

République Démocratique et Populaire d'Algérie  
Ministère de l'Enseignement Supérieur et de la Recherche  
Scientifique  
Université Abderrahmane Mira de Béjaïa  
Faculté des Sciences Exactes  
Département d'Informatique



## Mémoire de fin d'études

*En vue de l'obtention du diplôme de Master académique en Systèmes  
d'Information Avancés*

### Thème

*Chatbot pour e-commerce*

Présenté par :

**FERHAT Essaid**                      **BOUAGAL Naima**

Encadré par :

**Dr. EL BOUHISSI Houda**

Jury de soutenance :

<b>Dr. KHALED Hayette</b>	<b>Présidente du jury</b>
<b>Dr. OUZEGANE Redouane</b>	<b>Membre du jury</b>
<b>Dr. TAHAKOURT Zineb</b>	<b>Membre du jury</b>
<b>Dr. BENAI Yani Othmane</b>	<b>Membre du jury</b>

**Promotion 2024/2025**

*Au nom d'Allah, le Clément, le Miséricordieux*

---

## Remerciement

---

*À l'issue de ce travail, nous exprimons notre profonde gratitude et notre sincère reconnaissance à toutes les personnes qui ont contribué à sa réalisation.*

*Nous remercions d'abord **Allah, le Tout-Puissant**, pour la force, la patience et la persévérance qu'Il nous a accordées.*

*Nos plus vifs remerciements vont à **nos parents**, pour leur soutien moral, affectif et financier tout au long de ce parcours.*

*Nos sincères remerciements à **Mme Houda El Bouhissi**, notre encadrante, pour ses conseils avisés, sa disponibilité et sa confiance.*

*Nous remercions également les **membres du jury**, pour leur temps et leurs remarques enrichissantes.*

*Enfin, notre gratitude va à **l'équipe pédagogique du Département d'Informatique de l'Université de Béjaïa**, pour leur accompagnement tout au long de notre formation.*

---

## Dédicace

---

*Je dédie ce travail en premier lieu à **mes parents**, pour leur soutien indéfectible, leur amour inconditionnel et leurs sacrifices constants. Sans leur aide, leur soutien moral et leur encouragement constant, je n'aurais jamais pu parvenir à ce stade. Ce travail est le fruit de leur dévouement et de leurs prières. Vous êtes ma plus grande source d'inspiration.*

*À ma sœur **Samah Bouagal**,  
pour ta présence rassurante et ton soutien constant tout au long de ce parcours.  
Ton amour et tes encouragements m'ont permis de surmonter les obstacles et d'aller de l'avant.  
Ce travail est dédié à toi, qui as toujours cru en moi.*

*À **mes frères**,  
pour leur soutien moral et leur affection.  
Vous avez toujours été là pour me soutenir et me motiver, même dans les moments difficiles.  
Votre présence a été un pilier essentiel tout au long de ce parcours, et ce travail est aussi dédié à vous.*

*Je dédie également ce mémoire à **mes collègues de travail**, pour leur soutien, leur compréhension et leur aide précieuse. Vous avez facilité mon quotidien et m'avez permis de trouver un équilibre entre mes obligations professionnelles et académiques. Votre amitié et votre bienveillance ont été des forces importantes durant cette période.*

*Naïma.*

---

## Dédicace

---

*Je dédie ce travail en premier lieu à **mes parents**,  
pour leur amour infini, leur soutien sans faille  
et leur présence constante dans ma vie.  
Leur foi en moi, leur patience et leurs sacrifices  
m'ont permis de surmonter chaque étape de ce parcours.  
Ce mémoire vous est dédié,  
en hommage à tout ce que vous avez fait pour moi.*

***À ma famille,**  
dont l'amour et la solidarité m'ont toujours accompagné.  
Votre soutien moral, vos prières et vos encouragements  
ont été des forces essentielles dans l'accomplissement de ce travail.  
Vous êtes la raison pour laquelle  
j'ai pu poursuivre mes rêves et réaliser ce projet.*

*Je tiens également à dédier ce travail à **mes amis**,  
qui m'ont toujours encouragé à persévérer  
et à donner le meilleur de moi-même.  
Votre amitié et votre soutien moral ont été indispensables.  
Merci à vous tous d'avoir été là  
à chaque étape de cette aventure.*

*Essaid.*

# Table des matières

Liste des abréviations	9
<b>1 Généralités sur les Chatbots pour l'E-commerce</b>	<b>13</b>
1.1 Introduction	13
1.2 Historique et évolution des chatbots	14
1.3 Types de chatbots	15
1.3.1 Chatbots basés sur des regles	15
1.3.2 Chatbots basés sur l'intelligence artificielle	15
1.3.3 Chatbots hybrides	15
1.4 Applications de chatbot	17
1.5 Technologies utilisées dans les chatbots	19
1.5.1 Intelligence artificielle	19
1.5.2 Frameworks et outils de développement	21
1.6 Rôle des chatbots dans l'e-commerce	22
1.6.1 Support client automatisé	23
1.6.2 Recommandation de produits	23
1.6.3 Amélioration de l'expérience utilisateur	23
1.6.4 Évaluer la performance des chatbots	23
1.7 Limites et défis des chatbots	24
1.8 Conclusion	25
<b>2 Etat de l'art</b>	<b>26</b>
2.1 Revue de la littérature	26
2.2 Analyse comparative	33
2.3 Analyse critique	36
2.4 Discussion	37
2.5 Conclusion	37
<b>3 Approche proposée</b>	<b>39</b>
3.1 Introduction	39
3.2 Contributions	40

3.3	Approche proposée . . . . .	41
3.3.1	Collecte des données . . . . .	41
3.3.2	Traitement du texte avec BERT . . . . .	42
3.3.3	Gestion des données textuelles incomplètes ou bruitées . . . . .	42
3.3.4	Normalisation du texte . . . . .	42
3.3.5	Encodage sémantique avec BERT . . . . .	43
3.3.6	Calcul de similarité sémantique . . . . .	43
3.3.7	Recherche croisée texte-image avec CLIP . . . . .	44
3.3.8	Génération de réponses basées sur une base Q&A intelligente avec NLTK . . . . .	46
3.3.9	Génération de réponses conversationnelles avec GPT . . . . .	48
3.4	Conclusion . . . . .	50
<b>4</b>	<b>Implémentation et évaluation</b>	<b>52</b>
4.1	Introduction . . . . .	52
4.2	Description du Dataset . . . . .	53
4.2.1	Définition . . . . .	53
4.2.2	Aperçu du dataset . . . . .	53
4.2.3	Contenu du dataset . . . . .	53
4.2.4	Résumé statistique des données . . . . .	54
4.2.5	Graphiques de distribution des données . . . . .	55
4.3	Environnement de développement . . . . .	58
4.3.1	Environnement matériel . . . . .	58
4.3.2	Environnement logiciel . . . . .	59
4.4	Description de l'outil développé . . . . .	60
4.4.1	Page d'accueil du chatbot . . . . .	60
4.4.2	Recherche par texte . . . . .	61
4.4.3	Recherche par synonymes . . . . .	62
4.4.4	Recherche par image . . . . .	62
4.4.5	Gestion des requêtes générales . . . . .	63
4.4.6	Réponse intelligente (Fallback GPT) . . . . .	64
4.5	Évaluation . . . . .	65
4.5.1	Méthodologie . . . . .	65
4.5.2	Scénarios de test . . . . .	65
4.5.3	Présentation de la métrique et de son fonctionnement : . . . . .	66
4.5.4	Mise en œuvre de la métrique dans notre évaluation : . . . . .	66
4.5.5	Résultat obtenue : . . . . .	66
4.5.6	Interprétation des résultats : . . . . .	67
4.6	Conclusion . . . . .	68
	<b>Références</b>	<b>71</b>

# Liste des figures

3.1	<i>Approche proposée</i>	41
4.1	<i>Aperçu du dataset</i>	54
4.2	<i>Résumé statistique du dataset</i>	55
4.3	Nombre d'articles par saison.	56
4.4	Nombre d'articles selon la couleur dominante.	56
4.5	Répartition temporelle des articles.	57
4.6	Nombre d'articles par catégorie.	57
4.7	Nombre d'articles par genre cible.	58
4.8	Page d'accueil du chatbot avec réponse simple par NLTK	61
4.9	Résultat pour une recherche de chemise	61
4.10	Résultat pour une recherche de chemises	62
4.11	Sélection d'une image produit pour la recherche	63
4.12	Résultats de recherche par image avec score de similarité	63
4.13	Réponse explicative pour une requête liée à l'IA	64
4.14	Exemple de réponse via GPT	64
4.15	<b>(a)</b> Ensemble des requêtes par image et résultat obtenue.	67
4.16	<b>(b)</b> Ensemble des requêtes par texte et résultat obtenue.	67
4.17	Résultats obtenus pour les requêtes par image (a) et par texte (b).	67

# Liste des tableaux

1.1	Comparaison entre les chatbots basés sur des règles et ceux à IA . . . . .	17
2.1	Tableau comparatif . . . . .	35
2.2	Critères traités dans chaque article . . . . .	36

# Liste des abréviations

- AIML** Langage de Balises pour l'Intelligence Artificielle. 31
- CC-NEWS** Nouvelles de Common Crawl. 28
- CRM** Gestion de la Relation Client. 24
- CSAT** Score de Satisfaction Client. 24
- ERP** Planification des Ressources de l'Entreprise. 24
- FGA** Les réseaux à Attention Factorielle (Factorized Gated Attention).  
27
- GLUE** Évaluation Générale de la Compréhension du Langage. 28
- GS** Graphe de connaissances (Graph Structure). 32
- ITC** Apprentissage Contrastif Image-Texte. 30
- ITM** Correspondance Image-Texte. 30
- LM** Modélisation du Langage. 30
- LSA** Analyse Sémantique Latente. 31
- MED** Modèle Multimodal Encodeur-Décodeur. 30
- ML** Apprentissage Automatique. 19
- MLM** Modélisation de Langage Masqué. 29
- NSP** Prédiction de la Phrase Suivante. 28
- PCA** Analyse en Composantes Principales. 20
- RefCOCOg** Jeu de données d'expressions référentielles (version Google). 28
- RGPD** Règlement Général sur la Protection des Données. 24

---

**RoBERTa** Approche de Pré-entraînement Optimisée BERT. 28

**SAV** Service Après-Vente. 17

**SQuAD** Jeu de données de Réponses aux Questions de Stanford. 28

**SVM** Machine à Vecteurs de Support. 20

**TLN** Traitement du Langage Naturel. 21

**VITR** Transformateurs Vision avec apprentissage axé sur les relations.  
27

**VQA** Réponses aux Questions Visuelles. 30

**VSE $\infty$**  Incorporation Sémantique Visuelle infinie. 28

**VSRN++** Réseau de Raisonnement Sémantique Visuel amélioré. 28

**WGAN** Réseau Génératif Adversarial Wasserstein. 30

# Introduction générale

L'essor de l'intelligence artificielle (IA) a profondément transformé la manière dont les systèmes informatiques interagissent avec les utilisateurs. Dans le domaine du commerce électronique, cette évolution a permis d'offrir des expériences en ligne plus intelligentes, personnalisées et interactives. Parmi les innovations les plus marquantes figure le chatbot, un agent conversationnel capable de simuler un dialogue humain dans le but d'assister, de conseiller ou d'orienter les utilisateurs.

Cependant, malgré leur popularité, la plupart des chatbots actuels sur les sites e-commerce se limitent à des interactions textuelles simples. Dans un contexte où l'image joue un rôle clé — notamment dans le secteur de la mode — il devient indispensable d'intégrer la dimension visuelle aux systèmes d'assistance automatisée.

C'est dans ce cadre que s'inscrit ce travail, qui propose le développement d'un chatbot intelligent multimodal, capable d'interpréter des requêtes à la fois textuelles et visuelles. Pour cela, nous combinons plusieurs technologies d'IA complémentaires : des outils de traitement du langage naturel pour les questions simples, des modèles de représentation sémantique pour le croisement texte-image, ainsi que des générateurs de langage naturel pour la formulation des réponses. L'interaction suit une logique adaptative : réponse directe si la question est basique, recherche sémantique si la requête est plus complexe, puis reformulation ou génération automatique selon le cas. Le système sera intégré dans une application web construite avec Flask et évalué à l'aide du jeu de données Fashion Product Images (Small).

L'intégration de l'IA dans les plateformes e-commerce constitue aujourd'hui un levier stratégique essentiel pour améliorer l'expérience utilisateur et renforcer la compétitivité des entreprises. Les consommateurs recherchent des interactions fluides, rapides et pertinentes, sans avoir à explorer manuellement des catalogues interminables.

Les approches multimodales, en permettant au chatbot de traiter simultanément du texte et des images, ouvrent la voie à des échanges plus naturels et contextuels. Par exemple, un utilisateur peut exprimer son besoin sous

---

forme textuelle, tout en joignant une image du produit recherché. Grâce à la complémentarité des modèles employés, le système est en mesure d'identifier les éléments clés de la demande et de proposer des recommandations adaptées.

Ce projet relève un double défi : aligner le langage naturel sur les contenus visuels, et garantir des réponses à la fois pertinentes et rapides, dans une interface intuitive. Les bénéfices attendus sont multiples : gain de temps, navigation simplifiée, et satisfaction accrue grâce à une expérience utilisateur optimisée.

La méthodologie suivie s'articule autour de plusieurs étapes : étude de l'existant (chatbots, NLP, vision par ordinateur), identification des limites actuelles, définition du problème, conception de la solution, développement de l'application et évaluation des résultats selon des critères quantitatifs et qualitatifs.

Ce mémoire est structuré comme suit : le chapitre 1 présente les généralités sur les chatbots dans le e-commerce ; le chapitre 2 explore l'état de l'art des systèmes intelligents et multimodaux ; le chapitre 3 détaille l'approche proposée et son architecture technique ; le chapitre 4 décrit la mise en œuvre pratique et l'évaluation ; enfin, une conclusion de ce travail on proposant on des perspectives d'amélioration.

# Chapitre 1

## Généralités sur les Chatbots pour l'E-commerce

### 1.1 Introduction

Les chatbots sont devenus un outil incontournable dans le domaine de l'e-commerce. Ils permettent d'automatiser les interactions avec les clients, d'améliorer l'expérience utilisateur et d'optimiser le processus de vente. Grâce aux avancées en intelligence artificielle et en traitement du langage naturel, ces agents conversationnels sont capables de comprendre et de répondre aux demandes des utilisateurs avec une précision croissante.

L'intégration des chatbots dans les plateformes de commerce en ligne offre de nombreux avantages. Ils peuvent assurer un service client disponible 24h/24 et 7j/7, réduisant ainsi le temps d'attente des consommateurs. De plus, ils facilitent la personnalisation des recommandations de produits en analysant le comportement des utilisateurs et en proposant des offres adaptées à leurs besoins.

L'utilisation de chatbots permet également aux entreprises d'optimiser leurs coûts en réduisant la charge de travail des équipes de support client tout en améliorant la satisfaction des utilisateurs. Cependant, leur mise en place présente des défis tels que la compréhension des requêtes complexes, la gestion des erreurs et la nécessité d'une mise à jour continue pour suivre l'évolution des besoins des consommateurs.

Selon l'Oxford English Dictionary, un chatbot est défini comme « un programme informatique conçu pour simuler une conversation avec des utilisateurs humains, notamment via Internet » (Oxford English Dictionary, s.v. "chatbot").

Le mot *chatbot* vient de l'anglais «chat» qui désigne une discussion et

«bot» pour le robot. En français, on utilise les termes robot conversationnel ou agent de conversation virtuel.

Un chatbot est un logiciel capable de communiquer, ou de « discuter », avec un utilisateur humain dans des langues comme l'anglais. Des chatbots concurrents ont été évalués en fonction de leur capacité à tromper un juge lors d'une session de conversation à huis clos, dans le cadre de la préparation au concours annuel Loebner Prize [1].

La définition d'un chatbot adoptée pour ce document est la suivante : « un chatbot est un agent conversationnel destiné à simuler une interaction humaine en langage naturel dans le but de fournir une assistance ou un service à l'utilisateur. » [2]. Ce chapitre explore les fondements des chatbots, leurs types, les technologies impliquées ainsi que leur impact sur le secteur du commerce en ligne.

## 1.2 Historique et évolution des chatbots

En 1950, Alan Turing se demandait si un programme informatique pouvait parler à un groupe de personnes sans se rendre compte que leur interlocuteur était artificiel. Cette question, appelée *test de Turing*, est considérée par beaucoup comme l'idée génératrice des chatbots [3].

Le premier chatbot portant le nom d'ELIZA a été construit en 1966. ELIZA a simulé l'opération d'un psychothérapeute, renvoyant les phrases de l'utilisateur sous la forme interrogative [4]. Sa capacité à communiquer était limitée, mais elle a été une source d'inspiration pour le développement ultérieur d'autres chatbots [5]. ELIZA utilise la correspondance de modèles et un système de sélection de réponses basé sur des modèles [6]. Un inconvénient d'ELIZA est que ses connaissances sont limitées et qu'elle ne peut donc discuter que dans un domaine particulier de sujets. De plus, elle ne peut pas entretenir de longues conversations et ne peut pas apprendre ou découvrir le contexte de la discussion.

