

Descriptif détaillé et volume horaire des enseignements suivis depuis le début du cursus universitaire

KABONGO BUZANGU Ben

[M1 Données Apprentissage Connaissances \(2022-2023\)](#)

[M1 Data Science, Paris-Saclay \(2022-2023\)](#)

[L3 Informatique, Sorbonne Université \(2021-2022\)](#)

[L2 Informatique, Université de Caen Normandie \(2020-2021\)](#)

[L1 Informatique, Université de Caen Normandie \(2019-2020\)](#)

M1 Données Apprentissage Connaissances, Sorbonne Université (2022-2023)

<https://dac.lip6.fr/master/>

Projet Logiciel DAC

Janvier - Mai 2023

Travail sur sujet au choix proposé par les chercheurs du master.

Sujet : Are language models able to generate instructions for robots?

Projet PLDAC mené par Laure Soulier et Nicolas Thome.

Il vise à évaluer la capacité des modèles de langage neuronaux à générer des instructions pour des robots.

Le projet se concentre sur deux objectifs principaux : la première étape consiste à identifier les connaissances nécessaires pour générer des instructions à partir d'un jeu de données d'instructions, puis à les faire correspondre avec des ressources de connaissances. La deuxième étape consiste à évaluer si les modèles de langage intègrent ces ressources.

Les travaux précédents dans ce domaine montrent que les modèles de langage peuvent avoir des connaissances limitées et que leur capacité à générer des instructions pour des robots est donc incertaine.

Github : <https://github.com/BenKabongo25/m1-PLDAC>

Machine Learning

10 CM 2h, 10 TD 2h, 10 TP 2h = 60h

Introduction aux grandes thématiques de l'apprentissage statistique et de ses concepts fondamentaux avec un fort biais applicatif. Présentation des algorithmes usuels pour l'estimation de densité, la régression, la classification supervisée et non supervisée. Modèles et algorithmes de l'apprentissage statistique moderne et des réseaux de neurones, introduction aux outils mathématiques correspondants.

Contenu : Arbre de décision, Sélection de modèles, Estimation de densité, Classification bayésienne, Modèles linéaires, Descente de gradient, Perceptron, SVM, Réseaux de neurones, Apprentissage non supervisé, Théorie de l'apprentissage, Ensemble Learning, Processus Gaussien, Apprentissage par renforcement, Apprentissage sur les graphes

Projet : Réseaux de neurones

- Codage from scratch des réseaux de neurones : Auto-encodeurs, Convolution, etc.
- Github : <https://github.com/BenKabongo25/m1-machine-learning>

Recherche d'Information et Traitement Automatique de la Langue

10 CM 2h, 10 TP 4h = 60h

Présente les grandes problématiques posées autour du traitement de texte, et plus particulièrement pour la recherche d'information et le traitement automatique du langage. L'objectif de ce cours est donc de présenter et manipuler les principaux modèles pour analyser, synthétiser, exploiter/interroger et produire des documents.

Contenu :

- **Recherche d'information :**
 - Indexer et interroger une collection de documents -> Développer un moteur de recherche
 - Evaluer un moteur de recherche
 - Découvrir les avancées récentes dans le domaine sous l'angle du deep learning
- **Traitement automatique du langage naturel :**
 - Appréhender les modèles de classification et de fouille de texte (détection de sentiments, ...)

- Identifier la sémantique des éléments du texte (extraction de thèmes, représentations latentes et contextuelles, ...)
- Enrichir le texte grâce aux bases de connaissances
- Découvrir les avancées récentes dans le domaine sous l'angle du deep learning

TPs et Projets :

- Projet 1 : Bag-of-Words
- Projet 2 : Implémentation d'un papier de recherche sur la recherche d'information
 - Intitulé : **Offline Retrieval Evaluation Without Evaluation Metrics**
- Github : <https://github.com/BenKabongo25/m1-rital>

Modèles et Algorithmes Probabilistes et Statistiques pour l'Informatique

10 CM 2h, 10 TD 2h, 10 TP 2h = 60h

Vise à présenter les éléments de probabilités et de statistiques fondamentaux pour l'informatique. Au-delà du simple objectif d'exposer les notions classiques telles que les variables aléatoires, les moments ou les grandes familles de distributions, l'UE a pour but de montrer l'utilisation avancée de ces notions dans divers domaines de l'informatique tels que la bioinformatique, l'intelligence artificielle ou encore l'imagerie. Les techniques ainsi étudiées (chaînes de Markov, etc.) seront illustrées sur des applications provenant de ces différents domaines.

