



Inria

RAPPORT MISSION 4 : BILAN DE PROJET

Période du 25 Septembre 2023 au 25 Janvier 2024

Cycle MS2D 2022-2024 (Manager de Solutions Digitales et Data)

NOM DE L'ALTERNANT :

Ulrich KEMKA TAKENGNY

NOM L'ENCADRANT :

Mohamed EL AFRIT

NOM DE LA TUTRICE EN ENTREPRISE :

Frédérique SEGOND

Frédérique Segond

SOMMAIRE

Table des matières

1. Présentation de l'entreprise d'accueil	1
1.1. Inria : métiers et chiffres clés.....	1
1.2. Le service de l'apprenti : Mission Défense et Sécurité	1
1.3 Principales activités de l'apprenti	1
2. Présentation et déroulement du Projet Campaign :.....	2
2.1. Contexte et Objectifs	2
2.2. Planification et Méthodologie	2
2.2.1. Organisation du travail.....	2
2.2.2. Découpage du projet (lot1).....	3
2.3. Développement et Mise en Œuvre	4
2.3.1. Réalisation du projet.....	4
2.3.2. Défis techniques et scientifique.....	4
2.4. Description du projet en termes de ressources.....	5
2.4.1. Acteurs Impliqués	5
2.4.2. Méthodes et Outils Utilisés.....	5
3. Conduite du projet :.....	6
3.1. Pilotage du Projet.....	6
3.2. Gestion des réunions et suivi de l'avancement :	6
4. Bilan du projet :.....	7
4.1. Évaluation des objectifs initiaux et résultats obtenus	7
4.2. Stratégie et critères de qualité généraux	8
4.2.1. Approche de Tests et Assurance Qualité :	8
4.2.2. Critères de qualité généraux.....	9
5. Dimension économique	9
5.1. Budget prévisionnel du projet	9
5.2. Suivi du Budget au Cours du Projet	9
5.3. Indicateurs de Rentabilité	10
6. Conclusion.....	10

1. Présentation de l'entreprise d'accueil

1.1. Inria : métiers et chiffres clés.

Fondé en 1967, Inria, l'institut national de recherche en sciences et technologies du numérique, est un leader mondial dans son domaine. Avec 11 centres et plus de 20 antennes réparties stratégiquement, l'Institut emploie 2600 experts passionnés par l'avancement technologique et entrepreneurial. Inria se distingue par son engagement dans une multitude de domaines, allant de l'édition open source à la création de startups technologiques (Deeptech). L'Institut a joué un rôle clé dans la genèse de plus de 180 startups innovantes, reflétant son esprit pionnier et son expertise en matière de transfert technologique.

1.2. Le service de l'apprenti : Mission Défense et Sécurité

Lancée en mars 2020 et pilotée par Frédérique Segond, la Mission Défense et Sécurité est un axe stratégique d'Inria visant à renforcer la souveraineté numérique française et européenne. Cette mission rassemble une équipe spécialisée dans les projets de sécurité et de défense, jouant un rôle crucial dans la coordination et l'exécution des initiatives d'Inria dans ces domaines. En phase d'expansion dynamique pour accueillir de nouveaux projets d'envergure, l'équipe est en cours de croissance, cherchant à attirer des talents variés pour relever des défis technologiques et sécuritaires de plus en plus complexes

1.3 Principales activités de l'apprenti

En tant qu'apprenti sous la supervision du responsable informatique, je contribue activement au développement de l'outil pédagogique Intel Lab. Cette plateforme de simulation et d'expérimentation est essentielle pour faciliter la compréhension et la gestion de scénarios complexes. Mes responsabilités englobent le développement full stack, la réalisation de tests, l'intégration et la mise en production, tout en adoptant la méthode Agile pour assurer une gestion et un suivi de projet efficaces.

2. Présentation et déroulement du Projet Campaign :

2.1. Contexte et Objectifs

Le projet Campaign (Campagnes informAtionnelles Multimodales : imPAct, Identification, profilaGe et caractérisatioN) est un projet [RAPID](#) (Régime d'APpui à l'Innovation Duale), financé par l'Agence Innovation Défense et géré par [Storyzy](#). L'entreprise Storyzy est le porteur du projet, responsable de sa réalisation, de sa mise en œuvre et de son évaluation. Il se concentre sur l'identification et la caractérisation des campagnes informationnelles multimodales, en intégrant diverses modalités comme le texte, l'audio et les réseaux sociaux. L'objectif principal est de développer des outils et des méthodes pour analyser efficacement ces campagnes, en tenant compte de leurs aspects uniques et complexes.

