

Formation des formateurs en Sciences des données et IA 2025-2026

OBJECTIFS

Cette formation est ouverte à tout chercheur, enseignant-chercheur, doctorant et post-doctorant de l'Université PSL qui s'engage à créer au moins un cours reprenant les enseignements reçus :

- Proposer des modules adaptés aux différentes disciplines liées aux sciences des données et à l'intelligence artificielle.
- Les **participants peuvent choisir les modules** les plus adaptés à leurs besoins et à ceux de leur discipline.
- Chaque séance de 3H est qualifiée par niveau : débutant, intermédiaire, avancé
- Renforcement de la capacité des formateurs à dispenser une formation de qualité dans leur domaine respectif
- **Chaque participant bénéficiera d'un soutien à la création de nouveaux modules de formation** au sein de PSL à la suite du cours de formation

PROGRAMME

Cours de base

- Optimisation et apprentissage (18H)

Cours avancés

- SQL in practice, NoSQL and vector databases (9H)
- Web scraping (12H)
- Natural Language Processing (21H)
- Apprentissage par renforcement (18H)

[Retrouver le détail du programme](#)



Date et horaires :

[Retrouver le détail sur la page web du programme DATA](#)

Durée Totale de l'ensemble

des modules: 126h

Lieu : Estrapade, 16B rue de l'Estrapade, 75005 Paris

Formateurs : Muni Sreenivas Pydi, Nicolas Schreuder, Kimia Nadjahi, Bruno Chaves, Ana Basic, Aina Gari-Soler

Public concerné

chercheur, enseignant-chercheur, doctorant et post-doctorant de l'Université PSL

POUR VOUS INSCRIRE

1 créez votre compte - <https://psl-ecoleinterne-inscription.helvetius.net/sinscrire.php>

2 connectez-vous au portail de l'Ecole Interne PSL - <https://psl-ecoleinterne.monportail.psl.eu/pegasus/index.php>

3 faites votre demande en « ça m'intéresse » pour être recontacté à l'ouverture de la prochaine session ou inscrivez-vous à la date indiquée.

IA & Data Training for Academics 2025-2026

OBJECTIVE

This course is open to all researchers, teacher-researchers, doctoral students and post-doctoral students at PSL University who undertaketo create at least one course incorporating the lessons received :

- Main objective: to offer modules adapted to the various disciplines linked to data sciences and artificial intelligence; to strengthen the capacity of trainers to provide high-quality training in their respective fields.
- **Participants can choose the modules** best suited to their needs and those of their discipline.
- Each 3H session is qualified by level: beginner, intermediate,advanced
- **Researchers and teacher-researchers** who commit **to creating a new training module** within PSL at the end of their training will be eligible for **support**. Contact justine.petit@psl.eu@psl.eu for more information.

PROGRAM

Basic course

- Optimization and machine learning (18H)

Advanced courses

- SQL in practice, NoSQL and vector databases (9H)
- Web scraping (12H)
- Natural Language Processing (21H)
- Reinforcement learning (18H)

[Find the details of the program](#)



Date and times :

Find the details on the [DATA webpage](#)

Total duration of all modules: 126h

Place: Estrapade 16B rue de l'Estrapade, 75005 Paris

Trainers: Muni Sreenivas Pydi, Nicolas Schreuder, Kimia Nadjahi, Bruno Chaves, Ana Basic, Aina Gari-Soler

Target audience :

researchers, teacher-researchers, doctoral students and post-doctoral students at PSLUniversity

TO REGISTER

1 Create your account - <https://psl-ecoleinterne-inscription.helvetius.net/sinscrire.php>

2 log in to the PSL Internal School portal -

<https://psl-ecoleinterne.monportail.psl.eu/pegasus/index.php>

3Apply in "I'm interested" to be contacted again at the opening of the next session or register on the date indicated.

Basic courses - from 5:00 to 8:00 PM, Estrapade

— Optimization & machine learning (18h) with Kimia Nadjahi

6 lectures: from 9/02/2026 to 16/03/2026

- o Convexity (*Beginner*) 9/02/2026 & 16/02/2026
- o Standard algorithms for constrained or unconstrained problems (*Advanced*) 23/02/2026 & 02/03/2026
- o Large scale optimization (*Advanced*) 9/03/2026 & 16/03/2026

Advanced courses - from 5 to 8 pm, Estrapade

— Web scraping (12h) with Bruno Chaves

4 lectures: from 12/03/2026 and 2/04/2026

This lecture only requires basic knowledge of Python & SQL.

- o The basics 1 12/03/2026
- o The basics 2 19/03/2026
- o Advanced Topics and legal aspects 1 26/03/2026
- o Advanced Topics and legal aspects 2 2/04/2026

— SQL in practice, NoSQL and Vector (9h) with Bruno Chaves

3 lectures: from 09/04/2026 to 23/04/2026

This lecture only requires basic knowledge of Python & SQL.