Contenu : Probabilités, Statistiques, Maximum de vraisemblance, à posteriori, Expectation-Maximization, Tests d'hypothèse, Chaîne de Markov, Chaîne de Markov cachée, Échantillonnage, Régression Linéaire et Logistique.

TPs :

- Github : <https://github.com/BenKabongo25/m1-mapsi>

Modélisation Optimisation Graphes et Programmation Linéaire

10 CM 2h, 10 TD 4h = 60h

Introduit les graphes et la programmation linéaire comme outils de modélisation et de résolution de problèmes. Les étudiants apprendront à formaliser des problèmes comme des programmes linéaires, à résoudre par l'algorithme du Simplexe et à utiliser des méthodes générales de résolution de problèmes d'optimisation tels que la programmation dynamique et le branch and bound. Les travaux dirigés et les projets permettent aux étudiants de mettre en pratique ces notions sur des problèmes standard.

Contenu : Programmation linéaire, Algorithme du Simplexe, Graphes, Problèmes de flot, Problèmes de transport, Programmation dynamique

Projet : Optimisation équitable

- Résolution de problèmes d'optimisation soumis à des contraintes d'équité
- Github : <https://github.com/BenKabongo25/m1-mogpl-optim-equitable>
- Note : 19/20

Intelligence Artificielle et Manipulation Symbolique de l'Information

10 CM 2h, 10 TD 2h, 10 TP 2h = 60h

Permet aux étudiants de maîtriser les techniques d'intelligence artificielle et de manipulation symbolique de l'information. Il aborde les notions classiques d'IA ainsi que les avancées récentes telles que les algorithmes SAT et la programmation par ensembles réponse. Différentes approches d'utilisation d'information sont également présentées, telles que l'utilisation de l'IA pour la programmation de jeux et la planification de tâches avec différents formalismes.

Stockage et Accès aux Mégadonnées

10 CM 2h, 10 TD 2h, 10 TP 2h = 60h

Aborde les architectures de SGBD répartis. Stockage et indexation de données réparties. Réplication de données et cohérence des répliques. Transactions réparties et protocoles de validation répartie. Principes des bases de données parallèles, jointures parallèles, évaluation parallèle de requêtes avec MapReduce. Étude pratique des systèmes Dynamo DB, BigTable, Hbase, Calvin et Oracle NOSQL KVStore.

Logique et Représentations de Connaissances

10 CM 2h, 10 TD 2h, 10 TP 2h = 60h

Vise à donner des fondements logiques pour aborder la modélisation du raisonnement en IA, la représentation des connaissances en IA et en bases de données, ainsi que pour la spécification de programmes en algorithmique et en programmation. Le cours présentera également le formalisme des logiques de description pour la représentation des ontologies et initiera les étudiants à la programmation logique en leur enseignant les rudiments du langage PROLOG. Enfin, ce cours sera une introduction à la démonstration automatique de théorème et la théorie de la preuve, ce qui est utile pour le génie logiciel et l'IA.

Contenu : Logique des propositions, Prolog, Résolution, Représentation de connaissances, Graphes conceptuels, Logique de description, Logique modale, Logique épistémique, Logique temporelle, Réseau de Pétri, Intervalles de Allen.

Projet : Ecriture en Prolog d'un démonstrateur basé sur l'algorithme des tableaux pour la logique de description ALC

- Github : <https://github.com/BenKabongo25/m1-lrc-demonstrateur>

Modèles et Langage de Bases de Données Avancées

10 CM 2h, 10 TD 2h, 10 TP 2h = 60h

L'objectif de ce cours est de présenter et de comparer différents modèles et langages, la modélisation et l'interrogation de données complexes dans les applications modernes comme les réseaux sociaux, réseaux de capteurs, l'open data et le machine learning. En partant du modèle de données relationnelle, le cours introduit le relationnel-objet avec le langage SQL3, les standards XML et JSON pour la modélisation et l'interrogation d'informations semi-structurées et le standard RDF pour les données ouvertes et le web sémantiques.