2.2. Planification et Méthodologie

Le projet Campaign est structuré en cinq lots distincts, chacun ayant des objectifs précis alignés sur la mission générale du projet, qui vise à identifier et caractériser les campagnes informationnelles multimodales. Ces lots sont :

- **Identification des campagnes informationnelles multimodales** : Élimination des interférences sonores et conversion des fichiers vidéo et audio en texte.
- **Profilage et caractérisation des acteurs d'une campagne informationnelle** : Identification des campagnes orchestrées sur les plateformes sociales.
- **Evaluation et analyse d'impact d'une campagne informationnelle** : Jugement de l'impact et des méthodes des campagnes identifiées.
- **Démonstration et faisabilité** :
- **Evaluation et perspective** : Compréhension des origines et des motivations derrière les sources d'information.

2.2.1. Organisation du travail

Avant d'entrer dans le vif du projet, il a été mis en place la coordination du travail pour gestion administrative.

Ce travail est découpé en trois tâches et sera étendu sur la durée totale du projet (24 mois) :

- Tâche 1 : Rédaction de l'accord du consortium,
- Tâche 2 : Organisation des comités de suivi du projet,

Mission 4 : BILAN DE PROJET

- Tâche 3 : Identification des évènements important.

2.2.2. Découpage du projet (lot1)

Inria Défense et Sécurité travaille a travaillé en partenariat avec Storyzy sur le premier lot du projet. Nos objectifs à terme de ce lot sont :

- Débruitage et filtrage des fichiers vidéo et audio.
- Transcription de langues rares ou avec accents.
- Identification des locuteurs.
- Mise en place d'un pipeline de transcription capable de fonctionner avec des enregistrements audios présentant des accents, du bruit ou d'autres langues

La durée prévisionnelle de ce projet est de 24 mois. La réalisation de ces différentes taches est estimée en trimestre

Tableau 1: Découpage du lot en tâches

Taches					
Tache 1	Mise en place du filtre de débruitage du signal audio				
Tache 2	Evaluation sur un audio en langue française avec accent				
Tache 3	Evaluation en conditions réelle sur un audio de mauvaise qualité				
Tache 4	Evaluation réelles sur un audio en langue étrangère				

La chronologie du projet a été planifiée en phases correspondant aux tâches, avec des responsabilités spécifiques attribuées à chaque étape

L'approche semi-agile a été adoptée pour gérer le projet, avec des réunions de pilotage hebdomadaires pour assurer le suivi et l'adaptabilité du processus. Des réunions trimestrielles avec la DGA et Storyzy, ont permis de maintenir un alignement stratégique et de respecter le cahier des charges et les deadlines

2.3. Développement et Mise en Œuvre

2.3.1. Réalisation du projet

Nos objectifs sont centrés sur la capture et le traitement des signaux de parole afin d'ajouter une source d'information rapidement accessible par les analystes. Pour atteindre ces objectifs, nous avons envisagé une architecture qui repose sur les principes de modularité et d'indépendance des traitements. Cette architecture comprend un module d'adaptation de l'audio et le conditionnement du module de reconnaissance de la parole. De plus, nous prévoyons d'enrichir la transcription avec l'ajout de connaissances métier et d'évaluer la transcription en fonction des mots les plus pertinents pour l'identification des campagnes informationnelles.

