YNOV informatique Toulouse

# *Titre RNCP 39586*

*Ingénieur en science des données spécialisé en infrastructure data ou en apprentissage automatique.*

# *Rapport bloc 3*
# ELABORER ET PILOTER UN PROJET DATA

## *Élaborer et piloter le développement d'une plateforme data/IA destinée à tester la vulnérabilité et l'efficacité des contre-mesures face aux attaques par canaux auxiliaires*

## *(Side Channel Atacks)*

| ***Rédigé par*** | ***Responsabilité***<br>+ signature manuelle si pas de système de workflow |
|---|---|
| Maxence colvray | Apprenti datascientist |

# Table des matières

1 Description synthétique du projet ..... 3
1.1 Qu'est-ce qu'une attaque SCA ? ..... 3
1.2 A quel besoin peut répondre une plateforme data/IA pour tester des attaques SCA ? ..... 3
2 Le cadrage du projet (C3.1.1) ..... 4
2.1 Les problématiques ..... 4
2.2 Objectifs ..... 5
2.3 Les livrables ..... 5
2.4 Le cadre règlementaire ..... 6
2.5 Les contraintes ..... 6
2.6 Point de vigilance ..... 6
3 Plan de développement de la plateforme (C3.1.2 ; C3.1.3 ; C3.2.1 ; C3.2.2) ..... 6
3.1 Documentation du projet (C3.1.3) ..... 7
*3.1.1 QQOQCP ..... 7*
*3.1.2 Spécifications fonctionnelles ..... 7*
*3.1.3 Spécifications fonctionnelles ..... 8*
3.2 Dimensionnement du projet (C3.1.2) ..... 8
*3.2.1 Ressources humaines nécessaires ..... 8*
*3.2.2 Ressources matérielles et logistiques ..... 9*
*3.2.3 Chiffrage du projet ..... 9*
*3.2.4 Etude de faisabilité ..... 10*
3.3 Planification du projet (C3.2.1) ..... 10
*3.3.1 Stratégies macroscopique de développement ..... 11*
*3.3.2 Planning des sprints ..... 11*
*3.3.3 Matrice RCI ..... 12*
*3.3.4 Staffing ..... 13*
*3.3.5 Point de vigilance ..... 13*
3.4 Suivi de projet (C3.2.2) ..... 14
*3.4.1 Création du Backlog et sprint planning ..... 14*
*3.4.2 Gestion de l'avancement ..... 15*
4 Pilotage de l'équipe projet et outils de communication (C3.3.2) ..... 16
5 Plan de développement des compétences (C3.3.1) ..... 17
6 Mise en place d'un système de veille technologique et réglementaire (C3.4.1) ..... 18
7 Cas d'arbitrage externalisation pipeline de traitement des données (C3.3.3) ..... 19
8 Plan d'actions, éthique, confidentialité et responsabilité environnementale (C3.4.2) ..... 19
A.1 Annexes ..... 21
A.1.1 Terminologie ..... 21
A.1.2 Tableau ressources humaines nécessaires ..... 22
A.1.3 Schéma stratégies macroscopique de développement ..... 23
A.1.4 Planning des sprints ..... 24
A.1.5 Matrice RCI ..... 25

# 1 Description synthétique du projet

## 1.1 Qu'est-ce qu'une attaque SCA ?

Une attaque par canaux auxiliaires (SCA, pour Side Channel Attack) est une méthode de piratage pour laquelle on ne rentre pas dans le système, mais où m'on exploite les fuites physiques (comme les ondes électromagnétiques, la consommation électrique ou le temps d'exécution) pour en déduire des informations confidentielles. Ce projet concerne uniquement les SCA appliquées fuites d'ondes électromagnétiques, appelé aussi SCA EM. On retrouve 2 cas d'usage principaux :

**Cryptanalyse :**

Déchiffrage de clé de chiffrement (ex : AES, RSA, NTRU) en classifiant des traces EM généré par le processeur pendant les opérations de déchiffrement.

**Classification d'un trafic chiffré :**

Analyser un signal radio chiffré, provenant par exemple d'une borne wifi, d'un satellite, d'une radio, pour déduire des informations sur l'activité en cours. C'est-à-dire que même sans déchiffrer le contenu, on peut analyser des paternes permettant de classifier certaines caractéristiques du signal (ex : provenance, matériel utilisé, type de contenus etc.).

Plusieurs contre-mesures existent et ont besoin d'être testées pour palier à ce type d'attaques, blindage électromagnétique, capteurs de détection d'intrusion, filtrage et découplage ou contre-mesures logicielles comme le masquage ou le binding.

Pour mieux comprendre le principe des attaques par canaux auxiliaires assistées par intelligence artificielle, voici une vidéo explicative en français : https://www.youtube.com/watch?v=fmnXhsZk61Y

## 1.2 A quel besoin peut répondre une plateforme data/IA pour tester des attaques SCA ?

Ce projet s'applique aux entreprises qui ont besoin de tester, en laboratoire, la vulnérabilité et l'efficacité des contre-mesures appliquées à leur matériel hardware (ex : fabriquant de modules physiques cryptographiques, de système de télécommunication, de processeurs, de clé sans contacts ou encore de cartes bancaires).

Ici je me place en tant qu'entreprise de services qui livre une plateforme à des fabricants de matériel pour permettre aux ingénieurs électroniques de :

- **Collecter et stocker les données provenant de capteurs électromagnétiques** en temps réel en laboratoire.
- **Entrainer des modèles d'IA** permettant à l'utilisateur de réaliser des attaques SCA sur leur matériel, à partir de modèle entrainé avec les données collectées aux préalables, afin de tester la vulnérabilité et l'efficacité des contre-mesures.

La solution comprend des composants pour la collecte de signaux (via SDR, sondes EM, oscilloscopes), leur stockage (BDD OLTP et OLAP), ainsi que des modules IA pour entraîner et exécuter des attaques SCA (ex : récupération de clés AES/RSA, classification de signaux chiffrés). L'architecture repose sur une organisation en micro-services, avec une interface utilisateur, permettant aux ingénieurs de tester la robustesse de leurs équipements face à des attaques réalistes basé sur l'IA, tout en évaluant les stratégies de protection (binding, filtrage, blindage, etc.).

# 2 Le cadrage du projet (C3.1.1)

## 2.1 Les problématiques

Pour déterminer les problématiques techniques et opérationnelles, il est nécessaire d'avoir en tête ces 2 questions :

- Quels concepts techniques et composants mettre en œuvre pour capturer et exploiter les fuites physiques d'un système de communication crypté ?
- Comment aider les entreprises à tester de manière fiable et reproductible la vulnérabilité de leurs systèmes de communication et à évaluer l'efficacité des contre-mesures mises en place, dans un environnement de laboratoire ?

**Problématiques techniques :**

- Il est nécessaire de prendre en compte la collecte de données depuis des matériels de mesure et interface variés.
- Le format de données peut être multiple (série temporelle, spectrogramme, spectre de puissance, image etc.).
- Afin de collecter et tester des attaques il est essentiel de prendre en compte des notions de rapidité pendant la sauvegarde et la classification.
- Les attaques doivent être reproductibles ou évolutives pour un contexte donné, par conséquent les données collectées doivent être stockées sur le long terme.
- Les fabricants d'électronique ne possèdent pas principalement des compétences en informatique.

**Problématiques opérationnelles :**

- Certaines attaques SCA par IA ne sont pas encore bien connues ce qui renforce le fait de passer par une entreprise et un produit spécialisé externe.
- Actuellement le temps passé à tester de nombreuses contre-mesures dans des environnements volontairement changeant (bruit, isolation, sondes) est coûteux en temps.
- Le risque de subir une attaque augmente avec la réduction des composants, les attaques classiques sont plus dures à réaliser et les objets connectés sont plus nombreux.
- Le risque augmente également car l'IA permet, pour les attaquants, d'automatiser et d'augmenter la fiabilité des attaques SCA.

## 2.2 Objectifs




Afin de structurer efficacement la conception de la plateforme, j'applique la méthode SMART en décomposant le projet en 6 composants clés.

| Composants principaux | Spécifique | Mesurable | Atteignable/ réalisable | Temporellement défini |
|---|---|---|---|---|
| Collecte de données EM | Pipeline pour intégrer sonde EM+ osclilo et SDR | Affichage en flux avec délais < 10s | Intégration via protocole connue (pyVISA, SCPI) et librairies SDR | Livraison du module de collecte + affichage en BDD en 03/26 |
| BDD OLTP | Déployer une base pour traces EM brutes et métadonnées | Capacité d'insertion à 100k points/sec | API times-series database comme Influxdb | Livraison du module de collecte + affichage BDD en 05/26 |
| BDD OLAP | Construire un entrepôt de données structurés des series et les labéliser | Temps de requête < 1 s sur 10 000 traces ; agrégations précalculées | Outils compatibles (ex : ClickHouse, Influxdbv3, Airflow) | Livraison du module avec l'interface en 06/2026 |
| Entrainement des modèles IA | Intégrer un module pour entraîner des modèles CNN/LSTM avec des templates | Taux de précision sur modèle démo de 90% sur AES, RSA, détection type de contenus WIFI | Basé sur TensorFlow avec GPU | Modèles fonctionnels en 08/26 |
| Utiliser les modèle IA pour attaquer | Permettre aux utilisateurs de lancer des attaques SCA via UI | Inférence < 200 ms par trace ; résultat exportable (CSV/graph) | Modèles déployés en back-end API REST | Modèle fonctionnel en 10/26 |
| Interfaces UI | Interface BBD OLTP vers BDD OLAP, configuration modèles + résultats | Score UX ≥ 4/5 auprès de 5 testeurs externes avec plusieurs critères | Interface Streamlit | Version bêta testable en 012/2026 |

## 2.3 Les livrables

**Plateforme en micro-services :**
Livraison d'une plateforme modulaire, conteneurisée (avec Docker), découpée en micro-services couvrant les fonctionnalités principales incluant des templates de modèles IA préconfigurés pour faciliter l'entrainement des modèles.

**Modèles IA pré-entraînés de démonstration :**
Modèles IA prêts à l'emploi (RSA, AES, classification par type de contenu wifi) pour tester la plateforme immédiatement et illustrer les cas d'usage possibles.

**Service de création de modèles IA sur mesure(optionnel)**
Développement de modèles sur mesure très spécifique qui ne serait pas possible de réaliser avec les Template existants dans la plateforme.

**Documentation complète :**
Documentation fonctionnelle incluant, un guide d'utilisation de la plateforme ainsi que les bonnes pratiques de test et d'évaluation de contre-mesure. (Sans oublier une documentation technique interne pour les développeurs web, data ingénieur et data scientist).