En 1972, PARRY apparaît ; il s'agissait d'un patient atteint de schizophrénie [7]. PARRY est considéré comme plus avancé qu'ELIZA car il est censé avoir une « personnalité » et une meilleure structure de contrôle. Il définit ses réponses sur la base d'un système d'hypothèses et de « réponses émotionnelles » activées par le changement de poids dans les énoncés de l'utilisateur [8].

PARRY a été utilisé dans une expérience en 1979 lorsque cinq juges psychiatres ont interrogé par télétype un patient pour décider s'il était un programme informatique ou un véritable patient schizophrène. Les psychiatres ont donc posé dix diagnostics. Le premier psychiatre a posé deux diagnostics

corrects ; un autre en a donné deux incorrects. Le troisième considérait que les deux sujets étaient de vrais patients, et les deux autres diagnostiquaient que les deux sujets étaient des chatbots [9].

Cependant, l'échantillon de cinq psychiatres est petit et la signification des résultats n'est pas claire, car les personnes atteintes de schizophrénie présentent un certain degré d'incohérence dans leur discours. En général, PARRY est considéré comme un chatbot avec de faibles capacités en matière de compréhension du langage et d'expression des émotions. Il a également une faible vitesse de réponse et ne peut pas apprendre de la conversation.

## 1.3 Types de chatbots

### 1.3.1 Chatbots basés sur des règles

Il existe deux grandes catégories de chatbots : ceux basés sur des règles et ceux reposant sur IA. Les chatbots basés sur des règles fonctionnent à partir d'un ensemble de directives prédéfinies pour analyser les requêtes des utilisateurs et fournir des réponses ou exécuter des actions en conséquence. Leur capacité de réponse est limitée aux scénarios programmés, ce qui les rend moins adaptatifs que les chatbots utilisant l'IA. Toutefois, ils présentent l'avantage d'être plus simples à concevoir, moins coûteux à développer et plus faciles à déployer [10].

### 1.3.2 Chatbots basés sur l'intelligence artificielle

Les chatbots basés sur l'IA s'appuient sur des algorithmes d'apprentissage automatique pour analyser les requêtes des utilisateurs et générer des réponses adaptées. Grâce à leur capacité à apprendre des interactions précédentes, ils peuvent affiner leurs réponses au fil du temps, offrant ainsi plus de flexibilité et d'efficacité que les chatbots basés sur des règles. Toutefois, leur conception et leur déploiement requièrent des technologies avancées ainsi qu'une expertise approfondie [10].

### 1.3.3 Chatbots hybrides

Les chatbots hybrides sont des agents conversationnels qui combinent les approches basées sur des règles avec celles reposant sur l'intelligence artificielle. Cette architecture permet d'assurer une transition fluide entre des interactions automatisées simples et des traitements plus complexes nécessitant une analyse sémantique ou une intervention humaine.

En pratique, les requêtes courantes sont prises en charge à l'aide de règles prédéfinies, tandis que les demandes plus élaborées sont traitées grâce à des techniques d'apprentissage automatique ou transférées à un opérateur humain. Ce modèle hybride permet d'optimiser l'efficacité du service client tout en garantissant une certaine flexibilité et une personnalisation de l'expérience utilisateur [11].

Nous avons décrit ci dessus les différent types de chatbots, le tableau suivant illustre le différence entre les chatbots basés sur des règles et les chatbots IA[4].

Critères	Chatbots basés sur des règles	Chatbots à intelligence artificielle (IA)
<b>Simplicité</b>	Faciles à configurer et à programmer	Développement plus complexe nécessitant une expertise avancée
<b>Coût</b>	Moins coûteux à développer et à déployer	Plus onéreux en raison des technologies avancées requises
<b>Flexibilité</b>	Réponses fixes, adaptées aux scénarios prévus	Capables d'apprendre et de s'adapter aux nouvelles interactions
<b>Gestion des requêtes</b>	Conviennent aux tâches simples et répétitives	Peuvent traiter des requêtes complexes et variées
<b>Personnalisation</b>	Interaction limitée aux scénarios programmés	Expérience utilisateur dynamique et personnalisée
<b>Évolution</b>	Ne peuvent pas apprendre ni s'améliorer avec le temps	Améliorent leurs réponses en fonction des interactions passées
<b>Langage naturel</b>	Nécessite des mots-clés précis pour fonctionner	Comprennent et analysent le langage humain de manière plus naturelle
<b>Supervision</b>	Faciles à gérer et à surveiller	Demandent un suivi et des mises à jour régulières
<b>Confidentialité</b>	Peu de risques en matière de données sensibles	Posent des défis liés à la protection des données et à l'éthique

TABLE 1.1 – Comparaison entre les chatbots basés sur des règles et ceux à IA

## 1.4 Applications de chatbot

Les chatbots ont de nombreuses applications. Cette section couvre certains des domaines dans lesquels les chatbots sont utilisés.

- **Service Après-Vente (SAV)** : Lorsque l'utilisateur a des questions ou des préoccupations, il appelle généralement les numéros gratuits du service client. Ils doivent attendre longtemps pour leur tour. Même

pour les utilisateurs et la quatrième personne qui doit répondre à ces appels, cela devient épuisant. Il serait plus simple qu'un chatbot soit formé pour répondre à ces demandes. Les tâches répétitives ne prennent pas beaucoup de temps. De plus, les chatbots nécessitent peu de maintenance et peuvent être disponibles 24 heures sur 24. Ils ne s'ennuient également jamais [12].

- **Réseaux sociaux** : Nous allons utiliser ces bots sur les réseaux sociaux pour interagir avec les utilisateurs. Sur d'autres sites sociaux, les chatbots peuvent parfois être dangereux aussi. Par exemple, sur Instagram, des personnes pourraient créer de faux bots pour augmenter la popularité d'un utilisateur. Elles pourraient utiliser des chatbots pour simuler des "j'aime" et des commentaires. Des incidents similaires, où des personnes achètent des bots pour influencer l'opinion des gens, ont également été signalés sur Twitter. De plus, certains chatbots fabriquent des profils afin de fournir aux utilisateurs sur divers réseaux sociaux des comptes fictifs de followers [12].
- **Agences de voyage** : Les chatbots sont une fonctionnalité courante sur les sites web des agences de voyage. Ces bots aideront à localiser des restaurants et des motels avec des offres spécifiques. En plus d'aider à la réservation dans des hôtels et restaurants réputés, les bots s'occuperont de la recherche de billets pour différents types de transport, tels que les trains et les avions. L'expérience de réservation des utilisateurs sera améliorée et les bénéfices des agences augmenteront en conséquence [12].
- **Assistant personnel** : De plus, les bots peuvent servir d'assistants personnels. Parmi les assistants personnels bien connus, on trouve Google Assistant, Alexa et Siri. Ceux-ci peuvent aussi être utilisés pour le divertissement. En proposant des activités ou en faisant des recommandations personnalisées de musique, de comédie, d'actualités et autres, ils aident à maintenir l'intérêt des consommateurs. Ils peuvent également être utilisés pour recevoir des rappels d'alertes et pour suivre les activités quotidiennes des utilisateurs [12].
- **Amélioration des compétences** : De nombreux chatbots peuvent aider à améliorer les compétences en fournissant aux utilisateurs des systèmes de suggestions personnalisées basées sur leurs intérêts. Les intérêts des utilisateurs seront collectés et ils fourniront davantage d'instructions dans des domaines comme l'art numérique, la création d'affiches et le montage vidéo [12].
- **Secteurs bancaires** : Toutes les industries bancaires sont récemment passées aux plateformes en ligne. En conséquence, elles ont également intégré des chatbots sur leur site web pour permettre aux visiteurs de

profiter de toutes les fonctionnalités. De plus, les bots recommanderont aux consommateurs le type de compte qu'ils peuvent ouvrir, les intérêts associés à chaque type de compte, les intérêts de la banque et les options de prêts. Les utilisateurs bénéficieront de ne pas avoir à attendre dans de longues files d'attente à l'intérieur de la banque pour obtenir toutes les informations. Les utilisateurs sont limités à comprendre toutes les installations des secteurs bancaires en ligne. Cela réduira le temps des utilisateurs ainsi que la charge de travail du personnel bancaire [12].

## 1.5 Technologies utilisées dans les chatbots

### 1.5.1 Intelligence artificielle

L'IA est un domaine en pleine croissance qui a le potentiel de transformer de nombreux domaines de la société, y compris la santé. En tant que médecin, il est important de comprendre les bases de l'IA et ses applications potentielles en médecine. L'IA fait référence au développement de systèmes informatiques capables d'effectuer des tâches qui nécessitent généralement l'intelligence humaine, telles que la reconnaissance de modèles, l'apprentissage à partir de données et la prise de décisions. Cette technologie peut être utilisée pour analyser de grandes quantités de données sur les patients et identifier les tendances et les modèles qui peuvent être difficiles à détecter pour les médecins humains. Cela peut aider les médecins à gérer leur charge de travail plus efficacement et à fournir de meilleurs soins à leurs patients. Dans l'ensemble, l'IA a le potentiel d'améliorer considérablement la pratique de la médecine et d'améliorer les résultats pour les patients. Dans ce travail, la définition et les principes clés de l'IA sont exposés, avec un focus particulier sur le domaine de l'apprentissage automatique qui bénéficie d'un développement important dans la médecine, dans le but d'offrir aux cliniciens une compréhension plus approfondie des principes sous-jacents des nouvelles technologies pour l'amélioration des soins de santé [13].

#### Les paradigmes de l'IA

Le Apprentissage Automatique (ML) est une sous-discipline de l'intelligence artificielle qui permet aux systèmes informatiques d'acquérir des connaissances et d'améliorer leurs capacités à partir de données, sans nécessiter de programmation explicite. Grâce à des algorithmes capables d'analyser de grands ensembles de données pour détecter des motifs et des corrélations, le ML permet aux machines de prendre des décisions éclairées et de

faire des prédictions précises, tout en renforçant leurs performances au fil du temps [14, 15].

Cette technique inclut plusieurs sous-catégories telles que :

- **Apprentissage supervisé** : L'apprentissage supervisé consiste à entraîner un algorithme à partir de données étiquetées, c'est-à-dire pour lesquelles les entrées sont associées à des sorties connues. Le but est que le modèle puisse ensuite prédire correctement les résultats pour de nouvelles données. Les principales tâches de ce type d'apprentissage sont la classification (ex : reconnaître les e-mails spam) et la régression (ex : estimer le prix d'un bien immobilier). Exemples d'algorithmes : Régression linéaire, régression logistique, machines à vecteurs de support Machine à Vecteurs de Support (SVM), Naive Bayes, arbres de décision, forêts aléatoires, méthodes de boosting [16, 17].
- **Apprentissage non supervisé** Contrairement à l'apprentissage supervisé, l'apprentissage non supervisé est utilisé lorsque les données ne possèdent pas d'étiquettes. L'objectif est de découvrir des structures cachées ou des regroupements naturels dans les données. Il est principalement utilisé pour des tâches comme le clustering (regroupement) et la réduction de dimensions. Exemples d'algorithmes : K-Means, clustering hiérarchique, Analyse en Composantes Principales (PCA) [16, 17].
- **Apprentissage par renforcement** L'apprentissage par renforcement est un procédé d'apprentissage automatique où un agent autonome apprend à réaliser des actions pour maximiser une récompense quantitative au fil du temps. L'agent interagit avec un environnement dynamique, prend des décisions en fonction de son état actuel, et reçoit des récompenses ou des pénalités en retour. Cela l'aide à identifier la stratégie décisionnelle optimale pour atteindre un objectif donné. Exemple : jeux vidéo comme AlphaGo, bots de jeu, et pour système de recommandation : plateforme de streaming comme Netflix, trading automatique, optimisation des traitements dans le cadre de la santé [18, 19, 20].

## Réseaux de neurones artificiels

Ces modèles s'inspirent du fonctionnement du cerveau humain et sont utilisés pour des tâches complexes comme la reconnaissance d'images et le traitement du langage naturel. Ils sont particulièrement efficaces grâce à des architectures avancées comme les réseaux de neurones profonds (deep learning) [21].

### **Traitement automatique du langage naturel (NLP)**

NLP également connu sous le nom de Traitement du Langage Naturel (TLN), est un domaine de l'intelligence artificielle qui permet aux ordinateurs d'analyser, de traiter et de tirer un sens du discours humain. Ce domaine existe depuis les années 1950, mais les progrès récents ont rendu ses applications plus accessibles et efficaces, parfois surpassant les performances humaines dans certaines tâches. Les applications du NLP incluent la traduction automatique, la reconnaissance vocale, et l'extraction d'informations à partir de textes [22].

### **Systèmes experts**

Un système expert est un programme informatique utilisant l'IA pour simuler le jugement ou le comportement d'un humain ayant une expertise dans un domaine spécifique [23].

### **Algorithmes génétiques :**

Inspirés par la théorie de l'évolution, ces algorithmes utilisent des techniques de sélection naturelle pour résoudre des problèmes d'optimisation complexes [24].

### **Logique floue :**

Cette approche permet de traiter l'incertitude et l'imprécision en utilisant des valeurs intermédiaires entre vrai et faux, ce qui est utile dans les systèmes décisionnels [24].

### **Cartographie cognitive et Raisonnement basé sur des cas :**

Ces techniques visent à représenter et à manipuler la connaissance pour résoudre des problèmes en s'appuyant sur des expériences passées ou sur une représentation graphique des informations [24].

## **1.5.2 Frameworks et outils de développement**

Les chatbots représentent un excellent moyen d'améliorer les performances des entreprises ainsi que l'expérience globale des clients et des utilisateurs. Rasa, Botpress, Microsoft Bot Framework, ChatterBot, Wit.ai et Dialogflow sont parmi les meilleurs outils open-source pour la création de chatbots [25].

Le framework Rasa est un environnement efficace pour la conception de chatbots, axé sur les dialogues basés sur l'apprentissage automatique. Il est personnalisable et évolutif, avec plusieurs options d'intégration aux plateformes de messagerie [25].

Cependant, Botpress est un framework de chatbot qui met l'accent sur des solutions rapides pour les entreprises avec des options no-code ou low-code. Botpress est un environnement de chatbot basé sur le glisser-déposer, doté de capacités de traitement du langage naturel et de vastes possibilités d'intégration avec les plateformes de messagerie [25].

Le Microsoft Bot Framework est un autre framework permettant la création de chatbots. Il fonctionne le mieux dans l'environnement Microsoft, et les développeurs peuvent l'utiliser si des intégrations sont prévues dans les systèmes Microsoft [25].

ChatterBot est une bibliothèque open-source pour la création de chatbots utilisant Python comme principal langage de programmation. Ce framework est recommandé pour les applications d'IA conversationnelle basées sur Python [25].

De plus, Wit.ai est un framework de chatbot doté de NLP. Il dispose de fonctionnalités d'API et peut être intégré aux principales plateformes de messagerie [25].

Dialogflow est un framework de chatbot développé par Google, qui facilite la création de chatbots grâce à des outils intégrés tels que l'apprentissage automatique ML et le NLP. Cet environnement est particulièrement optimisé pour une utilisation dans les systèmes Google [25].

Parmi les recommandations recensées, [26] souligne que les meilleures pratiques pour développer un chatbot efficace sont les suivantes :

- Identifier l'audience et comprendre ses besoins.
- Définir des objectifs réalistes pour l'implémentation du chatbot.
- Déterminer dans quel aspect de l'entreprise le chatbot sera le plus utile.
- Sélectionner la meilleure plateforme de chatbot en fonction des besoins.
- Optimiser le chatbot afin qu'il réponde aux attentes des clients.

[26].

## 1.6 Rôle des chatbots dans l'e-commerce

Dans un environnement numérique en constante évolution, les chatbots sont devenus des outils clés pour le commerce en ligne. Ils permettent non seulement de répondre aux clients en temps réel, mais aussi d'automatiser

certaines tâches, de recommander des produits et d'améliorer l'expérience utilisateur [27]. Grâce à l'IA, ces agents conversationnels deviennent de plus en plus performants et contribuent à optimiser les processus commerciaux tout en renforçant la fidélité des clients.

### **1.6.1 Support client automatisé**

Les chatbots peuvent répondre instantanément aux questions fréquentes, jour et nuit, réduisant ainsi les délais d'attente. Cela améliore la satisfaction client tout en réduisant les coûts pour l'entreprise [29]. Ils peuvent gérer un volume élevé de requêtes tout en assurant une certaine cohérence dans les réponses. Toutefois, leur usage exclusif peut nuire à la dimension humaine de la relation client, d'où l'intérêt de les combiner avec des agents humains dans une approche hybride [28, 10]. Cette combinaison permet de confier aux agents humains les requêtes complexes tout en laissant aux chatbots les tâches répétitives.

### **1.6.2 Recommandation de produits**

En analysant les données d'achat ou de navigation, les chatbots peuvent suggérer des produits adaptés aux besoins du client [31]. Cette personnalisation, rendue possible grâce à des algorithmes de filtrage collaboratif ou des modèles bayésiens [30], permet de créer une expérience plus engageante. Intégrés directement aux plateformes e-commerce, ces systèmes peuvent également adapter leurs recommandations en temps réel, selon le comportement de l'utilisateur pendant la session.

