GPT-4o ouvre des perspectives intéressantes pour l'ingénierie des invites multimodales en combinant plusieurs types de données dans un même contexte de génération. Voici quelques axes d'exploration :

🔷 1. Analyse de texte manuscrit

- Utilisation de GPT-4o pour interpréter des notes manuscrites, via OCR combiné à une compréhension du contexte.
- Applications : transcription automatique, extraction d'informations clés, analyse de notes médicales ou de schémas manuscrits.

🔷 2. Analyse de graphiques et tableaux

- Capacité de GPT-4o à comprendre et interpréter des graphiques (courbes, histogrammes, camemberts, etc.).
- Extraction automatique d'informations pertinentes et génération de résumés.
- Génération de commentaires sur des tableaux financiers, scientifiques ou statistiques.

🔷 3. Analyse d'images

- Identification d'objets, de scènes et de contextes visuels.
- Explication de schémas techniques ou scientifiques.
- Applications en accessibilité (ex. : description d'images pour malvoyants).

🔷 4. Analyse audio et vidéo

- Compréhension et résumé de contenus audio (conférences, podcasts, réunions).
- Analyse de vidéos pour extraire du contenu clé (détection d'événements, reconnaissance faciale, sous-titrage automatique).
- Applications : automatisation de la veille médiatique, analyse de vidéos de surveillance, résumé de cours en ligne.

🔷 5. Applications en RAG multimodal

- Intégration dans un système RAG (Retrieval-Augmented Generation) capable de récupérer et d'analyser des documents contenant divers formats de données.
- Couplage avec FAISS pour indexer non seulement du texte, mais aussi des images et des représentations vectorielles de multimodalité.
- Amélioration des chatbots capables de répondre à des questions sur des documents techniques multimodaux.

📌 **Projet : Assistant Multimodal d'Analyse de Documents**

🔷 **Objectif** : Créer un assistant capable de traiter des documents contenant du texte, des images, des tableaux et des graphiques, et d'en extraire automatiquement des informations exploitables.
🔷 **Technologies** :

- **Ollama** avec **DeepSeek R1** pour le chatbot.
- **FAISS** pour l'indexation et la recherche vectorielle.
- **Nomic-Embed-Text** pour l'embedding de texte.
- **Tesseract OCR** ou **PaddleOCR** pour l'extraction de texte des images et documents scannés.
- **Matplotlib + OpenCV** pour analyser des graphiques et des tableaux.
- **LangChain** pour orchestrer l'ensemble et gérer les requêtes.

🛠️ **Étapes du projet**

1️⃣ Extraction et prétraitement des documents

- Développer un pipeline pour ingérer des documents PDF, images, ou fichiers Excel.
- Extraire le texte avec **PyMuPDF** pour les PDFs, **Tesseract OCR** pour les images.
- Extraire les tableaux avec **pandas + Camelot** pour les PDF, et **OpenCV** pour identifier des tableaux dans des images.
- Identifier et classer les types de contenus dans un document (texte, tableaux, images, graphiques).

2️⃣ Indexation des données

- Transformer le texte en embeddings avec **nomic-embed-text**.
- Stocker les embeddings avec **FAISS** pour permettre la recherche rapide.
- Ajouter un index pour les images et tableaux avec des représentations vectorielles basées sur **CLIP** (ou OpenAI Vision Embeddings si accessible).

3️⃣ RAG Multimodal : Recherche et Génération de Réponses

- Construire une interface où l'utilisateur pose des questions sur un document.

- Récupérer les passages pertinents avec **FAISS** et les donner en contexte à **DeepSeek R1**.
- Traiter les images et graphiques pour extraire les tendances et insights.
- Si un graphique est détecté, extraire les données et générer une interprétation textuelle.

4️⃣ Interface Utilisateur

- Interface en **Streamlit** ou en **Gradio** pour permettre aux utilisateurs d'uploader un document et de poser des questions.
- Afficher les réponses textuelles, ainsi que les graphiques ou tableaux pertinents.
- Option pour télécharger un résumé automatique du document.

📍 **Cas d'Usage**

✅ **Analyse de rapports financiers** → Extraire les chiffres clés, générer des résumés automatiques.
✅ **Audit de documents médicaux** → Comprendre des résultats d'analyses (tableaux + graphiques).
✅ **Traitement de documents scientifiques** → Résumer des articles contenant texte, équations et figures.
✅ **Assistance juridique** → Rechercher des clauses spécifiques dans des contrats et identifier des informations dans des PDF scannés.

🔷 **Pourquoi c'est faisable ?**

- Chaque brique technologique est bien documentée et dispose de bibliothèques open-source.
- Le projet est **modulaire** → tu peux commencer avec du texte, puis ajouter l'analyse des tableaux et images progressivement.
- L'indexation FAISS permet une recherche efficace sur de grands volumes de données.