**Formation des utilisateurs :**
Sessions de formation (à distance ou en présentiel) pour les équipes métiers, couvrant l'utilisation de la plateforme.

### 2.4 Le cadre règlementaire

Dans un premier temps, le cadre réglementaire de chaque déploiement s'inscrit dans le règlement interne de l'entreprise client avec le respect de leurs exigences en termes d'infrastructure et de cybersécurité.

La solution livrée au client doit leur permettre de respecter la norme internationale **ISO/IEC 17825:2024** qui concerne les méthodes de test pour la protection contre les attaques non-intrusives des modules cryptographiques.

L'**ANSSI** (Agence nationale de la sécurité des systèmes d'information) publie des guides et bonnes pratiques et recommande des contre-mesures pour les attaques par canaux auxiliaires.

### 2.5 Les contraintes

La première contrainte est normative car il est primordial de se mettre à la place de l'utilisateur pour qu'il puisse facilement corréler les résultats de tests avec les exigences normatives cités plus haut.

Les contraintes techniques majeurs sont :

- Acquérir des traces électromagnétiques ou des ondes radiofréquences suffisamment précise pour être exploités.
- Le système doit garantir une synchronisation temporelle précise entre les événements du matériel cible et l'acquisition du signal.
- La plateforme doit être capable d'enregistrer et de traiter en temps réel un grand volume de données.
- L'outil doit fournir des modèles d'intelligence artificielle préconfiguré avec des paramètres ajustables. L'utilisation de ces templates de modèle doit être suffisamment intuitif pour être utilisé par des ingénieur non-informaticien.

### 2.6 Point de vigilance

Voici les points de vigilances associés à chacune des contraintes citées plus haut :

- Les résultats affichés doivent être interprétables au regard des normes visées (ISO, FIPS, etc.), pas seulement techniques.
- Les signaux peuvent être pollués par du bruit, mal amplifiés, ou mal positionnés spatialement, rendant les traces inexploitées.
- La latence interne entre le déclencheur et l'enregistrement peuvent fausser les alignements temporels.
- Le buffer mémoire, la bande passante disque, ou les accès concurrents peuvent ralentir l'enregistrement des traces.
- Avec des modèles préconfigurés, si les interfaces ou paramètres sont trop abstraits ou mal guidés, l'ingénieur électronique ne les utilisera pas suffisamment

## 3 Plan de développement de la plateforme (C3.1.2 ; C3.1.3 ; C3.2.1 ; C3.2.2)

Le plan de développement présente l'ensemble des activités liées au développement de la plateforme permettant de collecter et stocker en temps réel des données en laboratoire, et d'entraîner des modèles d'IA personnalisables pour simuler facilement des attaques SCA et évaluer l'efficacité des contre-mesures sur du matériel réel.

Ce plan de développement couvre, les différents aspects du projet, du commencement à la livraison chez un premier client, en termes de **dimensionnement**, de **documentation**, de **planification** et **de suivi de projet**.

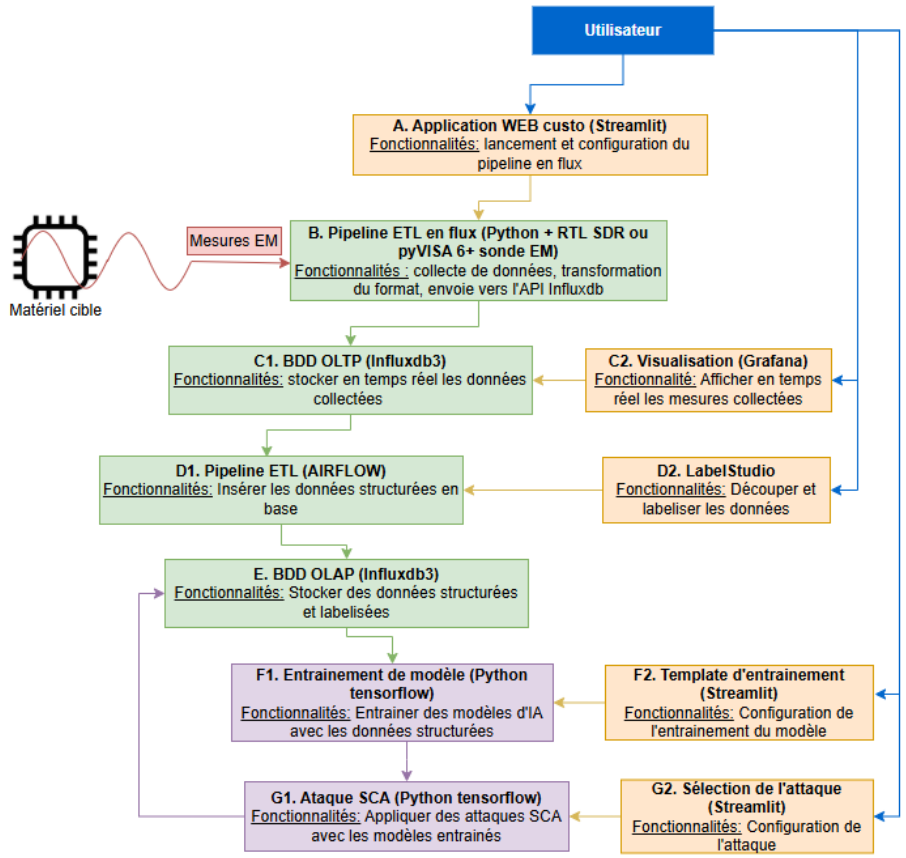
## 3.1 Documentation du projet (C3.1.3)

### 3.1.1 QQOQCP

| | |
|---|---|
| Qui ? | - Utilisateurs finaux : ingénieurs R&D electronique, fabricants de materiel sécurisé.<br>- Porteur du projet : une entreprise de services spécialisée en sécurité matérielle et IA.<br>- Parties prenantes : développeurs, data scientists, experts sécurité |
| Quoi ? | Développement d'une plateforme logicielle modulaire (micro-services) permettant :<br>- La collecte en temps réel de signaux électromagnétiques (EM, RF).<br>- L'entraînement de modèles IA pour simuler des attaques SCA.<br>- L'évaluation de l'efficacité des contre-mesures matérielles et logicielles |
| Où ? | - En environnement de laboratoire<br>- Chez les clients industriels (plateforme installée localement) |
| Quand ? | - Début du développement : début 2026<br>- MVP avec le client partenaire : mi 2026<br>- Version bêta complète : fin 2026 |
| Comment ? | - Par l'intégration de technologies connues (pyVISA, SDR, TensorFlow)<br>- Architecture modulaire conteneurisée (Docker + API REST)<br>- Interfaces utilisateur en Streamlit<br>- Documentation + formation à destination des clients |
| Pourquoi ? | - Répondre à la difficulté croissante de tester de nombreuses des attaques SCA<br>- Automatiser et fiabiliser les tests de sécurité physique via IA<br>- Réduire les coûts et le temps des tests en laboratoire<br>- Aider les industriels à anticiper les menaces side-channel sur leurs futurs produits |

### 3.1.2 Spécifications fonctionnelles

Ci-dessous le synoptique des composants de la plateforme, qui seront des micro-service conteneurisés, avec leurs fonctionnalités principales associés. En rouge les mesures physiques, en bleu les interactions utilisateurs, en jaune les interfaces, en vert le pipeline de données global (pipeline en flux et pipeline de données structurées) et en violet la partie attaque SCA par IA.



### 3.1.3 Spécifications fonctionnelles

Ci-dessous la liste des spécifications préliminaires pour chacun des composants :

| | |
|---|---|
| A. **Application Web Custom (Streamlit)** | Développée en Python 3.10 avec Streamlit ≥ 1.30 |
| | Choix du type de données collectées via la sonde EM + démarrage du pipeline (bibliothèques pyVISA pour oscilloscopes ou RTL-SDR + lib python pour puce RTL) |
| B. **Pipeline ETL en flux (Python + RTL-SDR / pyVISA)** | Sérialisation des trames via Apache Arrow (pyarrow Table) pour une transmission vectorisée et performante vers InfluxDB3 via API REST en mode Arrow. |
| | Données envoyées par lot de 10 000 à 50 000 lignes par payload, conformément aux recommandations InfluxDB v3. |
| C1. **BDD OLTP (InfluxDB v3)** | Utiliser une base en mode OLTP pour augmenter la rapidité de requête en lecture avec un schéma de données de type narrow. |
| C2. **Visualisation temps réel (Grafana)** | Temps d'affichage >15 s sur 3 graphes : signal brut, signal fenêtré, spectre (FFT). |
| | Permissions de modification des Dashboard interdite en lecture seule côté client. |
| D1. **Pipeline ETL batch (Airflow)** | Orchestration via DAG Python : extract_raw_data, structure_data, insert_olap. |
| | Logs stockés sur disque et accessibles via Airflow Web UI. |
| | Possibilité de config via dagRun.conf (utilisé par l'interface Streamlit) |
| D2. **LabelStudio (labellisation interactive)** | Front personnalisé avec TimeSeries pour découpage des signaux et labélisation. |
| | Export CSV ou JSON, consommation directe dans Airflow. |
| E. **BDD OLAP (InfluxDB v3)** | Stockage en Parquet via API InfluxDB v3 (DataFusion engine). |
| | Indexation avec tag qui défini le type de mesure dans des tables communes |
| F1. **Entraînement IA (Python + TensorFlow)** | Modèles CNN, LSTM, via TensorFlow 2 |
| | Pipeline : DataLoader → Preprocessor → ModelBuilder → Trainer → Saver. |
| F2. **Configuration entraînement IA (Streamlit)** | Formulaire UI pour sélection de modèle, hyperparamètres et dataset (menu dynamique depuis OLAP). |
| | Affichage des performances après entraînement depuis Tensorboard. |
| | Export du modèle au format .h5 + logs .json dans un répertoire. |
| G2. **Interface sélection attaque (Streamlit)** | Interface multi-étapes : sélection modèle → sélection traces → affichage signal → lancer attaque. |
| G1. **Attaque SCA (Python + TensorFlow)** | Chargement d'un modèle .h5 sélectionné depuis l'interface pour classification sur chaque trace. |
| | Affichage du résultat probabiliste de classification pour chaque trace. |