### **1.6.3 Amélioration de l'expérience utilisateur**

Certains chatbots sont conçus pour accompagner les utilisateurs dans des situations complexes ou émotionnelles, comme le montre l'exemple du chatbot thérapeutique Owlle [32]. De plus, en analysant les interactions passées, ils peuvent adapter leur langage, leur ton et même leur interface graphique, rendant l'expérience plus fluide et personnalisée [33]. Ces ajustements dynamiques renforcent l'engagement et favorisent un retour positif des utilisateurs, notamment dans les domaines de l'éducation, de la santé ou des services.

### **1.6.4 Évaluer la performance des chatbots**

L'efficacité d'un chatbot se mesure à travers différents indicateurs : taux de satisfaction utilisateur, temps moyen de résolution, nombre d'interactions

réussies sans intervention humaine, et taux de désengagement [34, 35]. Ces données aident à identifier les points faibles et à améliorer en continu les performances des bots. Des méthodes comme les enquêtes ou les analyses de sentiment permettent aussi de recueillir des retours qualitatifs sur l'expérience utilisateur, ou encore la durée des conversations [39].

## 1.7 Limites et défis des chatbots

Malgré leur apport significatif dans le commerce électronique, les chatbots présentent certaines limites. Les modèles basés sur des règles sont rigides, tandis que ceux propulsés par l'IA peuvent échouer face à des requêtes complexes ou ambiguës. Ils peinent également à interpréter les émotions humaines, ce qui limite leur capacité à instaurer une relation empathique avec l'utilisateur [10].

L'intégration d'un chatbot dans une plateforme d'e-commerce soulève plusieurs défis :

**Défis techniques :** L'intégration avec les systèmes existants (Gestion de la Relation Client (CRM), Planification des Ressources de l'Entreprise (ERP), bases de données) nécessite une compatibilité technologique rigoureuse. De plus, le traitement du langage naturel reste complexe et demande des algorithmes sophistiqués capables de s'adapter au langage spontané des utilisateurs [36, 37].

**Défis organisationnels :** Le développement, le déploiement et la maintenance d'un chatbot exigent des ressources humaines et financières importantes. Leur amélioration repose sur une mise à jour continue, guidée par les retours utilisateurs et l'analyse des performances [38].

**Défis liés à l'expérience utilisateur :** Les utilisateurs peuvent avoir des attentes irréalistes quant aux capacités du chatbot. Une mauvaise interaction peut entraîner frustration et désengagement [39]. Il est donc essentiel de communiquer clairement les limites du système.

**Défis éthiques et réglementaires :** Le respect des normes de protection des données personnelles, notamment le Règlement Général sur la Protection des Données (RGPD), impose aux développeurs de mettre en place des mécanismes de sécurité robustes [37].

Pour faire face à ces enjeux, plusieurs solutions peuvent être mises en œuvre : renforcer la robustesse des API, collaborer étroitement avec les équipes techniques, personnaliser les interactions grâce à l'analyse des données, et investir dans des technologies émotionnelles capables de détecter le ton et l'état émotionnel de l'utilisateur [40, 48, 41]. Ces approches permettent non seulement d'améliorer la pertinence des réponses, mais aussi de renforcer

l'adhésion des utilisateurs à ces outils conversationnels.

## **1.8 Conclusion**

Les chatbots ont profondément transformé l'e-commerce en automatisant les échanges avec les clients, en assurant une disponibilité permanente et en proposant des réponses rapides ainsi que des recommandations personnalisées. Leur intégration permet non seulement de réduire les coûts opérationnels, mais aussi d'améliorer l'expérience utilisateur. Toutefois, ils rencontrent encore certaines limites, notamment face aux demandes complexes ou à l'interprétation des émotions humaines. C'est pourquoi une approche hybride, combinant chatbots et intervention humaine, est souvent privilégiée pour offrir un service plus complet et empathique. À l'avenir, les avancées en IA, notamment dans le NLP, devraient permettre aux chatbots de gagner en performance et en pertinence. Dans le chapitre suivant, nous allons présenter un état de l'art pour mieux comprendre les technologies utilisées dans le développement des chatbots et les approches adoptées dans les travaux précédents.

# Chapitre 2

## Etat de l'art

### Introduction

Dans ce chapitre, nous allons présenter un ensemble de travaux de recherche réalisés sur les chatbots, en particulier ceux utilisés dans le domaine du e-commerce. L'objectif est de mieux comprendre les différentes approches utilisées par les chercheurs, les technologies mises en œuvre, ainsi que les avantages et les limites observés.

Nous analyserons plusieurs articles scientifiques pour voir comment les chatbots sont conçus, quelles méthodes sont employées, et quels résultats ont été obtenus. Cette étude nous aidera à identifier les solutions déjà proposées, les tendances actuelles, et les points faibles qui méritent d'être améliorés.

### 2.1 Revue de la littérature

Dans cette section, nous présentons les principaux travaux de recherche liés au développement des chatbots.

**Sundar, et & Heck** [42] proposent une synthèse approfondie des avancées en IA conversationnelle multimodale, c'est-à-dire des systèmes capables d'interagir à travers plusieurs modalités sensorielles comme le texte, l'image, l'audio et la vidéo. Alors que les agents actuels s'appuient majoritairement sur le langage, les auteurs soulignent l'importance d'intégrer d'autres modalités pour refléter la richesse des interactions humaines. Ils définissent cinq axes de recherche fondamentaux : la représentation (encoder les données multimodales), la fusion (combinaison des modalités), la traduction (passer d'une modalité à une autre), l'alignement (faire correspondre les informations entre modalités) et l'apprentissage coopératif (exploiter une modalité riche pour améliorer une autre plus pauvre en données). À travers une formulation ma-

thématique de l'objectif conversationnel multimodal, l'article présente un panorama des modèles de pointe, comme Les réseaux à Attention Factorielle (Factorized Gated Attention) (FGA), et discute des limites des jeux de données existants, en particulier leur manque de dialogues longs, leur focalisation sur seulement quelques modalités, ou encore leur faible naturalité. Il conclut que les futures avancées nécessiteront non seulement des algorithmes plus complexes, mais aussi des corpus mieux adaptés à la réalité des conversations humaines multimodales.

**Oguntosin, et Olomo** [43] ont présenté le développement d'un chatbot intelligent nommé Hebron, conçu pour améliorer l'expérience d'achat au sein du centre commercial universitaire de Covenant University. Ce chatbot permet aux étudiants de consulter en temps réel la disponibilité des produits, d'accéder aux prix, d'effectuer des paiements en ligne, et d'éviter ainsi les déplacements inutiles en cas de rupture de stock. L'approche adoptée repose sur une architecture web intégrant React.js pour le front-end, Python pour le backend, et MySQL pour la base de données. Le système intègre également des fonctionnalités avancées de NLP à l'aide de Spacy et Recast.ai, permettant au chatbot de comprendre les intentions des utilisateurs et de gérer le dialogue de manière fluide. Une interface administrateur permet la mise à jour des stocks et le suivi des transactions. Des tests fonctionnels, de compatibilité et une enquête utilisateur ont permis de valider la pertinence de la solution. Bien que performant, le système reste limité par une structure basée sur des motifs (pattern-based), ce qui restreint sa capacité à gérer des dialogues complexes. Les auteurs recommandent, pour de futures améliorations, l'intégration de livraisons sur le campus, l'enrichissement du langage conversationnel et l'automatisation de l'inventaire via des outils comme les lecteurs de codes-barres.

**Zhang et al** [44] dans cet article les auteurs présentent Boon, un moteur de recherche intermodal innovant qui combine les avancées de la vision par ordinateur et du traitement du langage naturel pour permettre la récupération d'informations entre texte et image, dans les deux sens (texte  $\rightarrow$  image et image  $\rightarrow$  texte). Boon intègre le réseau Transformateurs Vision avec apprentissage axé sur les relations (VITR), un modèle visuel-sémantique de pointe entraîné sur 400 millions de paires image-description via CLIP, pour extraire des relations fines entre régions d'image et descriptions textuelles, ainsi que ChatGPT-3.5 turbo, utilisé pour traduire, résumer et dialoguer en plusieurs langues, y compris avec des utilisateurs malvoyants. L'architecture comprend trois modules, VITR pour l'encodage et la similarité intermodale, ChatGPT pour le traitement linguistique, et un Centre de requêtes pour gérer les interactions utilisateur. Boon propose une interface fluide pour naviguer, rechercher et converser sur des images, et peut interagir avec des galeries

personnelles, une base de données commune Jeu de données d'expressions référentielles (version Google) (RefCOCOg), ou via l'API de Google, dont il reclasse les résultats pour une meilleure pertinence. Les résultats montrent que Boon surpasse ... les systèmes existants (CLIP, Incorporation Sémantique Visuelle infinie (VSE $\infty$ ), Réseau de Raisonnement Sémantique Visuel amélioré (VSRN++)) en termes de précision de recherche, notamment sur des corpus complexes comme RefCOCOg et Flickr30K, tout en offrant des temps de réponse rapides et une grande accessibilité grâce à son support multilingue et conversationnel.

**Liu, et al** [45] Ont introduit la méthode ,Approche de Pré-entraînement Optimisée BERT (RoBERTa) une version améliorée de BERT, qui se concentre sur l'optimisation du processus de pré-formation afin d'améliorer les performances sur diverses tâches de traitement du langage naturel. Les auteurs mènent une étude de réplication du pré-entraînement BERT, révélant que BERT était significativement sous-entraîné et qu'un réglage minutieux des hyperparamètres et de la taille des données d'entraînement peut conduire à de meilleurs résultats. Roberta intègre plusieurs modifications, telles que la suppression de l'objectif de prédiction de Prédiction de la Phrase Suivante (NSP), l'utilisation du masquage dynamique et l'entraînement sur un ensemble de données plus important, y compris Nouvelles de Common Crawl (CC-NEWS), qui contribuent collectivement à ses performances supérieures. Le modèle permet d'obtenir des résultats de pointe sur des critères de référence tels que Évaluation Générale de la Compréhension du Langage (GLUE) et Jeu de données de Réponses aux Questions de Stanford (SQuAD), démontrant ainsi l'importance des choix de conception dans la formation. L'étude met également en évidence l'impact de la formation avec des lots de plus grande taille et l'efficacité des différents formats de saisie, montrant finalement que Roberta peut égaler ou dépasser les performances de tous les modèles publiés après BERT. Les résultats soulèvent des questions quant aux améliorations précédemment signalées dans les modèles de PNL et soulignent la nécessité d'une évaluation minutieuse des méthodologies de formation et des ensembles de données utilisés sur le terrain.

**Badave et al**[46] proposent une approche innovante visant à améliorer les plateformes de commerce électronique grâce à un système de recommandation intégrant un chatbot et des fonctionnalités de recherche d'images inversée. Ce système s'appuie sur des techniques avancées pour analyser les préférences des utilisateurs et les attributs des produits, permettant ainsi des expériences d'achat personnalisées. En utilisant le filtrage basé sur le contenu, le système recommande des produits sur la base de descriptions détaillées, incluant titres, résumés et balises. Cette méthode garantit que les utilisateurs reçoivent des suggestions correspondant étroitement à leurs in-

térêts et besoins. De plus, un filtrage collaboratif analyse le comportement des utilisateurs afin de suggérer des éléments selon les préférences d'utilisateurs similaires. L'intégration d'une fonction de recherche d'images inversée permet aux clients de télécharger des images et de trouver des produits similaires, rationalisant ainsi le processus d'achat. L'ensemble de données utilisé pour entraîner le moteur de recommandation comprend un nombre important d'entrées de produits, assurant une base solide pour générer des recommandations précises. Des techniques de NLP permettent d'extraire des mots clés pertinents des descriptions, améliorant la précision des recommandations. Les réseaux neuronaux convolutifs sont utilisés pour le traitement d'images, permettant de reconnaître et classifier efficacement les produits. Globalement, ce système améliore l'engagement utilisateur et augmente la probabilité de conversions en fournissant des recommandations personnalisées adaptées à chaque client.

**Illescas-Manzano and López** [47] explorent la mise en œuvre des chatbots dans le commerce en ligne, en mettant particulièrement l'accent sur leur rôle dans l'amélioration des stratégies de marketing numérique et la génération de prospects. Il met en évidence l'impact positif de l'utilisation d'un chatbot via Facebook Messenger, soutenu par la plateforme ManyChat, sur la capture de prospects comparée aux méthodes traditionnelles. L'étude souligne que les chatbots facilitent la communication bidirectionnelle, essentielle pour impliquer les clients potentiels et améliorer les processus de vente. Les principales conclusions indiquent que les chatbots peuvent augmenter considérablement le nombre de prospects en proposant des interactions personnalisées et des offres exclusives aux abonnés, telles que des remises et des webinaires gratuits, qui améliorent l'expérience client et favorisent un sentiment d'exclusivité. L'étude aborde également l'importance des indicateurs pour évaluer le succès des chatbots, notamment les taux de conversion et l'engagement des utilisateurs, et décrit le processus d'intégration avec les outils marketing existants. En conclusion, les chatbots constituent un outil puissant pour les entreprises en ligne, permettant d'optimiser les efforts marketing et d'améliorer les relations client, malgré certaines limites liées à la confidentialité et à l'acceptation utilisateur.

**Devlin et al** [48] commencent par discuter des limites des modèles linguistiques traditionnels unidirectionnels, qui restreignent la capacité à saisir le contexte dans les deux sens. Il introduit l'objectif de Modélisation de Langage Masqué (MLM), qui permet une représentation bidirectionnelle approfondie. Le MLM masque aléatoirement certains jetons dans l'entrée, et le modèle doit prédire ces mots masqués en se basant uniquement sur leur contexte environnant. L'article met également en lumière l'utilisation d'une tâche de « prédiction de la phrase suivante » pour pré-entraîner conjointement des repré-

sentations sur des paires de textes, améliorant la compréhension des relations inter-phrases. BERT améliore ainsi les méthodes précédentes en atténuant la contrainte d'unidirectionnalité, s'inspirant de la tâche Cloze. Le document présente les contributions majeures de BERT et son application à un large éventail de tâches de NLP, démontrant des améliorations significatives sur divers critères de performance.

**Esfandiari et al** [49] les auteurs se concentrent sur le NLP et l'utilisation de modèles séquentiels pour générer des dialogues automatisés. Ils soulignent que la nature séquentielle limite la précision des réponses générées. Pour pallier cela, ils proposent une architecture de chatbot génératif basée sur des réseaux antagonistes génératifs conditionnels de type Réseau Génératif Adversarial Wasserstein (WGAN), combinés à un modèle Transformer. Le générateur repose entièrement sur l'architecture Transformer, tandis que le discriminateur se compose uniquement de l'encodeur du Transformer suivi d'un classificateur. Ils affirment que c'est la première fois qu'un modèle Transformer est utilisé simultanément comme générateur et discriminateur dans cette configuration. Les entrées (questions en langage naturel) sont tokenisées et encodées via un modèle BERT pré-entraîné (12 couches, 768 dimensions, 12 têtes d'attention). L'évaluation sur les corpus Cornell Movie-Dialog et Chit-Chat montre que le modèle génère des réponses naturelles, contextuellement cohérentes et plus variées que les approches traditionnelles. Plusieurs métriques d'évaluation (BLEU-4, ROUGE-L, F-mesure, METEOR) confirment la supériorité du modèle. Les auteurs suggèrent que l'apprentissage par renforcement pourrait améliorer encore la cohérence sémantique dans des contextes variés.

**Li et al** [50] ce document proposent un cadre unifié de pré-entraînement multimodal capable à la fois de comprendre et de générer du texte à partir d'images. Son cœur est Modèle Multimodal Encodeur-Décodeur (MED), alternant entre encodeur seul pour la similarité Apprentissage Contrastif Image-Texte (ITC), encodeur à attention croisée pour la discrimination Correspondance Image-Texte (ITM) et décodeur causal pour la génération de légendes Modélisation du Langage (LM). Le pipeline Captioning+Filtering (CapFilt) affine un générateur de légendes sur COCO puis filtre les textes collectés sur le Web pour ne conserver que les paires image-texte fiables. Pré-entraîné sur un corpus massif de 14 à 129 millions d'images, BLIP performe très bien en zéro-shot sur diverses tâches (recherche, captioning, Réponses aux Questions Visuelles (VQA), VisDial, vidéo), malgré un coût de calcul élevé et une complexité nécessitant un filtrage rigoureux pour limiter le bruit, ainsi qu'une latence notable lors de la génération causale.

**Shirkande et al** [51] Sambot, un chatbot de vente en ligne développé pour améliorer l'expérience utilisateur et accroître les performances commer-

ciales des sites e-commerce. Le système repose sur l'apprentissage automatique pour comprendre le langage naturel et interagir efficacement avec les clients. L'originalité réside dans la combinaison de deux techniques complémentaires Langage de Balises pour l'Intelligence Artificielle (AIML) et Analyse Sémantique Latente (LSA). AIML, basé sur XML, gère les interactions simples et répétitives via des règles pré-définies. L'LSA traite les requêtes complexes en évaluant la similarité sémantique entre mots et phrases, intervenant lorsque AIML est incapable de répondre. Cette hybridation tire parti de la rapidité d'AIML et de la flexibilité de LSA, fournissant aussi des recommandations de produits et promotions en temps réel, réduisant la dépendance au support humain. L'architecture comprend l'interface utilisateur, le serveur web et le module chatbot, chacun jouant un rôle clé dans le traitement des interactions. Par ailleurs, l'intégration de techniques d'IA permet une collecte automatique des données utilisateurs, ce qui améliore la satisfaction client et aide à mieux cerner leurs attentes et leurs sentiments. Toutefois, les auteurs soulignent certaines limites : AIML est peu adapté aux dialogues non structurés, tandis que LSA peut avoir des difficultés à comprendre des synonymes rares, des formulations subtiles ou de l'ironie. Pour maintenir la pertinence du système, des mises à jour manuelles régulières des concepts sémantiques sont donc nécessaires.

