

# Curriculum Vitae

## Antoine Lucas

Data Scientist & ML Engineer | Pharma · Cosmetics · Biotech · Paris, France

antoine.lucas.fra@gmail.com | linkedin.com/in/antoinelucasdata | github.com/antoinelucasfra

---

Data Scientist & ML Engineer with 5+ years building production-grade data solutions in regulated industries — pharma, cosmetics, and biotech. Expertise spans statistical modelling, Design of Experiments, applied ML (computer vision, NLP), and multi-omics analysis. Proven track record translating complex scientific questions into end-to-end systems: from study design and model development through to validated deployment and stakeholder training.

## Experience

Data Scientist & ML Engineer *Chanel Parfums Beauté R&D · via Astek / IT&M Stats*

Sep 2025 – present

- Designed and deployed computer vision models (texture classification, colour analysis) and NLP pipelines (olfactory descriptor mining from scientific literature) for fragrance and cosmetics formulation R&D
- Modelled sensory and physico-chemical properties with PLS regression, Gaussian processes, and Bayesian optimisation to accelerate ingredient screening cycles
- Built DoE workflows (mixture designs, response-surface methods) cutting required experiments per formulation cycle by 30–40%
- Engineered MLOps infrastructure: model versioning with DVC, drift monitoring, fully reproducible pipelines (Docker + GitHub Actions) in a GxP-adjacent environment
- Delivered R Shiny dashboards covering sensory scoring, spectral data visualisation, and interactive experiment planning for R&D scientists

Data Scientist — Internal & Consulting *Astek / IT&M Stats*

May – Sep 2025

- Developed reusable internal R packages and Quarto automated reporting templates, cutting analyst turnaround time by ~25% across consulting engagements
- Led internal upskilling programme on tidyverse, causal inference, and reproducible research for 8 statistical consultants
- Scoped and prototyped ML solutions for pre-sales client engagements in pharma and agri-food verticals

Data Scientist — Multi-omics *Abolis · Microbiome Studio · via Astek / IT&M Stats*

Dec 2024 – May 2025

- Built end-to-end multi-omics pipelines (metagenomics, metatranscriptomics, metabolomics) from raw sequencing reads to biological interpretation across 4 agro-industry studies
- Applied PERMANOVA, sparse PLS-DA, and co-occurrence network analysis to characterise 100+ microbial interactions and identify key functional drivers
- Packaged analyses as reproducible Quarto reports; integrated results into the client's production SaaS platform via a REST API

Data Scientist / Scrum Master *Sanofi R&D · via Astek / IT&M Stats*

Jul 2023 – Nov 2024

- Architected and delivered a full R Shiny platform (golem) managing pilot manufacturing chains for 100+ GxP clinical projects — 50+ daily active users across 3 departments
- Designed the data model, module architecture, and automated validation report generation to satisfy GxP audit-trail requirements (IQ/OQ/PQ on Posit Connect)
- Owned the full delivery lifecycle: user-story workshops → agile sprints → UAT → IQ/OQ documentation → production deployment
- Led a 6-person cross-functional team (data scientists, developers, QA); managed sprint planning and alignment with 5+ senior sponsors

Biostatistician L'Oréal R&D · via Astek / IT&M Stats

Oct 2021 – Jul 2023

- Delivered 100+ statistical analyses on ingredient efficacy (hydration, anti-ageing biomarkers) and molecular safety (QSAR, genotoxicity) across cosmetic and dermocosmetic pipelines
- Designed and analysed clinical and semi-clinical studies from protocol writing through regulatory-grade reports — consistent 2-week delivery cycle
- Built internal R tooling (survival analysis, mixed-effects models, automated Rmarkdown reporting) adopted as standard across the biostatistics team
- Mentored 2 junior analysts and 1 intern; co-authored methodology guides on non-inferiority testing and adaptive designs

## **Education**

Diplôme d'Ingénieur — Statistiques appliquées aux Sciences de la Vie Agrocampus Ouest

2021

Master — Mathématiques Appliquées · Statistiques Université Paris-Saclay

2021

## **Technical Skills**

Languages: R · Python · SQL · Bash

ML & Stats: scikit-learn · tidymodels · XGBoost · Bayesian optimisation · DoE · PLS · PERMANOVA · survival analysis

Apps & Viz: R Shiny (golem) · ggplot2 · Plotly · Quarto · Power BI

CV / NLP: PyTorch · OpenCV · Hugging Face · sentence-transformers

MLOps: Docker · GitHub Actions · Posit Connect · DVC · renv · uv

Domains: GxP · IQ/OQ/PQ · Clinical biostatistics · Multi-omics · Cosmetic R&D

## **Certifications**

Scrum Master — Agilbee (2023)

## **Languages**

French (native) · English (fluent, professional)