## 3.2 Dimensionnement du projet (C3.1.2)

### 3.2.1 Ressources humaines nécessaires

Ici nous utilisons un découpage en lots fonctionnels (composant A à G) pour calculer la charge du projet, ici chaque lot est découpé en taches techniques. Le tableau ci-dessous présente les taches détaillées, dont la charge a été estimée en recroisant les besoins avec les experts de chaque domaine, afin de réduire les incertitudes budgétaires et de mieux anticiper les coûts du projet. Chaque tache devra être finalisée par la rédaction d'un court PV de recette. J'ai également intégré les rôles Agile au développement de la plateforme où le rôle Scrum Master représente 15% de l'activité totale et l'activité Product Owner 15% également. Les activités Infra/cyber seront réparties tous aux longs du projet à travers différentes revues techniques. Au total le développement représente **232 semaines d'activités**, répartis sur plusieurs métiers (**3 développeurs web, 3 Data Engineer, 2 Data Scientist, 1 ingénieur infrastructure, 1 ingénieur cybersécurité, 2 PO** et **2 Scrum Master** chacun des 2 ayant un profil orienté électronique et l'autre informatique).

| | ID tache | Description de la tache | metier | durée |
|---|---|---|---|---|
| A | AT1 | Rédaction des specs detaillées | CTO | 1 |
| | | | dev web | 1 |
| | AT2 | Frontend de l'application | dev web | 3 |
| | AT3 | Backend avec declenchement du composant B | dev web | 3 |
| | ATT | Tests/validation | Testeur | 1 |
| B | BT1 | Rédaction des specs detaillées | CTO | 1 |
| | | | data ing | 1 |
| | BT2 | Script de collecte des données avec pyRTL-SDR et pyVisa | data ing | 2 |
| | BT3 | Transformation en formats exploitables | data ing | 2 |
| | BT4 | Insertion en base | data ing | 2 |
| | BTT | Tests/validation | Testeur | 1 |
| C1 | C1T1 | Rédaction des specs detaillées | CTO | 1 |
| | | | data ing | 1 |
| | C1T2 | Schéma de données partitionnement temporel et spatial | data ing | 1 |
| | C1T3 | Optimisation des schémas de tags et fields | data ing | 1 |
| | C1T4 | Batch writing et compression | data ing | 1 |
| | C1T5 | Requêtes optimisées et indexation | data ing | 2 |
| | C1TT | Tests/validation | Testeur | 1 |
| C2 | C2T1 | Rédaction des specs detaillées | CTO | 1 |
| | | | data sci | 1 |
| | C2T2 | Création des dashboards | data sci | 1 |
| | C2T3 | Création des requêtes | data sci | 2 |
| | C2TT | Tests/validation | Testeur | 1 |
| D1 | D1T1 | Rédaction des specs detaillées | CTO | 1 |
| | | | data ing | 1 |
| | D1T2 | Définition des DAGs | data ing | 2 |
| | D1T3 | Transformation et insertion des données dans Airflow | data ing | 2 |
| | D1TT | Tests/validation | Testeur | 1 |
| D2 | D2T1 | Rédaction des specs detaillées | CTO | 1 |
| | | | dev web | 1 |
| | D2T2 | Déploiement de LabelStudio connecté à la base OLAP | dev web | 2 |
| | D2T3 | Création de templates de tâches de labellisation | dev web | 5 |
| | D2TT | Tests/validation | Testeur | 1 |

| | ID tache | Description de la tache | metier | durée |
|---|---|---|---|---|
| E | ET1 | Rédaction des specs detaillées | CTO | 1 |
| | ET2 | | data ing | 1 |
| | ET3 | Modélisation du schéma de données analytique | data ing | 3 |
| | ET4 | Intégration des données transformées depuis l'ETL Airflow | data ing | 2 |
| | ETT | Tests/validation | Testeur | 1 |
| F1 | F1T1 | Rédaction des specs detaillées | CTO | 1 |
| | | | data sci | 1 |
| | F1T2 | Création de templates de modèles IA | data sci | 4 |
| | F1T3 | Implémentation des modèles IA CNN et LSTM | data sci | 8 |
| | F1TT | Tests/validation | Testeur | 1 |
| F2 | F2T1 | Rédaction des specs detaillées | CTO | 1 |
| | | | dev web | 1 |
| | F2T2 | Création du formulaire de configuration interactif | dev web | 3 |
| | F2T3 | Lancement de l'entraînement depuis l'interface | dev web | 2 |
| | F2TT | Tests/validation | Testeur | 1 |
| G2 | G2T1 | Rédaction des specs detaillées | CTO | 1 |
| | | | dev web | 1 |
| | G2T2 | Création d'une interface multi-étapes pour l'attaque SCA | dev web | 3 |
| | G2T3 | Intégration avec la base OLAP et les modèles exportés | data ing | 2 |
| | G2T4 | Lancement de l'inférence et affichage des résultats | dev web | 2 |
| | G2TT | Tests/validation | Testeur | 1 |
| G1 | G1T1 | Rédaction des specs detaillées | CTO | 1 |
| | | | data sci | 1 |
| | G1T2 | inférence à partir d'un modèle pré-entraîné | data sci | 2 |
| | G1T3 | Implémentation des métriques d'évaluation | data sci | 2 |
| | G1T4 | Export des résultats | data sci | 1 |
| | G1TT | Tests/validation | Testeur | 1 |
| Scrum master | SM | 15% de l'activité total | Scrum master | 55 |
| Product Owner | PO | 15% de l'activité total | Product owner | 55 |
| Infra | Infra | Assurer l'installation et les 2 déploiments de la plateforme | ing infra | 10 |
| Cyber | cyber | Garantir les exigences cybersécurtié de la plateforme | ing cyber | 12 |

### 3.2.2 Ressources matérielles et logistiques

**Déplacement**.
Des déplacements ponctuels sont prévus (1 à 2 jours/mois) pour les tests et l'accompagnement client. Le cout logistique est estimé à 5000€.

**Matériel hardware informatique.**
Les salariés de l'entreprise sont déjà équipés de matériel informatique, spécifiquement pour ce projet il faudra racheter 2 tours destinées aux Data Scientist ainsi que d'un serveur de préproduction permettant d'entrainer des modèles d'IA. Le cout estimé pour ces 3 ressources matériels est de 15000€.

**Matériel hardware de radio fréquence.**
L'analyse des émissions électromagnétiques nécessite du matériel RF spécifique. Un récepteur SDR (Software Defined Radio) comme le HackRF One. Pour les appareils émettant en haute fréquence (>3 Ghz) il faudra utiliser obligatoirement un oscilloscope ainsi que différents types de sonde électromagnétique. L'ensemble de ce type de matériel est estimé à 25 000€.

**Matériel hardware cible pour attaque SCA.**
Pour tester des attaques SCA, il faudra des cibles comme des microcontrôleurs (ex : STM32, ATmega), modules cryptographique embarquée et autre antenne de radio, le cout de ces équipements est estimé à 10 000€.

### 3.2.3 Chiffrage du projet

Le poste de dépense le plus volumineux est le cout humain liée au développement de la plateforme. L'ETP (équivalent temps plein) hebdomadaire est estimé à **3000 €/semaine**, ce qui inclut non seulement le temps de travail d'une ressource à temps plein, mais également les charges indirectes comme les frais RH, les équipements, les outils, les impôts, ainsi que les coûts de structure. Ici on utilise un ETP moyen identique pour chaque métier au sein de l'entreprise. A cela s'ajoute les ressources matériel et logistique cités plus haut, on doit également rajouter **un cout de formation de 10 semaines au total** (en moyenne à 5k€/semaine) répartis entre différents salariés afin de monter en compétence sur des domaines spécifique au projet. Afin de garantir les compétences clés en cryptanalyse et le respect des normes en vigueur, un **consultant spécialisé** interviendra pendant deux semaines (6k€/semaine). Pour terminer **une provision de 10 %** est intégrée au budget global afin de couvrir les risques et imprévus liés au projet.

| Poste de dépense | Chiffrage (en k€) |
|---|---|
| Cout humain (232 semaines) | 696 |
| Ressources matériel et logistique | 55 |
| Formations salarié (10 semaines) | 50 |
| Expert en cryptanalyse (2 semaines) | 12 |
| Provisions (10%) | 81 |
| Total | **894k€** |

### 3.2.4 Etude de faisabilité

Afin de correctement dimensionner le projet et de limiter les risques imprévus en termes de conception, un Proof Of Concept (POC) a été mené en amont.

Réalisé avec des apprentis développeurs, ce POC a permis de tester les premières briques techniques, ce constat a permis d'ajuster le chiffrage, d'identifier les compétences critiques et de renforcer les marges de sécurité dans la planification.

| Axe de faisabilité | Élément analysé | Résultat (défavorable, limité, favorable) | Commentaire |
|---|---|---|---|
| Technique | Collecte de données EM/RF avec SDR | favorable | Tests réalisés avec SDR et sondes EM, matériel compatible et librairies compatible. |
| | Collecte de données EM/RF avec oscilloscope | limité | Tests réalisés avec librairies Python compatibles mais maturité des librairies Python qui peuvent être partielles ou instables selon le modèle d'oscilloscope. |
| | Intégration technique UI | favorable | Interface Streamlit intuitive, adaptée aux développeurs python mais demande un temps de devloppement plus long que prévu. |
| Opérationnel | Organisation | limité | L'organisation du projet nécessite une meilleure isolation des métiers (IA, data ing, dev web), afin que chacun se concentre sur son domaine et soit plus autonome, ce qui est rendu possible grâce à l'architecture en microservices. |
| | Compétences internes | favorable | Une équipe pluridisciplinaire couvrant les domaines clés du projet (collecte RF, IA, dev web), renforcée ponctuellement par de la formation ciblée et l'appui d'un expert en cryptanalyse. |
| Économique | Competence interne | favorable | Les compétences clés sont présentes en interne, ce qui permet de réduire les coûts en limitant le recours à des prestataires externes |
| | Matériel | favorable | Le coût du matériel est bien maîtrisé et précisément ciblé |
| | Budget global | limité | Le budget global reste imprécis à ce stade, en raison de l'ampleur du projet et de plusieurs variables encore incertaines (charge réelle, ajustements techniques, inflation) |
| Légal | Conformité aux normes | favorable | Exigences ISO/IEC 17825:2024 et guides ANSSI claires permettant de se mettre facilment à la place des besoins utilisateur en termes de respect des normes. |
| Temporel | durée du projet | défavorable | La charge n'est pas estimable précisément à ce stade, le projet étant trop vaste ; elle devra être affinée avec les experts, d'autant plus dans un contexte d'inflation instable |
| | Risques techniques | Favorable | Risques bien identifiés lors du POC (bruit, latence, interface) avec mesures de réduction prévues. |