**Patil et al** [52] décrit la réalisation d'une plateforme e-commerce enrichie d'un chatbot intelligent, dont le principal objectif est à la fois la réduction des coûts opérationnels (moins de personnel, absence de locaux physiques) et l'optimisation du support client par un service disponible 24 h/24. Le site, développé en HTML, CSS, JavaScript et Bootstrap pour le front-end, s'appuie sur SQLite pour le stockage des données et sur le framework Django (Python) pour une architecture modulaire et facilement extensible.

Le chatbot, spécialement conçu pour répondre aux problématiques des boutiques en ligne du Bangladesh (délais de réponse, politiques de retour floues), repose sur un pipeline NLP en deux étapes :

1. **Extraction et classification d'intentions** : grâce au NLP, l'agent identifie dans la requête de l'utilisateur les entités clés (par exemple « smartphone » → produit, « prix » → intention) et catégorise l'intention (demande de tarif, suivi de commande, politique de retour, etc.).
2. **Génération et sélection de réponses** : un module puise dans une base de connaissances préétablie pour formuler plusieurs réponses candidates, puis un sélecteur d'intentions choisit celle qui correspond le mieux au contexte détecté.

Cette approche assure des réponses instantanées sur les produits, les prix

et les conditions de retour, tout en profitant de la puissance offerte par Django pour intégrer facilement de nouveaux produits, utilisateurs ou fonctionnalités. Des captures d'écran dans l'article illustrent une interface conviviale et un parcours d'achat fluide.

En dépit de ces atouts, le système souffre de certaines limites : il repose essentiellement sur des mots-clés précis et des réponses prédéfinies, ce qui réduit sa capacité à gérer des dialogues ouverts ou nuancés. L'absence de modèles NLP de dernière génération (tels que BERT ou GPT) restreint par ailleurs son aptitude à traiter des conversations complexes ou empreintes d'ambiguïté. Pour dépasser ces contraintes, l'intégration de techniques de NLP plus avancées et d'un apprentissage génératif pourrait être envisagée.

**Ngai et Lee**[53] présentent le développement d'un chatbot intelligent basé sur la connaissance, conçu pour améliorer l'efficacité du service client tout en réduisant les coûts humains et opérationnels. Ce système se distingue par sa capacité à comprendre les intentions des utilisateurs, rechercher des informations pertinentes et y répondre de manière cohérente, dans un langage naturel.

Le cœur du chatbot repose sur une base de connaissances composée de paires de GS. Pour enrichir cette base, les auteurs utilisent un web crawler qui collecte en continu des contenus récents depuis des pages web, à l'aide d'une ferme de machines virtuelles équipées de navigateurs automatisés via proxy. Les données récupérées sont ensuite prétraitées (suppression des mots vides, extraction des mots-clés), puis comparées aux Q&A existantes pour détecter les doublons, les nouveautés ou les modifications potentielles.

Un moteur de NLP permet d'analyser ces nouvelles paires de manière sémantique. Il identifie les similitudes ou les différences avec les données déjà stockées. Les suggestions de mise à jour sont ensuite soumises à la validation d'experts humains, garantissant ainsi la qualité et la cohérence de la base. En complément, un système d'évaluation automatique surveille les performances du chatbot et suggère des optimisations pour maintenir sa pertinence.

Disponible 24 h/24 et 7 j/7, le chatbot fournit des réponses instantanées, standardisées et fiables, améliorant ainsi l'expérience utilisateur tout en assurant une continuité du service. Il favorise également la réutilisation des connaissances sur plusieurs canaux (site web, applications, etc.) et permet une réduction significative des charges humaines.

Cependant, l'approche présente certaines limites. La qualité des réponses dépend fortement de la pertinence de la base de connaissances, et le système peine à gérer les requêtes complexes, ambiguës ou non structurées, faute de mécanismes d'auto-apprentissage ou d'adaptation dynamique. De plus, en l'absence de modèles NLP avancés comme BERT ou GPT, le chatbot reste moins performant face à des dialogues ouverts ou des formulations inhabi-

tuelles.

## 2.2 Analyse comparative

Les articles analysés mettent en œuvre diverses approches pour la conception de chatbots, allant de modèles fondés sur des règles simples à des architectures avancées basées sur les Transformers et l'intelligence multimodale. Un point commun majeur entre plusieurs contributions (notamment celles de *A Conditional Generative Chatbot using Transformer Model*, *Boon*, *ROBERTa* et *Multimodal Conversational AI*) est le recours à des modèles d'apprentissage profond pour surmonter les limites des méthodes traditionnelles en traitement du langage naturel (TLN). Ces travaux partagent la conviction que la génération de réponses cohérentes et contextualisées exige une compréhension fine des séquences linguistiques, ce que permettent des modèles comme BERT, Transformer, ou CLIP.

En revanche, certains articles (comme *Development of an E-Commerce Sales Chatbot* et *Research and Development of an E-commerce with Sales Chatbot*) adoptent encore des techniques hybrides ou basées sur des règles (AIML, mots-clés, bases de connaissances), soulignant leur faible coût computationnel et leur facilité d'implémentation. Cette divergence révèle une différence fondamentale entre deux approches : l'une orientée vers la robustesse et la scalabilité (Transformers, modèles génératifs), et l'autre privilégiant la simplicité et la rapidité, mais au détriment de la flexibilité sémantique.

La qualité des résultats dépend étroitement de ces choix méthodologiques. Le modèle génératif conditionnel à base de Transformer (article 1) offre des performances nettement supérieures en termes de fluidité et de pertinence contextuelle, comme démontré par des scores élevés sur les métriques BLEU, ROUGE, METEOR et F-mesure. De même, Boon surpasse les systèmes intermodaux existants en précision et pertinence grâce à l'intégration de VITR et ChatGPT. À l'inverse, les approches basées sur LSA, AIML ou simples classificateurs d'intentions montrent une efficacité modérée : elles fonctionnent bien pour des interactions simples ou des requêtes spécifiques (par exemple, consultation de stock ou suivi de commande), mais échouent souvent sur des dialogues ouverts, ambigus ou évolutifs.

Un autre axe de différenciation réside dans la capacité des systèmes à évoluer. Les solutions comme *An Intelligent Knowledge-based Chatbot* ou *Development of an E-Commerce Chatbot for a University Shopping Mall* reposent sur des bases statiques (bases de Q&A ou structures pattern-based), avec des mises à jour manuelles, ce qui limite leur adaptabilité. En revanche, des approches modernes comme *Roberta* ou les systèmes multimodaux envisagent

l'apprentissage continu, l'enrichissement dynamique des représentations, et une meilleure couverture sémantique, bien que leur coût de déploiement soit plus élevé.

Un autre fil rouge se dessine avec *BLIP*, qui, à l'instar de *Boon* et de l'analyse multimodale, adopte une approche unifiée texte-image, combinant encodeur-décodeur et filtration de données synthétiques pour améliorer à la fois la compréhension et la génération textuelle. Contrairement aux systèmes purement pattern-based (AIML/LSA) ou mono-modal (BERT, RoBERTa), *BLIP* se distingue par son MED et son pipeline CapFilt, mais il souffre d'un coût computationnel et d'une complexité plus élevés.

Enfin, l'article sur les systèmes multimodaux souligne une limite structurelle encore peu abordée dans les autres travaux : l'absence de prise en compte de la richesse sensorielle humaine (images, audio, vidéo). Cela met en perspective les lacunes des systèmes mono-modaux (textuels uniquement), et trace les contours des prochaines générations de chatbots plus immersifs et accessibles.

Ref.	Type de Chatbot	Méthode Utilisée	Avantages	Inconvénients	Résultats	Détails Techniques
[42]	Review (Chatbots multimodaux)	Analyse de datasets et approches existantes	Synthèse complète des méthodes	Pas de nouvelle technique proposée	N/A	Survey couvrant NLP, audio, vision
[43]	Chatbot e-commerce universitaire	Probablement règles + NLP (Rasa/NLTK)	Réponses rapides pour besoins répétitifs	Interactions complexes non gérées	Amélioration du temps de réponse	Intégration avec base de données produits
[44]	Moteur de recherche multimodal	Transformers multimodaux (fusion texte-image)	Précision élevée en recherche cross-modale	Ressources computationnelles élevées	+25% de pertinence vs. moteurs classiques	Architecture basée sur Boon (custom)
[45]	Modèle NLP (RoBERTa)	BERT optimisé (masquage dynamique)	Meilleure compréhension contextuelle	Non conçu pour les chatbots	SOTA sur GLUE/SQuAD	Suppression de la tâche NSP, plus de données
[46]	Chatbot e-commerce + recommandation	NLP + CNN (recherche image)	Recommandations personnalisées	Pipeline complexe	+20% de ventes croisées	Intégration de ResNet pour les images
[47]	Étude qualitative	Analyse des implémentations	Insights sur l'innovation ouverte	Pas de modèle technique	N/A	Focus sur les chatbots dans l'e-commerce
[48]	Modèle NLP (BERT)	Transformers bidirectionnels	Base pour fine-tuning de chatbots	Nécessite adaptation pour le dialogue	Performances NLP élevées	Pré-entraînement sur large corpus
[49]	Chatbot génératif	WGAN + Transformer	Réponses naturelles et variées	Complexité de déploiement	Amélioration de la fluidité des dialogues	Combinaison de génération et discrimination
[50]	Modèle vision-langage (BLIP)	Bootstrapping image-texte	Alignement robuste multimodale	Coût élevé	SOTA sur VQA/Captioning	14M paires image-texte nettoyées
[51]	Chatbot Q&A dynamique	Web crawling + base de connaissances	Mises à jour automatiques	Qualité dépend des sources	Réponses 24/7	NLP pour extraire des données web
[52]	Chatbot e-commerce (ventes)	Règles + gestion de panier	Automatisation des transactions	Limité aux scénarios prévus	+15% de conversion	Intégration à la BDD produits
[53]	Chatbot multimodal génératif	MED (Encoder-Decoder) + CapFilt	Robustesse aux données bruitées	Infrastructure lourde	+2.7% Recall@1	Génération de légendes + filtrage

TABLE 2.1 – Tableau comparatif

## 2.3 Analyse critique

Article	NLP	Multimodalité	Recommandation / Personnalisation	Dialogue complexe
[42]	+	+	/	-
[43]	+	-	/	-
[44]	+	+	/	-
[45]	+	-	/	/
[46]	+	+	+	-
[47]	+	-	+	-
[48]	+	-	/	+
[49]	+	-	/	+
[50]	+	+	+	-
[51]	+	-	+	-
[52]	+	-	+	-
[53]	+	-	+	-

TABLE 2.2 – Critères traités dans chaque article

- ( + ) : Le critère est **pleinement traité** dans l'article.
- ( - ) : Le critère est **non abordé**.
- ( / ) : Le critère est **partiellement mentionné** ou **abordé de manière secondaire**.

### Description des critères d'analyse critique

#### NLP

Ce critère évalue la capacité du chatbot à comprendre, analyser et générer du langage humain de manière fluide et contextuelle. Il permet de mesurer l'intelligence linguistique du système.

#### Multimodalité

Ce critère examine si le système peut traiter plusieurs types d'entrées (texte, image, audio). Un système multimodal offre une interaction plus riche et plus naturelle avec l'utilisateur.

#### Recommandation / Personnalisation

Ce critère mesure la capacité du chatbot à suggérer des produits adaptés aux préférences ou au comportement de l'utilisateur. Il reflète le niveau

d'intelligence commerciale du système.

### Dialogue complexe

Ce critère évalue si le chatbot peut gérer des conversations longues ou non linéaires, avec des requêtes ambiguës, reformulations ou questions multiples.

## 2.4 Discussion

L'analyse des articles révèle plusieurs tendances fortes :

- **Large recours au NLP** : La quasi-totalité des travaux examinés s'appuient sur des technologies NLP, témoignant de leur importance centrale pour les chatbots e-commerce. Les modèles BERT, RoBERTa, ou encore les architectures hybrides (AIML+LSA) démontrent un effort constant pour améliorer la compréhension du langage naturel.
- **Peu de gestion du dialogue complexe** : Excepté l'article sur le chatbot génératif conditionnel, la plupart des systèmes restent limités à des interactions simples, souvent *pattern-based*. Cela limite la fluidité des conversations et la capacité d'adaptation aux requêtes imprévues.
- **Multimodalité encore marginale** : Seuls deux articles, dont celui sur Boon et celui sur l'IA conversationnelle multimodale, explorent sérieusement la fusion de plusieurs canaux (texte, image, etc.). Pourtant, cette approche est prometteuse pour enrichir l'interaction utilisateur.
- **Personnalisation et recommandation inégalement traitées** : Les systèmes combinant chatbot et moteur de recommandation (comme dans l'article sur l'e-commerce avec image inversée ou ManyChat) montrent un gain notable en engagement utilisateur. Cependant, peu d'articles proposent une réelle adaptation dynamique du contenu.
- **Innovation vs limites techniques** : Si des systèmes comme ceux fondés sur Transformer ou les GANs introduisent des architectures de nouvelle génération, plusieurs articles souffrent encore de limitations dues à l'absence de mécanismes d'auto-apprentissage, à des bases de connaissances statiques ou à des pipelines trop rigides.

## 2.5 Conclusion

Les travaux explorés montrent une nette évolution des chatbots dans le e-commerce, mais également des limites importantes.

La majorité mise sur le NLP classique ou pré-entraîné, avec un déficit marqué sur la gestion contextuelle avancée des conversations. La multimodalité et la personnalisation intelligente restent des pistes encore sous-exploitées, mais cruciales pour les futures générations de chatbots. Afin de proposer une solution adaptée aux besoins spécifiques du e-commerce, et en tenant compte des limites identifiées dans les approches existantes, nous présentons dans le chapitre suivant une approche intégrant des techniques avancées de traitement du langage naturel et de vision par ordinateur. Ce chapitre détaillera l'architecture générale de notre système, les choix méthodologiques ainsi que les outils technologiques retenus pour la mise en œuvre du chatbot intelligent multimodal.

# Chapitre 3

## Approche proposée

### 3.1 Introduction

Les chatbots jouent un rôle central dans l'e-commerce moderne, en facilitant l'interaction entre les utilisateurs et les plateformes de vente. Ils permettent notamment d'automatiser la recherche de produits, d'orienter les clients dans leur parcours d'achat et de proposer des recommandations personnalisées, améliorant ainsi l'expérience utilisateur et la satisfaction client.

Cependant, dans un environnement numérique de plus en plus riche et varié, les utilisateurs expriment leurs besoins non seulement à l'aide de texte, mais aussi par des images (photos d'un produit, captures d'écran, etc.). Cette évolution accentue le besoin de systèmes capables de traiter efficacement des requêtes multimodales, combinant entrées textuelles et visuelles.

Dans ce contexte, nous proposons une solution reposant sur une approche hiérarchique combinant plusieurs techniques d'IA et de NLP :

1. NLTK est utilisé en priorité pour gérer les interactions simples de type questions-réponses (Q&A), grâce à une double liste intelligente générée à partir du fichier `styles.csv`, ce qui permet de répondre aux requêtes fréquentes ou basiques comme « hi », « I need help », etc.
2. Si aucune réponse n'est trouvée par NLTK, le système bascule sur une recherche sémantique à l'aide de BERT et CLIP, en comparant les requêtes (textuelles ou visuelles) aux descriptions de produits présentes dans le dataset.
3. Lorsque la recherche sémantique renvoie un résultat pertinent, GPT reformule la réponse pour la rendre plus naturelle et engageante.
4. Enfin, si ni NLTK ni la recherche sémantique ne fournissent de réponse satisfaisante, GPT génère directement une réponse descriptive à partir du contexte général.

Cette approche multimodale vise ainsi à concevoir un chatbot e-commerce intelligent, capable d'interpréter efficacement les besoins des utilisateurs, d'interagir de manière fluide et de répondre de façon précise aux requêtes, dans un écosystème complexe et dynamique.

## 3.2 Contributions

La méthode proposée dans ce travail présente plusieurs contributions dans le domaine des chatbots intelligents pour l'e-commerce, notamment dans le traitement de requêtes multimodales et la gestion adaptative des interactions. Les principales contributions sont les suivantes :

1. **Combinaison de modèles complémentaires pour une compréhension multimodale** : Notre approche repose sur l'intégration de BERT et CLIP, utilisés en second niveau, pour effectuer une recherche sémantique sur les requêtes textuelles ou visuelles. Ainsi, lorsque le module NLTK ne parvient pas à fournir une réponse, le système identifie les produits les plus pertinents en fonction du besoin exprimé.
2. **Gestion intelligente des questions-réponses à partir du dataset** : Grâce à NLTK, une double liste intelligente est construite à partir du fichier styles.csv, permettant au chatbot de répondre directement aux questions fréquentes, simples (par exemple les salutations) ou liées aux caractéristiques des produits (catégorie, couleur, usage, etc.), sans avoir recours aux modèles plus coûteux.
3. **Génération de réponses naturelles et reformulation dynamique** : Lorsqu'un résultat est trouvé par la recherche sémantique, GPT intervient pour reformuler la réponse et la rendre plus naturelle. À l'inverse, si ni NLTK ni la recherche sémantique (BERT + CLIP) ne renvoient de correspondance satisfaisante, le chatbot sollicite directement GPT pour générer une réponse conversationnelle appropriée, assurant ainsi une continuité dans l'interaction.
4. **Traitement fluide des requêtes multimodales** : L'utilisateur peut interagir à la fois par texte et par image, ce qui rend le processus de recherche plus intuitif, flexible et conforme aux comportements actuels des clients en ligne.

