



GUIDE D'ACCOMPAGNEMENT

PROJET IA

ANNÉE
2021 – 2022



GRANDES ECOLES DES MÉTIERS D'AVENIR

PRÉAMBULE

A l'attention de chaque étudiant et étudiante de l'IA School.

Ce guide est mis à votre disposition dans le but de vous accompagner efficacement dans la réalisation et le suivi de votre **projet IA**. **Pour sa bonne compréhension, il convient de vous référer aux consignes propres à votre niveau d'étude.**

Ce projet s'inscrit entièrement dans le programme pédagogique de vos années d'études à l'IA School et vous permet de valider, chaque année, 4 crédits ECTS. Il doit être un outil de suivi de votre montée en compétences et doit vous permettre le développement du socle de compétences nécessaires à l'exercice des métiers de la data.

Pour toute demande d'information, merci de contacter votre référente pédagogique.

Nous vous souhaitons une bonne réflexion et un bon travail.

L'équipe pédagogique.

TABLES DES MATIÈRES

PRÉAMBULE	1
TABLES DES MATIÈRES	2
LE PROJET	3
THÉMATIQUE DE L'ANNÉE 2021-2022	4
OBJECTIFS ET ATTENTES	5
RENDU	6
CONSIGNES	7
B1-B2 : Collecte et préparation de données	7
B. B3-M1 : PROJET D'ANALYSE ET DE VISUALISATION	10
C. M2 : CRÉATION D'UN MODÈLE	13
CALENDRIER DES RENDUS	16
SYNTHÈSE	17

1. LE PROJET

L'IA est l'un des deux axes de l'enseignement au sein de l'IA School.

Pour cette raison, il est l'un des projets majeurs de votre scolarité. **Une thématique globale est imposée (voir page suivante), mais les consignes diffèrent en fonction de votre année d'études afin de constituer une évolution progressive tout au long de votre scolarité (voir suite du guide).**

Le Projet vise à traiter des sujets autour des thématiques suivantes :

- Gestion de données
- Impact de l'IA sur la société
- Innovation

Le Projet IA a pour but d'approfondir vos réflexions d'année en année tout en utilisant les connaissances et compétences acquises au cours de votre scolarité. Il permet à chaque étudiant et étudiante de mettre en pratique les enseignements reçus.

Ce projet donne la possibilité de comprendre les enjeux auxquels les organisations seront confrontés dans l'avenir, tous secteurs d'activité confondus.

Ainsi, le projet IA comporte :

- Un thème commun à tous les niveaux d'étude et qui change chaque année
- Des consignes propres à chaque niveau d'étude

Pour bien appréhender ce qui est attendu de vous, il convient donc de bien prendre connaissance de cette problématique commune ainsi que des consignes propres à votre niveau d'étude.

→ VOTRE ENGAGEMENT

Le projet IA s'intègre dans chacune des organisations auprès desquelles nous évoluons. Il est ainsi de votre responsabilité de trouver un contexte pertinent et de savoir vulgariser votre travail de recherche, le rendre accessible à tous.

THÉMATIQUE DE L'ANNÉE 2021-2022

Vous devrez réaliser ce projet en lien avec le thème de l'année :

SMART CITIES

SMART

L'anglicisme "smart" a intégré notre vie depuis maintenant plusieurs années. Qu'il s'agisse de *smartphone*, *smart TV* ou encore de *smart contracts*, ce mot nous évoque forcément quelque chose. Mais saviez-vous que SMART est également un acronyme ? D'une part, dans le monde de l'informatique, SMART signifie "*Self-Monitoring, Analysis and Reporting Technology*". D'autre part, dans un contexte de gestion de projet où l'acronyme SMART permet de définir les indicateurs et objectifs suivants : Spécifique, Mesurable, Acceptable, Réaliste, Temporellement défini. En tout état de cause, le terme "smart" évoque toujours une évolution des habitudes.

CITIES

La ville est au cœur des préoccupations d'avenir. Ce sujet soulève diverses problématiques telles que, par exemple, le climat, la consommation des ressources, la digitalisation, les technologies, ou encore l'aménagement urbain. Il paraît désormais indéniable que nos villes doivent connaître une évolution, et que nous avons tous un rôle à jouer !