## 3.3 Planification du projet (C3.2.1)

J'ai choisi une approche Agile, car elle répond particulièrement bien aux caractéristiques de ce projet :

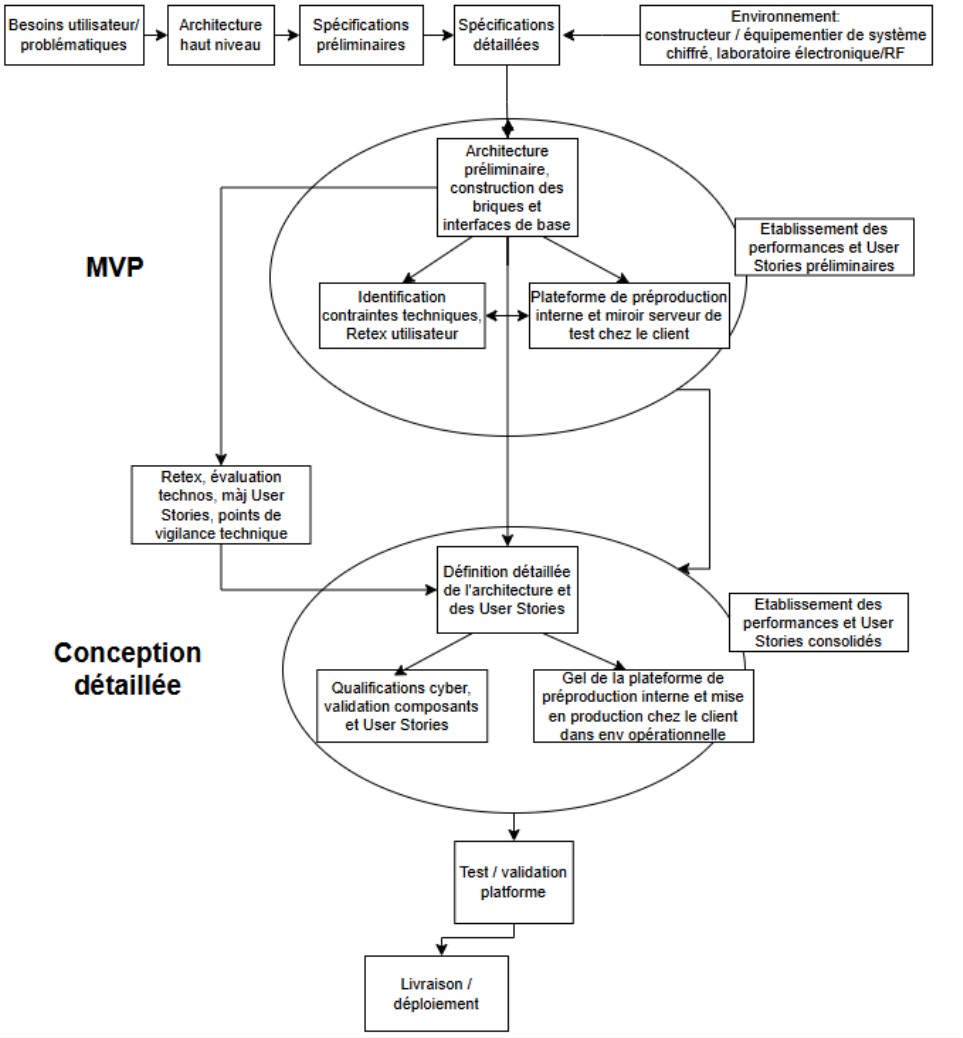
- Le système repose sur une architecture en micro-services, ce qui facilite naturellement l'application de méthodes Agiles, chaque service (collecte EM, IA, stockage, interface) peut évoluer de façon autonome, être développé, testé et livré indépendamment des autres. Cela favorise l'itération rapide et la réactivité aux retours terrain.
- Le projet est par ailleurs complexe avec de nombreuses hypothèses techniques concernant l'usage des fonctionnalités utilisateur. L'Agile permet donc d'adapter le développement des fonctionnalités aux besoins réels, souvent clarifiés à travers des démonstrations et tests utilisateurs.
- Enfin, l'équipe étant pluridisciplinaire (data, RF, IA, dev front/back), l'organisation en sprints, avec des rituels réguliers (daily, sprint review, sprint rétrospective), permet de maintenir une bonne coordination.

Dans ce cadre, j'ai choisi la méthode Scrum, une des approches Agile les plus utilisées. Elle repose sur des cycles courts appelés sprints, une priorisation continue du travail (backlog), et une interaction constante avec les parties prenantes. Scrum est particulièrement adapté à un projet de R&D complexe comme celui-ci, où les besoins et solutions évoluent au fil des expérimentations techniques et des retours utilisateurs.

### 3.3.1 Stratégies macroscopique de développement

Le projet sera découpé en 2 grandes phases, un MVP déployé chez le premier client partenaire, suivi d'une livraison complète sur ce même site. Cette approche itérative correspond à une intégration progressive chez le client, rendue nécessaire par la complexité technique du projet et la forte interaction entre les utilisateurs et la plateforme, qui impose d'ajuster le développement des fonctionnalités en fonction des usages réels.

L'intégration de nouvelles User Stories sera volontairement limitée avant le MVP, afin de garantir un périmètre maîtrisé, puis facilitée après validation du MVP, lorsque les fondations techniques et fonctionnelles seront consolidées.

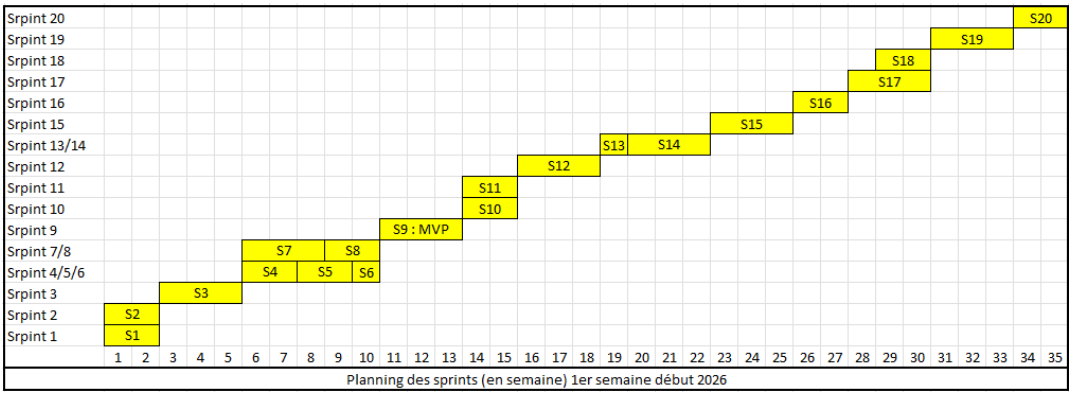



### 3.3.2 Planning des sprints

L'objectif est premièrement de découper les lots (ici appelés Epics dans un contexte Agile) de manière à trouver un compromis pour pouvoir **staffer les différents métiers en parallèle** (afin d'éviter les sous-charges ou surcharges de certains métiers**) tout en maintenant une cohérence entre les Epics**, c'est-à-dire travailler sur des Epics qui ont des interfaces communes et des User Stories interdépendantes durant un même sprint. A l'intérieur des sprints nous retrouverons des User Stories (qui sont les taches définies plus haut).

La durée des sprints va de 1 à 3 semaines. Des sprints plus longs réduisent la fréquence des retours et augmentent le risque de dérive. Ce planning démarre une fois la phase de réalisation des spécifications techniques détaillées qui ne sont pas intégrées dans un sprint à proprement parlé ici. La finalisation de chaque User Stories sera valider à travers un court PV de recette, une documentation global destiné aux utilisateurs sera réalisée au fur et à mesure du développement par le PO. Voici ci-dessous la liste des sprints ainsi que le planning des sprints qui débutera la première semaine de l'année 2026.

| Sprint | Epics / Lots | Description du sprint | Livrable / Objectif | Durée (en semaine) |
|---|---|---|---|---|
| Sprint 1 | B, C2 | Collecte EM/RF, Visualisation temps réel | Connexion SDR/Grafana opérationnelle | 2 |
| Sprint 2 | B, C1, C2 | Affichage temps réel, intégration SDR | Collecte stable, affichage synchro | 2 |
| Sprint 3 | A, B, C1, C2 | Stockage bruts, ETL initial finalisé | Données stockées dans BDD OLTP (Influxdb3) | 3 |
| Sprint 4 | C1, D1 | Insertion et automatisation batch | ETL fonctionnel pour les données brutes | 2 |
| Sprint 5 | C2, E | Structure analytique, modèle OLAP | OLAP prêt à agréger et interroger | 2 |
| Sprint 6 | F1, E | Requêtes optimisées, pré-agrégations | Modèle analytique testé | 1 |
| Sprint 7 | F1, F2 | Training IA CNN, config UI IA | Modèle CNN entraîné avec UI | 3 |
| Sprint 8 | F1, F2 | Training LSTM, visualisation résultats | Modèle LSTM + export résultats | 2 |
| Sprint 9 | Global | Déploiment du MVP au sein de l'entreprise partenaire | Commencer à récolter les premiers RETEX des utilisateurs finaux | 3 |
| Sprint 10 | D2, E | Labellisation et structuration | Dataset annoté, prêt pour entraînement | 2 |
| Sprint 11 | F1, C2 | Amélioration IA + stockage | Modèle IA affiné sur OLAP | 2 |
| Sprint 12 | G1, G2 | Attaque IA, interface sélection | Interface complète d'attaque SCA | 3 |
| Sprint 13 | G1, G2 | Affichage résultats attaque | Résultats IA affichés et exportables | 1 |
| Sprint 14 | C2, F2 | Réglages modèles + requêtes | Performances IA améliorées | 3 |
| Sprint 15 | G2, F2 | Finalisation UI attaque | Fonctionnement complet UI attaque | 3 |
| Sprint 16 | G1, F1 | Optimisation IA | Modèle IA optimisé | 2 |
| Sprint 17 | A, E | Dashboards + analyse avancée | Tableaux de bord analytiques disponibles | 3 |
| Sprint 18 | C1, D2 | Stabilisation labellisation + stockage | Système stable pour annotation & stockage | 2 |
| Sprint 19 | Global | Tests/ validation plateforme | Tests d'intégration validés | 3 |
| Sprint 20 | Global | Livraison + déploiment | Installation chez le client et démonstration aux utilisateurs | 2 |