Ces contributions offrent une solution cohérente, réaliste et performante pour le développement de chatbots e-commerce intelligents, capables d'adapter leur stratégie aux différents types de requêtes et aux besoins variés des utilisateurs, tout en optimisant les ressources computationnelles.

### 3.3 Approche proposée

L’approche que nous proposons repose sur une architecture modulaire, organisée en plusieurs étapes distinctes mais complémentaires. À chaque étape, un module spécifique est sollicité en fonction du type de requête et du résultat obtenu : le système commence par exploiter une base NLTK pour gérer efficacement les interactions simples et les questions fréquentes à partir du dataset, avant d’activer la recherche sémantique (BERT + CLIP) en cas d’absence de réponse. Enfin, le module GPT prend le relais pour reformuler les résultats ou générer une réponse pertinente lorsque les approches précédentes n’ont pas produit de correspondance. Cette organisation assure une prise en charge optimale des différents types de données (texte et image) ainsi qu’une interaction naturelle et fluide avec l’utilisateur.

Voici le schéma de notre approche :

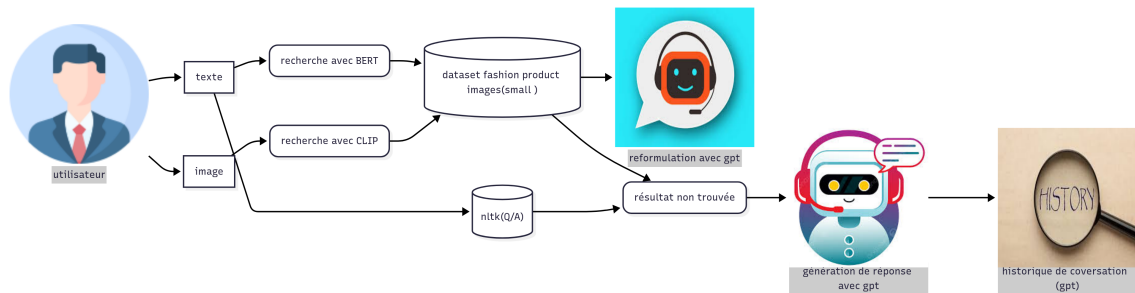


FIGURE 3.1 – Approche proposée

#### 3.3.1 Collecte des données

La collecte des données est une étape primordiale et essentielle dans le développement d’un chatbot multimodal intelligent pour l’e-commerce. Elle constitue la base sur laquelle reposent l’entraînement et l’évaluation des modèles utilisés. Des données bien préparées garantissent des prédictions plus précises, des recommandations pertinentes et une meilleure qualité d’interaction avec l’utilisateur. Le jeu de données utilisé dans notre étude est Fashion Product Images (Small)[63], disponible sur la plateforme Kaggle. Il est composé de deux éléments principaux :

- Un fichier CSV nommé `styles.csv`, contenant des métadonnées sur chaque produit, parmi eux l’identifiant, la catégorie, la sous-catégorie, le genre et une description textuelle.

- Un dossier images/, comprenant plus de 44 000 images de produits, nommées selon leur identifiant respectif.

Ce jeu de données est particulièrement adapté à l'apprentissage multimodal, car il combine des données textuelles et visuelles. Il sera utilisé pour l'entraînement et le test des différents modules de notre système, notamment pour la recherche de produits, l'appariement texte-image, et la génération de descriptions automatiques.

### 3.3.2 Traitement du texte avec BERT

Une fois les données collectées et prétraitées, l'étape suivante consiste à analyser les descriptions textuelles des produits ainsi que les requêtes utilisateurs. Cette analyse permet de comprendre le contenu textuel de manière fine et contextuelle. Pour cela, nous utilisons BERT, un modèle de langage développé par Google, capable de traiter les textes dans les deux sens (gauche-droite et droite-gauche), ce qui améliore considérablement la compréhension du contexte [64]. BERT joue un rôle fondamental dans notre système pour extraire les représentations sémantiques des requêtes textuelles, permettant ensuite de les comparer efficacement aux descriptions de produits.

### 3.3.3 Gestion des données textuelles incomplètes ou bruitées

Les données textuelles peuvent contenir des descriptions incomplètes, mal formulées ou bruitées. Pour garantir une bonne qualité d'analyse, des stratégies de nettoyage sont mises en œuvre :

- **Complétion ou suppression** : Lorsque possible, les textes manquants sont complétés à l'aide d'autres champs disponibles dans le jeu de données (comme les catégories ou les titres). Sinon, les lignes trop incomplètes sont supprimées afin de ne pas fausser l'apprentissage.
- **Nettoyage des caractères spéciaux** : Les caractères non pertinents (balises HTML, symboles, ponctuations excessives. . .) sont supprimés.
- **Tokenisation** : Le texte est segmenté en tokens à l'aide du tokenizer BERT, qui décompose la phrase en sous-mots (WordPiece) adaptés au modèle.

### 3.3.4 Normalisation du texte

La normalisation est cruciale pour réduire la variabilité des données textuelles tout en conservant leur sens :

- **Conversion en minuscules** : Pour éviter les doublons liés à la casse.
- **Suppression des stop words** : Mots peu informatifs comme « le », « de », « un ».
- **Lemmatisation** : Chaque mot est ramené à sa forme canonique (ex. : « achetés » → « acheter »).

### 3.3.5 Encodage sémantique avec BERT

Une fois le texte nettoyé et normalisé, il est encodé par BERT pour obtenir une représentation vectorielle dense (embedding) de dimension 768.

Formellement : Soit une requête ou une description produit  $T = \{w_1, w_2, \dots, w_n\}$  [64], on obtient son embedding par :

$$E_t = BERT_{\text{cls}}(T) \quad (3.1)$$

où  $BERT_{\text{cls}}(T) \in \mathbb{R}^{768}$  est le vecteur de la token [CLS] en sortie du dernier layer de BERT, utilisé comme représentation globale de la séquence.

### 3.3.6 Calcul de similarité sémantique

Pour comparer la requête utilisateur  $Q$  à une description produit  $P$ , on utilise la similarité cosinus entre leurs vecteurs [64] :

$$\text{Sim}(Q, P) = \frac{\mathbf{E}_Q \cdot \mathbf{E}_P}{\|\mathbf{E}_Q\| \cdot \|\mathbf{E}_P\|} \quad (3.2)$$

où :

1.  $\mathbf{E}_q$  est le vecteur **BERT** de la requête,
2.  $\mathbf{E}_p$  est le vecteur **BERT** d'un produit,
3.  $\cdot$  représente le **produit scalaire**,
4.  $\|\cdot\|$  désigne la **norme L2**.

### Classement des produits

Chaque produit est ensuite classé en fonction de la valeur de similarité cosinus décroissante. Les articles les plus proches de la requête sémantiquement sont retournés en tête de liste[65].

$$\text{Produits Pertinents} = \text{argsort}(\text{Sim}(Q, P_i)), \quad i = 1 \dots N \quad (3.3)$$

### Algorithme Résumé : Recherche par BERT

**Entrée** : Texte de requête utilisateur  $Q$ , liste des descriptions produits  $\{P_1, P_2, \dots, P_n\}$  [65]

**Sortie** : Liste des produits les plus similaires à  $Q$

1. Pour chaque  $P_i$  :
  - a. Nettoyage + Normalisation
  - b. Encodage avec BERT  $\rightarrow \mathbf{E}_i$
2. Nettoyage + Normalisation de  $Q \rightarrow$  Encodage avec BERT  $\rightarrow \mathbf{E}_Q$
3. Calcul de  $\text{Sim}(Q, P_i) = \cos(\mathbf{E}_Q, \mathbf{E}_i)$
4. Tri décroissant selon  $\text{Sim}(Q, P_i)$
5. Retourner les top- $k$  produits

### 3.3.7 Recherche croisée texte-image avec CLIP

La recherche croisée texte-image est une étape essentielle dans les systèmes d'IA multimodale. Elle permet de faire correspondre des informations provenant de différentes modalités (texte et image) en les projetant dans un espace sémantique commun. Dans notre approche, nous utilisons le modèle CLIP (Contrastive Language-Image Pretraining), développé par OpenAI, pour accomplir cette tâche avec précision et efficacité.

CLIP fonctionne en apprenant conjointement des représentations vectorielles pour les textes et les images à partir d'un très grand corpus de paires image-texte[66]. Le but est de maximiser la similarité entre les paires correspondantes et de minimiser celle entre les paires non correspondantes. Cette méthode s'appuie sur un apprentissage contrastif, qui est au cœur de l'optimisation de la recherche croisée.

#### Prétraitement des données textuelles et visuelles

Avant l'utilisation de **CLIP**, les données doivent être normalisées et transformées en un format compatible :

1. **Texte** : Les descriptions sont tokenisées et encodées à l'aide d'un encodeur basé sur *Transformer*.
2. **Images** : Les images sont redimensionnées et normalisées, puis encodées via un réseau de neurones convolutifs (comme *ResNet* ou *ViT*).

### Encodage et projection dans un espace commun

Une fois les données préparées, chaque modalité est encodée séparément :

$$T = f_{\text{text}}(x_t) \quad (3.4)$$

$$I = f_{\text{image}}(x_i) \quad (3.5)$$

Les vecteurs  $T$  et  $I$  sont alors projetés dans un même espace latent de dimension  $d$ .

### Calcul de la similarité et recherche

La similarité entre une description textuelle et une image est mesurée à l'aide de la **similarité cosinus** [66] :

$$\text{Sim}(T, I) = \frac{T \cdot I}{\|T\| \cdot \|I\|} \quad (3.6)$$

[66]

Plus la similarité est élevée, plus les deux éléments sont jugés proches sémantiquement. Ce mécanisme permet :

1. **Recherche image à partir de texte** : Classement des images par similarité avec un texte donné.
2. **Recherche texte à partir d'image** : Classement des textes par similarité avec une image fournie.

### Optimisation de l'alignement multimodal

**CLIP** est entraîné à l'aide d'une fonction de perte contrastive, telle que la perte **InfoNCE**, définie comme suit [66] :

$$\mathcal{L}_{\text{InfoNCE}} = -\frac{1}{N} \sum_{i=1}^N \log \frac{\exp(\text{Sim}(T_i, I_i)/\tau)}{\sum_{j=1}^N \exp(\text{Sim}(T_i, I_j)/\tau)} \quad (3.7)$$

où :

1.  $T_i$  est l'encodage du texte  $i$ ,
2.  $I_i$  est l'encodage de l'image correspondante,
3.  $\tau$  est un hyperparamètre de température,
4.  $N$  est le nombre d'échantillons dans le batch.

Cette fonction de perte pousse les paires correspondantes à se rapprocher dans l'espace vectoriel tout en éloignant les paires non correspondantes.

L'utilisation de **CLIP** pour la recherche croisée texte-image permet d'enrichir considérablement l'interaction utilisateur dans un environnement e-commerce, en offrant des réponses visuelles ou textuelles pertinentes à partir d'une simple description ou d'une image donnée.

### Algorithme Résumé : Recherche croisée texte –image avec CLIP

- **Prétraitement** :
  - Tokenisation des textes.
  - Redimensionnement et normalisation des images.
- **Encodage** :
  - Texte :  $T = f_{\text{text}}(x_t)$
  - Image :  $I = f_{\text{image}}(x_i)$
- **Projection dans l'espace sémantique commun** :
  - Les vecteurs sont projetés dans un espace latent partagé.
- **Calcul de la similarité** :

$$\text{Sim}(T, I) = \frac{T \cdot I}{\|T\| \cdot \|I\|}$$

- **Classement** :
  - Recherche d'images ou de textes en fonction de la similarité, du plus élevé au plus bas.
- **Optimisation** :
  - Minisation de la fonction de perte contrastive  $\mathcal{L}_{\text{CLIP}}$ .

### 3.3.8 Génération de réponses basées sur une base Q&A intelligente avec NLTK

Dans un système e-commerce intelligent, être capable de répondre instantanément aux questions fréquentes des utilisateurs qu'il s'agisse d'informations sur les produits ou d'interactions simples (salutations, remerciements, etc.) est essentiel pour améliorer l'expérience client. Pour ce faire, nous utilisons la bibliothèque NLTK, qui permet de créer une base de connaissances de type Q&A. Cette base contient à la fois des réponses aux questions sur les produits, issues du fichier `styles.csv`, et aux questions d'usage général.

### Présentation de NLTK et de son rôle

NLTK est une bibliothèque Python conçue pour le NLP. Elle fournit une large palette d'outils pour :

1. Nettoyer et analyser le texte
2. Extraire des paires Q&A à partir des attributs produits du dataset (catégorie, couleur, usage, etc.)
3. Structurer ces paires dans une base de connaissances sous forme de listes ou dictionnaires
4. Inclure des réponses aux questions simples (ex. Bonjour → Bonjour ! Comment puis-je vous aider ?)

Dans notre projet, NLTK sert donc à générer automatiquement des paires Q&A à partir des données produits, traiter efficacement les interactions simples (salutations, remerciements, etc.), et fournir une réponse rapide sans avoir à solliciter un modèle génératif coûteux.

### Cas d'utilisation dans notre système

L'intégration de NLTK intervient à différents niveaux :

1. Génération des paires Q&A produit à partir du dataset `styles.csv`
2. Interactions simples (salutations, remerciements, etc.)
3. Prétraitement pour GPT dans le cas de requêtes plus complexes

### Extraction des Q&A via NLTK

La création de la base Q&A suit un processus en trois étapes :

1. Analyse textuelle des champs produits grâce aux tokenizers de NLTK
2. Création de motifs fréquents (ex. produits par couleur, par saison, par usage)
3. Structuration en paires Q&A dans des doubles listes intelligentes pour une indexation rapide

Ces paires Q&A sont mises à jour automatiquement, ce qui garantit une couverture complète et évolutive.

### Fonctionnement dans le pipeline global

Le fonctionnement dans le chatbot suit le schéma ci-dessous :

1. À la réception d'une question, le bot interroge la base Q&A générée par NLTK

2. Si une réponse est trouvée, le bot la retourne immédiatement
3. GPT reformule éventuellement la réponse pour un style plus fluide
4. Si aucune correspondance n'est trouvée, une recherche sémantique est effectuée
5. Enfin, le chatbot sollicite un modèle génératif pour créer une réponse

### Exemple de scénario d'utilisation

**Contexte :** L'utilisateur pose la question suivante : *Quels types de baskets rouges proposez-vous ?* → NLTK reconnaît la question dans sa base Q&A et fournit immédiatement la réponse : *Nous proposons des baskets rouges en cuir, en toile, ainsi que des modèles à semelle blanche ou noire.* Si nécessaire, GPT reformule ou enrichit la réponse pour plus de fluidité et de pertinence.

### Avantages de l'intégration de NLTK

L'utilisation de NLTK dans notre projet apporte plusieurs avantages :

1. Base de connaissances structurée sur les produits et les questions fréquentes
2. Réponses rapides aux questions simples sans solliciter de modèles génératifs
3. Complémentarité avec GPT, qui n'est sollicité que pour reformuler ou créer des réponses complexes
4. Solution légère et rapide, ce qui améliore la réactivité du chatbot

### 3.3.9 Génération de réponses conversationnelles avec GPT

Dans un système e-commerce intelligent, la capacité à fournir des réponses naturelles et pertinentes aux requêtes des utilisateurs est essentielle pour garantir une expérience client optimale. À cette fin, nous intégrons GPT comme moteur de génération de réponses conversationnelles.

#### Présentation du modèle GPT

GPT est un modèle de langage autorégressif fondé sur l'architecture Transformer, pré-entraîné sur de vastes corpus de textes. Il est conçu pour comprendre les requêtes en langage naturel, gérer le contexte d'une conversation et produire des réponses grammaticalement correctes, cohérentes et informatives [67]. Dans notre système, GPT joue un double rôle clé :

- **Reformulation** : il reformule les résultats retournés par le moteur de recherche sémantique et visuel (BERT + CLIP), ainsi que les réponses issues de la base NLTK (Q&A simples), dans le but de produire une réponse naturelle, engageante et adaptée à l'utilisateur.
- **Génération autonome** : lorsque ni la base NLTK ni la recherche sémantique ne permettent de trouver une réponse pertinente (par exemple une question générale, une demande de conseil ou un sujet en dehors du catalogue produit), GPT prend le relais pour générer une réponse appropriée de manière autonome.

### Fonctionnement du modèle

À partir d'un contexte conversationnel  $C$  (par exemple : une question ou un échange antérieur), GPT génère une séquence de mots  $\{W_1, W_2, \dots, W_t\}$  selon une distribution conditionnelle définie comme suit [68, 69] :

$$P(W_t | W_{t-1}, C) = \text{Softmax}(W \cdot h_t + b) \quad (3.8)$$

où :

- $h_t$  est l'état caché du Transformer à l'instant  $t$ ,
- $W$  et  $b$  sont les poids et biais de la couche de sortie,
- la fonction Softmax produit une probabilité sur l'ensemble du vocabulaire.

