



# Détection Automatisée d'Exoplanètes par Analyse de Courbe de Lumière

Projet Pluridisciplinaire en Équipe - ING4 Groupe 553

ECE Paris

Simon GALLAIS, Charles De BLAUWE, Mathis LEITAO, Kamil BENJELLOUN,  
Oscar SCHWARTZ, Médéric ROLLAND

# Le Contexte "Big Data" Astronomique

---

Un volume gigantesque et encore sous-exploité

Plus de 5000 exoplanètes confirmées mais des millions de courbes de lumière restent à analyser

Des missions NASA qui changent tout

Kepler et TESS observent des centaines de milliers d'étoiles et produisent des téraoctets de données photométriques.

L'analyse humaine a atteint ses limites

Face à ce "Big Data" astronomique, l'analyse manuelle est devenue une tâche impossible, soulignant le besoin crucial d'automatisation.



# Problématique & Opportunité



## La Problématique Clé

Comment automatiser efficacement la détection et la caractérisation des exoplanètes à partir de signaux souvent très bruités ?

- Identifier les transits planétaires.
- Gérer la complexité des données brutes.

## L'Opportunité

Notre approche se concentre sur le développement d'un pipeline automatisé, explicable et performant, capable de fonctionner sur des machines standards.

## Le Verrou Technique

Distinguer un véritable transit planétaire d'un "faux positif" est un défi majeur. Les causes de confusion incluent :

- Étoiles binaires à éclipses.
- Bruit instrumental.
- Variations stellaires.

# État de l'Art & Choix Technique (Le Pivot)

## Deep Learning (CNN)

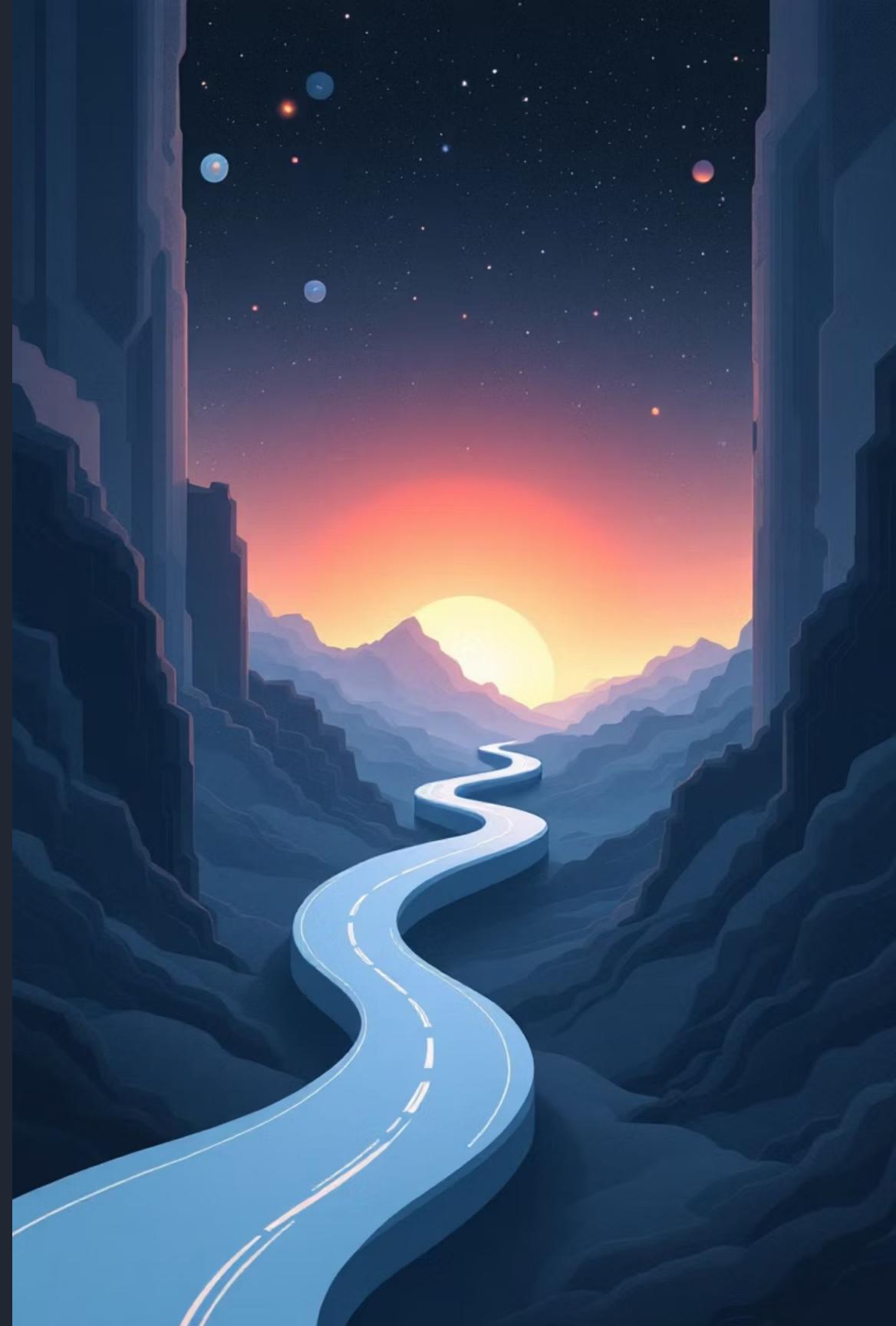
Analyse comparative des approches : Puissant mais lourd et coûteux en calcul, souvent perçu comme une "boîte noire" difficilement interprétable.

