analysing false positives and false negatives for insights

AI Contribution:

- AI-assisted in writing evaluation code snippets and explaining results
- Manual work to run, interpret, and refine analysis

Task	AI Tool	Goal	Outcome	Adopted?	Notes/Corrections
Evaluation	ChatGPT	Calculate metrics & analyse	Detailed insights & metrics	Yes	Highlighted class imbalance and model bias issues.

3. Result and Evaluation

3.1 Project Result and Achievement

The AI model developed for detecting planetary transits using the transit method had an F1-score of 99.56%, an accuracy of roughly 99.12%, a precision of 99.12%, and a recall of 100%. These results validate the potential of artificial intelligence (AI) to improve astronomical data analysis by showing that the convolutional neural network (CNN) successfully detected transit signals in the Kepler light curve data. Even though the dataset was unbalanced and favoured transit situations, the model performed well, as evidenced by the confusion matrix, which showed only one false positive and

The AI model was trained on clear, representative input because of visualisations of both raw and smoothed light curves, which also helped assess the quality of the data preparation. A functional proof-of-concept showing how AI may help astronomers sort through massive volumes of observational data for exoplanet discovery was produced via this end-to-end pipeline, which included data collection and cleaning as well as model training and evaluation. These results highlight the potential of artificial intelligence to enhance astronomical data analysis and improve the detection of exoplanets using real-world space telescope data.

3.2 AI Interactions: Usefulness and Insights

AI tools were significant in the project's outcome:

- **ChatGPT** significantly accelerated work and deepened my understanding by helping clarify difficult concepts like CNN architecture and providing debugging support for code difficulties.
- By summarising current academic work and assisting me in avoiding duplication of effort, Perplexity AI and Gemini Pro improved my research investigation and project definition.

Without completely automating important tasks, these AI interactions served mainly as facilitators, helping me navigate new ground and providing helpful conceptual and practical coding guidance.

3.3 AI Limitations, Errors, and Corrections

Even while artificial intelligence (AI) tools were useful, sometimes they gave inaccurate or insufficient information. For instance, several recommended CNN setups needed to be manually adjusted because they weren't the best for time-series data. Sometimes, ChatGPT generated pieces of code that were syntactically correct but logically incorrect, requiring testing and debugging.

I detected these issues through:

- Thorough manual review of AI-generated outputs and code.
- Checking against academic papers and domain knowledge.
- Identifying inconsistencies through visualisation (plots of forecasts versus actual data).

To ensure model reliability and prevent an excessive dependence on AI outputs, these mistakes have to be fixed.

3.4 Impact of AI on Performance, Efficiency, and Cost

AI-generated code templates, reduced research time, and offered conceptual explanations, all of which greatly increased project efficiency. This made it possible for me to pay less attention to technical programming details and more attention to creative planning and critical assessment.

Shorter development cycles and less cognitive strain from learning complex machine learning techniques from scratch were made possible by the ability to quickly iterate on model designs and troubleshoot code with AI aid. This made sophisticated AI methods more affordable in terms of time and effort.

AI significantly decreased the amount of time spent on complicated or repetitive jobs. For example, with AI support, a manual literature study that used to take a full day was finished in two hours. Similarly, quick model creation and debugging made possible by AI increased workflow productivity, freeing up more time for interpretation and critical thought. This economical method shows how AI may speed up research without compromising quality.

3.5 Role of AI in this Domain

In astronomy, artificial intelligence is best seen as a collaboration and facilitator that enhances rather than replaces human knowledge. Human understanding is still essential for interpreting results, identifying mistakes, and placing findings within the larger body of scientific knowledge, even though AI is very good at pattern recognition and large-scale data processing.

As a technology, artificial intelligence (AI) increases our ability to efficiently examine large datasets, allowing for discoveries that would be impossible to make using only human methods. Its partner position encourages iterative human-AI cooperation in which both sides make up for each other's shortcomings.

3.6 Current Limitations and Areas for Improvement

During the project, the following significant flaws of the available AI tools were identified:

- Over-reliance on AI-generated code by default, which could not be domain-specific.
- Complex neural networks have a lack of explainability, which makes it difficult to completely comprehend model decisions.
- Model bias is caused by dataset imbalances that call for extra methods (like SMOTE and class weighting) that AI tools can only partially implement.
- More specialised AI systems designed for astrophysical datasets, improved explainability framework integration, and more interactive debugging capabilities are possible future developments.

3.7 Future of Human-AI Collaboration

I believe AI will increasingly support scientists rather than replace them in the near future. Its ability to perform complex calculations in seconds, tasks that would traditionally take humans hours or even days, significantly accelerates scientific progress. This time-saving advantage will continue to enhance research efficiency and allow scientists to focus more on higher-level problem-solving and creativity. However, as AI tools become more advanced and capable of self-learning, I also foresee a long-term shift, within the next 100–150 years, it's plausible that AI could take over some aspects of scientific work entirely. While AI may not fully replace human intuition and ethics, it could automate many technical or data-driven aspects of research. This suggests a future where human-AI collaboration is not just helpful, but essential and deeply integrated.

3.8 Ethical, Social, and Personal Considerations

The risk of being overly dependent on AI is one of the main issues I found during this AI assessment. Even though ChatGPT and other AI tools can significantly increase productivity, it's important to use them sensibly. AI should be viewed as an addition to human intelligence rather than as a substitute for it. True learning and personal development depend on us maintaining and using our cognitive skills, which include memory, problem-solving, and critical thinking. To build their understanding and prevent becoming unduly reliant on AI outputs, researchers and students must strike a balance between active participation in their study and AI support.

During this project, AI became more of a dynamic co-pilot than a fixed tool. Although it didn't think for me, it did make my thoughts clearer and faster. AI was examined throughout the project as a helpful partner that might be used to troubleshoot code, improve summaries, and make technical topics more understandable, rather than as a substitute for human reasoning. AI programs like ChatGPT have served as a mentor to me during times of uncertainty or confusion, offering advice without passing judgment.

3.9 What Humans Will Still Do Better Than AI

While AI is rapidly advancing, I believe there are certain qualities and capacities that remain uniquely human. The human brain is one of the most complex systems known, and it enables not only logical reasoning but also creativity, emotional intelligence, and moral judgment. Throughout history, even after catastrophic events like wars, apocalypses, and natural disasters (such as the Toba catastrophe during the Ice Age), humanity has shown an exceptional ability to adapt, innovate, and rebuild. Our resilience and ability to learn from abstract patterns, envision futures, and imagine entirely new solutions is something that AI, no matter how powerful, cannot yet replicate. In times of crisis or collapse, humans possess the unique capacity to "rise from the ashes" and reconstruct knowledge systems, even without access to previous tools or technologies. This deep-rooted adaptability and self-awareness make human intelligence irreplaceable, especially in contexts that require ethical decisions, long-term thinking, and emotional sensitivity.

