

PROGRAMME DE FORMATION BACHELOR IA XP

Cahier des Charges de la MSPR « Intégrer des modèles d'IA »

COMPÉTENCES ÉVALUÉES :

- Analyser un modèle IA (données, performances, limites) en vue de son intégration.
- Concevoir une architecture d'intégration (flux, API, composants) cohérente avec le contexte.
- Implémenter un service backend d'inférence (API / script) robuste exploitant le modèle IA.
- Concevoir et développer une interface utilisateur permettant d'exploiter le modèle IA.
- Mettre en place des tests et corriger les dysfonctionnements de la solution IA.
- Déployer la solution IA et mettre en place un dispositif minimal de supervision.
- Prendre en compte la sécurité dans la solution IA.
- Documenter la solution et la présenter à des publics techniques et non techniques.

PHASE 1 : PRÉPARATION DE CETTE MISE EN SITUATION PROFESSIONNELLE RECONSTITUÉE

- **Durée de préparation** : 54 heures
- **Mise en œuvre** : Travail d'équipe constituée de 3 apprenants

PHASE 2 : PRÉSENTATION ORALE COLLECTIVE + ENTRETIEN COLLECTIF

- **Objectif** : mettre en avant et démontrer que les compétences visées par ce bloc sont bien acquises.
- **Moyen** : L'équipe utilise un support de présentation
- **Durée totale par groupe** : 50 mn se décomposant comme suit :
 - 20 mn de soutenance orale par l'équipe.
 - 30 mn d'entretien collectif avec le jury (questionnement complémentaire).
- **Jury d'évaluation** : 2 personnes (binôme d'évaluateurs) par jury

I - CONTEXTE

1 - Présentation générale

EnergiTech est une société anonyme fondée le 12 mars 2012 et filiale du groupe GreenPower Holding. Son siège social se situe à Paris. En plus du siège, l'entreprise possède trois sites opérationnels : un centre de production à Lyon, un parc solaire à Nantes et un hub dédié aux données à Strasbourg. Ces implantations permettent à EnergiTech de couvrir l'ensemble du territoire français tout en restant proche de ses principaux clients industriels et institutionnels.

2 - Business model

Le cœur de métier d'EnergiTech repose sur la production d'électricité issue de sources renouvelables – éolien, solaire et hydro-électricité – qu'elle revend à différents types de clients. Les grandes entreprises industrielles (par exemple AlimCo, une usine agro-alimentaire) signent des contrats d'achat d'énergie à long terme (Power Purchase Agreements) afin de sécuriser leur approvisionnement tout en réduisant leur empreinte carbone. Les gestionnaires de réseaux nationaux, tels que RTE, achètent quant à eux de l'énergie intermittente pour équilibrer le réseau. Enfin, les collectivités locales, soucieuses d'atteindre leurs objectifs de neutralité carbone, bénéficient d'offres dédiées.

Outre la vente d'électricité, EnergiTech propose des services de maintenance prédictive aux exploitants de ses parcs. Ces prestations sont facturées sous forme de contrats de service (SLA) et constituent une source de revenu récurrente. Un troisième levier économique réside dans la valorisation des agrégats de données anonymisées (mesures météorologiques, performances des turbines) vendues à des partenaires de recherche, toujours dans le strict respect du RGPD.

3 - Typologie des clients

Les clients d'EnergiTech se déclinent en quatre segments principaux :

- Industrie lourde – des acteurs comme AlimCo recherchent une énergie stable et un accompagnement pour réduire leurs coûts carbone.
- Gestionnaires de réseau – RTE utilise les données d'EnergiTech pour anticiper les fluctuations de production et assurer la continuité du service.
- Collectivités locales – la mairie de Lyon, par exemple, collabore avec EnergiTech afin de développer des projets d'énergie verte sur son territoire.
- Partenaires de recherche – MétéoFrance ou le CNRS accèdent à des séries temporelles de mesures environnementales pour leurs travaux scientifiques.

4 - Organisation interne

L'entreprise emploie environ 275 personnes réparties en plusieurs départements. La direction, les fonctions administratives et juridiques regroupent 25 salariés, tandis que la production et l'exploitation comptent 120 techniciens et ingénieurs chargés de la surveillance des parcs. Le service Recherche & Développement (R&D) rassemble 45 spécialistes, organisés autour de trois pôles : intelligence artificielle, optimisation énergétique et innovation produit. Un groupe de 30 experts assure la gouvernance des données, la mise en place des pipelines d'ingestion et le respect des exigences RGPD. Le département commercial et marketing, fort de 35 collaborateurs, gère la prospection, la négociation des contrats d'achat d'énergie et les relations clients. Enfin, le support et les services informatiques comptent 20 personnes dédiées à l'assistance technique et à la formation des utilisateurs.