Ci-dessus le planning des sprints. Certains sprints ont été volontairement planifiés en parallèle lorsque leurs périmètres et compétences sont hétérogènes. Cela permet d'accélérer l'avancement global du projet tout en optimisant la charge des équipes et en évitant les périodes creuses.
Ci-dessous le tableau des dépendances qui permet de visualiser les liens entre les différents sprints, afin d'anticiper les enchaînements critiques, planifier les livrables en cohérence, et limiter les risques de blocage ou de retard.

| sprint 1 | sprint 2 | sprint 3 | sprint 4 | sprint 5 | sprint 6 | sprint 7 | sprint 8 | sprint 9 | sprint 10 | sprint 11 | sprint 12 | sprint 13 | sprint 14 | sprint 15 | sprint 16 | sprint 17 | sprint 18 | sprint 19 | sprint 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| / | / | sprint 2 | sprint 3 | sprint 4 | sprint 5 | sprint 3 | sprint 7 | sprint 6 & 9 | sprint 9 | sprint 9 | sprint 10 | sprint 12 | sprint 13 | sprint 14 | sprint 15 | sprint 16 | sprint 16 | sprint 17 & 18 | sprint 19 |

### 3.3.3 Matrice RCI

Le modèle RCI est adapté ici car le projet fonctionne en mode agile, sans validation hiérarchique complexe.Les rôles R, C et I sont attribués par métier (Dev Web, Data Ingénieur, Data Scientist) et non par personne individuelle.

| Matrice RCI | | | | |
|---|---|---|---|---|
| Sprint | Description | Dev Web | Data Ing | Data Scientist |
| Sprint 1 | Collecte EM/RF, Visualisation temps reel | I | R | C |
| Sprint 2 | Affichage temps reel, integration SDR | R | R | C |
| Sprint 3 | Stockage bruts, ETL initial finalise | R | R | C |
| Sprint 4 | Insertion et automatisation batch | C | R | I |
| Sprint 5 | Structure analytique, modele OLAP | I | R | C |
| Sprint 6 | Requetes optimisees, pre-agregations | I | C | R |
| Sprint 7 | Training IA CNN, config UI IA | C | I | R |
| Sprint 8 | Training LSTM, visualisation resultats | C | I | R |
| Sprint 9 | Deploiment du MVP | R | R | R |
| Sprint 10 | Labellisation et structuration | R | R | C |
| Sprint 11 | Amelioration IA + stockage | I | R | I |
| Sprint 12 | Attaque IA, interface selection | R | I | R |
| Sprint 13 | Affichage resultats attaque | R | C | R |
| Sprint 14 | Reglages modeles + requetes | C | I | R |
| Sprint 15 | Finalisation UI attaque | R | C | C |
| Sprint 16 | Optimisation IA | I | C | R |
| Sprint 17 | Dashboards + analyse avancee | R | C | C |
| Sprint 18 | Stabilisation labellisation + stockage | I | R | C |
| Sprint 19 | Tests/ validation plateforme | R | R | R |
| Sprint 20 | Livraison + deploiement | R | R | R |

### 3.3.4 Staffing

Voici le Staffing détaillé, réparti par sprint et affecté aux membres de l'équipe selon leur métier. Ce tableau montre quelle personne est responsable de chaque sprint et sur quelle période elle intervient.

| | Sprint 1 | Sprint 2 | Sprint 3 | Sprint 4 | Sprint 5 | Sprint 6 | Sprint 7 | Sprint 8 | Sprint 9 | Sprint 10 | Sprint 11 | Sprint 12 | Sprint 13 | Sprint 14 | Sprint 15 | Sprint 16 | Sprint 17 | Sprint 18 | Sprint 19 | Sprint 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Data Ing 1 | * | | * | | | * | | | * | | | * | * | | | * | * | | * | * |
| Data Ing 2 | * | | * | * | | * | | * | * | * | * | * | * | | * | | * | * | * | * |
| Data Ing 3 | * | | * | * | | * | * | | * | | | | * | | | * | | * | * | * |
| Data Sci 1 | | * | * | * | | | | * | * | * | | | * | | | | * | * | * | * |
| Data Sci 2 | | * | * | | * | * | | | * | * | | | * | * | | | | | * | * |
| Dev web 1 | * | * | | | * | * | | | * | * | * | | * | | * | | * | * | * | * |
| Dev web 2 | | | | * | * | | * | | * | | * | * | * | | | * | | | * | * |
| Dev web 3 | | | | | * | | | | * | | * | * | | * | | * | | * | * | * |

Les ingénieurs infra et cybersécurité seront mobilisés en fonction des besoins ponctuels tout au long du projet, sans répartition fixe.
Les Product Owners et le Scrum Master, quant à eux, seront présents en continu sur l'ensemble des sprints pour assurer le pilotage et la coordination.

### 3.3.5 Point de vigilance

- Un point de vigilance important concerne la charge élevée sur les Data Scientists, qui sont moins nombreux (2 personnes) mais responsables de nombreux sprints critiques liés à l'IA. Cela peut entraîner un retard si une charge imprévue apparaît sur les phases de training, optimisation et attaque IA. Il faudra anticiper en priorisant les tâches et en envisageant un soutien ponctuel si nécessaire, éventuellement via un sous-traitant. Cependant, cette option impliquerait de mobiliser une partie du budget de provision, ce qui doit être soigneusement évalué en fonction des marges disponibles.

- Nous avons dans notre équipe une personne présentant un handicap léger de type trouble de l'attention. Ce trouble se manifeste principalement par une difficulté à rester concentré sur une tâche pendant de longues périodes, avec une tendance à se disperser ou à perdre temporairement le fil des activités en cours. Cela n'impacte pas ses capacités cognitives ni sa compréhension des enjeux techniques, mais peut rendre plus difficile la gestion de tâches complexes sous pression ou très interdépendantes. Afin de favoriser son intégration et de lui offrir un cadre de travail adapté, nous avons choisi de l'orienter vers des sprints à faible impact bloquant, avec des délais plus souples et des missions qui permettent une certaine autonomie. Cette organisation permet de préserver son efficacité tout en évitant une surcharge mentale inutile. Elle s'inscrit dans une logique d'équipe inclusive, où chacun peut contribuer à hauteur de ses compétences, dans un environnement respectueux et adapté.

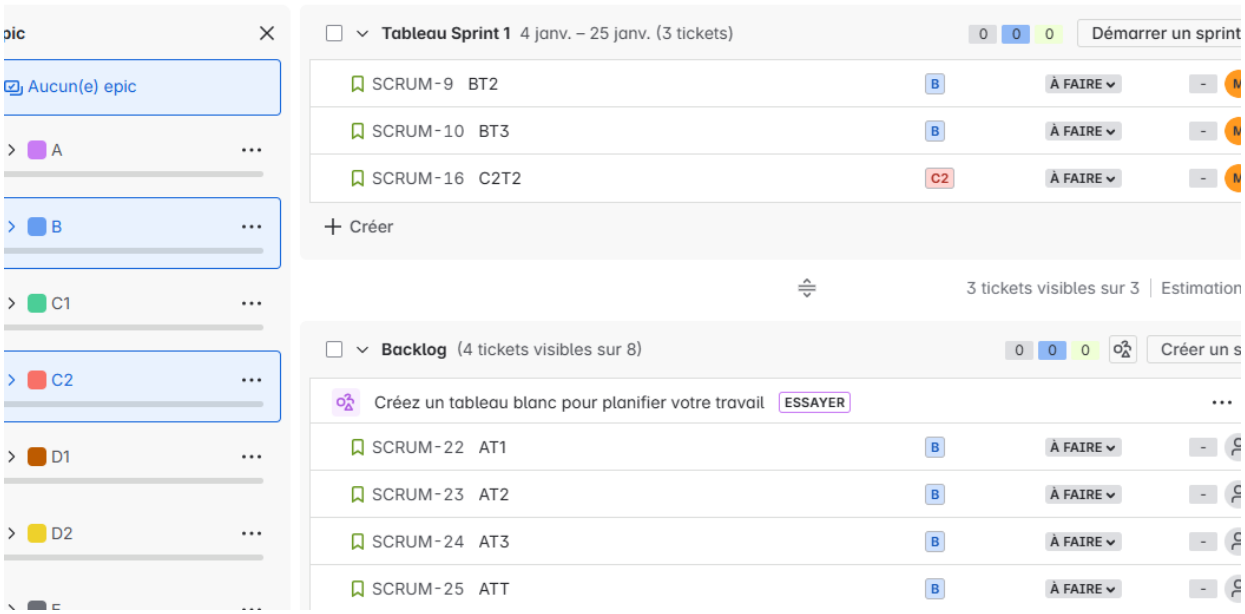
## 3.4 Suivi de projet (C3.2.2)

### 3.4.1 Création du Backlog et sprint planning

Comme outils de pilotage de projet j'ai décidé d'utiliser Jira. C'est un logiciel qui est particulièrement bien adapté au suivi de projets Agile, car il permet de structurer facilement le Backlog, de gérer les sprints, et d'assurer un suivi clair des tâches, des responsabilités et de l'avancement.

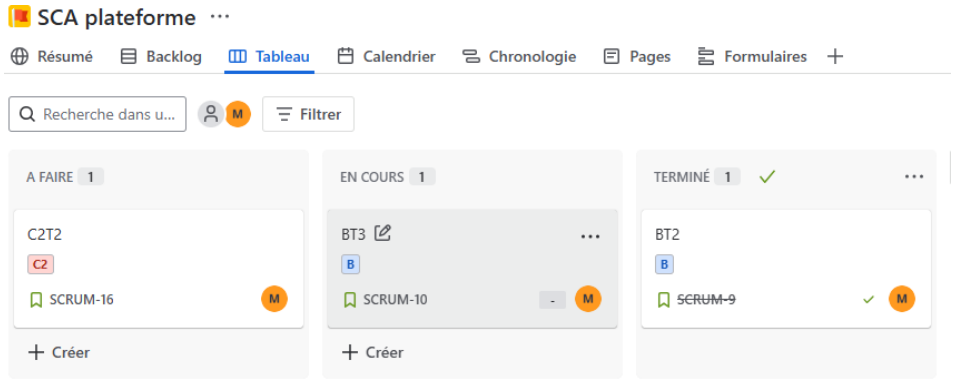
Pour de démarrer le suivi de projet avec Jira en mode Agile, j'ajoute toutes mes User Stories dans le Backlog ensuite je les associes aux épics ce qui est relativement facile ici car j'ai déjà défini que les épics correspondent aux composant principaux de la plateforme et les User Stories sont les tâches associées (ex : la User Stroy « BT2 » fait partie de l'Epic B).
Dans une approche Agile, on planifie généralement au maximum 2 à 3 sprints à l'avance, afin de conserver la flexibilité nécessaire pour s'adapter aux retours utilisateurs et aux imprévus techniques identifiés lors des Sprint Planning, en revanche les Epics eux sont planifiés à long terme c'est ce qui nous a permis de créer précédemment le planning global des sprints.