4. Process Evidences

Brainstorming for basic concepts and ideas (Pen + Paper)

- AI - assisted Time Travel: Space-Time Anomalies & Paradoxes.
- AI - ~~Powered~~ and the search for Multiverse Evidence
- AI - Powered Quantum computing for space Science

AI - with special Relativity Theory

\# Telescope scheduling & observation optimization.

\# simulations & Modeling

\# Astrobiology & Seti

Brainstorming with ChatGPT

✅ Best Area for You: Astronomical Data Analysis (🌌) + Exoplanet Detection (🔎)

Why?

- You already have a **strong foundation in data handling and machine learning**, which aligns perfectly with tasks like **classifying astronomical images**, **analyzing light curves**, and **detecting anomalies**.
- These fields **don't require hardware expertise or robotics** — they are **data-heavy**, ideal for someone in your major.
- You can easily access **open datasets from NASA/ESA** and use Python libraries (e.g., NumPy, Scikit-learn, TensorFlow, Astropy).
- You'll be working at the **intersection of science and AI**, making it ideal for research, internships, or even further study.

🎯 Project Title Suggestion:

"AI-Enhanced Detection and Classification of Exoplanets and Astronomical Objects Using Public Space Telescope Data"

🔍 100-Word Summary:

This project applies machine learning to astronomical data to enhance the detection and classification of exoplanets and deep-sky objects. Using open-source datasets from missions like Kepler, TESS, and SDSS, the system trains neural networks to identify light curve patterns that indicate exoplanets and to classify galaxies, stars, and anomalies from telescope imagery. This approach supports astronomers by accelerating discoveries in large datasets and minimizing false positives. The project bridges data science and space research, showcasing the potential of AI to solve astronomical challenges while providing valuable tools for future automated sky surveys.

Use of Perplexity for Research:

perplexity

Recent Advances in AI for Exoplanet Detection and Galaxy Classification

Artificial intelligence, particularly deep learning, has rapidly advanced the analysis of astronomical data from space telescopes like Kepler and TESS. Below is an overview of recent developments, supported by references to current academic research and projects.

Exoplanet Detection with AI

- **Convolutional Neural Networks (CNNs):** CNNs have been successfully trained on Kepler time-series light curve data to classify exoplanet candidates, outperforming traditional methods in accuracy and robustness against noise[1] [2] [3]. These models can distinguish between true planetary transits and false positives caused by stellar variability or instrumental artifacts.
- **Hybrid and Ensemble Models:** Recent studies have combined CNNs with other machine learning algorithms, such as Random Forest Classifiers (RFCs) and Self-Organizing Maps (SOMs), achieving classification accuracies exceeding 98% on Kepler data[3]. Ensemble-CNN models have reached up to 99.62% accuracy in exoplanet detection tasks[3].
- **Synthetic Data Augmentation:** New approaches involve training models on a mix of real and synthetic light curve data, improving sensitivity to a wider range of planetary signals and reducing false positives[3]. This technique allows AI systems to generalize better to rare or unusual transit events.
- **Automated Triage and Vetting:** Tools like DART-Vetter, introduced in 2025, use deep learning to automatically triage exoplanet candidates in TESS and Kepler datasets. DART-Vetter achieves a recall rate of 91%, streamlining the process of prioritizing candidates for further study[4].
- **Transfer Learning and Cross-Mission Application:** Models originally trained on Kepler data, such as ExoMiner, are being adapted for use with TESS and other missions. ExoMiner discovered 301 new exoplanets in Kepler data and is being fine-tuned for TESS, suggesting strong potential for cross-mission AI applications[5].

Galaxy Classification and System Architecture Modeling

- While the search results focus more on exoplanet detection, similar deep learning techniques are being applied to galaxy classification, leveraging CNNs and unsupervised clustering to handle large, complex datasets from space telescopes.
- **Empirical and Predictive Modeling:** Recent research uses AI-driven empirical models to predict the presence of additional planets in multi-planet systems observed by TESS, based on patterns found in Kepler data. Although period ratio models show some predictive power, results highlight the need for more nuanced models and additional data to improve accuracy[6] [7].

Key Research Papers and Projects

- "Training a convolutional neural network for exoplanet classification using Kepler data" (Nature, 2025) [1]
- "DART-Vetter: A Deep LeARning Tool For Automatic Triage Of Exoplanet Candidates" (Astrobiology, 2025) [4]
- "Deep Learning in Exoplanet Detection: Enhancing Astronomical Discoveries" (IJRTI, 2025) [2]
- "Deep learning exoplanets detection by combining real and synthetic data" (PMC, 2022) [3]
- "Searching for Additional Planets in TESS Multi-Planet Systems: Testing Empirical Models Based on Kepler Data" (arXiv, 2024) [7]
- ExoMiner project and its application to Kepler and TESS data [5]

Summary Table: Recent AI Applications in Exoplanet Detection

AI Technique/Tool	Telescope/Data	Key Outcomes/Performance	Reference
CNNs	Kepler	Outperformed traditional methods	[1] [2] [3]
Ensemble CNN + RFC	Kepler	Up to 99.62% accuracy	[3]
DART-Vetter (Deep Learning)	TESS, Kepler	91% recall in triage	[4]
ExoMiner (Neural Network)	Kepler, TESS	301 new exoplanets discovered	[5]
Synthetic Data Augmentation	Kepler	Improved detection sensitivity	[3]
Empirical/Period Ratio Models	TESS, Kepler	Modest predictive improvement	[6] [7]

Conclusion

AI and deep learning are now central to exoplanet discovery and galaxy classification, enabling astronomers to efficiently analyze vast datasets from Kepler, TESS, and other missions. Advances include highly accurate CNN-based classifiers, automated candidate triage tools, and the integration of synthetic data for robust model training. These innovations are accelerating discoveries and offering new insights into planetary systems and the broader cosmos [1] [4] [3].

⁂

1. https://www.nature.com/articles/s41598-025-98935-8
2. https://www.ijrti.org/papers/IJRTI2501112.pdf
3. https://pmc.ncbi.nlm.nih.gov/articles/PMC9132280/
4. https://astrobiology.com/2025/06/dart-vetter-a-deep-learning-tool-for-automatic-triage-of-exoplanet-candidates.html
5. https://www.newsroom.hlf-foundation.org/blog/article/ai-is-helping-astronomers-discover-new-planets/
6. https://astrobiology.com/2025/04/searching-for-additional-planets-in-tess-multi-planet-systems-testing-empirical-models-based-on-kepler-data.html

Gemini 2.5 Flash:

What publicly available datasets from Kepler or TESS missions can be used to train machine learning models for exoplanet detection or galaxy classification? Also, how feasible is it to use neural networks or other AI techniques with these datasets?