- o SQL in practice 09/04/2026
- o NoSQL databases 16/04/2026
- o Vector databases 23/04/2026

— Natural Language Processing (21h) with Aina Gari-Soler

7 lectures: from 23/03/2026 to 11/05/2026

- o Background (refresh): basic concepts from Machine Learning and Neural Networks 23/03/2026
- o Word vectors: Latent Semantic Analysis (LSA), word2vec and GloVe 30/03/2026
- o Language Models: from N-gram models to Recurrent Neural Networks 13/04/2026
- o Sequence-to-sequence modelling and Machine Translation (MT) 20/04/2026
- o Attention, self-attention and Transformer-based language models 27/04/2026
- o Pretraining and Finetuning large language models 04/05/2026
- o Understanding ChatGPT 11/05/2026

— Reinforcement learning (18h) with Ana Basic

6 lectures: from 21/05/2026 to 25/06/2026

- o Introduction to reinforcement learning 21/05/2026
- o Markov decision processes and dynamic programming 28/05/2026
- o Bandit algorithms 04/06/2026
- o Reinforcement learning: stochastic approximation, algorithms, actor-critic, model-based 11/06/2026 & 18/06/2026
- o Reinforcement with approximation 25/06/2026

SYLLABUS - DETAILED DESCRIPTION OF EACH CLASS

Optimization and machine learning

18 hours, with Kimia Nadjahi

The objective of this module is to give an overview of the main optimization algorithms, with a focus on their application to statistical learning.

— **Lectures 1 (beginner) and 2 (beginner): convexity**

- Introduction and convexity
- Convex optimization
- Duality

— **Lectures 3 (advanced) and 4 (advanced): standard algorithms for constrained or unconstrained problems**

- Unconstrained optimization: gradient descent, Newton's method
- Constrained optimization: interior-point methods

— **Lectures 5 (advanced) and 6 (advanced): large scale optimization**

- Advanced first-order methods
- Stochastic algorithms

Web scraping

12 hours, with Bruno Chaves

— **Lectures 1 and 2: the basics**

- Python, HTML and CSS: the basics
- XPATH and CSS selectors
- Scraping with http queries
- Scraping with browser emulation
- Practice

— **Lectures 3 and 4: advanced topics and legal aspects**

- Public and private APIs (Application Programming Interface)
- Hidden APIs
- Identifying/bypassing blocking points
- Recurring and/or large-scale scraping
- The legal context of research data: discussion with a data officer
- Practice

SQL in practice, NoSQL and vector databases

9 hours, with Bruno Chaves

— **Lecture 1: SQL in practice**

- Understanding SQL in practice: SQL structures and research questions
- SQL in the era of ChatGPT
- Hands-on lab with SQLite (using Python)

— **Lecture 2: NoSQL databases**

- Overview of NoSQL databases
- Types of NoSQL databases and use cases
- Hands-on lab with MongoDB (using Python)

— **Lecture 3: Vector databases**

- Overview of vector databases and their use cases
- Basics of vector similarity search
- Hands-on lab: Implementing a vector database for a machine learning application

Natural language processing

21 hours with Aina Gari-Soler

Natural Language Processing (NLP) is a subfield of artificial intelligence, which studies principles and algorithms that allow an artificial system to understand and process human natural language. In this course, students will learn about the latest advancements in using deep learning for NLP. They will gain both conceptual understanding and practical skills to build and use their own models for various purposes.

— **Lecture 1: Background (refresh)**

- Basic concepts from machine learning and neural networks
- Exercise: Basics of Scikit-Learn and PyTorch

— **Lecture 2: Word vectors**

- Latent Semantic Analysis (LSA)
- word2vec and GloVe
- Exercise: GloVe implementation

— **Lecture 3: language models: from N-gram models to Recurrent Neural Networks**

- Exercise: LSTM-based Language Models

— **Lecture 4: sequence-to-sequence modelling and Machine Translation (MT)**

- Exercise: English-French Translation

— **Lecture 5:**

- Attention
- Self-attention
- Transformer-based language models
- Exercise: natural-language generation with transformers

— **Lecture 6: pretraining and finetuning large language models**

- Exercise: tutorial on Hugging Face

— **Lecture 7: Understanding chatGPT**

Reinforcement learning

18 hours, with Ana Basic

— **Lecture 1: introduction to reinforcement learning**

— **Lecture 2: Markov decision processes and dynamic programming**

- Finite horizon, infinite horizon (discounted, optimal stopping, stochastic shortest path)
- Algorithms: value iteration, policy iteration, linear programming