Contenu : Modélisation des données, SQL, SQL3, XML, XPath, XQuery, RDF, SPARQL, NoSQL, JSON

Ingénierie du Logiciel

10 CM 2h, 10 TD 2h, 10 TP 2h = 60h

L'UE présente l'analyse d'un domaine métier, la mise en place de procédures de validation par le test, la conception orientée composant et service, la conception détaillée à l'aide de patrons de conception, ... l'ensemble s'intégrant dans une méthodologie de développement qui définit des étapes, des rôles pour les intervenants, des livrables... Les travaux dirigés permettent d'appliquer la démarche présentée en cours sur des exemples choisis ; en TME les étudiants sont organisés en groupes de 6 à 8 et jouent le rôle d'une SSII répondant à un appel d'offre, en appliquant la même démarche. L'UE s'adresse à un public large, mais nécessite un peu d'expérience en orienté -objet notamment pour travailler avec la notation UML.

Projet : Conception d'un logiciel de conception de Livre dont vous êtes le héros

- Github : https://github.com/Alex-XIA1/IL_PROJET

Anglais

10 Cours 3h = 30h

M1 Data Science, Paris-Saclay (2022-2023)

<https://www.universite-paris-saclay.fr/formation/master/informatique/m1-data-science>

Mathematics for Data Science

21h

Content :

- **Linear algebra**
 - Linear map
 - Vectors and matrices, vector-matrix operations
 - Matrix inversions, diagonalization, determinants, eigen values
 - Resolution of systems of linear equations
 - Linear regression (multi-dimensional) with pseudo-inverse
 - norm, scalar products, basic inequalities (triangular inequality
 - Linear transforms such as Fourier Transform and convolutions
- **Functional analysis**
 - Hilbert spaces
 - Limits, bounds, convergence, uniform convergence
- **Calculus**
 - Derivatives and integrals
 - Jacobian, Hessian matrices
 - Linear operators (differential, integral)
 - Simple differential and integral equations

Deep Learning

21h

Content :

- Introduction to neural networks
- Efficient training of neural networks
- Convolutional neural networks
- Generative neural networks (Variational auto-encoders, Generative auto-encoders, Adversarial auto-encoders).

TPs : <https://github.com/BenKabongo25/m1-deep-learning>

Hand of Machine Learning

21h

Content : The objective of this course is to provide the essential bases in machine learning or machine learning: the main families of models and the associated algorithms (inference and learning). In particular, will be treated:

- The theoretical bases of learning
- Probabilistic models and Bayesian learning
- Linear models and SVM
- General framework: representation, loss function, generalization and over-learning, entropy, supervised / unsupervised

Project : Kaggle competition : To Be or Not To Be

- Github : <https://github.com/BenKabongo25/m1-HoML-to-be-challenge>

Anglais

30h

Fundamental Principles of Machine Learning

21h

Content : Gradient descent, Regressions, Perceptron, Overfitting, Regulation, PCA, SVM, Kernels, linear algebra

Applied Stats

21h

Content :

- Descriptive statistics and well-known distributions
- Point and interval estimation
- Frequentist estimation vs Bayesian estimation (Maximum likelihood and maximum a posteriori)
- Distribution of maximum likelihood estimates
- Hypothesis tests (T-test, chi-square, ANOVA)
- Least squares method, linear regression
- Multivariate distributions, linear discriminant analysis.

Information Retrieval

21h

In this module, we will discuss the main concepts and tools of information retrieval.

We will explore the different phases of the construction of the flagship application of this domain, the search engine: collection and cleaning of documents, detection of spam, indexing and implementation of information retrieval models. This will enable us to build a search engine that uses large volumes of information and responds to information needs in a relevant way.