5 - Chiffres clés (2024)

En 2024, EnergiTech a réalisé un chiffre d'affaires d'environ 210 millions d'euros, issu de la vente d'électricité, des services de maintenance et de la monétisation des données. La capacité installée totale s'élève à 125 MW, répartie entre 45 MW d'éolien (site de Lyon), 30 MW de solaire (Nantes) et 50 MW d'hydroélectricité. Les capteurs installés sur les différentes installations génèrent près de 12 téraoctets de données chaque année, couvrant les mesures de production, les paramètres environnementaux et les logs de supervision. L'effectif global de l'entreprise s'établit à 275 salariés, dont 45 travaillent au sein du service R&D.

6 - Conformité et enjeux réglementaires

Toutes les données collectées sont traitées conformément au Règlement Général sur la Protection des Données (RGPD). Les flux provenant des capteurs sont anonymisés afin d'éliminer toute possibilité d'identification personnelle. Un registre des traitements a été mis en place et fait l'objet d'audits annuels. EnergiTech détient également la certification ISO27001 depuis 2023, attestant de la mise en œuvre d'un système de management de la sécurité de l'information. Par ailleurs, l'entreprise respecte la directive européenne sur l'énergie, notamment les obligations de transparence relatives à la production et à la consommation d'énergie.

7 - Vision stratégique (2025-2030)

La feuille de route d'EnergiTech prévoit de doubler sa capacité installée pour atteindre 250 MW d'ici 2028, tout en automatisant 80 % des opérations de maintenance grâce à l'introduction de modèles d'intelligence artificielle capables de prédire les pannes. L'objectif ultime est d'atteindre la neutralité carbone sur l'ensemble de ses sites d'ici 2030, en combinant optimisation de la production, stockage d'énergie et amélioration continue des processus grâce à l'analyse de données.

II – CAHIER DES CHARGES

Dans le prolongement des travaux menés sur la collecte et le stockage des données (MSPR TPRES53), le service R&D d'EnergiTech a développé plusieurs modèles de maintenance prédictive pour les éoliennes.

Ces modèles utilisent notamment :

- des mesures de capteurs (vitesse du vent, vibrations, température, puissance délivrée, etc.),
- des historiques d'incidents / pannes,
- des informations liées aux interventions des techniciens (planification, durée, compétences).

Les data scientists du hub de Strasbourg fournissent désormais :

- un jeu de données prêt à l'emploi (échantillon de données réelles ou simulées),
- un modèle IA déjà entraîné, sous la forme d'un artefact réutilisable (fichier modèle + rapport de performance).

Important – notion d'inférence :

Par inférence, on entend la capacité du modèle IA à produire une prédiction à partir de nouvelles données d'entrée (par exemple : probabilité de panne dans les 7 jours pour une éolienne).

Objectif principal

EnergiTech souhaite mettre à disposition des techniciens de maintenance et des ingénieurs R&D une application simple permettant :

- d'interroger le modèle IA via un service backend d'inférence,
- de visualiser le résultat de la prédiction via une interface utilisateur,
- d'assurer un minimum de sécurité (authentification, gestion de rôles, clé API côté backend),
- de disposer d'une traçabilité des appels (logs, nombre de requêtes, erreurs) pour supervision.

Enjeux pour EnergiTech

Ce projet doit permettre :

- de réduire les arrêts non planifiés des éoliennes,
- d'optimiser la planification des interventions (priorités, affectation des techniciens),
- de renforcer la confiance dans l'utilisation de l'IA en rendant le système transparent, sécurisé et documenté,
- de tester une première intégration d'un modèle IA dans le SI, avant un éventuel déploiement plus large.

III – LES BESOINS EXPRIMÉS

EnergiTech confie à votre équipe la conception et la réalisation d'une première version fonctionnelle (MVP) d'un système d'intégration de modèle IA pour la maintenance prédictive.

1. Analyse du modèle IA et de ses limites (niveau pragmatique)

Les data scientists mettent à disposition :

- un jeu de données d'évaluation,
- un rapport de performance pour un ou plusieurs modèles candidats (par exemple :
 - Modèle A : classification « panne dans les 7 jours / pas de panne »,
 - Modèle B : régression « temps estimé avant panne »).