2. OBJECTIFS ET ATTENTES

Bien que la thématique soit commune, les objectifs et attentes du Projet IA évoluent en fonction de votre niveau d'étude. Il convient donc de vous référer à l'année qui vous correspond (ici et pour la suite du guide) :

B1/B2 - CRÉATION D'UN OUTIL DE SAISIE DE DONNÉES

- Comprendre l'impact de l'intelligence artificielle dans notre société
- Développer un outil de récolte de données
- Être capable de justifier l'intérêt de votre outil
- Faire le lien entre les données exploitées et la problématique business

B3/M1 - PROJET D'ANALYSE ET DE VISUALISATION

- Réfléchir aux enjeux liés à l'intelligence artificielle sur la société
- Être en mesure de définir une problématique et proposer des réponses
- Être capable d'investir ses compétences techniques
- Faire des choix techniques et les justifier
- Comprendre les limites des technologies utilisées et proposer des solutions pour les contourner

M2 - CRÉATION D'UN MODÈLE D'APPRENTISSAGE

- Appréhender les enjeux liés à l'intelligence artificielle sur la société
- Être en mesure de définir une problématique et y répondre
- Être capable d'investir ses compétences techniques
- Faire des choix techniques et les justifier
- Faire preuve de professionnalisme en travaillant en équipe

3. RENDU

Le Projet IA fait l'objet d'une évaluation écrite pour **tous les étudiants**. En outre, une soutenance orale complète cette évaluation pour les étudiants de **Bachelor 1, de Bachelor 2 et de Mastère 1.**

Rendu écrit et oral

La restitution du projet IA se fera impérativement au travers un document écrit en respectant les modalités du tableau ci-dessous et le plan proposé.niveau d'études :

Niveau	Evaluation
B1	<ul style="list-style-type: none">• Rapport écrit (5-10 pages) comprenant un tableau Excel (100 lignes et 10 à 15 colonnes)• Soutenance de 10 minutes avec un support de présentation
B2	<ul style="list-style-type: none">• Rapport écrit (5-10 pages) comprenant un tableau Excel (200 lignes et 10 à 15 colonnes)• Soutenance de 10 minutes avec un support de présentation
B3	<ul style="list-style-type: none">• Rapport écrit (10-15 pages)• Un notebook
M1	<ul style="list-style-type: none">• Rapport écrit (10-15 pages)• Un notebook• Soutenance de 15 minutes avec un support de présentation
M2	<ul style="list-style-type: none">• Support libre présentant le projet (écrit, ppt, vidéo etc...)• Un notebook

4. CONSIGNES

A. B1-B2 : Collecte et préparation de données

Une base de données sert à collecter des informations et les organiser de manière à ce qu'elles soient lisibles et consultées par un utilisateur ou un programme. La base de données est traitée, gérée et stockée et sert à centraliser et structurer les informations en un seul endroit pour pouvoir les exploiter.

Avant de traiter et analyser des données, celles-ci doivent être collectées.

En Bachelor 1 et Bachelor 2, votre travail est de créer une interface qui va venir récupérer des données : un outil de saisie des données.

Vous êtes libre de créer l'outil de votre choix mais nous vous imposons certaines modalités :

B1	<ul style="list-style-type: none">• Un travail individuel (un écrit de 5-10 pages)• Une problématique liée au thème• La restitution des données dans un fichier Excel d'au moins 100 lignes et 10 variables (colonnes)• Un cadre théorique qui respecte le plan
B2	<ul style="list-style-type: none">• Un travail individuel (un écrit de 5-10 pages)• Une problématique liée au thème• La restitution des données dans un fichier Excel d'au moins 200 lignes et 10 variables (colonnes)• Des variables calculées, c'est-à-dire les indicateurs déductibles de vos résultats.• Des axes d'analyse• Un cadre théorique qui respecte le plan

Pour définir quelles données vous allez récolter, il faut, avant tout, définir une problématique à laquelle votre travail va répondre.

Cela revient donc à travailler en plusieurs étapes :

- ❖ Définir une **problématique** liée au thème
- ❖ Sélectionner les différents **éléments pertinents** (les variables) permettant de répondre à cette problématique
- ❖ Créer votre outil et collecter les **données**
- ❖ Réaliser un **écrit** (5-10 pages) à partir des données récoltées en expliquant dans quelle mesure vos données permettraient de répondre à la problématique.
Le plan suivant doit vous permettre de comprendre nos attentes :

I. Résumé

Résumé du projet en 20 lignes maximum

II. Introduction

Présentation du sujet et de la problématique.