Bases de données : Optimisation et transactions

42h

Content :

- **Optimisation :** Introduction, Gestion de la mémoire et stockage, index de hachage, index B, Implémentation de SQL via les opérateurs relationnels, Plan d'exécution.
- **Transactions :** Transaction, Mécanisme et protocoles de Contrôle de Concurrence, Journalisation et reprise sur panne, Maintenance d'applications et tuning des transactions

Distributed Systems for Massive Data Management

21h

Content : Regular expressions, Docker, MongoDB, Hadoop, NoSQL, Redis, Kafka, Graph Stores

Large Scale Distributed Data Processing

21h

Content : Hadoop, Spark, Big Data streaming, Elasticsearch, GraphX, TensorFlow, Numpy, Pandas, Jax.

Intelligence artificielle Logique et Contraintes

42h

Le cours introduit le concept de problème de satisfaction de contraintes et donne quelques principes pouvant guider la formalisation de tels problèmes en les illustrant sur des cas concrets. L'accent est mis sur le cas des problèmes à domaines finis. Il présente différentes notions de cohérence locale ainsi que des algorithmes permettant de les établir, afin de pouvoir transformer automatiquement de tels problèmes en des problèmes plus simples. Il présente également les méthodes de recherche exhaustives permettant de trouver, énumérer les solutions de tels problèmes.

L3 Informatique, Sorbonne Université (2021-2022)

<https://sciences.sorbonne-universite.fr/formation/offre-de-formation/licences/licences-generales-l2-l3/licence-dinformatique>

Projet Recherche : Combiner Apprentissage par Renforcement et Méthodes Évolutionnaires : Algorithme CEM

Janvier - Mai 2022

Encadrés par M. Olivier Sigaud, ce projet de recherche a consisté en la conception d'un nouvel algorithme qui fusionne les réseaux de neurones, l'apprentissage par renforcement et les méthodes évolutionnaires.

Github : <https://github.com/labicquette/CEM-ERL>

Sciences de données

10 CM 1h45, 10 TD 1h45, 10 TP 1h45 = 52.5h

Contenu : Introduction à l'intelligence artificielle et la science des données, Visualisation des données, Apprentissage supervisé, Régression linéaire, K-NN, Perceptron, Noyaux, Méthodes d'évaluation, Descente de gradient, Fonctions coûts, Régularisation, Arbre de décision, Random Forest, Bagging, Apprentissage non supervisé, Clustering, K-Means

TPs et Projets :

- Github: <https://github.com/BenKabongo25/l3-lu3in026-sciences-de-donnees>

Intelligence Artificielle et Jeux

10 CM 1h45, 10 TD 1h45, 10 TP 1h45 = 52.5h

Contenu :

- Décision collective et optimisation
 - Principe de stabilité, équité, efficacité, manipulabilité
 - Problèmes des mariages stables
 - Modélisation et méthodes de résolution
 - Dynamiques des systèmes multi-agents
 - Émergence des comportements collectifs
 - Comportements réactifs et passage sur des robots réels
- Méthodes de l'IA
 - Recherche d'heuristique dans des graphes
 - Algorithmes de jeux : Minimax, alphabeta, A*, ...
 - Apprentissage par renforcement

Projets :

- Problème des mariages stables
 - https://github.com/BenKabongo25/l3-lu3in025-ia-et-jeux/tree/main/Projet_1
- Projet jeu Colonel Blotto
 - https://github.com/BenKabongo25/l3-lu3in025-ia-et-jeux/tree/main/Projet_2/projet-spatial-blotto-benkabongo25
- Projet robot : apprentissage par renforcement et systèmes multi-agents
 - https://github.com/SU-LU3IN025-fev2022/projet-robot-groupe_3_ben_kabongo

Statistiques pour l'Informatique

10 CM 1h45, 10 TD 1h45, 10 TP 1h45 = 52.5h

Contenu :

- **Probabilités et statistiques descriptives** : Dénombrement, Probabilités conditionnelles, Indépendance, Variables aléatoires, Espérance, variance et lois des grands nombres, Théorème central limite.
- **Statistiques inférentielles** : estimation ponctuelle, intervalle de confiance, tests d'hypothèses, tests d'indépendance, tests d'ajustements, chaînes de Markov, simulations.

