

Phase 1 - Rapport de Projet PFE

Phase 1 - Rapport de Projet PFE

Titre du projet PFE :

Développement d'un système intelligent de validation d'images d'interventions fibre optique à l'aide de modèles multimodaux et d'agents RAG

Paragraphe 1-2 : Contexte du problème et description du domaine

Dans le cadre de l'expansion de la fibre optique, Canal Plus Telecom souhaite optimiser la qualité des installations techniques réalisées par ses techniciens. Actuellement, la validation de la conformité des interventions repose sur des vérifications manuelles, générant un risque d'erreurs, de lenteur de traitement et d'insatisfaction client. Le projet propose de créer un système intelligent capable d'analyser automatiquement les images d'intervention pour valider ou identifier les éventuelles non-conformités (installation aérienne, souterraine, façade, présence de matériel obligatoire, etc.).

Le domaine abordé est à l'intersection de l'intelligence artificielle appliquée à la vision par ordinateur et des systèmes d'agents conversationnels multimodaux. Ce projet contribue à la transformation digitale du contrôle qualité opérationnel et s'inscrit dans les objectifs stratégiques de fiabilisation et d'automatisation des processus terrain.

Paragraphe 3-4 : Revue de la littérature et état de l'art

La validation automatique d'images repose historiquement sur les réseaux de neurones convolutifs (CNN) tels que ResNet et EfficientNet. Récemment, les modèles transformateurs visuels (ViT) et les modèles multimodaux comme CLIP ou GPT-4 Vision offrent une capacité inégalée d'analyse et de compréhension contextuelle des images. Pour la génération de connaissances augmentées, la

Phase 1 - Rapport de Projet PFE

technique de pointe actuelle repose sur le RAG (Retrieval-Augmented Generation), combinant modèles d'élargissement sémantique (embedding) et bases vectorielles (pgvector, Pinecone).

Des frameworks comme LangChain et Spring AI permettent aujourd'hui d'orchestrer facilement ces modèles dans des agents complexes capables de répondre à des requêtes multimodales (texte, image, audio). L'utilisation de GPT-4o pour le traitement temps réel de l'audio renforce l'émergence d'agents intelligents interactifs dans les applications industrielles.

Paragraphes 5-6 : Idée du projet, données et sources

L'idée du projet est de développer une plateforme intelligente à trois volets :

1. Analyse automatique de qualité d'installation à partir d'images d'intervention via GPT-4 Vision.
2. Accès à la documentation interne Canal Plus Telecom à l'aide d'un agent RAG basé sur pgvector.
3. Interaction en temps réel par voix ou texte avec des agents multimodaux utilisant GPT-4o.

Les données sources sont constituées des images d'interventions réelles collectées par les techniciens, des bases documentaires internes (procédures, FAQ, guides) ainsi que des prompts utilisateurs (texte/audio). Une base PostgreSQL équipée de pgvector stockera les embeddings des documents pour le moteur de recherche intelligent.

Paragraphes 7-8 : Méthodologie et étapes proposées

La solution sera construite en utilisant un backend Spring Boot avec Spring AI et Langchain4j pour orchestrer les agents IA. Le frontend Angular offrira les interfaces de soumission d'images, de prompts texte et audio, et d'affichage des résultats. Les étapes clés incluent : collecte et prétraitement des données, intégration des APIs OpenAI (GPT-4 Vision, GPT-4o), construction du moteur RAG, développement d'agents spécifiques et déploiement d'une base vectorielle

Phase 1 - Rapport de Projet PFE

PostgreSQL.

Architecture Globale du Système

Frontend Angular: Upload image / prompt texte / audio. Visualisation des résultats IA (analyse image, audio, documents). WebRTC pour prompts vocaux.

Backend Spring Boot: API REST pour communication frontend/backend. Orchestration IA avec Spring AI + Langchain4j.

Base de données PostgreSQL avec extension pgvector.

Agents IA spécialisés: analyse image, RAG, dialogue audio, génération d'images.

Services externes: OpenAI APIs, transcription audio Whisper.

Paragraphe 9-10 : Conclusion et innovation

Le projet présente une forte valeur ajoutée car il fusionne des technologies de pointe dans un contexte industriel concret. Il permettra de réduire le temps de validation des interventions, d'améliorer la qualité perçue des installations et de fiabiliser les audits internes. L'usage combiné de l'analyse multimodale, de la recherche augmentée (RAG) et de l'interaction vocale positionne ce projet à la pointe de l'innovation opérationnelle.

En outre, il contribuera à la transformation des processus de formation et d'onboarding, en fournissant aux nouveaux techniciens un accès immédiat aux connaissances clés grâce aux agents intelligents. Ce projet démontre ainsi une application tangible et innovante de l'intelligence artificielle multimodale au service de la qualité industrielle.

Bibliographie

1. Vaswani et al., "Attention Is All You Need", 2017.

Phase 1 - Rapport de Projet PFE

2. Dosovitskiy et al., "An Image is Worth 16x16 Words: Vision Transformer (ViT)", 2020.
3. Radford et al., "Learning Transferable Visual Models From Natural Language Supervision (CLIP)", 2021.
4. OpenAI, "GPT-4 Technical Report", 2023.
5. OpenAI, "Introducing GPT-4o", 2024.
6. Spring.io, "Spring AI Documentation", 2024.
7. LangChain, "Building Applications with LLMs", 2024.
8. Johnson, "pgvector: AI embeddings in PostgreSQL", 2023.
9. Meta AI, "Segment Anything Model (SAM)", 2023.
10. Bommasani et al., "On the Opportunities and Risks of Foundation Models", 2021.