III. Démarche

Quelles données avez-vous cherchées ? Comment allez-vous les récolter ?

IV. Résultats

Présenter de façon sommaire vos résultats

V. Interprétation

Comment votre jeu de données peut répondre à la problématique

VI. Conclusion

→MODALITES D'EVALUATION DU RENDU ÉCRIT

B1

- **Pertinence** du choix de votre sujet par rapport au thème imposé
- Pertinence des **variables**
- **Style d'écriture** (grammaire, tournure de phrases, orthographe)
- **Respect du plan imposé**
- Capacité à **prendre du recul** sur votre travail pour le mettre en perspective face aux enjeux de la société de demain.

B2

- **Pertinence** du choix de votre sujet par rapport au thème imposé
- Pertinence des **variables**
- **Style d'écriture** (grammaire, tournure de phrases, orthographe)
- **Respect du plan imposé**
- Capacité à **prendre du recul** sur votre travail pour le mettre en perspective face aux enjeux de la société de demain.
- Capacité à **intéresser le lecteur**.

→MODALITES D'EVALUATION DU RENDU ORAL

La soutenance doit permettre au jury d'évaluer votre **bonne compréhension du sujet** et votre **capacité à réfléchir aux domaine(s) d'application de l'intelligence artificielle**.

Les éléments attendus lors de cette soutenance sont :

- Une **introduction** du sujet et de la problématique
- Une **synthèse à l'oral de votre projet**
- Une brève **conclusion**

L'objectif de cet oral n'est pas de lire ce que vous avez rédigé, mais de restituer oralement de manière claire, concise et précise votre travail.

⚠ **Nous vous rappelons qu'une présentation orale s'accompagne toujours d'un support visuel (présentation Power Point, site internet, vidéo, autres visuels...).**

B. B3-M1 : PROJET D'ANALYSE ET DE VISUALISATION

L'analyse et la visualisation des données sont des étapes fondamentales dans le cycle de vie de la donnée. Ces deux phases sont étroitement liées et permettent la compréhension et la représentation des données en les rendant exploitables et accessibles.

En Bachelor 3 et en Mastère 1, vous allez recueillir et traiter des données afin de soumettre des recommandations pertinentes. Votre mission a pour objectif de donner vie aux données.

Vous êtes libre quant aux choix des outils/plateformes utilisés mais ils devront être justifiés, voici cependant les modalités imposés :

B3	<ul style="list-style-type: none">• Un travail individuel (notebook + écrit de 10-15 pages)• Une problématique liée au thème• Une source de données minimum• Le stockage de la base de données créée avec SQL• Un Dashboard (outil au choix)• Des indicateurs d'analyse / KPIs• Un cadre théorique qui respecte le plan
M1 DA et DS	<ul style="list-style-type: none">• Un travail individuel ou en groupe (notebook + écrit de 10-15 pages)• Une problématique en lien avec le thème• Des données récoltées via du scrapping, de deux sources différentes minimum• Un Dashboard (outil au choix)• Des indicateurs d'analyse définis avec du code (R ou Python au choix) / KPIs• Un cadre théorique qui respecte le plan

Cela revient donc à travailler en plusieurs étapes :

- ❖ Définir une **problématique** liée au sujet

- ❖ **Collecter** les données souhaitées

Vous êtes libres quant aux sources de données mais il est important que celles-ci soient fiables, structurées et justifiées comme par exemples : <https://opendata.paris.fr> ou <https://www.kaggle.com/>. Pour les étudiants en mastère 1, nous vous rappelons qu'il faut collecter les données via deux sources minimum, en faisant du webscrapping.

- ❖ **Préparer** et **analyser** les données

Il s'agit pour vous de nettoyer les données brutes et de les structurer. Cela vous permet de mettre en évidence les doublons, les données incomplètes et/ou incorrectes. Il faut standardiser les données pour limiter les risques. Vous devrez ensuite les transformer en informations exploitables ; les convertir les données sous une forme plus lisible (graphiques, tableaux, cartes, documents de texte, etc...) et réaliser un Dashboard.