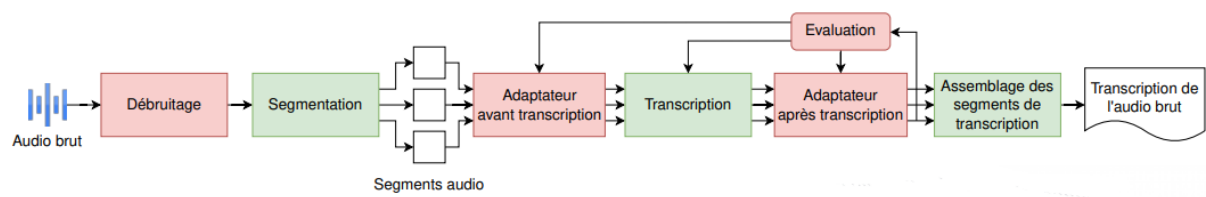


Figure 1: Architecture du pipeline

2.3.2. Défis techniques et scientifique

Le projet Campaign a rencontré plusieurs défis techniques et scientifiques significatifs, notamment :

- La gestion de **langues rares et peu documentées**, nécessitant des approches spécifiques pour la transcription et l'analyse.
- La diversité **des sources multimodales** telles que la radio et les réseaux sociaux, exigeant une capacité d'adaptation et de traitement varié.
- La présence **de bruits variés**, y compris le bruit de capture, le bruit ambiant et de transmission, qui complexifient l'analyse audio.
- La difficulté liée à **la superposition de sons**, comme la musique de fond qui interfèrent avec la clarté de la parole.
- La nécessité de gérer **des accents spécifiques**, qui pose des défis uniques en termes de reconnaissance vocale.
- Le traitement de **vocabulaires spécifiques**.
- L'évolution constante des **spécificités linguistiques**, nécessitant une mise à jour continue des modèles et approches.

Ces défis ont requis des solutions innovantes et adaptatives pour garantir l'efficacité du projet.

2.4. Description du projet en termes de ressources

2.4.1. Acteurs Impliqués

Le projet a impliqué une équipe interdisciplinaire. En plus de l'équipe de Storyzy, nous avons comme membre :

- Frédérique, la directrice de la Mission Défense et Sécurité
- Emilie, gestionnaire du budget et des aspects administratifs
- Jean-François, Directeur de recherche et chercheur référent
- Jonas, responsable de projet, garantissant le respect des délais et la qualité du travail
- Heiko, Ingénieur, en charge de la mise en place du pipeline de transcription
- Pierre, en charge de l'amélioration des modèles et de l'évaluation des résultats

2.4.2. Méthodes et Outils Utilisés

Plusieurs outils et méthodes ont été utilisés :

Pour le Développement du pipeline et le modèle, :

- Le modèle Whisper d'OpenAI pour la reconnaissance vocale
- Poetry comme outil de gestion de dépendances
- Scikit-learn et SpeechBrain pour le traitement des données et la reconnaissance vocale

Pour l'Assurance Qualité du code :

- GitLab pour le versioning et la collaboration
- La bibliothèque unittest pour les tests unitaires
- Pylint pour le linting de code
- Coverage report pour le suivi de la couverture des tests

Pour la Gestion Budgétaire :

- Outils collaboratifs comme SharePoint
- Suivi par un chargé de contrat de recherche (CCR) via l'ERP ECSAIE

Cette combinaison de ressources humaines et techniques a permis de relever les défis techniques du projet, tels que le traitement des langues peu dotées et l'adaptabilité aux différentes sources de bruit et accents.

3. Conduite du projet :

3.1. Pilotage du Projet

Le pilotage du projet "Campaign" a été orchestré suivant une méthodologie semi-agile, permettant une grande réactivité et adaptabilité face aux exigences changeantes du projet. Le responsable de projets informatiques, Jonas, a joué un rôle clé dans la surveillance du respect des échéances et la coordination des différentes équipes. Des outils de gestion de projet et des tableaux de bord ont été mis en place pour permettre une visualisation en temps réel de l'avancement et des ressources

3.2. Gestion des réunions et suivi de l'avancement :

Les réunions du projet Campaign sont organisées, avec des agendas préétablis et des buts définis pour garantir une contribution active et la collaboration de toute l'équipe. Les réunions se concentrent sur le partage des avancées, la résolution collective des problèmes et la planification future. Les procès-verbaux rédigés sur un SharePoint et distribués après les réunions assurent une documentation et un suivi consciencieux des résolutions et des tâches futures, qui sont cruciales pour la mémoire organisationnelle du projet.

Les réunions au sein de l'équipe Défense et Sécurité ont lieu chaque semaine, et celles entre l'Inria Défense et Sécurité et ses partenaires ont lieu tous les trimestres.