Votre mission :

- interpréter les métriques fournies (précision, rappel, F1-score, RMSE, etc.),
- identifier les forces et limites du modèle (zones de mauvaise performance, déséquilibre des classes...),
- proposer un choix argumenté du modèle à intégrer, selon le besoin métier et les contraintes d'exploitation,
- expliciter dans la documentation les hypothèses d'usage du modèle (ce pour quoi il est adapté... et ce pour quoi il ne l'est pas).

On ne vous demande pas de réentraîner le modèle, mais de l'analyser et de l'intégrer.

2. Conception de l'architecture d'intégration

Vous devez concevoir une architecture séparée incluant au minimum :

- un frontend (interface utilisateur) simple,
- un backend d'inférence, exposant une API ou un point d'entrée script bien défini,
- un composant modèle IA (chargement du modèle, fonction d'inférence),
- une base de données ou stockage léger pour :
 - les logs d'appels au modèle,
 - éventuellement quelques informations sur les utilisateurs / techniciens.

Les livrables devront comporter :

- un diagramme d'architecture décrivant les composants et les flux,
- une description textuelle des principaux scénarios d'usage (ex. : « un technicien se connecte, sélectionne une éolienne, lance une prédiction... »).

3. Implémentation du service backend d'inférence

Le backend est au cœur de la mission. Il doit :

- être développé de préférence en Python (fortement recommandé, mais d'autres langages sont acceptés si justifiés),
- exposer un ou plusieurs endpoints permettant d'envoyer les données nécessaires au modèle et de récupérer la prédiction,
- intégrer le chargement du modèle IA fourni (ou l'appel à un composant dédié),
- valider les données d'entrée (types, valeurs manquantes, cohérence de base),
- gérer les erreurs de façon robuste (logs, réponse claire côté API).

L'API devra être documentée (contrat d'interface) :

- format des requêtes (champs attendus, types),
- format des réponses (structure, codes de retour, gestion des erreurs).

4. Interface utilisateur

Vous devez développer une interface utilisateur simple (web ou autre technologie équivalente) permettant de :

- saisir ou sélectionner les paramètres d'entrée nécessaires à la prédiction (ex. : identifiant de l'éolienne, certains indicateurs de capteurs...),
- lancer une inférence via le backend,
- afficher clairement le résultat (ex. probabilité de panne, classe de risque, temps estimé avant panne).

Aucune librairie spécifique n'est imposée : une simple application web (HTML/CSS/JS, templates, etc.) suffit, à condition que le flux utilisateur soit clair.

5. Tests et correction des dysfonctionnements

Vous devez mettre en place :

- des tests unitaires sur les fonctions critiques (chargement du modèle, prétraitement, logique métier),
- des tests d'intégration vérifiant le bon fonctionnement de bout en bout (appel API → inférence → réponse),
- un rapport de tests ou, a minima, des captures d'écran / traces d'exécution des tests.

Les éventuels dysfonctionnements identifiés devront être corrigés et documentés (bugs connus restant éventuellement non corrigés).

6. Déploiement local et supervision minimale

La solution devra :

- être déployable en local (sur le poste d'un développeur ou d'un technicien),
- être accompagnée d'un guide de déploiement pas-à-pas,
- mettre en place un dispositif minimal de supervision, par exemple :
 - logs structurés (date, utilisateur, endpoint, statut, temps de réponse),
 - un compteur ou une petite page / endpoint de healthcheck indiquant que le service est opérationnel.

7. Sécurité

Vous devez :

- mettre en place une authentification et une gestion de rôles minimale (par exemple, rôle Technicien vs Manager),
- protéger l'API d'inférence par au moins une clé API ou un token,

IV – LES LIVRABLES

À l'issue de la mission, EnergiTech attend un ensemble de productions tangibles, chacune clairement documentée et prête à être exploitées par les équipes de data-science et la direction technique d'EnergiTech.

Chaque livrable doit être présenté sous forme de fichiers numériques organisés dans un répertoire unique nommé `Projet_Integration_modele_IA` et livré sous forme d'archive compressée (.zip) afin de garantir la reproductibilité du travail.

