

Jane Grace Davis Ngo Massing Soun

Technique de l'informatique

Groupe 04118

Projet 3

Présenté à M Nicolas Payre

Département D'informatique

Exploration de nouvelles technologies

Cegep de Sherbrooke

9 December 2025

Rapport de Projet : Application IA Multimodale

I- Introduction

Dans le cadre de ce projet d'exploration technologique, j'ai développé une application web interactive basée sur l'intelligence artificielle. Le but était d'explorer en profondeur des technologies modernes d'IA qui n'ont pas été couvertes en classe, notamment les modèles multimodaux et les interfaces Web interactives. Ce projet inclut trois fonctionnalités principales : analyse d'images, analyse de texte et tableau de bord.

Ce rapport présente les motivations, les outils, la méthodologie, les résultats, les limites et les perspectives du projet. Il décrit aussi l'utilisation d'outils tels que VS Code et GitHub Copilot, qui ont contribué à accélérer le développement.

II- Contexte et motivations

L'intelligence artificielle multimodale devient extrêmement importante dans l'industrie. Les systèmes capables de comprendre à la fois le texte et les images sont utilisés dans plusieurs domaines : médias, robotique, analyse de documents, sécurité, etc. La motivation principale derrière ce projet était de construire un outil simple mais puissant exploitant ces modèles modernes.

Plutôt que d'utiliser un modèle de classification basique, j'ai choisi BLIP, un modèle vision-langage de pointe capable de générer une description textuelle complète d'une image. Cela donne une forme d'intelligence plus humaine et représente un vrai saut technologique comparé aux modèles classiques.

III- Technologies utilisées

Ce projet utilise plusieurs technologies récentes :

- Streamlit : framework moderne pour créer des applications Web en Python.
- HuggingFace Transformers : bibliothèque offrant des milliers de modèles IA.
- BLIP (image-to-text) : modèle capable de décrire une image en langage naturel.

- Pipeline NLP (sentiment analysis) : analyse la tonalité d'un texte.
- Torch: backend nécessaire pour exécuter les modèles IA.
- VS Code : environnement de développement.
- GitHub Copilot : assistant intelligent ayant aidé à générer du code et accélérer le développement.

Ces technologies représentent l'état de l'art dans le développement d'applications IA modernes.

IV- Architecture de l'application

L'application est divisée en trois modules principaux :

Module 1 : Analyse d'images

L'utilisateur téléverse une image. Le modèle BLIP génère alors automatiquement une phrase décrivant cette image. Cela va au-delà d'une simple liste d'objets et permet une interprétation plus riche.

Module 2 : Analyse de texte

L'utilisateur entre un texte, qui est analysé par un pipeline NLP afin de déterminer s'il est positif, négatif ou neutre. Ce type d'analyse est très utilisé en entreprise (analyse de commentaires, satisfaction client, etc.).

Module 3 : Tableau de bord

Un module extensible qui servira prochainement à afficher des graphiques, des statistiques et des historiques d'analyses.

V- Méthodologie

La méthodologie suivie dans ce projet comprend :

1. Recherche et sélection des modèles IA.
2. Création d'un environnement virtuel Python.
3. Installation des dépendances : Streamlit, Torch, Transformers, Timm.
4. Développement du module d'analyse d'images.
5. Développement du module d'analyse de texte.
6. Mise en place de la navigation et de l'interface.
7. Débogage lié aux environnements virtuels et importations.
8. Tests d'utilisation.
9. Finalisation du rapport.

La démarche a été itérative, chaque module ayant été testé indépendamment avant intégration.

VI- Résultats obtenus

- BLIP génère des descriptions cohérentes même pour des images complexes.
- Le modèle NLP donne des analyses de sentiment fiables.
- L'application fonctionne entièrement en local sans serveur externe.
- L'interface Streamlit est fluide et réactive.

Des tests sur plusieurs images ont permis de valider le comportement du modèle. Les descriptions générées étaient quelques fois pertinentes.

VII- Limites du projet

Le projet présente toutefois certaines limites : Les performances IA sont limitées par l'absence de GPU, BLIP peut mal interpréter des images stylisées ou trop abstraites, le tableau de bord n'a pas encore de données réelles et l'application ne sauvegarde pas encore les analyses.

Ces limites ouvriront la porte à des améliorations futures importantes.

VIII- Améliorations futures

- Ajouter un historique des analyses et les stocker dans une base SQLite.
- Générer automatiquement un rapport PDF après chaque analyse.
- Intégrer un modèle multimodal plus puissant comme LLaVA.
- Ajouter un module d'analyse audio ou vidéo.
- Permettre le déploiement de l'application dans le cloud.
- Ajouter des graphiques dans le tableau de bord.

IX- Conclusion

Ce projet m'a permis de découvrir et maîtriser des technologies modernes comme Streamlit, Transformers, BLIP et GitHub Copilot. Il démontre l'intégration réussie d'une intelligence artificielle multimodale dans une application Web simple et efficace. L'expérience acquise servira de base à de futurs projets IA plus ambitieux.

QUELQUES IMAGES DE L'APPLICATION WEB

Navigation

Aller à :

- ☐ Analyse d'Images
- ☒ Analyse de Texte
- ☐ Tableau de Bord



Analyse de Texte avec IA


Écris un texte

Je suis enervée

Analyser

Sentiment : **POSITIVE** (score : 0.96)



Deploy 

Navigation

Aller à :

- ☒ Analyse d'Images
- ☐ Analyse de Texte
- ☐ Tableau de Bord



Analyse d'Images avec IA

Choisis une ou plusieurs images



Drag and drop files here

Limit 200MB per file • JPG, JPEG, PNG

Browse files

Navigation

Aller à :

- ☐ Analyse d'Images
- ☐ Analyse de Texte
- ☒ Tableau de Bord



Tableau de bord des analyses

Voici l'historique d'utilisation :

	type	input	result
0	texte	bonjour!	POSITIVE
1	image	119_desequilibre.png	scale, weighing machine
2	image	cd489fd9-031a-4864-9b0d-c1d3a7f71107.jpeg	church, church building
3	texte	Je suis énervée	POSITIVE

Graphique du nombre d'analyses par type :