## Machine Learning Classique

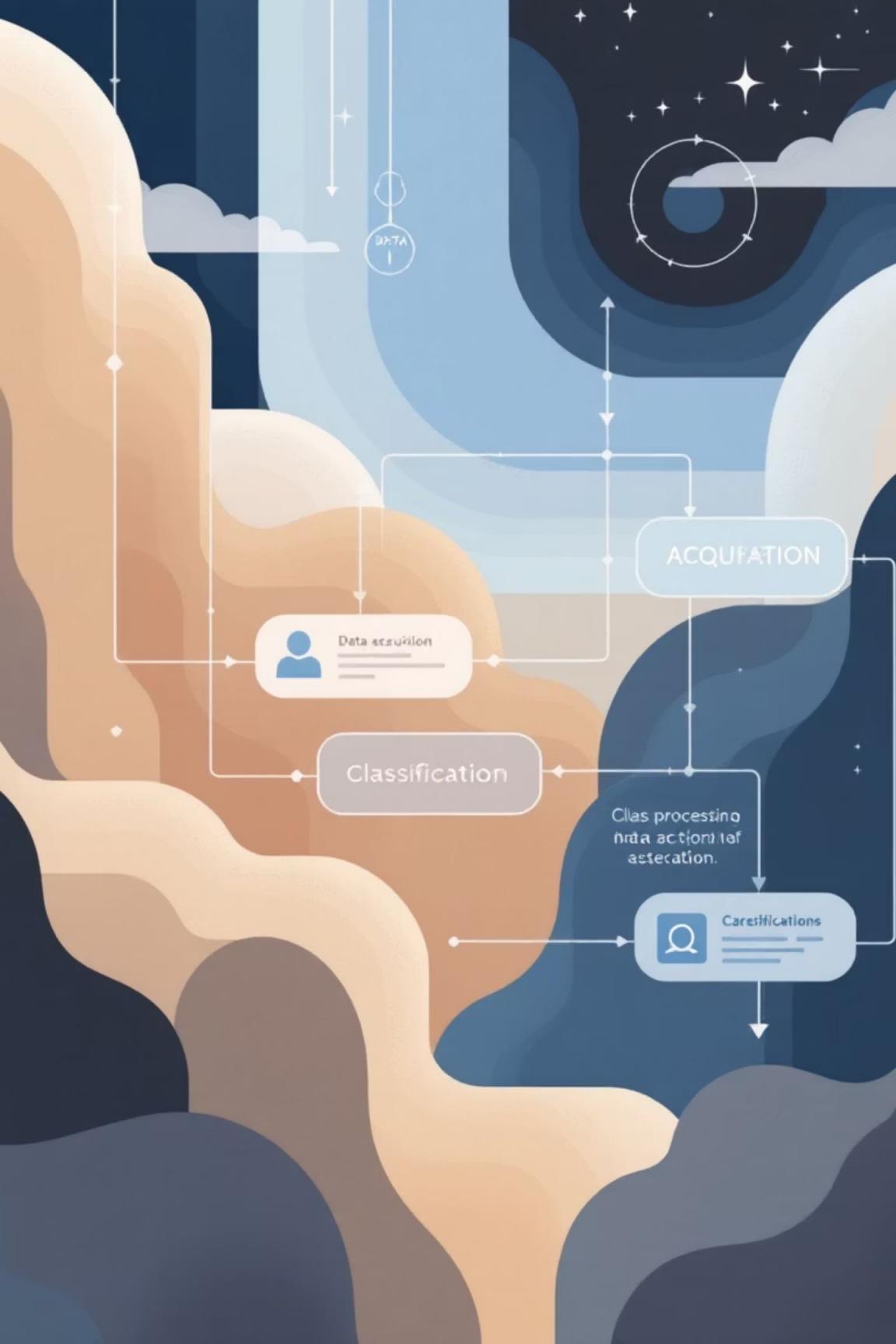
Moins gourmand en ressources, offrant une meilleure explicabilité. Notre choix s'est porté sur cette voie pour un compromis optimal.