- ❖ **Exploiter** les données

Mettre en place des indicateurs de suivi. Les étudiants en mastère 1 doivent le faire avec du code (choix des technologies libre mais justifié)

- ❖ Réaliser un **écrit** (10-15 pages)

À partir des données récoltées, il faut expliquer dans quelle mesure l'exploitation de ces données permettrait de répondre à la problématique. Le plan suivant doit vous permettre de comprendre nos attentes :

I. Résumé

Résumé du projet en 20 lignes maximum

II. Introduction

Présentation du sujet et de la problématique.

III. Démarche

Quelles données avez-vous cherchées ? Comment les avez-vous récoltées ?

IV. Résultats

Présenter de façon sommaire vos résultats

V. Interprétation

Comment votre jeu de données peut répondre à la problématique

VI. Conclusion

→ MODALITES D'EVALUATION DU RENDU ÉCRIT

B3

- **Pertinence** du choix de votre sujet par rapport au thème imposé
- Pertinence des **variables**
- **Style d'écriture** (grammaire, tournure de phrases, orthographe)
- Capacité à **replacer votre travail technique dans le contexte de votre problématique.**

M1

- **Pertinence** du choix de votre sujet par rapport au thème imposé
- Pertinence des **variables**
- **Style d'écriture** (grammaire, tournure de phrases, orthographe)
- Capacité à **replacer votre travail technique dans le contexte de votre problématique.**

→ MODALITES D'EVALUATION DU RENDU ORAL

La soutenance doit vous permettre de :

- présenter brièvement le **contexte** et la **problématique**.
- et d'**expliquer en détails votre démarche et son applicabilité.**

Cette restitution orale doit prendre la **forme d'un pitch**, où l'importance est donnée à l'intelligibilité que vous allez donner à votre projet. **L'objectif est de vous adresser au jury comme s'il s'agissait d'une équipe mixte, c'est-à-dire pas uniquement technique.**

⚠ **Nous vous rappelons qu'une présentation orale s'accompagne toujours d'un support visuel (présentation Powerpoint, site internet, vidéo, autres visuels...).**

C. M2 : CRÉATION D'UN MODÈLE D'APPRENTISSAGE

En Mastère 2, vous allez suivre le cycle du travail d'un data scientist et d'un data analyst dans son intégralité.

Après avoir récupéré et analysé les données, vous allez les interpréter. Pour cela, l'étape de modélisation est fondamentale.

À partir des objectifs du projet, vous allez choisir un algorithme d'apprentissage, entraîner votre modèle de manière à ce qu'il soit le plus performant possible. Au terme de cette étape, vos résultats doivent vous permettre d'orienter la prise de décision et ainsi répondre à votre problématique.

Le but de cette phase est de construire un modèle d'apprentissage qui soit une bonne approximation du phénomène réel qu'on cherche à modéliser.

Appliquer un modèle d'apprentissage signifie faire des choix en fonction (machine learning, modélisation de série temporelle, deep learning), déterminer les bonnes métriques (la fiabilité des modèles), définir les algorithmes pertinents.

Vous êtes libre quant aux choix des outils/platformes utilisés mais ils devront être justifiés, voici cependant les modalités imposés :

M2 DA et DS	<ul style="list-style-type: none">• Un travail en équipe• Une problématique en lien avec le thème• Un Web service• Un modèle d'apprentissage• Une interprétation• Un cadre théorique qui respecte le plan (support libre + notebook)
------------------------	---

Nous vous conseillons vivement de former une équipe mixte (Data Analyst + Data Scientist). Cela vous permettra de travailler dans des conditions semblables à celles que vous trouverez en entreprise.

Cela revient donc à travailler en plusieurs étapes :

- ❖ Définir une **problématique** liée au sujet
- ❖ **Collecter** et **stocker** les données. Pour cela, vous devez utiliser les techniques de web scrapping python et alimenter votre base de données (SQL ou MongoDB au choix).
Vous êtes libres quant au choix des sources ainsi que des outils mais toutes ces décisions doivent être **justifiées** (quel périmètre justifie l'outil ?) ;
- ❖ **Analyser** les données
Vous devez les transformer en informations exploitables ; les convertir les données sous une forme plus lisible (graphiques, tableaux, cartes, documents de texte, etc...) et faire les tests nécessaires.
- ❖ **Modéliser**
Vous devez créer un modèle d'apprentissage, le justifier, définir des algorithmes et déterminer les bonnes métriques.
- ❖ **Visualiser**
Vous devez créer un web service qui va permettre de présenter les résultats obtenus.
- ❖ **Interpréter** vos résultats
Vous devez réaliser un écrit (entre 5 et 10 pages) où, à partir de vos résultats, vous préconisez une stratégie. C'est ici que vous allez développer des compétences et connaissances métiers. C'est ici que vous allez mettre en avant votre plus value, au delà de vos capacités techniques.
Le plan suivant doit vous permettre de comprendre nos attentes :