Mission 4 : BILAN DE PROJET

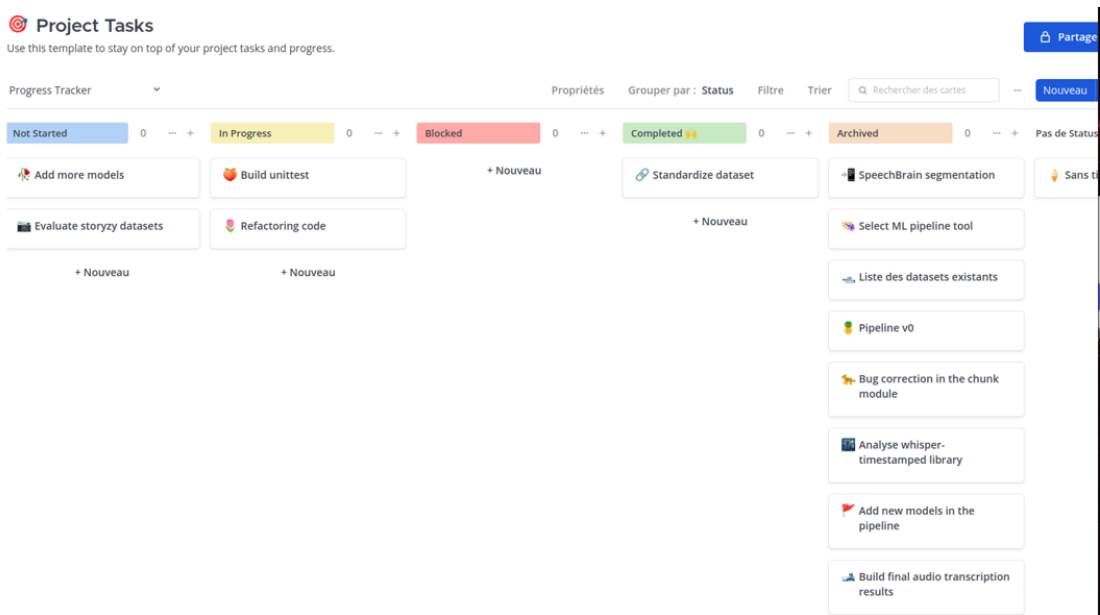


Figure 2: Etat d'avancement du développement du pipeline

4. Bilan du projet :

4.1. Évaluation des objectifs initiaux et résultats obtenus

Le projet Campaign a pour objectif l'identification et la caractérisation des campagnes informationnelles sur plusieurs modalités (texte, audio, réseaux sociaux). Le premier lot, sur lequel notre équipe s'est concentrée, visait la capture et le traitement de signaux de parole, facilitant ainsi l'accès à des informations cruciales pour les analystes et permettant l'identification de tendances au sein des médias locaux.

Tableau 2: Résultats attendus du lot 1

Livrables	Responsables	Délais prévisionnels	Retard
Document de l'état de l'art du nettoyage du signal audio	Inria	$T_0 + 3$	$T_{prev} + 6$
Document d'évaluation du comportement du filtre en conditions réelles sur les trois cas d'usage	Inria & Storyzy	$T_0 + 21$	$T_{prev} + 6$

Tableau 3: Planning prévisionnel du Lot 1 :

Tâches \ Année - Trimestres	AN1-T1	AN1-T2	AN1-T3	AN1-T4	AN2-T1	AN2-T2	AN2-T3	AN2-T4
1. Mise en place du filtre de débruitage du signal audio								
2. Evaluation sur un audio en langue française avec accent								
3. Evaluation en conditions réelle sur un audio de mauvaise qualité								
4. Evaluation réelles sur un audio en langue étrangère								

Malgré des progrès significatifs, comme le développement d'un pipeline de transcription robuste et la création d'outils de reconnaissance vocale pour des conditions d'utilisation réelles, le projet a subi des retards substantiels. Ces contretemps ont été principalement dus à des difficultés dans le recrutement de personnel qualifié, entraînant un décalage de deux trimestres par rapport au planning initial.