Notre choix s'est porté sur le **Machine Learning Classique (XGBoost)** couplé à une ingénierie des fonctionnalités (Feature Engineering) rigoureuse.

Cette approche offre une précision similaire (~94% AUC) tout en garantissant une explicabilité supérieure et un coût de calcul minimal.



# Analyse Fonctionnelle : Notre Solution



## 1. Acquisition

Utilisation des APIs NASA MAST et de la bibliothèque Lightkurve pour collecter les données photométriques brutes.

## 2. Prétraitement

Application de techniques de "detrending" (nettoyage) et de "folding" (repliement) pour isoler le signal de transit.

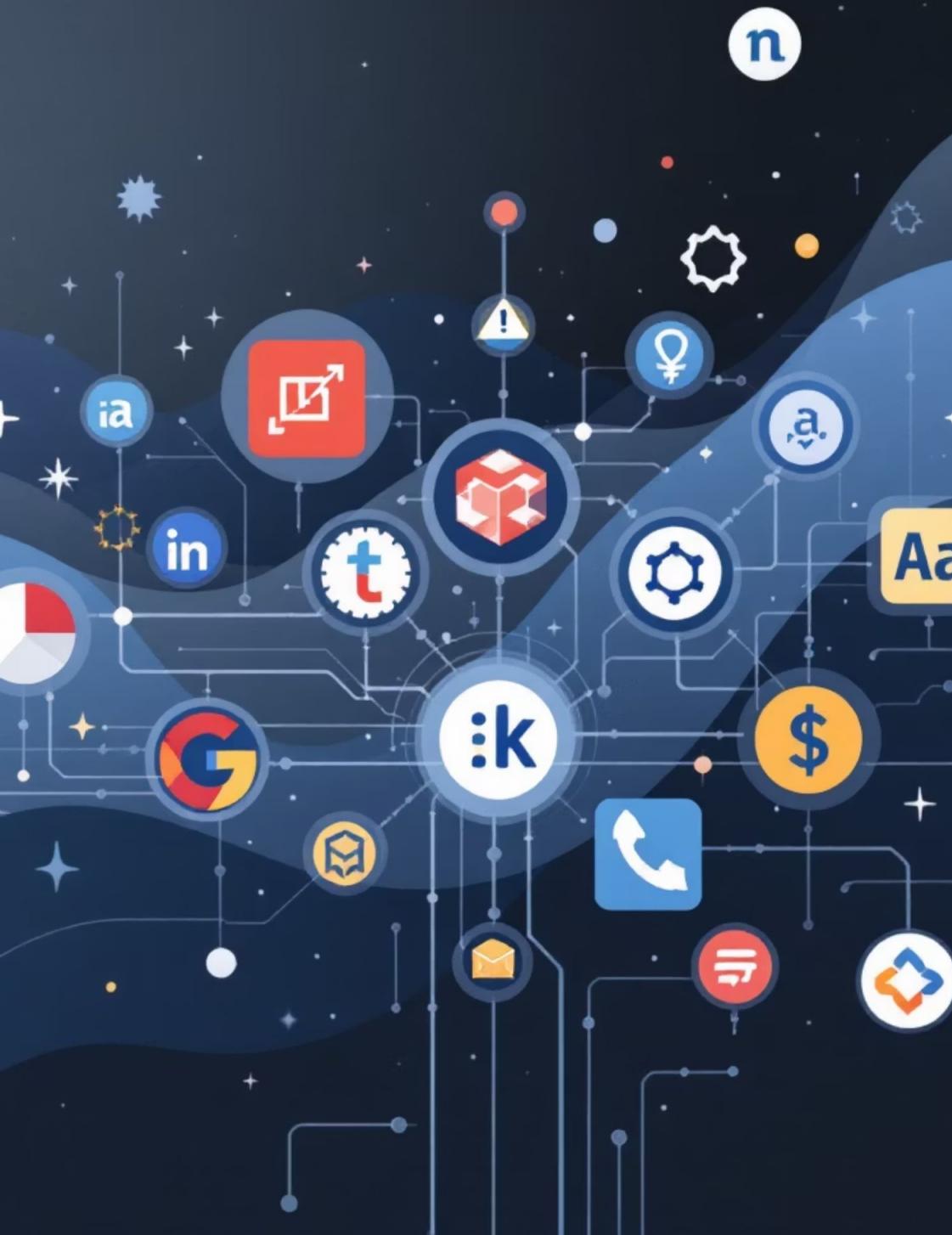
## 3. Analyse Intelligente

Transformation de la courbe en données mathématiques (Feature Engineering avec *TSFRESH*) puis classification automatique par notre modèle *XGBoost*.

## 4. Visualisation

Affichage des résultats sur une interface Web interactive (React) pour permettre la validation finale par un astronome.

# Architecture Technique & Outils



## Stack Backend & Astronomie

Un socle Python robuste pour le traitement de données massives. Nous utilisons les bibliothèques spécialisées Lightkurve et Astropy pour communiquer avec les serveurs de la NASA et nettoyer les signaux stellaires.