VII. Résumé

Résumé du projet en 20 lignes maximum

VIII. Introduction

Présentation du sujet et de la problématique.

IX. Démarche (collecte, analyse, modélisation, gestion de projet)

Quelles données avez-vous cherchées ? Comment les avez-vous récoltés

Quels outils ? Pourquoi ?

Comment avez-vous mis en place la gestion de ce projet ? Pourquoi ?

X. Résultats

Présenter de façon sommaire vos résultats

XI. Interprétation

Comment vos résultats peuvent répondre à la problématique

XII. Conclusion

Prendre du recul par rapport à l'ensemble du travail réalisé, le remettre en questions, proposer des pistes d'amélioration.

→ MODALITES D'EVALUATION DU RENDU ÉCRIT


M2

- **Pertinence** du choix de votre sujet par rapport au thème imposé,
- Capacité à **analyser** la problématique
- Capacité à **justifier** votre démarche, les problèmes que vous avez rencontrés et les **solutions** mises en place
- **Performance du modèle**
- **Capacité à prendre du recul** par rapport à l'ensemble des décisions prises.
- Capacité à présenter le travail à une équipe mixte, c'est-à-dire **le rendre intelligible** à un lecteur qui n'a pas les capacités techniques.

5. CALENDRIER DES RENDUS

Classes	Date butoir du rendu écrit	Dates de la soutenance orale
Mastère 1 ^{ère} année – M1	20 mai 2022	Du 23 au 31 mai 2022*
Mastère 2 ^{ème} année – M2		Pas d'oral en M2
Mastères en rentrée décalée - M1 et M2		Pas d'oral
Bachelor 1 ^{ère} année – B1		Du 23 au 31 Mai 2022*
Bachelor 2 ^{ème} année – B2		Du 23 au 31 Mai 2022*
Bachelor 3 ^{ème} année – B3		Pas d'oral en B3

**Les passages de soutenances orales ont lieu lors des partiels du second semestre.*

 **ATTENTION** : cela ne signifie pas que les projets doivent être rendus à la date limite. Une date limite signifie simplement qu'il s'agit de la date maximum de rendu des dossiers. Il est demandé de procéder au rendu du dossier au plus tôt.

6. SYNTHÈSE

→ Rendu écrit : synthèse

Classe	Contenu du projet IA	Modalités d'évaluation écrite	Date de rendu écrit	Travail de groupe ou individuel
B1	Création d'un outil de collecte de données	Rapport écrit comprenant <u>de 5 à 10 pages Notebook</u>	20 mai 2022	Individuel
B2				
B3	Analyse et visualisation via un Dashboard	Rapport écrit <u>de 5 à 10 pages Notebook</u>	20 mai 2022	Individuel
M1			20 mai 2022	Groupe max 3 étudiants de même niveau
M1 RD			20 mai 2022	
M2	Création d'un modèle d'apprentissage	<u>Support libre</u> (document écrit, ppt, vidéo) <u>Notebook</u>	20 mai 2022	Groupe max 3 étudiants de même niveau
M2 RD			20 mai 2022	

→ Rendu oral : synthèse

Classe	Contenu de la restitution orale	Modalités d'évaluation orale	Date de soutenance orale
B1	Introduction du sujet et de la problématique Synthèse à l'oral de votre démarche et relation au thème Conclusion	10 minutes de présentation, 5 minutes de questions-réponses, 5 minutes de feedback	Du 23 au 31 Mai 2022
B2			
B3	Pas d'oral		
M1	Introduction : présentation du cadre et de la problématique Synthèse de la démarche, des résultats et de son applicabilité	15 minutes de présentation, 5 minutes de questions-réponses, 5 minutes de feedback	Du 23 au 31 Mai 2022
M2	Pas d'oral		