Projets :

- Github: <https://github.com/BenKabongo25/I3-lu3in005-stats-et-info>
- **Projet 1 : Exploration, exploitation**
 - Jeux : Puissance 4, Bandits manchots
 - Algorithmes : UCB, Greedy, e-Greedy, Monte Carlo Tree Search
 - Note : 18/20
- **Projet 2 : Classification probabiliste**
 - Maximum de vraisemblance, maximum à postériori, naives bayes
 - Note : 18/20
- **Projet 3 : Chaîne de Markov**
 - Etude de propagation d'une maladie au sein d'une population
 - Note : 17.6/20

Algorithmique II

10 CM 1h45, 10 TD 1h45, 10 TP 1h45 = 52.5h

Contenu : Preuve et complexité des algorithmes, Programmation récursive, Théorème maître, Introduction aux graphes, Parcours de graphes, Algorithmes gloutons, Algorithme de Dijkstra, Algorithme de Prim, Programmation dynamique

Projet : Programmation dynamique

Réseaux

10 CM 1h45, 10 TD 1h45, 10 TP 1h45 = 52.5h

Contenu : Protocoles réseaux : Couche physique, Ethernet, IP (IPv4), ARP, TCP, UDP, DNS,, DHCP, HTTP

Projet : Création d'un analyseur de protocoles réseau

- Le programme prend en entrée un fichier texte formaté. Le fichier doit contenir des caractères hexadécimaux décrivant une ou plusieurs trames Ethernet.
- Github: <https://github.com/BenKabongo25/Ducky>
- Présentation du projet : <https://youtu.be/-XJrLu-kKGE>
- Note : 18.4/20

Systèmes de Gestion de Bases de Données

10 CM 1h45, 10 TD 1h45, 10 TP 1h45 = 52.5h

Contenu : Bases de données relationnelles, Algèbre relationnelle, Evaluation des requêtes, Optimisation des requêtes, Index, Arbre B+, Transactions, Contrôle de concurrence, Dépendance fonctionnelles, formes normales, Trigger, Vues.

Technologies Web

10 CM 1h45, 10 TP 3h30 = 52.5h

Contenu : HTML5, CSS3, JavaScript, React, NodeJs, Express, MongoDB, SQL

Projet : Création d'un réseau social à l'image de Twitter

- Github: <https://github.com/BenKabongo25/Ducky>
- Note : 20/20

Architecture des ordinateurs

10 CM 1h45, 10 TD 1h45, 10 TP 1h45 = 52.5h

Contenu : Algèbre de Boole, Programmation en Assembleur MIPS

Fondements des systèmes d'exploitation

10 CM 1h45, 10 TD 1h45, 10 TP 1h45 = 52.5h

Contenu : Système d'exploitations, ordonnancement des tâches

Orientation et Insertion Professionnel

10 Cours 3h = 30h

Définition du projet professionnel par les étudiants qui au cours du semestre essaient d'en apprendre un peu plus sur un métier au choix, notamment en organisant des interviews avec des experts du métier.

Métier choisi : **Data scientist**

Anglais

10 Cours 3h = 30h

L2 Informatique, Université de Caen Normandie (2020-2021)

<https://uniform.unicaen.fr/catalogue/formation/licences/5615-licence-informatique?s=trouver-sa-formation&r=>

Mathématiques discrètes

25h TD = 25h

Contenu : Logique propositionnelle, Fonctions mathématiques, Preuves et Raisonnement : raisonnement par récurrence, etc.

- Note : 20/20

Calcul Scientifique

12.5h CM, 12.5h TP = 25h

Contenu : Calcul scientifique avec numpy, matplotlib, scipy, sympy

TPs :

- Github : <https://github.com/BenKabongo25/l2-calcul-scientifique>
- Note : 20/20

Structures algébriques

15h CM, 35h TD = 50h

Contenu : Structures algébriques : groupes, anneaux, corps

- Note : 20/20

Algorithmique

15h CM, 25h TD, 10h TP = 50h

Contenu : Structures de données : files, piles, listes chaînées, listes doublement chaînées, arbres binaires, arbres binaires de recherche

Projet : Arbre binaire de recherche

- Github: <https://github.com/BenKabongo25/l2-binary-tree>
- Note : 20/20

Introduction à la Programmation Orientée Objet

15h CM, 35h TP = 50h

Contenu : Java, classes, objets, héritage, polymorphisme, exceptions, etc.