## Moteur d'Intelligence Artificielle

Le cœur de notre innovation. L'outil TSFRESH extrait les caractéristiques mathématiques du signal, qui sont ensuite analysées x XGBoost.

## Interface de Visualisation

Une application Web interactive développée en React.js. Elle permet à l'astronome de visualiser les courbes de lumière, de voir le verdict de l'IA en temps réel et de valider manuellement les candidats exoplanètes.

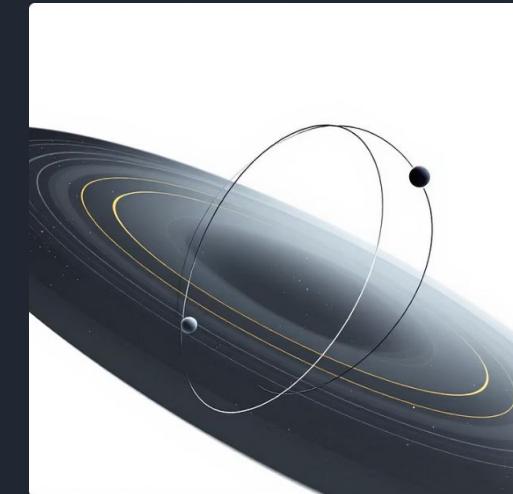
# Focus sur le Prétraitement

Le prétraitement est une étape cruciale pour révéler le signal de transit planétaire, souvent masqué par le bruit stellaire.



## Courbe Brute (Bruitée)

Le signal de transit est noyé dans le bruit intrinsèque de l'étoile et les artefacts instrumentaux, rendant la détection difficile.



## Courbe Nettoyée (Clair)

Après "blanchiment" du signal, le transit en "U" apparaît clairement, facilitant la détection automatique.

Cette étape essentielle permet d'optimiser la qualité des données pour le modèle de classification.