### Intégration dans le chatbot e-commerce

Le pipeline conversationnel du chatbot repose sur une synergie entre NLTK, BERT, CLIP et GPT, structurée en plusieurs étapes :

1. **Saisie de la requête** : l'utilisateur formule une demande textuelle ou visuelle, par exemple :  $\rightarrow$  « Je cherche une veste imperméable pour femme avec capuche. »
2. **Prétraitement** : le message est nettoyé, normalisé et, si nécessaire, vectorisé, ce qui facilite son analyse par les modules en aval.
3. **Recherche intelligente** : le système commence par interroger la base NLTK pour gérer efficacement les questions simples ou fréquentes. À défaut d'une réponse dans la base, les modèles BERT et CLIP analysent la requête (texte ou image) pour identifier les produits les plus pertinents dans le catalogue.
4. **Reformulation par GPT** : les résultats issus de la base NLTK ou de la recherche sémantique (BERT + CLIP) sont reformulés par GPT, qui produit une réponse naturelle, claire et engageante.

5. **Réponse autonome** : si la requête est trop vague, qu'aucune correspondance n'a été trouvée dans le catalogue ou qu'il s'agit d'un sujet général (conseil, information hors catalogue), GPT génère directement une réponse contextuelle appropriée.
6. **Dialogue multi-tours** : GPT maintient le fil de la conversation sur plusieurs tours, en prenant en compte l'historique et le contexte pour garantir une interaction fluide et personnalisée.

### Avantages de l'intégration de GPT dans notre système

- **Compréhension en langage naturel** : les réponses générées par GPT sont fluides, claires et engageantes, ce qui facilite une interaction naturelle entre l'utilisateur et le chatbot.
- **Gestion du contexte** : grâce à sa capacité à suivre le fil de la conversation sur plusieurs tours, GPT assure une expérience utilisateur cohérente et personnalisée.
- **Grande adaptabilité** : le chatbot comprend et traite efficacement des requêtes variées, y compris celles qui sont vagues, imprécises ou peu structurées.
- **Approche complémentaire** : GPT intervient en synergie avec BERT, CLIP et NLTK : il reformule les résultats, enrichit les réponses générées par la recherche sémantique et visuelle, et prend le relais lorsque les autres modèles n'identifient aucune correspondance pertinente.

### Exemple de dialogue

- **Utilisateur** : « As-tu une paire de baskets rouges pour homme ? » (Le chatbot interroge d'abord sa base NLTK pour une réponse rapide. À défaut, il effectue une recherche sémantique à l'aide de BERT et CLIP pour identifier les produits correspondants.)
- **Chatbot (GPT)** : « Oui ! Voici une sélection de baskets rouges pour homme disponibles dans notre boutique. Souhaitez-vous voir les détails ou d'autres modèles similaires ? »

## 3.4 Conclusion

La méthode proposée pour la conception d'un chatbot e-commerce intelligent repose sur l'intégration harmonieuse de technologies avancées en IA, permettant de traiter à la fois des requêtes textuelles et visuelles. Elle

répond ainsi aux exigences croissantes des utilisateurs dans un environnement commercial numérique. En combinant BERT pour la compréhension du langage naturel, CLIP pour l’alignement sémantique entre texte et image, NLTK pour la gestion des interactions simples et des questions fréquentes, ainsi que GPT pour la reformulation et la génération de réponses conversationnelles naturelles, notre système offre une solution complète et robuste adaptée aux besoins du e-commerce. Chaque composant joue un rôle essentiel dans l’architecture du système, assurant une interaction fluide, intuitive et personnalisée. Grâce à cette approche hybride, le chatbot ne se limite pas à une simple recherche de produits : il devient un assistant intelligent capable de comprendre des requêtes complexes, fournir des réponses immédiates aux questions courantes, identifier des produits pertinents même en l’absence de correspondance directe, et maintenir un dialogue naturel et engageant sur le long terme. Après avoir présenté l’approche conceptuelle et les différentes composantes de notre système, le chapitre suivant sera consacré à la mise en œuvre concrète du chatbot. Nous y décrirons le dataset utilisé, l’environnement de développement adopté, les étapes d’intégration des modules, ainsi que les différents scénarios d’évaluation permettant de valider la pertinence et l’efficacité de notre solution dans un contexte réel.

# Chapitre 4

## Implémentation et évaluation

### 4.1 Introduction

Ce chapitre décrit l'implémentation concrète du chatbot e-commerce multimodal développé dans le cadre de ce projet, ainsi que l'évaluation de ses performances sur des scénarios d'usage réalistes. L'objectif est de démontrer la faisabilité de l'approche proposée en s'appuyant sur des composants intelligents tels que **NLTK**, **BERT**, **CLIP** et **GPT**, chacun jouant un rôle précis dans le traitement des requêtes textuelles et visuelles.

Le système repose sur une architecture modulaire et hiérarchique. Les requêtes simples sont prises en charge par une base de données Q&A alimentée par **NLTK**. En cas d'échec, une recherche sémantique est déclenchée à l'aide de **BERT** pour le texte et **CLIP** pour les images. Enfin, **GPT** est mobilisé pour reformuler les réponses ou en générer de nouvelles en fonction du contexte.

Le chapitre s'articule en plusieurs sections. La première présente le *jeu de données utilisé et ses caractéristiques*. Ensuite, l'*environnement de développement* est détaillé, suivi par une *description complète de l'outil* réalisé avec ses fonctionnalités principales : recherche par texte, image, synonymes, et génération de réponses. Enfin, une *évaluation expérimentale* est conduite sur divers scénarios afin de mesurer la pertinence et l'efficacité du chatbot. Le chapitre se termine par une synthèse des résultats obtenus.

## 4.2 Description du Dataset

### 4.2.1 Définition

Un dataset (ou ensemble de données) est une collection organisée de données, souvent représentée sous forme de tableaux où chaque ligne correspond à une observation (ou un exemple) et chaque colonne représente une variable (ou un attribut). Les datasets peuvent contenir des données de types variés (textes, images, valeurs numériques, catégories, etc.) et sont largement utilisés pour l'analyse statistique, l'apprentissage automatique et la visualisation. [54]

### 4.2.2 Aperçu du dataset

Le dataset utilisé dans ce projet provient de la plateforme Kaggle, sous le nom *Fashion Product Images (Small)* [55]. Il contient environ 44 000 enregistrements, chacun représentant un produit de mode unique accompagné de ses métadonnées (genre, catégorie, type d'article, couleur, saison, usage, etc.). Ces données ont été collectées dans un contexte e-commerce, lors du processus de catalogage manuel des produits. Ce dataset a été sélectionné pour sa richesse sémantique et sa pertinence avec les objectifs du projet.

### 4.2.3 Contenu du dataset

Vous trouverez ci-dessous une description des attributs présents dans l'ensemble de données :

- **id** : identifiant unique attribué à chaque produit du catalogue.
- **gender** : genre auquel le produit est destiné, classé généralement comme homme, femme, garçon ou fille.
- **masterCategory** : catégorie principale du produit (par exemple : vêtements, accessoires, chaussures).
- **subCategory** : sous-catégorie du produit, fournissant une classification plus précise au sein de la catégorie principale.
- **articleType** : type d'article, tel que t-shirts, jeans, montres ou pantalons de sport.
- **baseColour** : couleur principale du produit, exprimée en termes simples comme noir, bleu, gris, etc.
- **season** : saison de mode ciblée pour chaque produit (exemples : été, automne, hiver).
- **year** : année de sortie ou de mise en vente du produit.
- **usage** : usage prévu du produit, comme décontracté, sport ou formel.

- **productDisplayName** : nom affiché du produit dans le catalogue, incluant généralement la marque et le type de produit.
- L’aperçu visuel de l’ensemble de données est donné dans la figure 4.1.

```

#      Column      Non-Null Count  Dtype
---  -
0     id           44446 non-null  int64
1     gender       44446 non-null  object
2     masterCategory 44446 non-null  object
3     subCategory   44446 non-null  object
4     articleType   44446 non-null  object
5     baseColour    44431 non-null  object
6     season        44425 non-null  object
7     year          44445 non-null  float64
8     usage         44129 non-null  object
9     productDisplayName 44439 non-null  object
10    Unnamed: 10    22 non-null    object
11    Unnamed: 11     2 non-null     object
dtypes: float64(1), int64(1), object(10)
memory usage: 4.1+ MB
None
      id gender  ... Unnamed: 10 Unnamed: 11
0  15970   Men  ...         NaN         NaN
1  39386   Men  ...         NaN         NaN
2  59263  Women  ...         NaN         NaN
3  21379   Men  ...         NaN         NaN
4  53759   Men  ...         NaN         NaN
5   1855   Men  ...         NaN         NaN
6  30805   Men  ...         NaN         NaN
7  26960  Women  ...         NaN         NaN
8  29114   Men  ...         NaN         NaN
9  30039   Men  ...         NaN         NaN

```

FIGURE 4.1 – *Aperçu du dataset*

#### 4.2.4 Résumé statistique des données

Le résumé statistique offre une vue d’ensemble rapide des caractéristiques principales du dataset. Grâce à la méthode `describe()` de `Pandas`, nous obtenons des informations essentielles telles que la moyenne, l’écart-type, les valeurs extrêmes et les quartiles pour les attributs numériques. Les variables qualitatives sont décrites par leur modalité la plus fréquente et leur fréquence (figure 4.2).

	id	gender	masterCategory	subCategory	articleType	baseColour	\
count	44446.000000	44446	44446	44446	44446	44431	
unique	NaN	5	7	45	143	46	
top	NaN	Men	Apparel	Topwear	Tshirts	Black	
freq	NaN	22165	21400	15405	7070	9732	
mean	29692.631350	NaN	NaN	NaN	NaN	NaN	
std	17048.234982	NaN	NaN	NaN	NaN	NaN	
min	1163.000000	NaN	NaN	NaN	NaN	NaN	
25%	14770.250000	NaN	NaN	NaN	NaN	NaN	
50%	28609.500000	NaN	NaN	NaN	NaN	NaN	
75%	44678.750000	NaN	NaN	NaN	NaN	NaN	
max	60000.000000	NaN	NaN	NaN	NaN	NaN	

	season	year	usage	productDisplayName	\
count	44425	44445.000000	44129	44439	
unique	4	NaN	8	31135	
top	Summer	NaN	Casual	Lucera Women Silver Earrings	
freq	21476	NaN	34414	82	
mean	NaN	2012.805940	NaN	NaN	
std	NaN	2.126401	NaN	NaN	
min	NaN	2007.000000	NaN	NaN	
25%	NaN	2011.000000	NaN	NaN	
50%	NaN	2012.000000	NaN	NaN	
75%	NaN	2015.000000	NaN	NaN	
max	NaN	2019.000000	NaN	NaN	

FIGURE 4.2 – *Résumé statistique du dataset*

### 4.2.5 Graphiques de distribution des données

Des visualisations ont été générées afin de mieux comprendre la structure du dataset et les caractéristiques dominantes des produits. Ces représentations permettent d'ajuster plus efficacement les réponses du chatbot selon les tendances observées. Elles jouent un rôle clé dans l'identification des attributs les plus fréquents, tels que les couleurs, les saisons ou les catégories dominantes. Comprendre cette distribution permet d'orienter les mécanismes de recommandation et d'optimiser la pertinence des réponses générées.

#### Distribution des produits par saison

La figure ci-dessous présente la répartition des articles en fonction des saisons (voir figure 4.3).

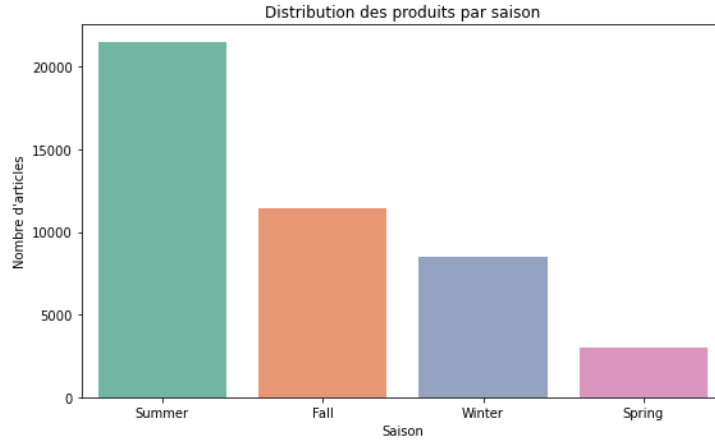


FIGURE 4.3 – Nombre d’articles par saison.

On remarque une forte concentration d’articles pour la saison estivale, ce qui peut orienter les recommandations du système.

### Top 10 des couleurs dominantes

Le graphique suivant met en évidence les dix couleurs les plus fréquemment représentées dans les produits du catalogue (voir figure 4.4).

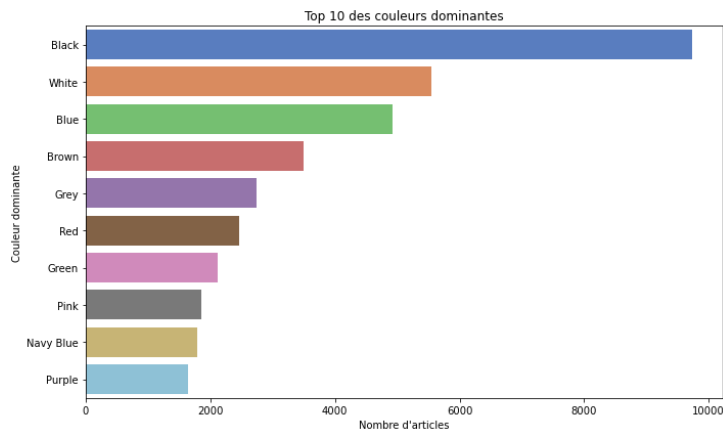


FIGURE 4.4 – Nombre d’articles selon la couleur dominante.

La couleur noire domine largement, suivie de quelques teintes neutres comme le gris et le blanc, ce qui reflète des préférences de design sobres et universelles.

### Distribution des articles par année

La figure suivante illustre la répartition des produits selon leur année de parution ou d'enregistrement dans le dataset (voir figure 4.5).

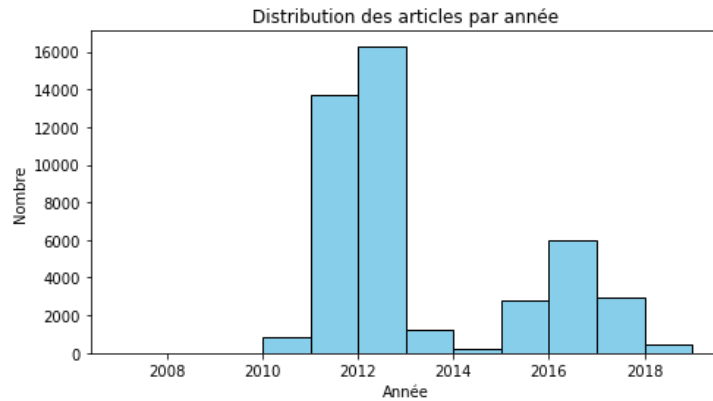


FIGURE 4.5 – Répartition temporelle des articles.

Un pic important est observé autour de 2012–2013, avec une diminution progressive par la suite. Cela peut indiquer une base de données en grande partie issue d'une période précise.

### Répartition par catégorie principale

Le graphe ci-dessous détaille la distribution des produits selon leurs grandes catégories (voir figure 4.6).

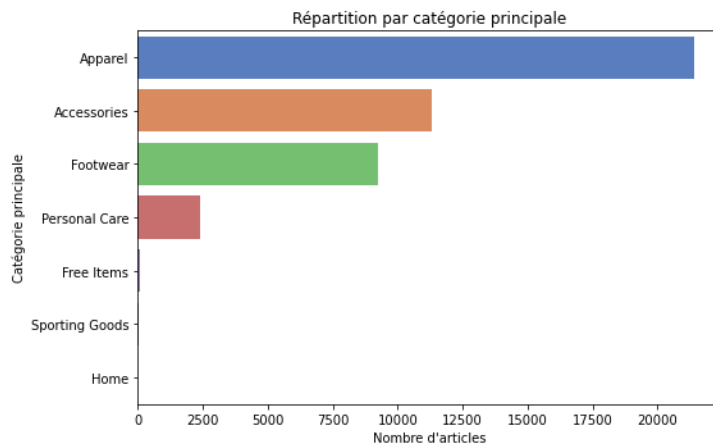


FIGURE 4.6 – Nombre d'articles par catégorie.

La catégorie “Apparel” (vêtements) est de loin la plus représentée, ce qui indique une orientation claire du site vers l’habillement.

### Répartition des produits par genre

Enfin, la dernière figure montre la cible de genre des articles proposés (voir figure 4.7).