Vient ensuite ce qu'on appelle le sprint planning, c'est à dire choisir les US qui feront partie du Sprint 1. Comme défini précédemment l'objectif du sprint 1 est de connecter les données provenant d'un SDR (équipement hardware radio qui veut dire software define radio) avec Grafana, on avait défini que ce sprint regroupe une partie de l'Epic B et C2. De ce fait on va retrouver dans le sprint 1 trois US provenant de l'Epic B et C2 comme on peut le voir ci-contre :

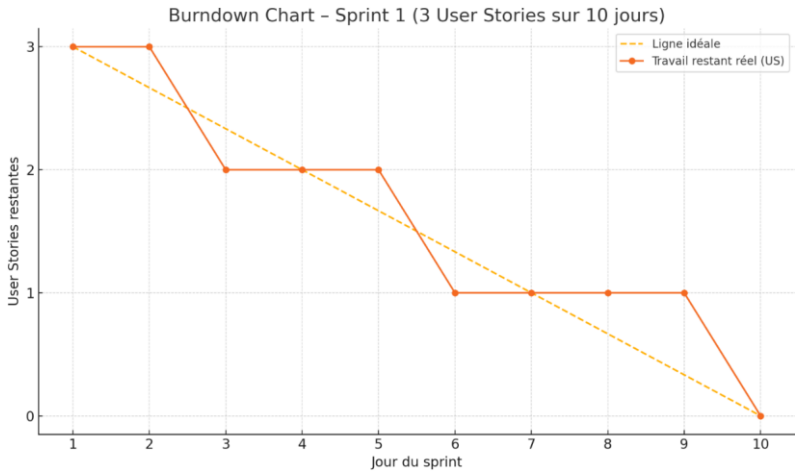


### 3.4.2 Gestion de l'avancement

Pour suivre l'avancement des sprints j'utilise principalement 2 méthodes. Premièrement le Sprint Board qui permet tout simplement de définir le statut de l'US (en cours, à faire ou terminé), ci-dessous voici le Sprint Board du sprint 1 :




Ensuite, j'utilise le graphique d'avancement ci-dessous (burndown chart) associé au sprint board afin de suivre en temps réel la progression des tâches et de m'assurer que le rythme de développement reste cohérent avec les objectifs fixés. Ce graphique présente le graphique d'avancement du Sprint 1, avec un suivi de l'avancement basé sur 3 User Stories planifiées sur 2 semaines (10 jours). La courbe permet de visualiser la progression réelle par rapport à l'avancement idéal, en identifiant les moments de blocage.




En parallèle du suivi d'avancement j'utilise un suivi de budget simple sur Excel, ci-dessous le suivi budgétaire le jour 6 du sprint 1. On observe que l'équipe a complété environ 2 User Stories sur les 3 prévues, soit les deux tiers du sprint, ce qui indique que l'avancement est conforme aux attentes à ce stade et que l'on est même légèrement en avance, d'où l'écart de -600€ sur le prévisionnel.

| Suivi budgetaire sprint 1 Jour 6 | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Tâche | Description | Métier | Jours Plannifiés | Jours Réalisés | TJM (€) | Coût estimé (€) | Coût réel (€) | Ecart (€) | % Consommation |
| BT2 | Script de collecte des données avec pyRTL-SDR et pyVisa | DATA ING | 10 | 8 | 600 | 6000 | 4800 | 0 | 80% |
| BT3 | Transformation en formats exploitables | DATA ING | 10 | 6 | 600 | 6000 | 3600 | 0 | 60% |
| C2T2 | Creation des dashboards | DATA SCI | 5 | 6 | 600 | 3000 | 3000 | -600 | 80% |

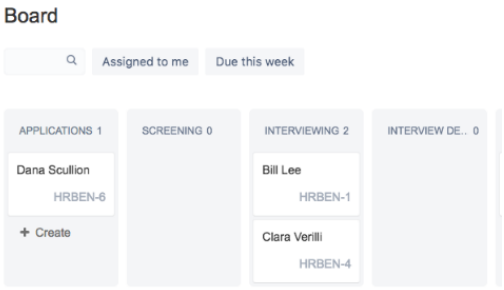
# 4 Pilotage de l'équipe projet et outils de communication (C3.3.2)

Le pilotage du projet s'appuie sur une organisation Agile structurée autour de sprints et de composants techniques indépendants (collecte EM, IA, stockage, interface...). Cela permet d'affecter les missions en fonction des expertises spécifiques de chaque membre de l'équipe notamment data scientist, data ingénieur et développeur web.

La répartition de la charge de travail est équilibrée entre les différents profils métier grâce à l'outil Jira, qui centralise les tâches, à qui elles seront attribuées et défini l'ordre des priorités ainsi que l'avancement. Pour rappel les responsabilités sont précisées à travers une matrice RACI par sprint.

Plusieurs outils de communication collaboratifs ont été utilisés :

- **Jira** facilite le pilotage de l'équipe projet en permettant la création de tickets, incidents ou demandes particulières. (Par exemple Organiser et inviter l'équipe projet aux cérémonies Scrum ou créer un ticket s'il y a eu un problème de répartition des charges) Il permet aussi évidemment de créer un Board (exemple ci-contre) pour visualiser le Staffing des tâches assignées à chaque membre.




- **Slack** et **Teams** sont utilisés comme outils de communication instantanée pour faciliter les échanges rapides au sein de l'équipe projet. Ils permettent de centraliser les discussions par canal thématique (technique, organisation, support) et d'assurer une meilleure réactivité dans la coordination quotidienne.

- **GitHub** est utilisé pour le versionnage du code, la revue collaborative des développements, et la gestion des branches par fonctionnalité. **Google Drive** complète cet écosystème en centralisant les documents partagés (spécifications, comptes rendus, livrables), assurant ainsi une accessibilité fluide et un archivage structuré.

- **Les rituels Agiles**, en plus du Sprint Planning, deux autres cérémonies Scrum sont utilisées lors du pilotage de projet, le Daily Scrum et le Sprint Review:
  Le Daily Scrum est une réunion courte et quotidienne durant laquelle chaque membre de l'équipe partage ce qu'il a accompli la veille, ce qu'il prévoit de faire aujourd'hui et les éventuels blocages rencontrés. Cela permet à l'équipe de rester synchronisée, d'ajuster rapidement si besoin, et de détecter les obstacles en temps réel. Le Sprint Review a lieu en fin de sprint. L'équipe présente le travail terminé (souvent sous forme de démo) au Product Owner et aux parties prenantes. L'objectif est de montrer la valeur livrée recueillir des retours et ajuster le backlog si nécessaire.

- Pour rappel un membre de l'équipe présente un handicap léger de type trouble de l'attention, se traduisant par des difficultés à maintenir sa concentration sur des tâches longues ou complexes sous pression. Pour favoriser son intégration, il a été orienté vers des missions à faible impact bloquant, avec davantage d'autonomie et des délais plus souples. Son poste de travail a également été aménagé. Il est placé coté fenêtre et non coté couloir afin d'éviter les zones de passage pour limiter les perturbations, et des ajustements ont été faits sur la luminosité et le contraste de l'écran afin d'améliorer son confort visuel.

# 5 Plan de développement des compétences (C3.3.1)

Les compétences clés mobilisées dans le cadre de ce projet sont variées et couvrent plusieurs domaines techniques et technologiques de pointe. Plus précisément, le projet a nécessité une expertise en collecte de signaux radiofréquences et électromagnétiques (RF/EM), en traitement et analyse de données massives (big data), en conception et développement d'interfaces utilisateurs adaptées aux besoins métiers, ainsi qu'en entraînement et déploiement de modèles d'intelligence artificielle (IA) appliqués à des problématiques complexes telles que la cryptanalyse ou la détection de signaux.

Afin d'identifier les forces internes et les axes d'amélioration, une grille d'analyse des compétences a été élaborée. Celle-ci a permis de dresser un état des lieux des savoir-faire déjà présents au sein de l'équipe projet, en les croisant avec les compétences requises pour chaque composant technique identifié (collecte RF/EM, pipeline de traitement temps réel, modules IA, conformité aux standards SCA). Cet outil d'évaluation a constitué la base d'un plan structuré de montée en compétences, conçu en concertation avec les membres de l'équipe technique et les ressources humaines.

Ce plan de montée en compétences repose sur plusieurs actions ciblées, dont les principales sont :

- Une formation spécialisée en cryptanalyse par intelligence artificielle, destinée aux data scientists impliqués dans les aspects analytiques du projet. Cette formation vise à renforcer leur capacité à traiter des données chiffrées ou bruitées à l'aide d'algorithmes IA avancés.

- Un accompagnement technique sur les normes de sécurité SCA (Side Channel Analysis) et la norme ISO 17825, afin d'assurer que les développements respectent les standards internationaux en matière de sécurité des circuits intégrés face aux attaques par canaux auxiliaires.

