



# Théo Joly

ÉTUDIANT M2 DATA SCIENCE & BIG DATA  
- RECHERCHE STAGE DÈS AVRIL 2026 -



## FORMATION

sept. 2024 - juil. 2026

**MASTER EN GESTION DES DONNÉES ET EXTRACTION DE CONNAISSANCES À LARGE ÉCHELLE**

Université Paris-Saclay, Versailles

2024

**LICENCE EN INFORMATIQUE**

Université Paris-Saclay, Versailles

2021

**BACCALAURÉAT GÉNÉRAL - MATHÉMATIQUES, NUMÉRIQUE ET SCIENCES INFORMATIQUES**

Lycée Jules Vernes, Limours

## PROFIL



theo.joly@theojoly.fr



Île-de-France



Permis B - Véhiculé

## OBJECTIF

Passionné par l'ia et la data, je cherche à faire de l'un de ces deux centres d'intérêts mon métier.

## COMPÉTENCES

**Data Science & IA** : Python (Expert), R, Machine Learning, LLMs (GenAI), Graphes (NetworkX), Recherche Opérationnelle.

**Big Data & Data Engineering** : SQL Avancé (PL/SQL), NoSQL, Hadoop Ecosystem, Elastic Stack (ELK), Web Sémantique (RDF/SPARQL), ETL (Talend).

**Développement & Outils** : Rust, Java, Docker, Git/GitHub, Linux (Bash).

**Langues** : Français (Langue maternelle), Anglais (Technique).

## CENTRES D'INTÉRÊT

- Culture Japonaise** : Apprentissage de la langue, mangas et projet de voyage.
- Aménagement Paysager** : Rénovation d'espaces verts et plantation d'arbres.
- Cuisine** : Gastronomie (source d'inspiration de mon projet SQL).

## EXPÉRIENCE PROFESSIONNELLE

**MARAÎCHER**

Ferme Lafouasse | juil. 2025

Travail en équipe, adaptation à des conditions difficiles (canicule, tâches physiques intenses), gestion de l'imprévu et découverte des réalités du monde agricole.

## PROJETS ACADÉMIQUES & RÉALISATIONS

**Exploration Hybride de Data Lakes (M2 DataScale)** : Conception d'un moteur de recherche sémantique innovant combinant LLMs (GenAI) et Graphes de Connaissances. L'outil permet d'identifier automatiquement les tables pertinentes (Union/Joinure) au sein de lacs de données massifs pour enrichir les analyses.

**Application Client pour l'INRAE (Projet Livré)** : Développement et déploiement d'une application Windows (Python) générant des séquences aléatoires sous contraintes d'unicité stricte. Outil utilisé en production pour la réalisation de la série de podcasts « Les Voix de la Recherche ».

**Data Engineering & Architecture SQL** : Double compétence en gestion de la donnée : Création d'un **pipeline ETL** complet sous Talend (audit qualité, nettoyage, consolidation) et architecture d'une base de données complexe en **PL/SQL** (Triggers, Procédures stockées) pour un réseau social.

**Algorithmique sur Graphes & Recherche Opérationnelle (TER)** : Étude avancée sur la structure des réseaux : Comparaison d'heuristiques (**Tabu Search**, **Recuit Simulé**) pour l'optimisation de l'isolation virale dans des graphes orientés. Réimplémentation critique de l'algorithme **PageRank** sur données massives.