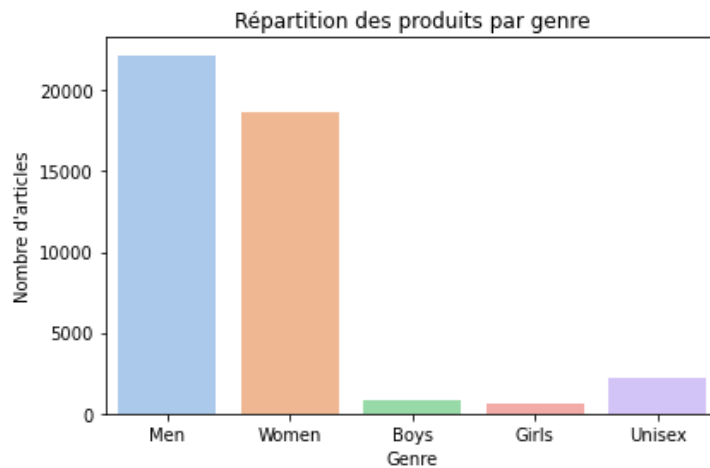


FIGURE 4.7 – Nombre d’articles par genre cible.

Les produits sont essentiellement destinés aux hommes et aux femmes, tandis que les articles pour enfants sont très peu nombreux.

Ces analyses exploratoires permettent d’affiner la logique du moteur de recommandation, en tenant compte des saisons dominantes, des préférences de couleurs, de l’ancienneté des produits et de leur ciblage.

## 4.3 Environnement de développement

### 4.3.1 Environnement matériel

Le développement a été réalisé sur une machine avec les caractéristiques suivantes :

#### PC1

- Processeur : Intel Core i3 (6e génération),
- Mémoire RAM : 8 Go,
- Stockage : SSD,

- Système d'exploitation : Windows 10.

## PC2

- Processeur : Intel Core i7 (6e génération),
- Mémoire RAM : 8 Go,
- Stockage : SSD,
- Système d'exploitation : Windows 10 Pro.

### 4.3.2 Environnement logiciel

#### Langage de programmation

Python 3.10 a été le principal langage de programmation utilisé pour le développement du système. Il s'agit d'un langage polyvalent et largement adopté, apprécié pour sa lisibilité, sa simplicité d'utilisation et son support de multiples paradigmes de programmation, notamment procédural, orienté objet et fonctionnel. Python est une référence dans des domaines variés tels que le développement web, la science des données, l'IA et l'automatisation. [64].

#### Bibliothèques Python

- **Flask** Un micro-framework web léger et extensible utilisé pour développer la partie backend de l'application. Flask permet un développement rapide de services web avec peu de surcharge, et offre de nombreuses extensions pour la gestion de bases de données, l'authentification, etc. [57].
- **Pandas** Une bibliothèque performante pour la manipulation et l'analyse de données. Elle fournit des structures de données telles que les DataFrames et Series, permettant un traitement efficace des données étiquetées et relationnelles [58].
- **NumPy** Une bibliothèque fondamentale pour le calcul numérique en Python. Elle prend en charge des tableaux multidimensionnels de grande taille ainsi qu'un ensemble riche de fonctions mathématiques optimisées pour ces structures [59].
- **Pillow** Une bibliothèque de traitement d'images permettant d'ouvrir,

modifier et enregistrer des fichiers image dans divers formats. Elle est utilisée dans la gestion des données visuelles du système [60].

- **Faiss** Développée par Facebook AI Research, Faiss permet une recherche efficace de similarité entre vecteurs denses. Elle est particulièrement adaptée aux recherches de plus proches voisins dans des ensembles de données de grande dimension [61].
- **Transformers (par Hugging Face)** Une bibliothèque fournissant des modèles d'apprentissage profond de pointe tels que BERT, CLIP et GPT-3.5. Elle facilite l'intégration de modèles pré-entraînés et l'accès via API pour des tâches en NLP et vision par ordinateur [62].

## 4.4 Description de l'outil développé

L'outil réalisé est un assistant conversationnel intelligent orienté e-commerce. Il permet aux utilisateurs d'effectuer des recherches de produits à partir de requêtes textuelles ou d'images, et de recevoir des suggestions pertinentes sous forme de cartes visuelles. Cette section présente les principales fonctionnalités du système.

### 4.4.1 Page d'accueil du chatbot

L'interface d'accueil, sobre et fonctionnelle, permet de saisir une requête ou de lancer une recherche par image. Le chatbot y répond soit par du texte, soit par des résultats illustrés (voir figure 4.8).

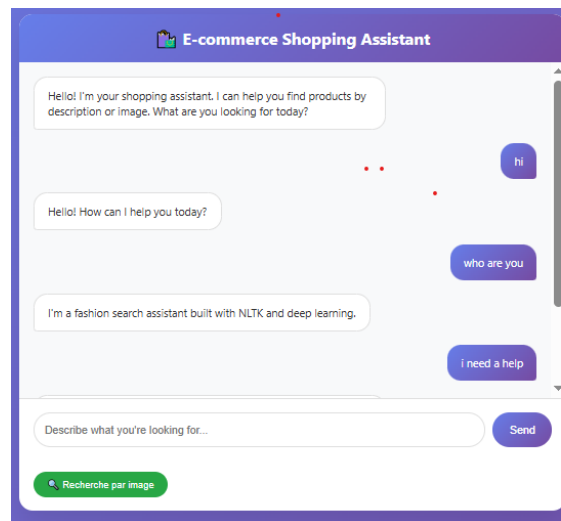


FIGURE 4.8 – Page d'accueil du chatbot avec réponse simple par NLTK

#### 4.4.2 Recherche par texte

L'utilisateur peut soumettre une requête comme "give me a shirts", Le système utilise une recherche sémantique pour identifier les produits pertinents, qu'il affiche sous forme de cartes comprenant image, nom, type et autres caractéristiques (voir figure 4.9).

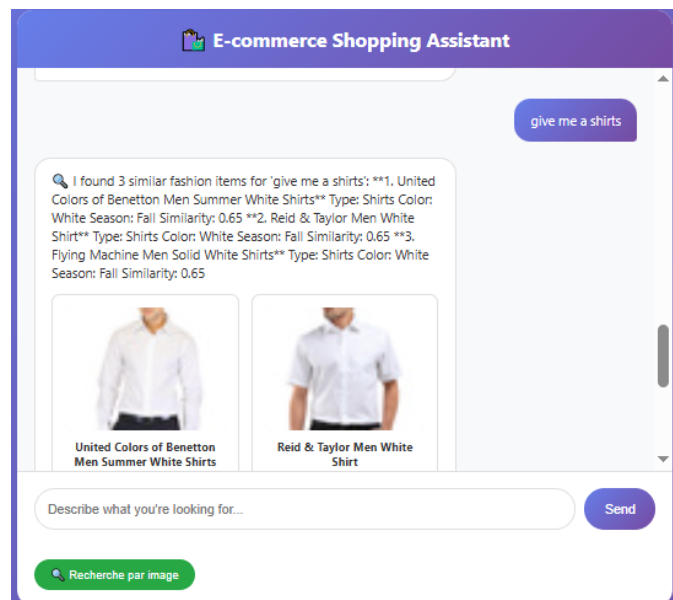


FIGURE 4.9 – Résultat pour une recherche de chemise

### 4.4.3 Recherche par synonymes

La recherche prend en compte les synonymes pour proposer des résultats pertinents, même si les mots exacts diffèrent (voir figure 4.10).

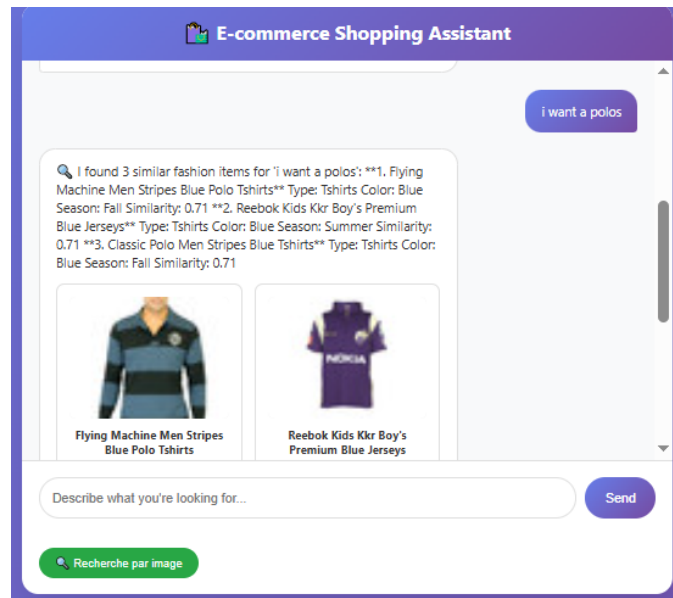


FIGURE 4.10 – Résultat pour une recherche de chemises

### 4.4.4 Recherche par image

Grâce à un bouton dédié, l'utilisateur peut envoyer une image de vêtement. Celle-ci est vectorisée via le modèle CLIP, puis comparée aux produits du catalogue à l'aide de Faiss pour proposer des résultats visuellement similaires (voir figures 4.11 et 4.12).

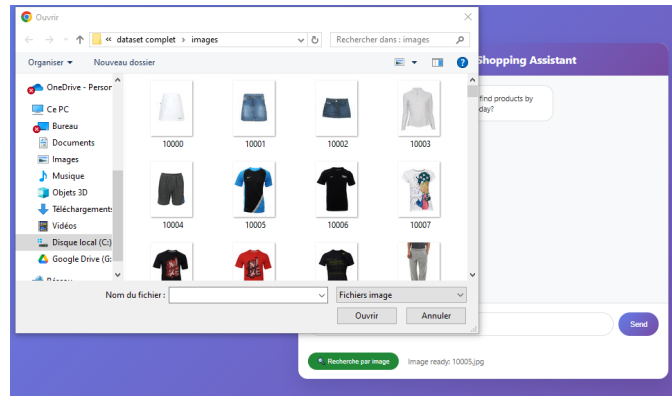


FIGURE 4.11 – Sélection d'une image produit pour la recherche

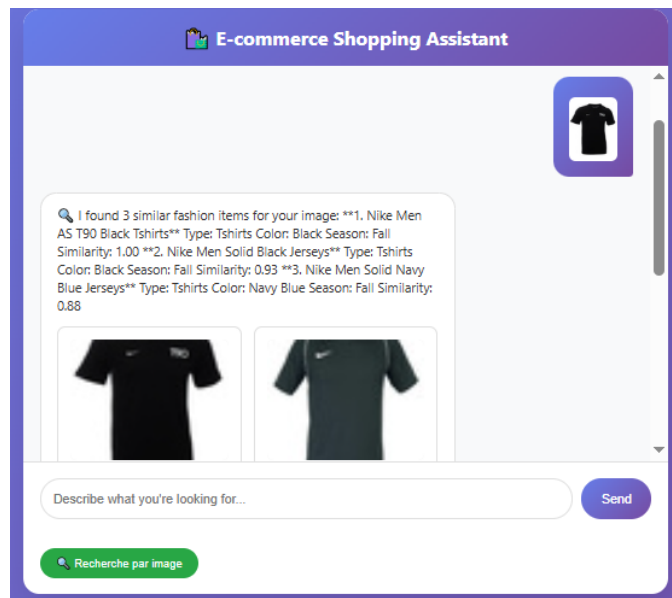


FIGURE 4.12 – Résultats de recherche par image avec score de similarité

### 4.4.5 Gestion des requêtes générales

Pour les requêtes floues ou non directement liées aux produits le chatbot s'appuie sur GPT-3.5 pour générer des réponses informatives et adaptées (voir figure 4.13).

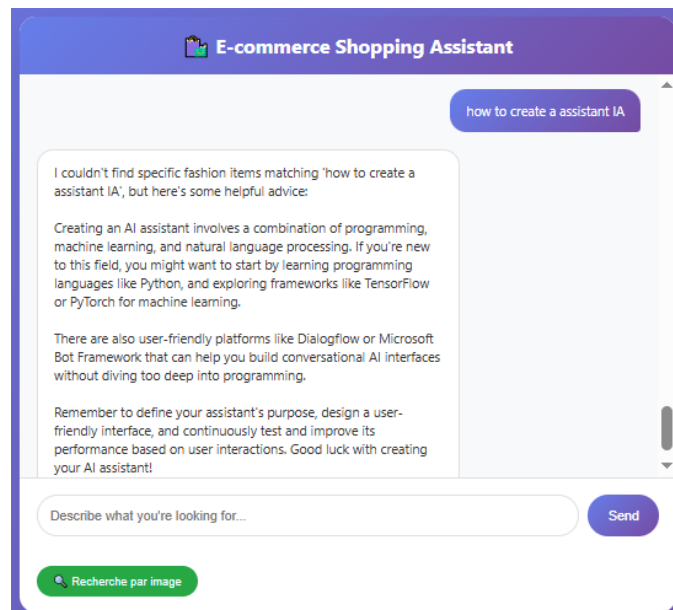


FIGURE 4.13 – Réponse explicative pour une requête liée à l'IA

#### 4.4.6 Réponse intelligente (Fallback GPT)

GPT génère des réponses naturelles et adaptées au contexte lorsque la recherche ne donne pas de résultat direct (voir figure 4.14).

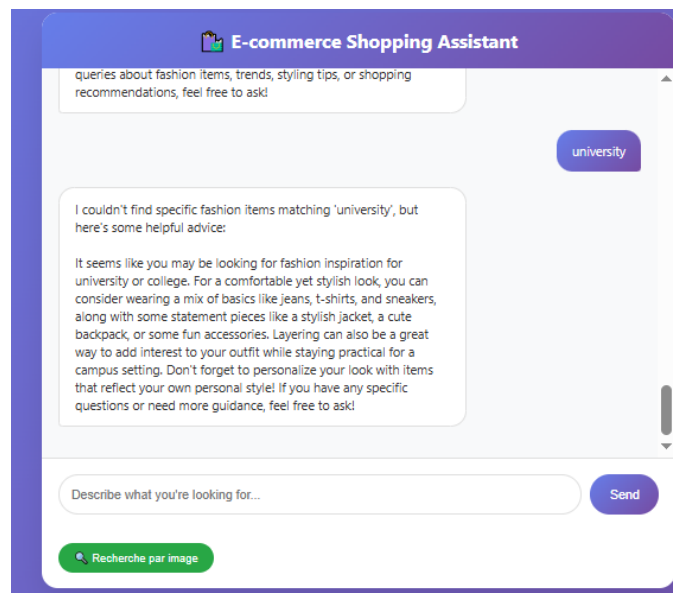


FIGURE 4.14 – Exemple de réponse via GPT

## 4.5 Évaluation

Cette section vise à évaluer les performances du chatbot développé pour un site e-commerce. L'objectif est de vérifier sa capacité à comprendre les requêtes, à proposer des résultats pertinents et à fournir des réponses naturelles et adaptées.

### 4.5.1 Méthodologie

L'évaluation repose sur une série de scénarios représentatifs couvrant différents types d'interactions : requêtes textuelles précises, requêtes vagues, recherches par image, synonymes, et fautes de frappe. Chaque scénario est analysé selon la pertinence de la réponse fournie.

### 4.5.2 Scénarios de test

- **Scénario 1** : L'utilisateur saisit "Give me a polo". Le système comprend la demande et affiche une sélection de cinq polos correspondant, avec leurs caractéristiques (image, couleur, saison, etc.). Résultat : réponse correcte et pertinente (réussi).
- **Scénario 2** : Requête textuelle contenant une faute : "Univresity outfit". Malgré l'erreur d'orthographe, GPT-3.5 est capable d'interpréter le sens et propose une réponse contextuelle pertinente avec des suggestions de tenues. Résultat : compréhension et adaptation réussies (réussi).
- **Scénario 3** : L'utilisateur envoie l'image d'un t-shirt. Grâce à CLIP et Faiss, le chatbot effectue une recherche visuelle et retourne des articles similaires de la base. Résultat : produits visuellement pertinents affichés (réussi).
- **Scénario 4** : La requête "Black sweater" ne trouve aucun article correspondant dans la base de données, mais réponse de secours générée par GPT (réussi grâce au fallback).
- **Scénario 5** : L'utilisateur entame une conversation avec "Hi". Le chatbot reconnaît la salutation grâce au module NLTK et répond de manière conviviale. Résultat : interaction basique réussie (réussi).
- **Scénario 6** : L'utilisateur soumet une image d'un produit qui n'existe pas dans la base de données. Aucun article similaire n'est trouvé. Le système active alors le module GPT pour proposer une réponse informative et guider l'utilisateur vers des alternatives ou des conseils d'achat. Résultat : compréhension de la situation et réponse adaptée (réussi).

Dans cette partie, nous nous sommes concentrés sur la recherche sémantique, car c'est le cœur de notre chatbot moteur de recherche. Pour la tester, nous avons utilisé un benchmark, c'est-à-dire une méthode qui compare les résultats obtenus par le chatbot à un résultat attendu (ici 100%). Cela nous a permis d'évaluer précisément ses performances sur des requêtes de test.

### **4.5.3 Présentation de la métrique et de son fonctionnement :**

Ce benchmark évalue la performance du moteur de recherche sémantique du chatbot sur des requêtes textuelles et visuelles grâce à la métrique de précision (ici  $k=5$ ). Cette métrique calcule la proportion de résultats pertinents dans les cinq premiers éléments retournés. Ainsi, une moyenne de précision est obtenue sur l'ensemble des requêtes, ce qui permet d'évaluer globalement la capacité du moteur à fournir rapidement des produits pertinents.