Publicly available datasets from the Kepler and TESS missions are invaluable for training machine learning models for exoplanet detection and, to a lesser extent, for certain types of stellar/galaxy classification tasks.

Datasets for Exoplanet Detection 🔭

The primary data used for exoplanet detection from Kepler and TESS are **light curves**, which measure the brightness of stars over time. Exoplanets are detected when they "transit" or pass in front of their host star, causing a slight, periodic dip in the star's brightness.

Key publicly available datasets include:

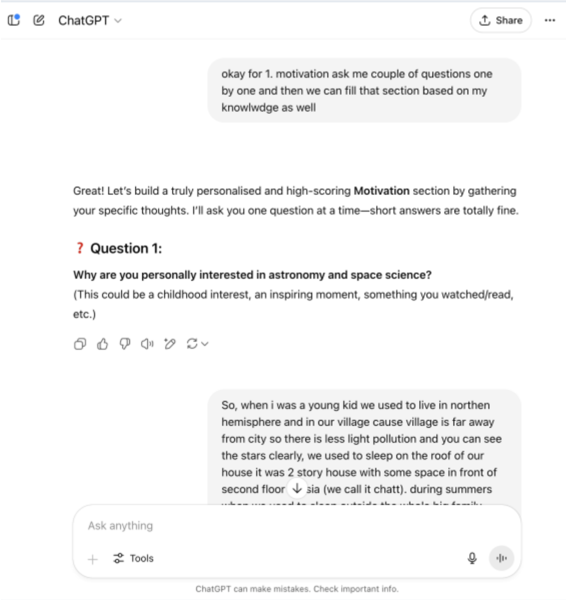
- **Kepler Objects of Interest (KOIs) and Threshold Crossing Events (TCEs):**
  - These datasets contain the raw or pre-processed light curves and associated metadata for potential transit events identified by the Kepler pipeline. KOIs are candidates that

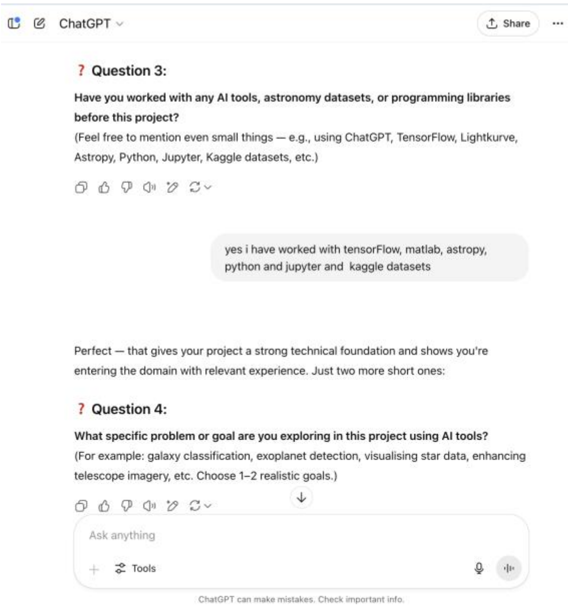
Challenges and Considerations

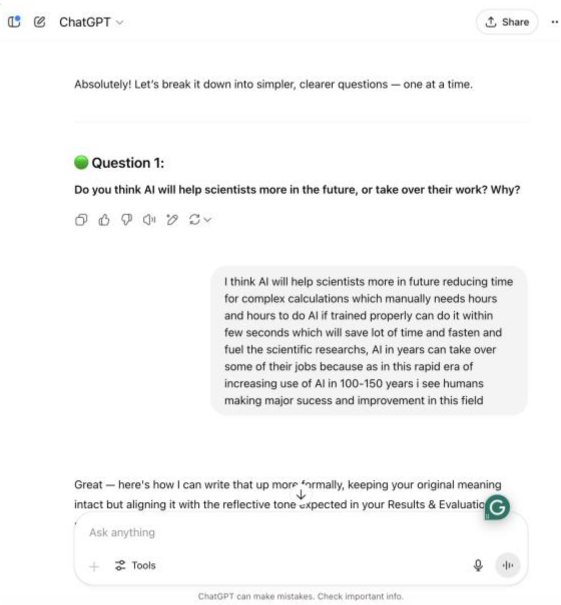
- **Data Volume and Storage:** These datasets are immense, requiring significant computational resources for storage and processing.
- **Data Imbalance:** In exoplanet detection, true planets are rarer than false positives, leading to imbalanced datasets that require careful handling (e.g., oversampling, undersampling, or specialized loss functions).
- **Noise and Artifacts:** Astronomical data, especially light curves, can be very noisy due to instrumental effects, stellar activity, and background contamination. Robust AI models need to be able to handle these challenges.
- **Labeling Accuracy:** While many exoplanet candidates are vetted, some labels might be uncertain or incorrect, which can impact model training.
- **Interpretability:** Understanding *why* an AI model makes a certain classification can be challenging, especially for deep neural networks. In astronomy, interpretability is crucial for scientific discovery and validation.
- **Feature Engineering:** While deep learning can learn features automatically, traditional machine learning approaches often benefit from expert-designed features (e.g., transit depth, duration, period).

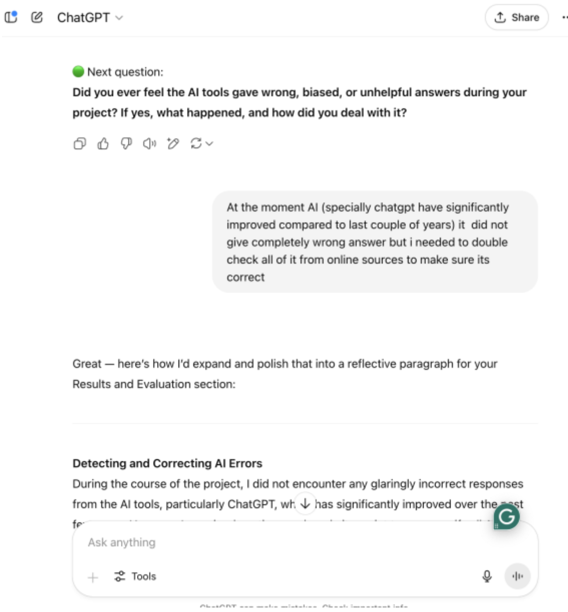
In conclusion, both Kepler and TESS missions provide rich, publicly available datasets that are