Projet : Conception de petits jeux en Java

- Jeux : Nim, TicTacToe
- Github: <https://github.com/BenKabongo25/l2-fil-rouge>
- Note : 19/20

Conception Logicielle

15h CM, 35h TP = 50h

Conception d'un logiciel en Java.

Projet : Création d'un logiciel de rendu 3D par lancer de rayons (raytracing)

- Github: <https://github.com/BenKabongo25/lancer-de-rayons>
- Note : 19.23/20

Sécurité et aide à la décision

15h CM, 35h TP = 50h

Contenu : Introduction à l'intelligence artificielle, algorithmes minimax, alphabeta, programmation linéaire, graphes

Projet : Jeu d'infection : Minimax et alphabeta

- Github: <https://github.com/BenKabongo25/l2-jeu-infection>
- Note : 20/20

Programmation avancée

15h CM, 35h TP = 50h

Contenu : Autres paradigmes de programmation, programmation fonctionnelle avec Haskell, programmation avancée en Java.

Projet : Création d'un jeu de bataille navale en Java

- Github: <https://github.com/BenKabongo25/l2-bataille-navale>
- Note : 17.5/20

Données, web et multimédia

10h CM, 15h TP = 25h

Contenu : Introduction aux données web, à la recherche d'information (TF-IDF), à la reconnaissance des formes

Bases de données

7.5h CM, 7.5h TD, 10h TP = 25h

Contenu : Algèbre relationnelle, Bases de données relationnelles, SQL

- Note : 19.6/20

Systèmes logiques, combinatoire et séquentielle

10h CM, 9h TD, 6h TP = 25h

Contenu : Algèbre de Boole, systèmes logiques, portes logiques, systèmes combinatoires, circuits, bascules, etc.

Systèmes

10h CM, 15h TP = 25h

Contenu : Introduction aux systèmes d'exploitation : shell, bash, linux, grep, awk, etc.

Réseaux

10h CM, 7.5h TD, 7.5h TP = 25h

Contenu : Protocoles réseaux : couches physique, Ethernet, IP.

Outils Informatiques pour la communication

10h CM, 15h TP = 25h

Contenu : Latex, Beamer pour la rédaction des documents scientifiques.

Anglais

25h TD = 25h

L1 Informatique, Université de Caen Normandie (2019-2020)

<https://uniform.unicaen.fr/catalogue/formation/licences/5615-licence-informatique?s=trouver-sa-formation&r=>

Algèbre linéaire 1

50h

Contenu : Systèmes d'équation, pivot de Gauss, vecteurs, matrices, ...

Algèbre linéaire 2

50h

Contenu : Espace vectoriel, théorème du rang, applications linéaires, matrices d'applications linéaires, injectivité, surjectivité, bijectivité, ...

Outils de calculs, probabilités et statistiques

50h

Contenu : introduction aux probabilités et statistiques, événements, univers, probabilité, variable aléatoire, lois des variables aléatoires discrètes et continues, espérance, variance, ...

- Note : 20/20

Introduction à la programmation

50h

Contenu : Python, C.

Projet : Jeu de domino

- Github : <https://github.com/BenKabongo25/l1-domino-number>

Technologies web 1

50h

Contenu : HTML5, CSS3

Projet : réalisation d'un site web en HTML5 et CSS3

Technologies web 2

50h

Contenu : HTML5, CSS3, Javascript

Projet : réalisation d'un site web en HTML5, CSS3 et Javascript

Conception Logicielle 1

50h

Projet : Création d'un jeu vidéo en Python

- Librairie : Pygame

Logique

50h

Contenu : logique et raisonnements, calcul scientifique, logique propositionnelle, introduction aux bibliothèques python numpy, sympy, scipy, matplotlib

Anglais

25h