- Des modules de formation en ligne (E-learning) sur des technologies spécifiques, telles que l'utilisation d'InfluxDB version 3 pour la gestion des séries temporelles, la classification automatisée d'images de spectres, ainsi que la numérisation et l'interprétation de signaux RF dans un contexte de surveillance ou d'analyse.

| Activité | Metier | Objectif | Type de formation |
|---|---|---|---|
| Cryptanalyse IA | DATA SCI | Maitriser l'analyse de fuites avec CNN/LSTM | Formation spécialisée présentielle |
| Normes de sécurité SCA | Tous les membres technique | Comprendre la norme ISO17825 | Acompagnement documentaire + webinaire |
| Nouvelles technologies | DATA ING/ DATA SCI | Prendre en main InfluxdbV3, classification de spectrogramme | E-learning + tutoriels |

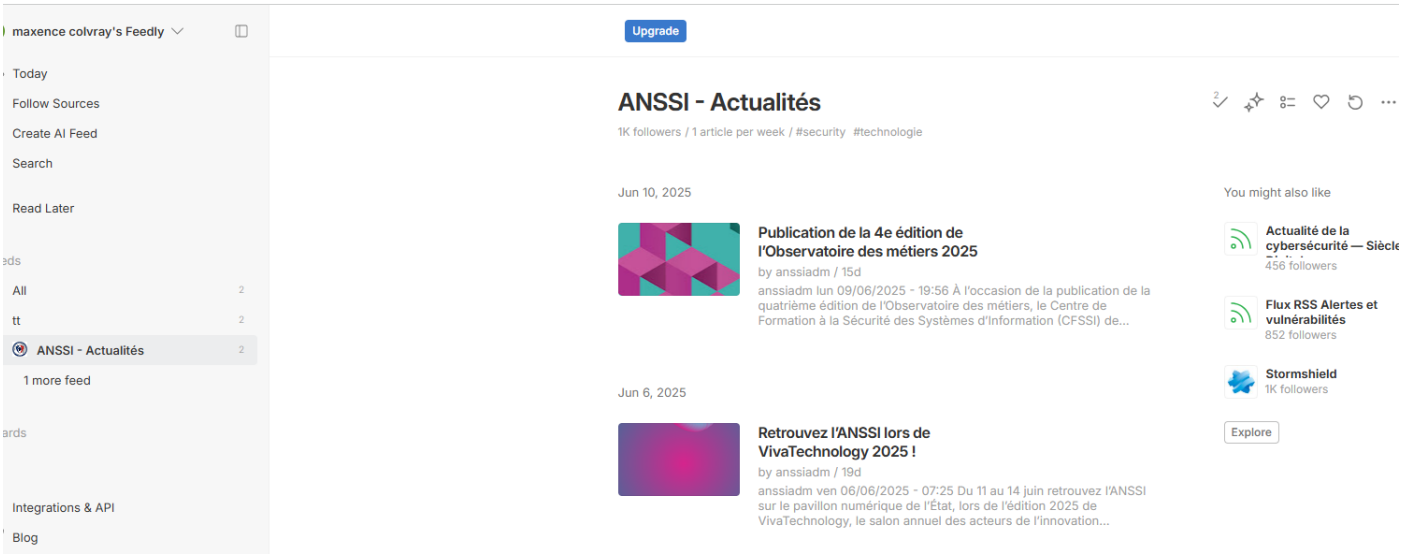
# 6 Mise en place d'un système de veille technologique et réglementaire (C3.4.1)

Dans le cadre du projet, une veille technologique continue a été mise en place afin de suivre les évolutions dans trois domaines clés : la cryptanalyse SCA (Side Channel Analysis), l'intelligence artificielle appliquée à la cybersécurité, et les normes de protection des systèmes, telles que l'ISO/IEC 17825. Cette veille joue un rôle stratégique, car elle permet non seulement d'orienter les choix techniques en temps réel, mais aussi d'anticiper les impacts réglementaires, les nouvelles vulnérabilités identifiées et les bonnes pratiques en matière de sécurité des systèmes embarqués.

La méthodologie adoptée repose principalement sur l'utilisation de l'outil **Feedly**, largement reconnu dans les milieux techniques pour sa capacité à centraliser et automatiser la surveillance d'un large éventail de sources fiables. Les flux agrègent des contenus issus de plateformes scientifiques (IEEE, ArXiv), d'organismes de référence (ANSSI, CNIL), ainsi que de blogs spécialisés, permettant ainsi une couverture à la fois académique, industrielle et opérationnelle.

Cette démarche est complétée par une participation régulière à des webinars spécialisés, en particulier sur les thématiques SCA et IA appliquée à la cybersécurité, afin de rester en contact direct avec les retours d'expérience d'experts du domaine. En parallèle, des newsletters professionnelles (CNIL, ANSSI, ENISA) permettent de rester informé des évolutions réglementaires et des publications récentes en matière de conformité et de protection des données.

Chaque vendredi après-midi, un créneau de deux heures est réservé aux membres de l'équipe pour effectuer de la veille technologique sur les thématiques clés du projet. Ci-dessous, le tableau de bord Feedly dédié, incluant les actualités de l'ANSSI.



# 7 Cas d'arbitrage externalisation pipeline de traitement des données (C3.3.3)

Une incertitude est apparue concernant la réalisation des pipelines de données (pipeline en flux temps réel vers la base OLTP pour acquisition de traces RF/EM et pipeline batch Airflow vers la base OLAP).
Ces deux composants sont critiques pour la performance de la plateforme, mais nécessitent des compétences pointues (ETL, optimisation, monitoring temps réel) qui ne sont pas pleinement disponibles en interne pour l'instant.

Options envisagées :

- Internalisation complète, au risque d'un retard dû au manque de disponibilité et montée en compétence trop lente.
- Sous-traitance globale des deux pipelines à une société spécialisée, avec un coût élevé et une moindre maîtrise interne sur le long terme.
- Arbitrage mixte, recruter un data ingénieur pour les pipelines temps réel (OLTP), domaine où la réactivité et la proximité avec le matériel sont cruciales et sous-traiter la partie Airflow / OLAP, plus structurée, externalisable sans perte de cohérence

**Décision retenue :**
L'option 3 a été choisie car elle concilie maîtrise interne sur la couche sensible du projet, tout en optimisant le coût et les délais pour le traitement batch. Cette décision a été prise à l'issue d'une réunion tripartite RH/technique/pilotage projet. Ce réajustement permet de sécuriser les livrables critiques du MVP en garantissant la qualité et le respect des jalons mais permettra tout de même de monter en compétence en interne concernant la conception de pipeline en flux.

# 8 Plan d'actions, éthique, confidentialité et responsabilité environnementale (C3.4.2)

Dans le cadre du projet, un plan d'actions spécifique a été mis en œuvre afin d'intégrer de manière proactive les dimensions éthiques, juridiques et environnementales liées à la gestion de données sensibles. Ce plan s'inscrit dans une démarche responsable, en cohérence avec les principes de la science des données appliquée à des contextes sensibles tels que la radiofréquence (RF), la cryptanalyse, ou encore l'intelligence artificielle.

**Confidentialité et protection des données**
Le traitement de données issues de signaux électromagnétiques ou radiofréquences (EM/RF) soulève des enjeux importants en matière de confidentialité, notamment en lien avec les métadonnées associées à ces signaux. Pour répondre à ces exigences :

Une anonymisation systématique des métadonnées critiques (telles que les horodatages précis, les identifiants de source matérielle ou de capteurs) a été mise en place dans les bases de données OLTP/OLAP, limitant ainsi les risques de corrélation ou d'identification indirecte.

Un guide de bonnes pratiques pour une IA éthique est disponible incluant les recommandations de la CNIL et de l'ANSSI. Il encadre l'utilisation des modèles d'IA afin de garantir la transparence, la non-discrimination, la supervision humaine, et la traçabilité des décisions algorithmiques, ces bonnes pratiques sont maintenant réglementés à travers l'IA ACT européen.

**Responsabilité environnementale**
Conscients de l'impact environnemental des traitements de données à grande échelle, nous avons également pris des mesures visant à optimiser la consommation énergétique :

Le recours à des containers légers et à des micro-services permet de réduire significativement l'empreinte carbone des déploiements, tout en assurant la scalabilité et la portabilité de la plateforme. Ces choix techniques s'alignent avec les objectifs de sobriété numérique promus dans le cadre des démarches RSE.

**Déploiement et maîtrise des coûts**
Les actions prévues sont intégrées dans la phase MVP du projet, plus précisément entre les sprints 4 et 8, afin de garantir leur prise en compte dès les premières versions testables de la plateforme.
Le coût global de ce plan reste maîtrisé, représentant moins de 5 % du budget total, grâce à :

- L'utilisation prioritaire de solutions open source ;
- Une implication ciblée des ingénieurs pour limiter la charge projet ;
- Une revue de sécurité restreinte mais ciblée, focalisée sur les points de vigilance identifiés.

**Résultats attendus**
Les bénéfices attendus de ce plan d'actions sont multiples :

- Une meilleure conformité réglementaire, notamment vis-à-vis du RGPD, des recommandations de l'ANSSI et des normes ISO 17825 ;
- Une réduction des risques juridiques liés à la gestion des données sensibles ;
- Un renforcement de la confiance des partenaires institutionnels et industriels, grâce à une meilleure transparence et à une gouvernance claire des flux de données.

# A.1 Annexes

## A.1.1 Terminologie

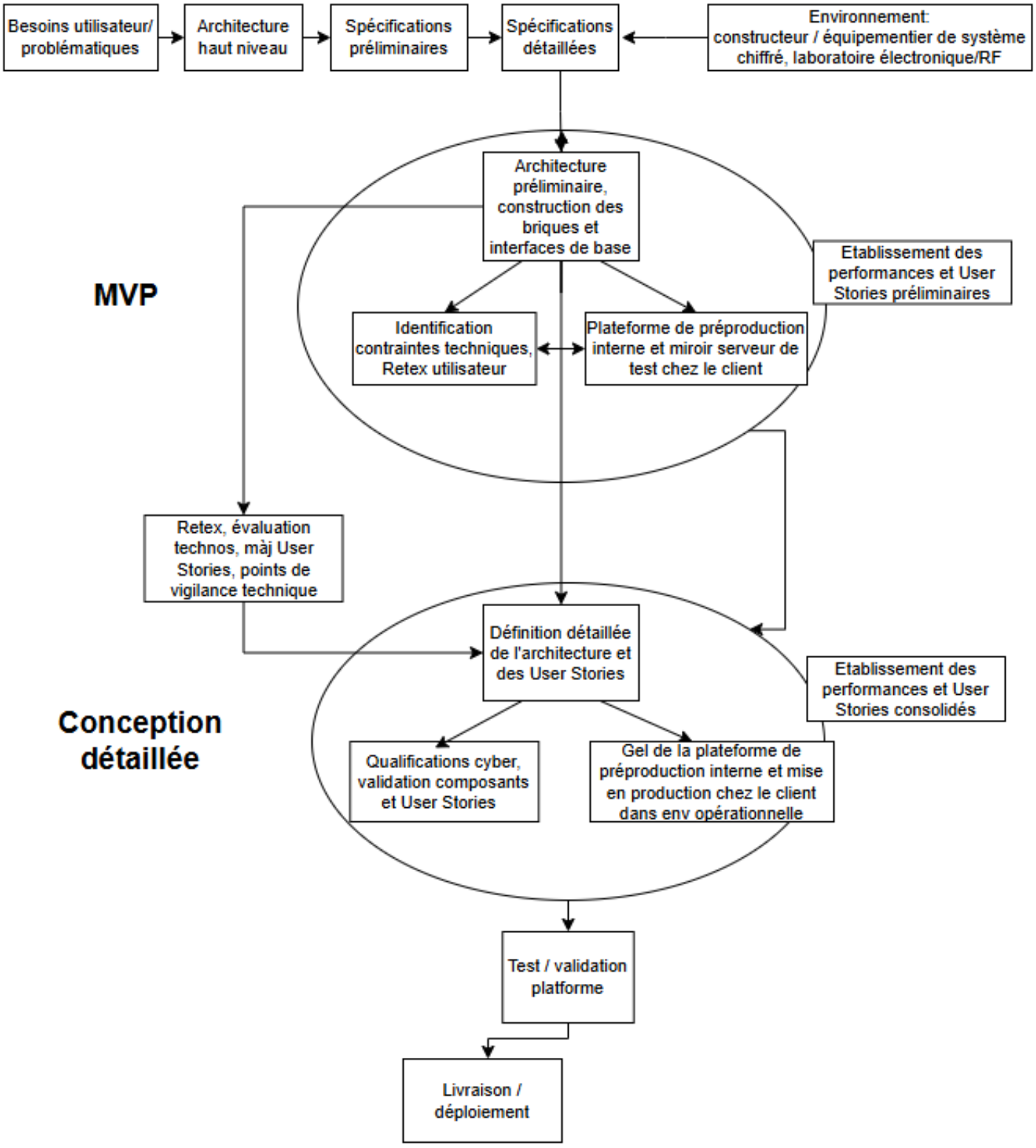
| | |
|---|---|
| SCA | Side Channel Attack |
| EM | Electromagnétique |
| BDD OLTP | Base de données Online Transactional Processing |
| BDD OLAP | Base de données Online Analytical Processing |
| RF | Radiofréquence |
| CTO | Chief Technology Officer |
| PO | Product Owner |
| SM | Scrum Master |
| SDR | Software Define radio |
| US | User story |
| DATA SCI | Data Scientist |
| DATA ING | Data Engineer |
| CNIL | Commission nationale de l'informatique et des libertés |
| ANSSI | Agence nationale de la sécurité des systèmes d'information |