Les livrables majeurs sont :

- **Documentation technique détaillée incluant**
 - Rappel du contexte et des objectifs,
 - Description de l'architecture,
 - Description de l'API d'inférence,
 - Description du modèle IA (rôle, données d'entrée/sortie, limites),
 - Dispositif de supervision (logs, healthcheck, métriques simples),
 - Choix techniques justifiés.
- **Guide utilisateur (manuel pour un technicien / manager EnergiTech) :**
 - Procédure de connexion,
 - Scénarios d'usage (lancer une prédiction, interpréter le résultat).
- **Note / rapport sur les risques éthiques et de sécurité :**
 - Analyse des risques,
 - Impacts possibles en cas de mauvaise prédiction,
 - Mesures de mitigation.
- **Schéma d'architecture (diagramme)**
 - Composants principaux,
 - Flux de données entre frontend, backend, modèle, stockage/logs.
- **Scripts / documentation de déploiement**
 - Instructions claires pour installer les dépendances,
 - lancer le backend,
 - lancer l'interface,
 - vérifier que l'ensemble fonctionne.
- **Code source**
 - du service backend d'inférence
 - Code source de l'interface utilisateur
- **Jeu de tests + rapport de tests**
 - Tests unitaires et d'intégration,
 - preuves d'exécution (sortie de tests, capture d'écran, rapport).
- **Fichiers de configuration documentés**
 - Explication des variables (clé API, URL, paramètres techniques...).

La documentation, les guide et les note doivent être rédigés dans un style professionnel, avec une mise en page claire (titres, sous-titres, paragraphes justifiés). Il servira de livrable officiel à présenter lors de l'oral et constituera la base documentaire pour les équipes de production.

En complément des livrables, l'équipe projet devra préparer un support de présentation destiné à la soutenance finale devant le client (public technique). Ce support devra synthétiser les principaux éléments du travail réalisé : démarche suivie, difficultés rencontrées, solutions mises en place, résultats obtenus et perspectives.

Il est important de souligner que l'évaluation de cette MSPR repose sur la **combinaison des trois éléments suivants** :

- la **qualité du travail** réalisé au cours du projet,
- la pertinence et l'exhaustivité des **livrables remis**
- et la capacité de l'équipe à **présenter, justifier et valoriser** ce travail lors de la soutenance orale.

Les équipes devront donc **s'assurer que la soutenance reflète bien l'ensemble des compétences attendues**, en démontrant à la fois la maîtrise technique et la capacité à communiquer efficacement auprès d'un client professionnel.

V – RESSOURCES FOURNIES

Afin de permettre à l'équipe projet de mener à bien la mission, EnergiTech mettra à disposition un ensemble de ressources techniques et documentaires. Ces éléments constituent la base de travail commune et devront être exploités et enrichis par les apprenants.

1. Assistance et périmètre

Dans le cadre de ce projet pédagogique, l'équipe projet n'aura aucun contact direct EnergiTech. Le cahier des charges constitue la seule expression officielle du besoin. Toute demande de clarification devra être traitée avec l'encadrant pédagogique, jouant le rôle du client.

2. Webographie

Les apprenants ont à disposition la plateforme LinkedIn Learning et la Bibliothèque ENI.

Les liens ci-dessous sont suggestions de ressources. Ils ne sont pas exhaustifs : les apprenants sont encouragés à chercher d'autres sources.

2.1 Modèles IA & évaluation :

- Scikit-learn – Site officiel

Documentation et exemples pour les modèles de classification/régression et les métriques.

<https://scikit-learn.org/scikit-learn.org>

2.2 APIs backend & services d'inférence :

- FastAPI – Tutoriel (FR)

Guide utilisateur pour créer des API en Python (endpoints, requêtes/réponses, doc automatique).

<https://fastapi.tiangolo.com/fr/tutorial/fastapi.tiangolo.com>

(Les apprenants peuvent choisir un autre framework, mais ce tutoriel donne une bonne base.)

2.3. Tests unitaires & d'intégration :

- pytest – Site officiel

Documentation pour écrire et organiser des tests automatisés en Python.

<https://docs.pytest.org/> (via <https://pypi.org/project/pytest/>) PyPI

- Introduction à pytest (FR)

Présentation en français du framework pytest avec exemples.

<https://he-arc.github.io/livre-python/pytest/index.html> he-arc.github.io

2.4. Sécurité applicative (OWASP) :

- OWASP Top 10 (FR)

Document de référence sur les 10 risques de sécurité les plus critiques pour les applications web.

<https://owasp.org/Top10/fr/> OWASP Foundation

- Article de vulgarisation OWASP Top 10 (FR)

Explications et exemples concrets des vulnérabilités et solutions.

<https://www.pirates.fr/top-10-owasp-decryptage-du-top-10-owasp-et-solutions-pratiques/> PIIRATES

2.5. RGPD & protection des données

- Guide RGPD du développeur – CNIL

Guide officiel pour intégrer le RGPD dans les développements web et applicatifs.