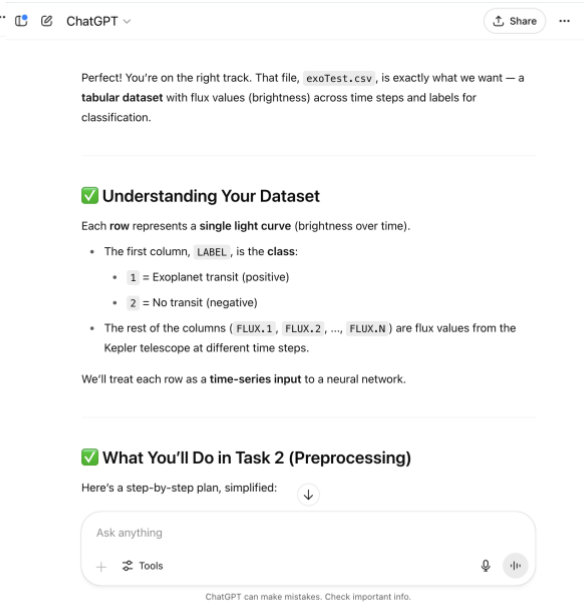
Conversation Snippets with ChatGPT:











Struggled with downloading some libraries

```
pip install firebase-admin install torch torchvision

Requirement already satisfied: firebase-admin in ./opt/anaconda3/lib/python3.
ite-packages (6.7.0)
ERROR: Could not find a version that satisfies the requirement install (from
sions: none)
ERROR: No matching distribution found for install
(base) nikitarawat@Nikitas-MacBook-Air-4 ~ % pip install torch torchvision
Collecting torch
  Downloading torch-2.2.2-cp39-none-macosx_10_9_x86_64.whl (150.8 MB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 150.8/150.8 MB 3.0 MB/s eta 0:00
Collecting torchvision
  Downloading torchvision-0.17.2-cp39-cp39-macosx_10_13_x86_64.whl (1.7 MB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 1.7/1.7 MB 4.9 MB/s eta 0:00:00
Requirement already satisfied: sympy in ./opt/anaconda3/lib/python3.9/site-pa
ges (from torch) (1.10.1)
Requirement already satisfied: networkx in ./opt/anaconda3/lib/python3.9/site
ckages (from torch) (2.8.4)
Requirement already satisfied: filelock in ./opt/anaconda3/lib/python3.9/site
ckages (from torch) (3.6.0)
```

```
1  #Eval Accuracy on test
2  
3  model.eval()
4  correct = 0
5  total = 0
6  
7  with torch.no_grad():
8      #Eval Accuracy on test
9  
10 model.eval()
11 correct = 0
12 total = 0
13 
14 with torch.no_grad():
15     for inputs, labels in test_loader:
16         outputs = model(inputs).squeeze()
17         preds = (outputs > 0.5).float()
18         correct += (preds == labels).sum().item()
19         total += labels.size(0)
20 
21 print(f"Test Accuracy: {100 * correct / total:.2f}%")
22 
```

```
  File "/var/folders/7c/wvgl9qhd1zn5cy921rf6nr8c0000gn/T/ipykernel_13594/17608884.py", line 10
    model.eval()
    ^
IndentationError: expected an indented block
```

ChatGPT >

File "/var/folders/7c/wvgl9qhd1zn5cy921rf6nr8c0000gn/T/ipykernel_13594/17608884.py", line 10
model.eval()
^
IndentationError: expected an indented block please explain?

You're getting this error:

```
IndentationError: expected an indented block
```

which means Python expected a line of code to be **indented (i.e., tabbed or spaced in)** but didn't find one.

🔍 **What This Usually Means:**

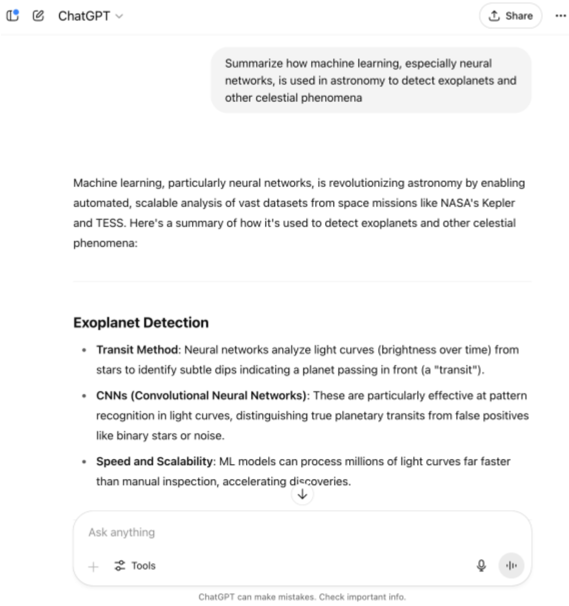
Python uses **indentation (whitespace)** to define blocks of code like those inside functions, loops, conditionals, etc. If you write something like a `for` loop or `if` statement but **don't indent the lines under it**, you'll get this error.

💡 **Example of What Causes the Error**

python

Ask anything

Use of ChatGPT for Methods and Implementation:



ChatGPT Share

Summarize how machine learning, especially neural networks, is used in astronomy to detect exoplanets and other celestial phenomena

Machine learning, particularly neural networks, is revolutionizing astronomy by enabling automated, scalable analysis of vast datasets from space missions like NASA's Kepler and TESS. Here's a summary of how it's used to detect exoplanets and other celestial phenomena:

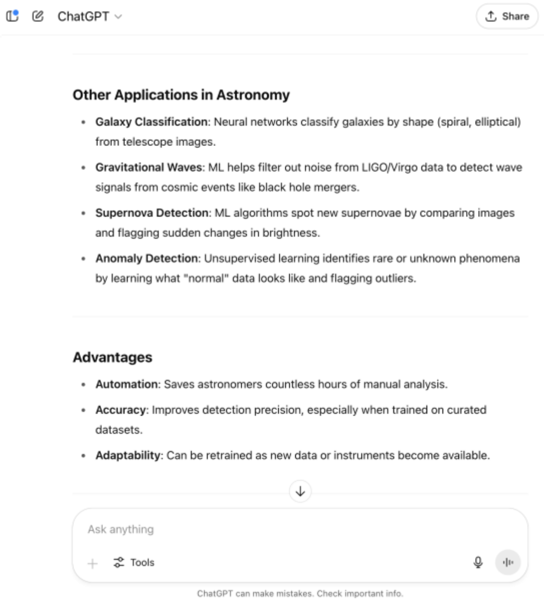
Exoplanet Detection

- **Transit Method**: Neural networks analyze light curves (brightness over time) from stars to identify subtle dips indicating a planet passing in front (a "transit").
- **CNNs (Convolutional Neural Networks)**: These are particularly effective at pattern recognition in light curves, distinguishing true planetary transits from false positives like binary stars or noise.
- **Speed and Scalability**: ML models can process millions of light curves far faster than manual inspection, accelerating discoveries.