## A.1.2 Tableau ressources humaines nécessaires

| | ID tache | Description de la tache | metier | durée |
|---|---|---|---|---|
| A | AT1 | Rédaction des specs detaillées | CTO | 1 |
| | AT2 | Frontend de l'application | dev web | 1 |
| | AT3 | Backend avec declenchement du composant B | dev web | 3 |
| | ATT | Tests/validation | Testeur | 1 |
| B | BT1 | Rédaction des specs detaillées | CTO | 1 |
| | BT2 | Script de collecte des données avec pyRTL-SDR et pyVisa | data ing | 2 |
| | BT3 | Transformation en formats exploitables | data ing | 2 |
| | BT4 | Insertion en base | data ing | 2 |
| | BTT | Tests/validation | Testeur | 1 |
| C1 | C1T1 | Rédaction des specs detaillées | CTO | 1 |
| | C1T2 | Schéma de données partitionnement temporel et spatial | data ing | 1 |
| | C1T3 | Optimisation des schémas de tags et fields | data ing | 1 |
| | C1T4 | Batch writing et compression | data ing | 1 |
| | C1T5 | Requêtes optimisées et indexation | data ing | 2 |
| | C1TT | Tests/validation | Testeur | 1 |
| C2 | C2T1 | Rédaction des specs detaillées | CTO | 1 |
| | C2T2 | Création des dashboards | data sci | 1 |
| | C2T3 | Création des requêtes | data sci | 2 |
| | C2TT | Tests/validation | Testeur | 1 |
| D1 | D1T1 | Rédaction des specs detaillées | CTO | 1 |
| | D1T2 | Définition des DAGs | data ing | 1 |
| | D1T3 | Transformation et insertion des données dans Airflow | data ing | 2 |
| | D1TT | Tests/validation | Testeur | 1 |
| D2 | D2T1 | Rédaction des specs detaillées | CTO | 1 |
| | D2T2 | Déploiement de LabelStudio connecté à la base OLAP | dev web | 2 |
| | D2T3 | Création de templates de tâches de labellisation | dev web | 5 |
| | D2TT | Tests/validation | Testeur | 1 |

| | ID tache | Description de la tache | metier | durée |
|---|---|---|---|---|
| E | ET1 | Rédaction des specs detaillées | CTO | 1 |
| | ET2 | Modélisation du schéma de données analytique | data ing | 1 |
| | ET3 | Intégration des données transformées depuis l'ETL Airflow | data ing | 3 |
| | ET4 | Tests/validation | Testeur | 2 |
| | ETT | | | 1 |
| F1 | F1T1 | Rédaction des specs detaillées | CTO | 1 |
| | F1T2 | Création de templates de modèles IA | data sci | 1 |
| | F1T3 | Implémentation des modèles IA CNN et LSTM | data sci | 4 |
| | F1TT | Tests/validation | Testeur | 8 |
| F2 | F2T1 | Rédaction des specs detaillées | CTO | 1 |
| | F2T2 | Création du formulaire de configuration interactif | dev web | 1 |
| | F2T3 | Lancement de l'entraînement depuis l'interface | dev web | 3 |
| | F2TT | Tests/validation | Testeur | 2 |
| G2 | G2T1 | Rédaction des specs detaillées | CTO | 1 |
| | G2T2 | Création d'une interface multi-étapes pour l'attaque SCA | dev web | 1 |
| | G2T3 | Intégration avec la base OLAP et les modèles exportés | data ing | 3 |
| | G2T4 | Lancement de l'inférence et affichage des résultats | dev web | 2 |
| | G2TT | Tests/validation | Testeur | 2 |
| G1 | G1T1 | Rédaction des specs detaillées | CTO | 1 |
| | G1T2 | inférence à partir d'un modèle pré-entraîné | data sci | 1 |
| | G1T3 | Implémentation des métriques d'évaluation | data sci | 2 |
| | G1T4 | Export des résultats | data sci | 2 |
| | G1TT | Tests/validation | Testeur | 1 |
| Scrum master | SM | 15% de l'activité total | Scrum master | 55 |
| Product Owner | PO | 15% de l'activité total | Product owner | 55 |
| Infra | Infra | Assurer l'installation et les 2 déploiments de la plateforme | ing infra | 10 |
| Cyber | cyber | Garantir les exigences cybersécurtié de la plateforme | ing cyber | 12 |

### A.1.3 Schéma stratégies macroscopique de développement



## A.1.4 Planning des sprints

| Sprint | Epics / Lots | Description du sprint | Livrable / Objectif | Durée (en semaine) |
|---|---|---|---|---|
| **Sprint 1** | B, C2 | Collecte EM/RF, Visualisation temps réel | Connexion SDR/Grafana opérationnelle | 2 |
| **Sprint 2** | B, C1, C2 | Affichage temps réel, intégration SDR | Collecte stable, affichage synchro | 2 |
| **Sprint 3** | A, B, C1, C2 | Stockage bruts, ETL initial finalisé | Données stockées dans BDD OLTP (influxdb3) | 3 |
| **Sprint 4** | C1, D1 | Insertion et automatisation batch | ETL fonctionnel pour les données brutes | 2 |
| **Sprint 5** | C2, E | Structure analytique, modèle OLAP | OLAP prêt à agréger et interroger | 2 |
| **Sprint 6** | F1, E | Requêtes optimisées, pré-agrégations | Modèle analytique testé | 1 |
| **Sprint 7** | F1, F2 | Training IA CNN, config UI IA | Modèle CNN entraîné avec UI | 3 |
| **Sprint 8** | F1, F2 | Training LSTM, visualisation résultats | Modèle LSTM + export résultats | 2 |
| **Sprint 9** | Global | Déploiment du MVP au sein de l'entreprise partenaire | Commencer à récolter les premiers RETEX des utilisateurs finaux | 3 |
| **Sprint 10** | D2, E | Labellisation et structuration | Dataset annoté, prêt pour entraînement | 2 |
| **Sprint 11** | F1, C2 | Amélioration IA + stockage | Modèle IA affiné sur OLAP | 2 |
| **Sprint 12** | G1, G2 | Attaque IA, interface sélection | Interface complète d'attaque SCA | 3 |
| **Sprint 13** | G1, G2 | Affichage résultats attaque | Résultats IA affichés et exportables | 1 |
| **Sprint 14** | C2, F2 | Réglages modèles + requêtes | Performances IA améliorées | 3 |
| **Sprint 15** | G2, F2 | Finalisation UI attaque | Fonctionnement complet UI attaque | 3 |
| **Sprint 16** | G1, F1 | Optimisation IA | Modèle IA optimisé | 2 |
| **Sprint 17** | A, E | Dashboards + analyse avancée | Tableaux de bord analytiques disponibles | 3 |
| **Sprint 18** | C1, D2 | Stabilisation labellisation + stockage | Système stable pour annotation & stockage | 2 |
| **Sprint 19** | Global | Tests/ validation plateforme | Tests d'intégration validés | 3 |
| **Sprint 20** | Global | Livraison + déploiment | Installation chez le client et démonstration aux utilisateurs | 2 |

## A.1.5 Matrice RCI

| Matrice RCI | | | | |
|---|---|---|---|---|
| Sprint | Description | Dev Web | Data Ing | Data Scientist |
| Sprint 1 | Collecte EM/RF, Visualisation temps reel | I | R | C |
| Sprint 2 | Affichage temps reel, integration SDR | R | R | C |
| Sprint 3 | Stockage bruts, ETL initial finalise | R | R | C |
| Sprint 4 | Insertion et automatisation batch | C | R | I |
| Sprint 5 | Structure analytique, modele OLAP | I | R | C |
| Sprint 6 | Requetes optimisees, pre-agregations | I | C | R |
| Sprint 7 | Training IA CNN, config UI IA | C | I | R |
| Sprint 8 | Training LSTM, visualisation resultats | C | I | R |
| Sprint 9 | Deploiment du MVP | R | R | R |
| Sprint 10 | Labellisation et structuration | R | R | C |
| Sprint 11 | Amelioration IA + stockage | I | R | I |
| Sprint 12 | Attaque IA, interface selection | R | I | R |
| Sprint 13 | Affichage resultats attaque | R | C | R |
| Sprint 14 | Reglages modeles + requetes | C | I | R |
| Sprint 15 | Finalisation UI attaque | R | C | C |
| Sprint 16 | Optimisation IA | I | C | R |
| Sprint 17 | Dashboards + analyse avancee | R | C | C |
| Sprint 18 | Stabilisation labellisation + stockage | I | R | C |
| Sprint 19 | Tests/ validation plateforme | R | R | R |
| Sprint 20 | Livraison + deploiement | R | R | R |