<https://www.cnil.fr/fr/guide-rgpd-du-developpeur> CNIL

- Guide RGPD du développeur – version complète (GitHub)

Version détaillée avec fiches et exemples de code.

<https://lincnil.github.io/Guide-RGPD-du-developpeur/> Lincnil+1

- Guide “Sécurité des données personnelles” – CNIL (PDF)

Mesures de sécurité à mettre en œuvre pour protéger les données personnelles (édition 2024).

<https://www.cnil.fr/fr/guide-de-la-securite-des-donnees-personnelles-nouvelle-edition-2024> CNIL+1

2.6. Interfaces web simples (HTML / CSS / JS)

- Apprendre le développement web – MDN (FR)

Parcours complet pour débuter en HTML, CSS, JavaScript.

https://developer.mozilla.org/fr/docs/Learn_web_development MDN Web Docs

- Votre premier site web – MDN (FR)

Tutoriel pas à pas pour créer une première page web.

https://developer.mozilla.org/fr/docs/Learn_web_development/Getting_started/Your_first_website MDN Web Docs

- Commencer avec le HTML – MDN (FR)

Bases de la syntaxe HTML et structure d'une page.

https://developer.mozilla.org/fr/docs/Learn_web_development/Core/Structuring_content/Basic_HTML_syntax MDN Web Docs+1

3. Jeu de données (exemple)

Nom du fichier : energiTech_maintenance_sample.csv

Contenu : historique de mesures issues de capteurs d'éoliennes et d'informations de maintenance (vitesse du vent, niveau de vibration, température, puissance délivrée, indicateur de maintenance, survenue de panne, délai avant panne, identifiants pseudonymisés de techniciens, dates d'intervention planifiée, etc.).

Usage attendu : ce fichier sert de référence pour comprendre le contexte métier, la nature des données utilisées par les modèles IA et illustrer les cas d'usage de l'application. Il n'est pas demandé de réentraîner

4. Modèle IA de classification (fichier modèle entraîné)

Nom du fichier : model_classification.pkl

Objet : modèle supervisé permettant de prédire la survenue d'une panne dans les 7 prochains jours (sortie binaire : panne / pas de panne).

Usage attendu : c'est le modèle principal recommandé pour cette MSPR. Vous devez l'analyser à partir du rapport de performance fourni, puis l'intégrer dans votre service backend d'inférence (chargement du fichier modèle, appel à la fonction de prédiction, gestion des entrées/sorties et des erreurs).

5. Modèle IA de régression (fichier modèle entraîné)

Nom du fichier : model_regression.pkl

Objet : modèle supervisé estimant le temps restant avant une panne (en nombre de jours).

Usage attendu : ce modèle est fourni comme option complémentaire. Vous pouvez l'étudier, comparer ses performances à celles du modèle de classification et, si vous le jugez pertinent, proposer une intégration dans votre application. Vous restez cependant libres de ne déployer que le modèle de classification si vous estimez que la régression n'apporte pas de valeur suffisante dans l'état actuel. un modèle à partir de ce jeu de données

6. Rapport de performance du modèle

Nom du fichier: rapport_performance_modeles_EnergiTech.pdf

Contenu :

- description synthétique du jeu de données utilisé ;
- résultats détaillés pour le modèle de classification (accuracy, précision, rappel, F1-score, matrice de confusion, commentaires) ;
- résultats détaillés pour le modèle de régression (MAE, RMSE, R^2 , commentaires) ;
- recommandation du service R&D d'EnergiTech sur le choix du modèle à privilégier.

Usage attendu : ce rapport est la base de votre analyse. Vous devez vous appuyer sur ces éléments pour :

- identifier les forces et limites de chaque modèle,
- justifier votre choix de modèle à intégrer dans l'application,
- expliciter, dans votre documentation, les conditions d'utilisation et les limites du modèle retenu.

Les apprenants devront analyser ces éléments et choisir le modèle à intégrer.

Consignes importantes :

Vous ne réentraînez pas les modèles IA dans le cadre de cette MSPR : l'évaluation porte sur votre capacité à analyser, intégrer, sécuriser et exposer un modèle existant dans une solution applicative complète (backend d'inférence + interface utilisateur + tests + supervision + documentation).

Vous êtes libres de :

- n'intégrer que le modèle de classification,
- ou proposer un dispositif permettant de choisir entre plusieurs modèles, si cela reste cohérent avec le temps et le niveau de la MSPR.

Toute hypothèse ou adaptation (par exemple : choix d'un seul modèle, simplification d'un format de sortie) doit être explicitement justifiée dans votre documentation technique.