— **Lecture 3: bandit algorithms**

- Epsilon greedy and the EXP3 algorithms
- Contextual bandits, bandits with expert advice and the EXP4 algorithm
- Stochastic bandits, Upper Confidence Bound (UCB) algorithm
- Thompson sampling

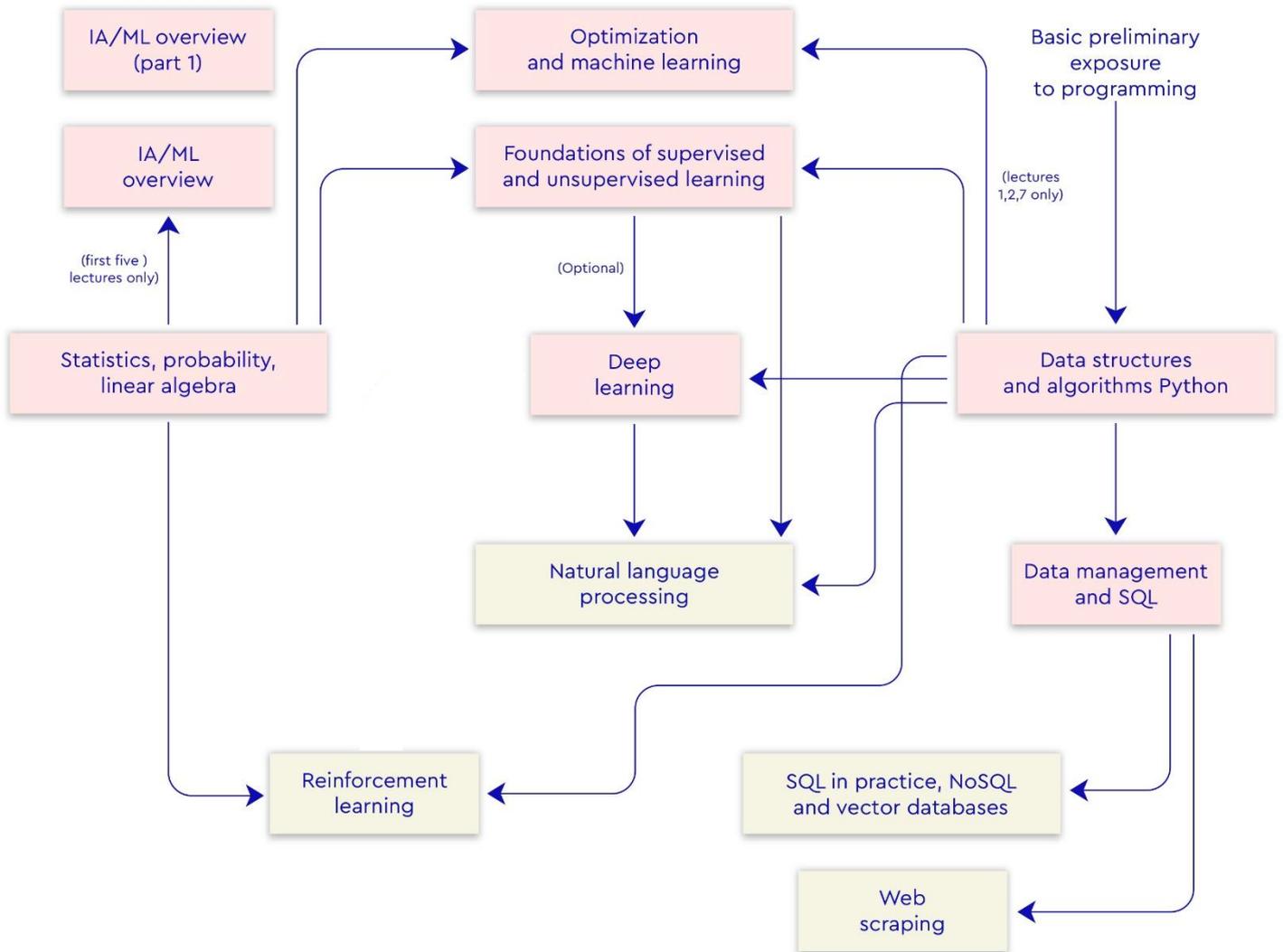
— **Lectures 4 and 5: from dynamic programming to reinforcement learning**

- Introduction to stochastic approximation

SYLLABUS - DETAILED DESCRIPTION OF EACH CLASS

- Algorithms: TD-learning, SARSA, Q-learning
- Actor-critic
- Model-based reinforcement learning, upper confidence reinforcement learning (UCRL)
 - **Lecture 6: reinforcement learning with approximation**
- From linear to Deep Neural Networks (DNN)

Prerequisite



How to read this picture ?

Basic courses are in pink. Advanced courses are in light green.

An arrow of the form **A** → **B** indicates that, to follow course B, it is necessary to be comfortable with the content of course A.