La transparence dans l'identification des causes de ces retards est cruciale pour l'ajustement des stratégies et la prévention de tels écarts dans les phases subséquentes du projet.

4.2. Stratégie et critères de qualité généraux

4.2.1. Approche de Tests et Assurance Qualité :

Le développement du projet obéit à certains critères définis en amont dans le cahier de charges.

Quelques critères et outils d'aides utilisés :

- **Tests Unitaires et Fonctionnels** : Utilisation de la librairie *unittest* pour garantir la solidité des composants individuels et des fonctionnalités globales du système.
- **Documentation et Style de Code** : Application de *Pylint* pour maintenir un style de code cohérent et *autodocstring* pour documenter clairement les fonctions.
- **Tests de Performance et d'Évolution** : Évaluation régulière des performances du système pour s'assurer qu'il ne régresse pas par rapport aux versions précédentes et comparer les améliorations apportées par les nouveaux modèles.

4.2.2. Critères de qualité généraux

Dans le cadre de cette partie du projet, les critères principaux de réussites du premier lot sont :

- **Pertinence des Informations** : Assurer la remontée d'informations pertinentes et précises pour une recherche spécifique.
- **Système de Reconnaissance Vocale** : Développer un système de reconnaissance vocale fiable et fonctionnel en conditions réelles, capable de s'adapter à divers scénarios.
- **Performance de Reconnaissance** : Prioriser la performance de reconnaissance sur les mots et phrases les plus pertinents pour le contexte de l'enquête.

5. Dimension économique

NB : Pour des raisons de confidentialité, nous ne pouvons pas fournir de valeurs, même approximatives.

5.1. Budget prévisionnel du projet

Le projet Campaign dispose d'un budget prévisionnel détaillé réparti par lots et catégories de dépenses. Voici le tableau récapitulatif :

Tableau 4: Tableau de Budget Prévisionnel

Lot		Coût global HT > 1 000 000 €				% du Coût global du Projet
		Personnel	Equipements & matériels	Missions	Sous-traitance	
0. Coordination/Gestion	Storyzy	0	0	0	0	0
	Inria	0	0	0	0	
1 Identification des campagnes info...	Storyzy	X	X	X	X	35
	Inria	X	X	X	X	
2 à 5	Storyzy	X	X	X	X	65
	Inria	X	X	X	X	
Projet global (%)		78	12	8	2	100

5.2. Suivi du Budget au Cours du Projet

Le suivi budgétaire est réalisé en collaboration avec toutes les parties prenantes, garantissant la transparence et l'alignement avec les objectifs du projet.

Tableau 5: Tableau de Suivi du Budget

Catégorie	Prévu (€HT)	Réel (€HT)	Écart (€HT)
Personnel	X	X	X
Missions			X
Equipements & matériels	X	X	X
Sous-traitance	X	X	X

5.3. Indicateurs de Rentabilité

Les indicateurs de rentabilité sont évalués en fonction de l'impact technologique et de l'intégration des recherches dans des applications pratiques. L'efficacité coût-bénéfice est un facteur clé dans l'évaluation du succès économique du projet

6. Conclusion

Le projet Campaign se distingue par son approche innovante en matière de reconnaissance vocale et d'identification des campagnes informationnelles multimodales. Les critères de qualité définis et une gestion budgétaire prudente ont été des facteurs clés dans l'avancement du projet. Malgré les défis inhérents à tout projet de recherche et développement, l'équipe a démontré une adaptabilité et des compétence technique et managériale qui ont permis de surmonter les obstacles rencontrés.

Le suivi régulier et la méthodologie agile semi-structurée ont facilité une gestion efficace, tandis que les tests unitaires et de performance ont assuré le développement d'un produit fiable et pertinent. La collaboration entre Storizy et Inria, soutenue par des financements ciblés, a favorisé la création de solutions qui pourront potentiellement transformer les méthodes de surveillance et d'analyse informationnelle.

Le projet Campaign représente non seulement un succès technique mais aussi un modèle de gestion de projet interdisciplinaire et inter-institutionnel. Les enseignements tirés de cette expérience enrichiront les futures initiatives, et la technologie développée promet de contribuer significativement à la sécurité et à l'analyse informationnelle à l'échelle nationale et internationale.