# Planning & Avancement

Phase 1 (Sept-Nov)  
État de l'art et choix de  
l'architecture. **Terminé.**

Phase 3 (Fév-Mars)  
Développement du modèle ML et  
de l'interface Web. **À venir.**

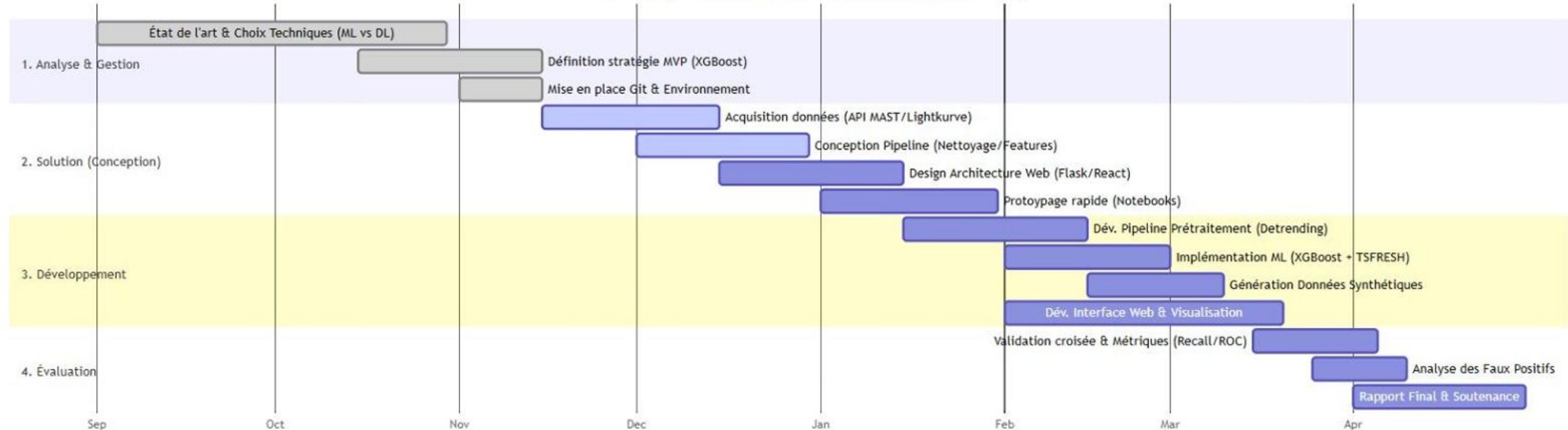


Phase 2 (Déc-Jan)  
Conception du pipeline et  
prototypage sur Notebooks.  
**En cours.**

Phase 4 (Avril)  
Validation des résultats et  
réécriture du rapport final. **À venir.**

# Planning

Planning : Détection d'Exoplanètes (Sep - Avr)

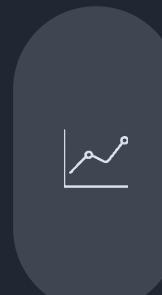


# Matrice RACI

Tâches	R	A	C	I
Pipeline de traitement	Simon	Oscar, Mathis	Tous	Tous
Modèle ML	Oscar	Mathis, Simon	Tous	Tous
Interface web	Médéric	Charles	Tous	Tous
Sécurité	Kamil	Kamil	Tous	Tous
Architecture code	Charles	Médéric	Tous	Tous
Documentation	Mathis	Tous	Tous	Tous

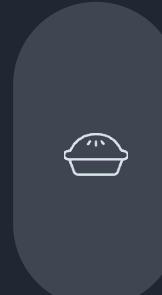
# Vision Finale : Interface Utilisateur Intuitive

Notre objectif est de rendre l'analyse d'exoplanètes accessible même aux non-experts, grâce à une interface claire et interactive.



## Visualisation de la Courbe de Lumière

Représentation graphique interactive des données brutes et prétraitées, permettant une inspection visuelle immédiate.



## Probabilité de Planète (Score de Confiance)

Affichage clair du score de confiance attribué par le modèle, indiquant la probabilité qu'il s'agisse d'une exoplanète.



# Conclusion & Prochaines Étapes

## Stratégie Validée

Notre approche par Machine Learning Classique est robuste et nos outils sont en place pour la prochaine phase de développement.

## Prochaine Étape Immédiate

Implémentation du script de "Detrending" et premier entraînement du modèle XGBoost avec les données prétraitées.

Nous sommes déterminés à créer une solution pragmatique qui démocratise la recherche d'exoplanètes et ouvre de nouvelles perspectives pour les astronomes amateurs et professionnels.

