

SOUTENANCE - L'IA Pero

Moteur de Recommandation de Cocktails par IA Semantique

Adam Beloucif & Amina Medjdoub *Mastere Data Engineering et IA - EFREI Paris Pantheon-Assas Universite RNCP40875 - Bloc 2 Tutrice : MALAEB Sarah*

SLIDE 1 - Presentation du Projet

Problematique

Comment recommander des cocktails personnalisés en comprenant les préférences exprimées en langage naturel ?

Contexte Metier

- **Utilisateurs cibles** : Amateurs de cocktails, bartenders, curieux
- **Besoin** : Outil intelligent dépassant les filtres traditionnels

Pitch

"Un moteur de recommandation IA sémantique basé sur SBERT et assisté par l'IA Generative via RAG pour découvrir des cocktails personnalisés"

SLIDE 2 - Analyse du Besoin Utilisateur

Persona Cible

Caracteristique	Description
Profil	Amateur de cocktails, 25-45 ans
Objectif	Decouvrir de nouvelles recettes
Contrainte	Budget variable

Scenarios d'Usage

1. **Decouverte** : "Je veux quelque chose de frais et fruité"
2. **Precision** : "Un mojito avec une touche tropicale"

Contraintes

- Interface intuitive (texte libre + dropdowns)
- Temps de reponse < 3 secondes
- Cout API maitrise

SLIDE 3 - Methodologie et Organisation

Approche Agile/Kanban

Sprint 1 : Backend (API + Database)
Sprint 2 : Intégration (Front + Cache)
Sprint 3 : Interface (Stylé UI)
Sprint 4 : Optimisations + Tests

Repartition des Taches

Adam	Amina
Backend RAG	Interface UI
SBERT/Embeddings	Design Speakeasy
Optimisation perf	Tests E2E

Outils

- **Git/GitHub** : Versioning code
- **VS Code** : Developpement
- **Streamlit Cloud** : Deploiement

SLIDE 4 - Architecture Globale



SLIDE 5 - Pipeline IA

Etapas du Pipeline

1. **Collecte** : Texte libre + Budget dropdown
2. **Pre-traitement** : Normalisation + enrichissement
3. **Guardrail SBERT** : Similarite cosinus (seuil 0.35)
4. **Embeddings** : all-MiniLM-L6-v2 (384 dimensions)
5. **Scoring** : Classement Top-N par similarite
6. **RAG Gemini** : Generation recette personnalisée

Focus Guardrail

Seuil	Comportement
< 0.35	Rejete (hors-sujet)
>= 0.35	Accepte (domaine cocktail)

SLIDE 6 - Referentiel de Donnees

Sources

Source	Volume	Format
Genere	600 cocktails	CSV
Kaggle	1000 cocktails	CSV enrichi
Ingredients	61 profils	JSON

Schema de Donnees

```
{
  "name": "Mojito",
  "description_semantique": "Cocktail frais...",
  "ingredients": ["60ml Rhum", "Menthe", "Citron vert"],
  "instructions": ["Piler la menthe...", "..."],
  "taste_profile": {
    "Douceur": 3.5,
    "Acidite": 4.0,
    "Amertume": 1.5,
    "Force": 3.0,
    "Fraicheur": 5.0
  }
}
```

SLIDE 7 - Implementation Technique

Modele SBERT

Parametre	Valeur
Modele	all-MiniLM-L6-v2
Dimensions	384
Latence	~50ms
Cout	Zero (local)

Strategie RAG

- **Prompt** : Persona "Barman Speakeasy 1920s"
- **Fallback** : 5 modeles Gemini en cascade
- **Cache** : MD5 hash → reponse instantanee

Code (extrait)

```
# Guardrail sémantique
similarity = cosine_similarity(query_emb, keywords_emb)
if max(similarity) < 0.35:
    return {"status": "error", "message": "Hors-sujet"}
```

SLIDE 8 - Interface Utilisateur

Theme Speakeasy (Annees 1920)

- Palette : Or (#D4AF37) + Noir (#0D0D0D)
- Typographie : Playfair Display
- Ambiance : Musique jazz optionnelle

Composants

Element	Fonction
Input texte	Envie libre
Dropdown	Budget (4 niveaux)
Sidebar tabs	Filtres / Recherche / Stats
Radar Chart	Profil gustatif 7D
Export	Telechargement recette

SLIDE 9 - Demonstration Live

Scenario 1 : Requete Valide

Input : "Un cocktail frais et tropical" **Budget :** Modere

Resultat attendu :

- Recette generee avec nom unique
- Ingredients + Instructions
- Radar chart profil gustatif

Scenario 2 : Guardrail

Input : "Quelle heure est-il ?"

Resultat attendu :

- Message d'erreur
- "Le barman ne comprend que les cocktails!"

SLIDE 10 - Resultats et Performance

Metriques

Metrique	Valeur
Cocktails indexes	600+
Precision guardrail	>95%
Temps recherche SBERT	~50ms
Amelioration perf	40-60x

Comparaison Avant/Après

Operation	Avant	Après
Recherche SBERT	2-3s	50ms
Cache hit	N/A	5ms

SLIDE 11 - Limites et Ameliorations

Limites Actuelles

- Dataset 600 cocktails (extensible)
- Rate limit API Gemini (15 req/min)
- SBERT non fine-tune domaine

Pistes d'Amelioration

Amelioration	Impact
Fine-tuning SBERT	+20% pertinence
Vector DB	Scalabilite
Chatbot guide	Meilleure UX
Multilingue	Reach elargi

SLIDE 12 - Conclusion

Ce que nous avons realise

- Pipeline NLP complet (embeddings → similarite → RAG)
- Interface utilisateur immersive

- Optimisations de performance significatives
- Tests automatisés E2E

Compétences RNCP Bloc 2 Validées

Compétence	Démonstration
Conception IA	Pipeline SBERT + RAG
Implementation	Backend Python + Frontend
Optimisation	Cache, précomputation 40-60x
Gouvernance	Guardrail, éthique, traçabilité

Merci !

Questions ?

Adam Beloucif & Amina Medjdoub Mastere Data Engineering et IA - EFREI Paris Pantheon-Assas Université Tutrice : MALAEB Sarah