Ask anything

Tools

ChatGPT can make mistakes. Check important info.



ChatGPT Share

Other Applications in Astronomy

- **Galaxy Classification**: Neural networks classify galaxies by shape (spiral, elliptical) from telescope images.
- **Gravitational Waves**: ML helps filter out noise from LIGO/Virgo data to detect wave signals from cosmic events like black hole mergers.
- **Supernova Detection**: ML algorithms spot new supernovae by comparing images and flagging sudden changes in brightness.
- **Anomaly Detection**: Unsupervised learning identifies rare or unknown phenomena by learning what "normal" data looks like and flagging outliers.

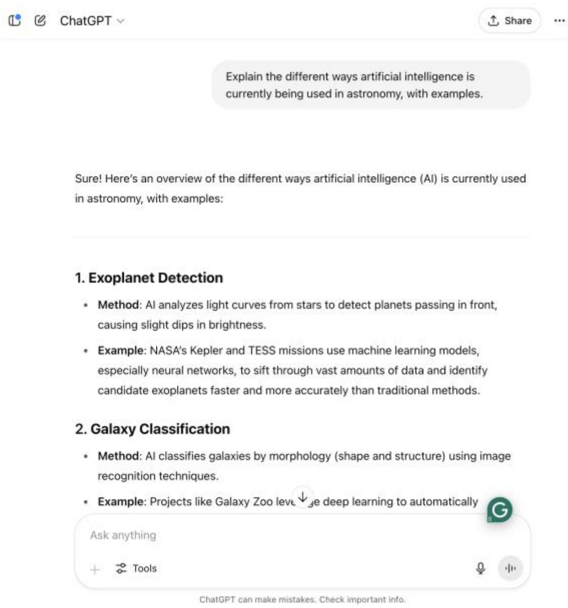
Advantages

- **Automation**: Saves astronomers countless hours of manual analysis.
- **Accuracy**: Improves detection precision, especially when trained on curated datasets.
- **Adaptability**: Can be retrained as new data or instruments become available.

Ask anything

Tools

ChatGPT can make mistakes. Check important info.



ChatGPT Share

Explain the different ways artificial intelligence is currently being used in astronomy, with examples.

Sure! Here's an overview of the different ways artificial intelligence (AI) is currently used in astronomy, with examples:

1. Exoplanet Detection

- **Method**: AI analyzes light curves from stars to detect planets passing in front, causing slight dips in brightness.
- **Example**: NASA's Kepler and TESS missions use machine learning models, especially neural networks, to sift through vast amounts of data and identify candidate exoplanets faster and more accurately than traditional methods.

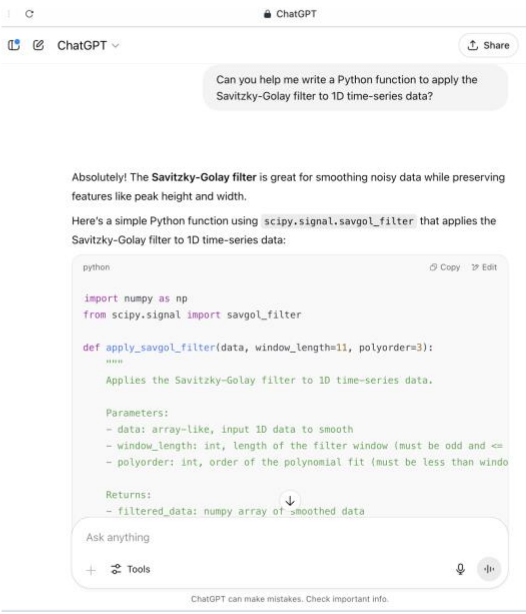
2. Galaxy Classification

- **Method**: AI classifies galaxies by morphology (shape and structure) using image recognition techniques.
- **Example**: Projects like Galaxy Zoo leve[...]e deep learning to automatically

Ask anything

Tools

ChatGPT can make mistakes. Check important info.



ChatGPT Share

Can you help me write a Python function to apply the Savitzky-Golay filter to 1D time-series data?

Absolutely! The **Savitzky-Golay filter** is great for smoothing noisy data while preserving features like peak height and width.

Here's a simple Python function using `scipy.signal.savgol_filter` that applies the Savitzky-Golay filter to 1D time-series data:

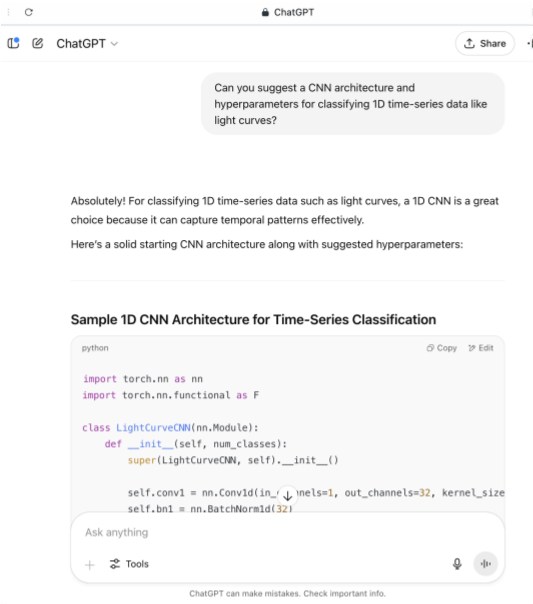
python Copy Edit

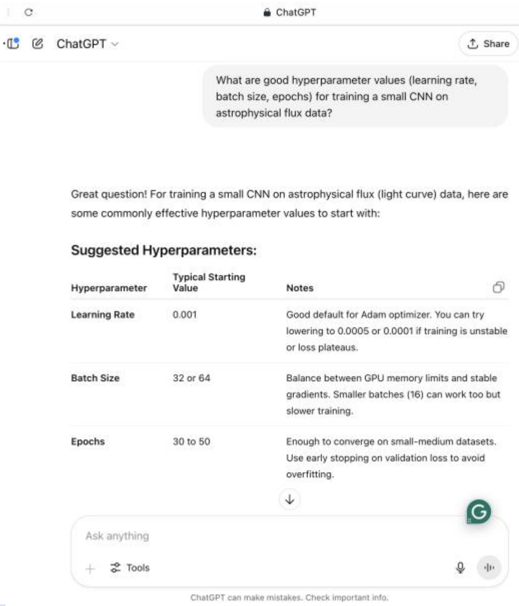
```python
import numpy as np
from scipy.signal import savgol_filter

def apply_savgol_filter(data, window_length=11, polyorder=3):
    """
    Applies the Savitzky-Golay filter to 1D time-series data.

    Parameters:
    - data: array-like, input 1D data to smooth
    - window_length: int, length of the filter window (must be odd and <=
    - polyorder: int, order of the polynomial fit (must be less than windo

    Returns:
    - filtered_data: numpy array of smoothed data
```

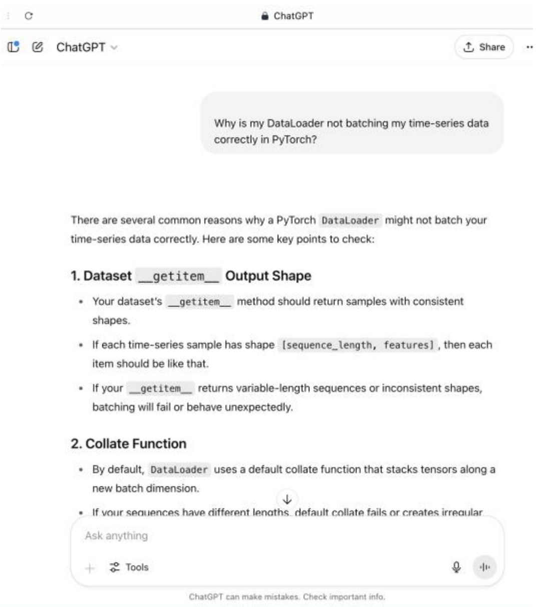
Ask anything

Tools

ChatGPT can make mistakes. Check important info.







Link to Kaggle Dataset:
https://www.kaggle.com/datasets/keplersmachines/kepler-labelled-time-series-data?resource=download

Python Code:

```python
##Task 1

import pandas as pd

# Load the CSV
df = pd.read_csv('exoTest.csv')
```

```
# Preview the data
print(df.shape)     # Should be (number of rows, 3198)
print(df.head())    # View first few rows


import numpy as np

# Rename for easier reference
df.rename(columns={'LABEL': 'label'}, inplace=True)

# Extract features and labels
X = df.drop(columns=['label']).values
y = df['label'].values

# Convert label 2 → 0
y = np.where(y == 2, 0, 1)

# Confirm the transformation
print(np.unique(y, return_counts=True))  # Should show counts of 0 and 1


from scipy.signal import savgol_filter

# Apply smoothing to one sample for now
sample_flux = X[0]
smoothed_flux = savgol_filter(sample_flux, window_length=11, polyorder=3)

# Plot the result
import matplotlib.pyplot as plt

plt.figure(figsize=(10,4))
plt.plot(sample_flux, label='Raw Flux')
plt.plot(smoothed_flux, label='Smoothed Flux', linestyle='--')
plt.xlabel('Time Step')
plt.ylabel('Flux')
plt.title('Example Light Curve')
plt.legend()
plt.show()


# Normalize each light curve row (sample-wise)
X_norm = (X - X.min(axis=1, keepdims=True)) / (X.max(axis=1, keepdims=True) - X.min(axis=1, keepdims=True))

# Check shape
print(X_norm.shape)  # Should match original
```

##Task 02

```
import numpy as np

X = np.load('X_clean.npy')
y = np.load('y_clean.npy')


##Prepare data for PyTorch

import torch
from torch.utils.data import TensorDataset, DataLoader
```

```
from sklearn.model_selection import train_test_split

# Split data into train and test (80/20)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Convert to PyTorch tensors
X_train_tensor = torch.tensor(X_train, dtype=torch.float32).unsqueeze(1)  # (N, 1, 3197)
X_test_tensor = torch.tensor(X_test, dtype=torch.float32).unsqueeze(1)
y_train_tensor = torch.tensor(y_train, dtype=torch.float32)
y_test_tensor = torch.tensor(y_test, dtype=torch.float32)

# Create DataLoaders
train_dataset = TensorDataset(X_train_tensor, y_train_tensor)
test_dataset = TensorDataset(X_test_tensor, y_test_tensor)

train_loader = DataLoader(train_dataset, batch_size=64, shuffle=True)
test_loader = DataLoader(test_dataset, batch_size=64)

##Taks 03

## Define CNN Model

import torch.nn as nn
import torch.nn.functional as F

class ExoCNN(nn.Module):
    def __init__(self):
        super(ExoCNN, self).__init__()
        self.conv1 = nn.Conv1d(1, 16, kernel_size=5)
        self.pool = nn.MaxPool1d(2)
        self.conv2 = nn.Conv1d(16, 32, kernel_size=5)
        self.fc1 = nn.Linear(32 * 796, 64)
        self.fc2 = nn.Linear(64, 1)

    def forward(self, x):
        x = self.pool(F.relu(self.conv1(x)))  # (N, 16, ~1596)
        x = self.pool(F.relu(self.conv2(x)))  # (N, 32, ~796)
        x = x.view(x.size(0), -1)
        x = F.relu(self.fc1(x))
        return torch.sigmoid(self.fc2(x))     # Sigmoid for binary output


##Train the Model

model = ExoCNN()
criterion = nn.BCELoss()
optimizer = torch.optim.Adam(model.parameters(), lr=0.001)

# Training loop
for epoch in range(5):  # Increase to ~20 for better accuracy
    model.train()
    running_loss = 0.0
    for inputs, labels in train_loader:
        outputs = model(inputs).squeeze()
        loss = criterion(outputs, labels)

        optimizer.zero_grad()
        loss.backward()
```

```
        optimizer.step()

        running_loss += loss.item()
    print(f"Epoch {epoch+1} Loss: {running_loss/len(train_loader):.4f}")

#Eval Accuracy on test

model.eval()
correct = 0
total = 0

with torch.no_grad():
    for inputs, labels in test_loader:
        outputs = model(inputs).squeeze()
        preds = (outputs > 0.5).float()
        correct += (preds == labels).sum().item()
        total += labels.size(0)

print(f"Test Accuracy: {100 * correct / total:.2f}%")

## Task 4: Evaluating and Testing
import torch
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, confusion_matrix
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np

# Make sure your model is in evaluation mode
model.eval()

# Disable gradient calculation for inference
with torch.no_grad():
    # Predict on test data
    y_pred_probs = model(X_test_tensor).squeeze()  # output probabilities
    y_pred = (y_pred_probs >= 0.5).int().numpy()   # threshold at 0.5 to get labels

y_true = y_test_tensor.int().numpy()

# Calculate metrics
accuracy = accuracy_score(y_true, y_pred)
precision = precision_score(y_true, y_pred)
recall = recall_score(y_true, y_pred)
f1 = f1_score(y_true, y_pred)

print(f"Test Accuracy: {accuracy:.4f}")
print(f"Precision: {precision:.4f}")
print(f"Recall: {recall:.4f}")
print(f"F1-score: {f1:.4f}")

# Confusion Matrix
cm = confusion_matrix(y_true, y_pred)
sns.heatmap(cm, annot=True, fmt='d', cmap='Blues')
plt.xlabel('Predicted')
plt.ylabel('Actual')
plt.title('Confusion Matrix')
plt.show()
```

Outputs:

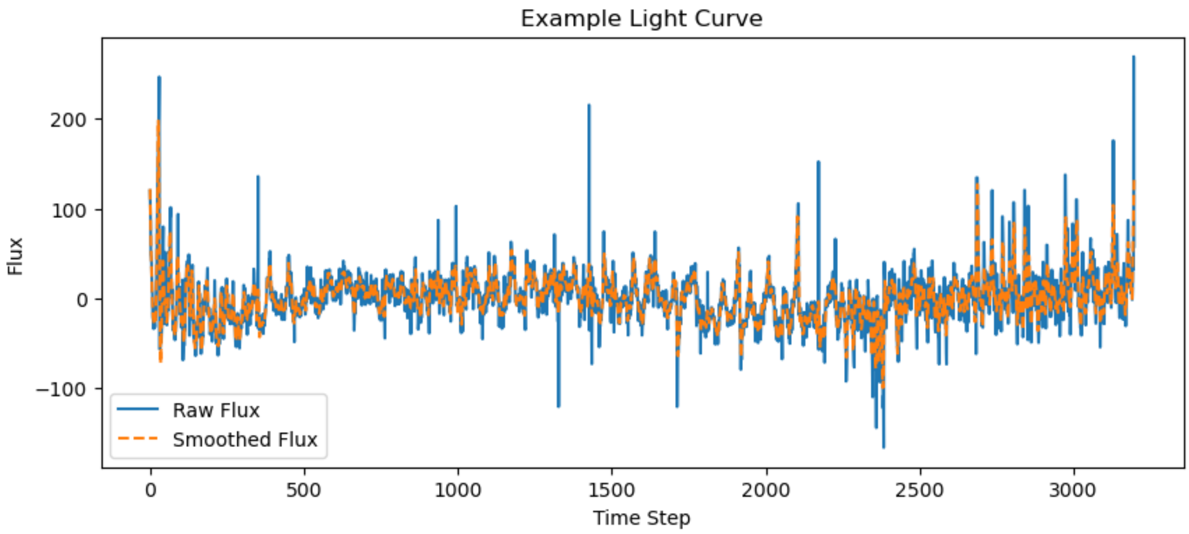


Figure 1: *Light curve showing raw flux (blue) and smoothed flux (orange) over time. Smoothing reduces noise to highlight underlying brightness variations.*

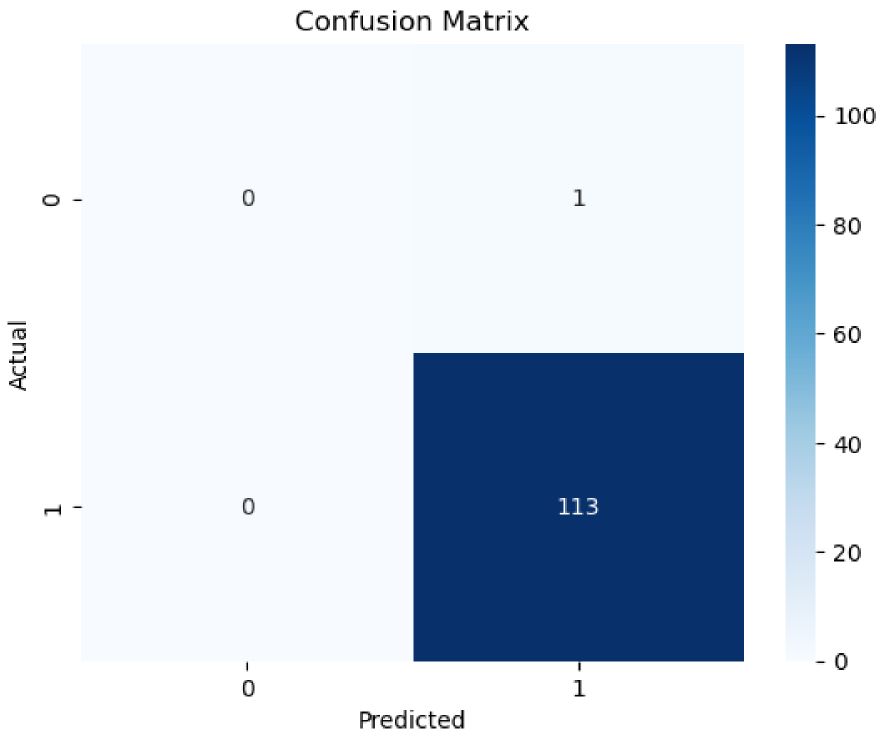


Figure 2: *Confusion matrix showing model predictions. The model correctly predicted 113 positive cases but misclassified 1 negative case as positive. No true negatives or false negatives observed.*

Part 2: Feedback on AUT AI Aspirations: Learning and Teaching

1. **Why did I choose this aspiration**

   My academic experience and professional development are directly impacted by the desire of learning and teach, which is why I selected this aspiration. As a student navigating a frequently changing educational environment, I understand how important artificial intelligence (AI) can be in enhancing traditional methods of teaching. AI's capacity to streamline difficult subjects, offer individualised learning assistance, and facilitate 24/7 accessibility is highly compatible with my objective of becoming an expert in data science and astronomy, fields where a thorough conceptual knowledge is important(OpenAI, 2023).

   Many classrooms contain more than 50-60 students, making it challenging for every student to have their questions answered in the restricted amount of class time, even though teachers can provide explanations and answer queries. Professors may lack the time and resources necessary to adequately address every single question. By giving students instant, individualised support and enabling them to freely ask questions and go over explanations as frequently as necessary, AI can help close this gap(OpenAI, 2023; Holmes et al., 2019).

   AUT's goal of promoting creative, inclusive, and successful education is supported by the advancement of AI integration in learning and teaching(AUT, 2025). In addition to improving access to education through individualised support that adjusts to a range of learning demands, society gains from graduates who are more equipped for the AI-driven workforce.

2. **Feedback on Aspiration**

   AUT's "Learning and Teaching" focused area outlines an ambitious and progressive vision for using AI in curriculum development, teaching, evaluation, and individualised learning(AUT, 2025). Though there are some challenges with implementation, inclusion, and practicality, overall, it is straightforward and ambitious. There are several holes and restrictions in the plan that need more consideration, even while the intended future state correctly emphasises curriculum-wide AI integration and the growth of AI literacy among staff and students.

   One of the most commendable elements is the aim to provide AI-powered personalised learning. This has real potential to bridge learning gaps for students who hesitate to ask questions in large classes, whether due to social anxiety, cultural norms, or fear of judgment. Even though professors can help clarify doubts, it is not feasible for them to attend to every individual in classes of 50–60 students. Time constraints and workload make it hard to ensure that no student is left behind. In such cases, AI tutors or chatbots can provide non-judgmental, instant, and repeated help, enabling students to progress at their own pace(OpenAI, 2023).

The focus on discipline-specific AI knowledge for both staff and students is another advantage. This is in line with the fact that different professions use different AI tools and effects; for example, students studying data science might use generative models like GitHub Copilot (GitHub, n.d.), while nursing students might use diagnostic AI tools. The aim, however, lacks precise procedures for large-scale implementation, and it runs the risk of overburdening already overworked academic staff.

A further possible issue is the excessive dependence on AI tools without enough critical literacy instruction. AI should not be a crutch, but a co-pilot. If students just use AI to finish assignments or decipher content, they can miss out on chances to practice their research, writing, and thinking abilities. Despite mentioning "minimum standards of AI knowledge," the ambition ought to more clearly encourage AI literacy, which encompasses ethical awareness, critical thinking, and decision-making boundaries (Holmes et al., 2019).

Although it makes sense technologically to include AI into already-existing platforms like Canvas or Panopto, this raises concerns about tool stability, staff training, and cost. Particularly when handling sensitive student data, not all AI systems are developed or safe enough for use in academic settings. Clearer procedures for data protection and evaluating AI tools would be beneficial to the plan, particularly when working with foreign providers that might not follow local privacy laws or cultural norms.

3. A Possible Implementation Plan

Step 1: Introduce AI Orientation Modules Across All Papers

- **Action**: Create short, engaging video modules introducing AI tools (e.g., ChatGPT, Grammarly, GitHub Copilot) relevant to the course.
- **Timing**: Shown at the start of each semester or paper.
- **Goal**: Build baseline AI literacy, reduce hesitation in using tools, and clarify ethical use.
- **Stakeholders**: Course coordinators, lecturers, student support teams.
- **Feasibility**: Highly feasible; can be templated and reused across papers.

Step 2: Establish Student Support for AI Use

- **Action**: Develop peer-led AI support groups or discussion forums (via Discord or Canvas) where students share best practices, issues, or insights.

- **Stakeholders**: Students, learning advisors.
- **Resources Needed**: Platform integration (e.g., Canvas), moderation, AI-literate student mentors.

Step 3: Staff Upskilling and AI Tool Access

- **Action**: Implement a professional development programme for lecturers that includes hands-on training on AI tools, ethical considerations, and discipline-specific applications.
- **Tools**: ChatGPT, GitHub Copilot, Grammarly, AI-enhanced Canvas/Panopto features.
- **Stakeholders**: AUT Centre for Learning and Teaching, IT services.
- **Feasibility**: Moderate—requires dedicated resources and scheduling but critical for long-term cultural shift.

Step 4: Provide Equitable Tech Access and Guidelines

- **Action**: Ensure all students have access to the AI tools via university licenses and provide culturally responsive usage guides (e.g., for te reo Māori).
- **Stakeholders**: IT department, Māori student services.
- **Resources**: University-wide tool licensing and culturally safe AI use frameworks.

Step 5: Evaluate Progress and Impact

- **Action**: Collect ongoing feedback via short surveys from students and staff after each semester, and track engagement and confidence levels.
- **Metrics**:
  - Student comfort/confidence with AI tools
  - Feedback on the usefulness of modules
  - Staff engagement with AI training
  - Academic performance trends
- **Stakeholders**: Learning analytics team, academic quality committee.

Reference:

Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., Allam, S. S., Allende Prieto, C., An, D., Anderson, K. S. J., Anderson, S. F., Annis, J., Bahcall, N. A., Barkhouser, R. H., Bastian, S., Berlind, A., Blanton, M. R., Bochanski, J. J., Brewington, H. J., Brinchmann, J., Brinkmann, J., Brunner, R. J., & Budavári, T. (2009). The seventh data release of the Sloan Digital Sky Survey. *The Astrophysical Journal Supplement Series, 182*(2), 543–558. https://doi.org/10.1088/0067-0049/182/2/543

Auckland University of Technology. (2025, April). *A2* [Artificial Intelligence]. Canvas. https://canvas.aut.ac.nz/courses/19452/assignments/156803

Borucki, W. J., Koch, D., Basri, G., Batalha, N., Brown, T., Caldwell, D., Caldwell, J., Christensen-Dalsgaard, J., Cochran, W. D., DeVore, E., & Dunham, E. W. (2010). Kepler planet-detection mission: Introduction and first results. *Science*, *327*(5968), 977–980. https://doi.org/10.1126/science.1185402

GitHub. (n.d.). *GitHub Copilot*. https://github.com/features/copilot

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign. https://curriculumredesign.org/wp-content/uploads/AI-in-Education-CCR-2019.pdf

Lintott, C. J., Schawinski, K., Slosar, A., Land, K., Bamford, S., Thomas, D., Raddick, M. J., Nichol, R. C., Szalay, A., Andreescu, D., & Vandenberg, J. (2008). Galaxy Zoo: Morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. *Monthly Notices of the Royal Astronomical Society, 389*(3), 1179–1189. https://doi.org/10.1111/j.1365-2966.2008.13689.x

OpenAI. (2023). *ChatGPT* (Mar 14 version) [Large language model]. https://chat.openai.com/

Shallue, C. J., & Vanderburg, A. (2018). Identifying exoplanets with deep learning: A five-planet resonant chain around Kepler-80 and an eighth planet around Kepler-90. *The Astronomical Journal, 155*(2), 94. https://doi.org/10.3847/1538-3881/aa9e09

Wright, J. T., Marcy, G. W., Howard, A. W., Johnson, J. A., Morton, T. D., Fischer, D. A., Reffert, S., Foreman-Mackey, D., Isaacson, H., Valenti, J. A., Anderson, J., Piskunov, N., Lind, K., Brewer, J. M., Petigura, E. A., Latham, D. W., Cargile, P., Hebb, L., Weiss, L. M., … Fabrycky, D. C. (2015). The search for exoplanets: Progress and prospects. *Annual Review of Astronomy and Astrophysics*, *53*, 409–447. https://doi.org/10.1146/annurev-astro-082214-122246