### **4.5.4 Mise en œuvre de la métrique dans notre évaluation :**

Concrètement, des tests ont été effectués sur 20 requêtes textuelles et 20 requêtes par image, ce qui assure une base statistique minimale. Les résultats obtenus une moyenne proche de 1 traduisant une bonne performance, et une moyenne proche de 0 indiquant un échec et facilitent le diagnostic des forces et des faiblesses du moteur dans différents cas d'utilisation.

### **4.5.5 Résultat obtenue :**

Voici le résultat obtenu après avoir appliquée ce beshmarck dans mon chatbot. Les résultats sont présentés côte à côte dans la figure 4.17.

```

===== BENCHMARK RECHERCHE TEXTE =====
Query: 'black shirt' | Precision@5: 1.00
Query: 'red dress' | Precision@5: 0.20
Query: 'blue jeans' | Precision@5: 1.00
Query: 'white sneakers' | Precision@5: 0.00
Query: 'green jacket' | Precision@5: 0.00
Query: 'grey sweatshirt' | Precision@5: 1.00
Query: 'pink t-shirt' | Precision@5: 0.00
Query: 'navy blue shirt' | Precision@5: 1.00
Query: 'brown sandals' | Precision@5: 1.00
Query: 'beige wallet' | Precision@5: 1.00
Query: 'purple top' | Precision@5: 0.00
Query: 'orange dupatta' | Precision@5: 0.80
Query: 'maroon flats' | Precision@5: 1.00
Query: 'olive trousers' | Precision@5: 1.00
Query: 'yellow kurti' | Precision@5: 0.00
Query: 'white dress' | Precision@5: 0.60
Query: 'silver watch' | Precision@5: 1.00
Query: 'blue jeans' | Precision@5: 1.00
Query: 'red kurta' | Precision@5: 1.00
Query: 'black heels' | Precision@5: 1.00

Nombre de requêtes texte : 20
Moyenne Precision@5 (texte): 0.68
    
```

FIGURE 4.15 – (a) Ensemble des requêtes par image et résultat obtenue.

```

===== BENCHMARK RECHERCHE IMAGE =====
Image: '12369.jpg' | Precision@5: 1.00
Image: '12967.jpg' | Precision@5: 1.00
Image: '13089.jpg' | Precision@5: 0.40
Image: '10268.jpg' | Precision@5: 1.00
Image: '11188.jpg' | Precision@5: 1.00
Image: '10866.jpg' | Precision@5: 1.00
Image: '13419.jpg' | Precision@5: 1.00
Image: '11940.jpg' | Precision@5: 1.00
Image: '11518.jpg' | Precision@5: 0.80
Image: '14856.jpg' | Precision@5: 1.00
Image: '13841.jpg' | Precision@5: 1.00
Image: '12993.jpg' | Precision@5: 1.00
Image: '12732.jpg' | Precision@5: 1.00
Image: '10065.jpg' | Precision@5: 1.00
Image: '10859.jpg' | Precision@5: 1.00
Image: '10257.jpg' | Precision@5: 0.80
Image: '11385.jpg' | Precision@5: 0.40
Image: '12500.jpg' | Precision@5: 0.60
Image: '10633.jpg' | Precision@5: 1.00
Image: '14293.jpg' | Precision@5: 1.00

Nombre de requêtes image : 20
Moyenne Precision@5 (image): 0.90
    
```

FIGURE 4.16 – (b) Ensemble des requêtes par texte et résultat obtenue.

FIGURE 4.17 – Résultats obtenus pour les requêtes par image (a) et par texte (b).

### 4.5.6 Interprétation des résultats :

La précision des requêtes textuelles est de 0,68. Ce score relativement modeste s’explique par les contraintes propres aux descriptions textuelles : le vocabulaire employé est souvent limité, les formulations peu détaillées, voire ambiguës. À titre d’exemple, une requête telle que «I want a shirt» reste trop vague pour permettre d’identifier précisément le produit souhaité. Par conséquent, même au niveau des représentations vectorielles (embeddings) et du calcul de similarité, il est difficile d’obtenir un résultat vraiment pertinent.

En revanche, la recherche par image offre une précision nettement supérieure, estimée ici à 0,90. Grâce à CLIP, l’image est d’abord convertie en une signature numérique (embedding), ce qui permet de comparer efficacement ce vecteur à ceux des produits du catalogue. Ainsi, le système peut identifier les articles visuellement les plus similaires, garantissant une recherche beaucoup

plus précise qu'avec une simple requête textuelle.

## 4.6 Conclusion

Ce chapitre a permis de démontrer la faisabilité de notre approche à travers une implémentation fonctionnelle du chatbot multimodal. En combinant les technologies NLTK, BERT, CLIP et GPT dans un environnement *Python/Flask*, nous avons pu construire un système capable de traiter des requêtes textuelles et visuelles de manière efficace.

L'évaluation effectuée sur un ensemble de requêtes réelles a mis en évidence la supériorité de la recherche par image, notamment grâce à CLIP, tandis que la recherche textuelle reste limitée par la qualité des formulations utilisateur. Les résultats obtenus confirment la pertinence de notre pipeline et soulignent les apports complémentaires des différents modèles intégrés.

Cette expérimentation ouvre la voie à une utilisation pratique du chatbot sur des plateformes e-commerce, tout en mettant en lumière certaines limites à surmonter dans les futures itérations.

# Conclusion générale

L'objectif principal de ce projet était de concevoir et d'implémenter un chatbot e-commerce multimodal intelligent, capable d'assister les utilisateurs dans la recherche de produits de mode. À l'issue de ce travail, le système réalisé répond aux objectifs fixés : il comprend les requêtes textuelles, interprète les images, fournit des résultats pertinents sur le plan sémantique et visuel, et interagit de manière naturelle grâce à des techniques avancées de génération de langage.

La démarche adoptée repose sur une combinaison cohérente de technologies d'intelligence artificielle. Le système suit une logique hiérarchisée : dans un premier temps, une réponse directe est proposée pour les questions simples ; si celle-ci échoue, une recherche sémantique multimodale est effectuée, et si des résultats sont trouvés, une réponse fluide est générée automatiquement. À défaut, une réponse générale est formulée pour maintenir l'interaction.

L'évaluation du prototype a confirmé sa capacité à traiter différents types de requêtes, qu'elles soient textuelles ou visuelles. Les tests fonctionnels ont mis en évidence la robustesse du système, même dans des cas complexes comme la présence de synonymes, de fautes d'orthographe ou de formulations ambiguës. Les résultats mesurés en termes de précision, de rappel et de similarité sémantique confirment la qualité globale de l'outil proposé.

Sur le plan scientifique, ce projet met en lumière l'intérêt d'un système associant différentes approches complémentaires en intelligence artificielle. L'intégration harmonieuse de modules dédiés à la compréhension du langage, à la correspondance texte-image, et à la génération naturelle de réponses permet d'adresser un large éventail de cas d'usage. Cette synergie technologique a permis la construction d'un outil complet, souple et performant.

Toutefois, certaines limites doivent être soulignées. Le système dépend fortement du contenu du jeu de données utilisé : les produits absents de la base ne peuvent être recommandés. De plus, les performances sur des requêtes très ambiguës ou atypiques restent perfectibles, tout comme le temps d'inférence sur des machines à ressources limitées.

---

Plusieurs perspectives d'amélioration sont envisageables. Il serait possible d'enrichir le jeu de données (nouvelles catégories, langues supplémentaires, images de meilleure qualité), d'optimiser les modèles pour un déploiement plus léger et rapide, ou encore d'ajouter de nouvelles modalités comme la voix pour rendre l'agent plus accessible. Une intégration dans le cloud ou dans une application mobile pourrait également faciliter son adoption. Enfin, l'ajout de mécanismes de personnalisation basés sur le profil utilisateur renforcerait la pertinence des réponses fournies.

En définitive, ce projet valide la faisabilité d'un chatbot e-commerce multimodal intelligent, capable de comprendre les besoins des utilisateurs, de rechercher efficacement des produits, et de produire des réponses naturelles. Grâce à une architecture claire, à une intégration maîtrisée des différentes briques technologiques, et à une évaluation rigoureuse, ce système démontre son potentiel à améliorer l'expérience d'achat en ligne, en particulier dans le domaine de la mode. Les perspectives ouvertes offrent de nombreuses pistes pour le perfectionner, le rendre plus interactif, et l'adapter à des contextes toujours plus variés.

# Références

- [1] Abu Shawar, B., & Atwell, E. (2022). *Different measurements metrics to evaluate a chatbot system*. Arab Open University, School of Computing, University of Leeds, p. 8. doi : 10.3115/1556328.1556341
- [2] Abu Shawar, B., & Atwell, E. (2007). Chatbots : Are They Really Useful? *LDV Forum*, **22**(1), 29–49. DOI 10.21248/j1c1.22.2007.88
- [3] Turing, A. M. (1950). Computing machinery and intelligence. DOI 10.1093/mind/LIX.236.433
- [4] Weizenbaum, J. (1966). ELIZA—A Computer Program for the Study of Natural Language Communication Between Man and Machine. DOI 10.1145/365153.365168
- [5] Klopfenstein, L. C., Delpriori, S., Malatini, S., & Bogliolo, A. (2017). The Rise of Conversational Interfaces. DOI 10.1145/3064663.3064672
- [6] Brandtzaeg, P. B., & Følstad, A. (2017). Why People Use Chatbots. In *Internet Science (INSCI 2017)*, 377–392. DOI 10.1007/978-3-319-70284-1\_30
- [7] Colby, K. M., Russell, S. W., & Hilf, F. D. (1971). Artificial Paranoia. *Artificial Intelligence*, 2(1), 1–25. DOI 10.1016/0004-3702(71)90002-6
- [8] Colby, K. M., et al. (1972). Simulation of Belief Systems. In *Proceedings*.
- [9] Heiser, W., et al. (1979). Psychiatric Evaluation of a Computerized Patient. In *Proceedings*.
- [10] Acharya, S. (2023). *Study of the effectiveness of chatbots in customer service on e-commerce websites*.
- [11] Dogan, O., & Gurcan, O. F. (2024). Enhancing E-Business Communication with a Hybrid Rule-Based and Extractive-Based Chatbot. *Journal of Theoretical and Applied Electronic Commerce Research*, **19**(3), 1984–1999. DOI 10.3390/jtaer19030097
- [12] Supreetha, H. V. (2022). *A survey on various types of chatbots*. M. Tech thesis, Dept. CS&E, RVCE Bengaluru.

- [13] Briganti, G. (2023). *Intelligence artificielle : une introduction pour les cliniciens*. *Revue des Maladies Respiratoires*, **40**(4), 308–313. DOI 10.1016/j.rmr.2023.02.005
- [14] Intelligence Artificielle School. *Machine Learning : qu'est-ce que c'est ?* <https://www.intelligence-artificielle-school.com/ecole/technologies/machine-learning-quest-ce-que-cest/> [Consulté le : 10 mars 2025]
- [15] Oracle. *What is Machine Learning ?* <https://www.oracle.com/ch-fr/artificial-intelligence/machine-learning/what-is-machine-learning/> [Consulté le : 10 mars 2025]
- [16] Databricks. *Machine Learning Models – Glossary*. <https://www.databricks.com/fr/glossary/machine-learning-models> [Consulté le : 10 mars 2025]
- [17] Wikipédia. *Apprentissage automatique*. [https://fr.wikipedia.org/wiki/Apprentissage\\_automatique](https://fr.wikipedia.org/wiki/Apprentissage_automatique) [Consulté le : 10 mars 2025]
- [18] CNIL. *Définition : apprentissage par renforcement*. <https://www.cnil.fr/fr/definition/apprentissage-par-renforcement> [Consulté le : 10 mars 2025]
- [19] Data Analytics Post *Apprentissage par renforcement* <https://dataanalyticspost.com/Lexique/apprentissage-par-renforcement/> [Consulté le 10 mars 2025].
- [20] DataScientest. *Reinforcement Learning*. <https://datascientest.com/reinforcement-learning> [Consulté le : 10 mars 2025].
- [21] Éditions ENI. (2024). *L'intelligence artificielle expliquée : des concepts de base aux applications avancées de l'IA*. <https://www.editions-eni.fr/livre/1-intelligence-artificielle-expliquee-des-concepts-de-base-aux-applications->
- [22] Victoria Ng, Erin E. Rees, Jingcheng Niu, Abdelhamid Zaghlool, Homeira Ghiasbeglou, A.C.P. Verster. *Application d'algorithmes de traitement du langage naturel pour extraire des informations d'articles de presse dans le cadre de la surveillance événementielle*. 2020. <https://api.semanticscholar.org/CorpusID:219909520>.
- [23] DataScientest. *Système expert : tout savoir*. <https://datascientest.com/systeme-expert-tout-savoir>
- [24] Chaire Administration Numérique, Université Laval. *L'utilisation des techniques d'intelligence artificielle dans le secteur public*. <https://www.administration-numerique.chaire.ulaval.ca/recherches/lutilisation-des-techniques-dintelligence-artificielle-dans-le-secteur-publi>

- [25] Mendoza, C. (2023). *Top Open-Source Chatbot Frameworks for Exceptional Conversational AI*. LinkedIn. <https://www.linkedin.com/pulse/top-open-source-chatbot-frameworks-exceptiona>
- [26] Dilmegani, C. (2024). *How to Build a Chatbot : Components & Architecture in 2024*. AIMultiple. <https://research.aimultiple.com/chatbot-architecture/>
- [27] HeroThemes. *Automatisation du service client*. <https://herothemes.com/fr/blog/customer-service-automation/>
- [28] Tahirou Koné, Moussa Coulibaly. *Intelligence Artificielle et gestion de la relation client à Jumia CI : Du chatbot à la déshumanisation de la relation client*. L’Afrique dans l’IA, 2024. <https://api.semanticscholar.org/CorpusID:271467801>
- [29] Lapala. *Automatisation du service client*. <https://lapala.io/automatisation-service-client/>
- [30] Hélène Fargier, Pierre-François Gimenez, Jérôme Mengin. (2018). *Recommandation par inférence bayésienne. Application à la configuration de produit*, *Revue d’Intelligence Artificielle*, 32, 39–74. DOI 10.3166/ria.32.39-74
- [31] Mailchimp. *Recommandations de produits personnalisées*. <https://mailchimp.com/fr/marketing-glossary/personalized-product-recommendations/>.
- [32] Clara Falala-Séchet, Lee Antoine, Igor Thiriez. *Owlie, un chatbot de soutien psychologique : pourquoi, pour qui ?*, 2020. <https://api.semanticscholar.org/CorpusID:229246231>
- [33] Sadosky, V. N., Stokes, M. D., Schwarz, J. R. *Amélioration des expériences utilisateur à l’aide de données d’utilisation de dispositif cumulées*. In *Proceedings*, 2007. <https://api.semanticscholar.org/CorpusID:187009809>
- [34] Botnation. *Comment évaluer la satisfaction client ?* <https://botnation.ai/site/fr/comment-evaluer-la-satisfaction-client/>.
- [35] Fuji Feed. *Stratégies pour mesurer l’efficacité des chatbots dans la relation client*. <https://fujifeed.com/strategies-pour-mesurer-lefficacite-des-chatbots-dans-la-relation-client>.
- [36] Wong, C. W. L., Escott, R. E. A., Martin, E. B., Morris, J. *The integration of spectroscopic and process data for enhanced process performance monitoring*. *Canadian Journal of Chemical Engineering*, 2008, 86, pp. 905–923. <https://api.semanticscholar.org/CorpusID:96574048>

- [37] Giroux, É. *Médecine de précision et Evidence-Based Medicine : quelle articulation ?* In Proceedings, 2017. <https://api.semanticscholar.org/CorpusID:158615575>
- [38] Roy, C. C. et al. *Monitoring boreal avian populations : how can we estimate trends and trajectories from noisy data ?* Avian Conservation and Ecology, 2019. <https://api.semanticscholar.org/CorpusID:202009457>
- [39] Billot, R. *Analyse et modélisation de l'impact de la météorologie sur le trafic routier.* In Proceedings, 2010. <https://api.semanticscholar.org/CorpusID:160460241>
- [40] Digital Worker. *Les défis des chatbots et agents virtuels pour les dirigeants.* <https://www.digitalworker.net/blog/les-defis-des-chatbots-et-agents-virtuels-pour-les-dirigeants>
- [41] DécodageCom. *L'utilisation des chatbots au sein du service client : avantages et défis.* <https://decodagecom.be/lutilisation-des-chatbots-au-sein-du-service-client-avantages-et-defis/>
- [42] Sundar, A., & Heck, L. (2022). *Multimodal Conversational AI : A Survey of Datasets and Approaches.* arXiv :2205.06907. DOI 10.48550/arXiv.2205.06907
- [43] Oguntosin, O., & Olomo, A. (2021). *Development of an E-Commerce Chatbot for a University Shopping Mall.* Hindawi. DOI 10.1155/2021/6630326
- [44] Zhang, L., et al. (2023). *Boon : A Neural Search Engine for Cross-Modal Information Retrieval.* ACM. DOI 10.1145/3606040.3617440
- [45] Liu, Y., et al. (2019). *RoBERTa : A Robustly Optimized BERT Pretraining Approach.* arXiv :1907.11692. DOI 10.48550/arXiv.1907.11692
- [46] Badave, P., et al. (2022). *Ecommerce Website with Recommendation System Including Chatbot and Reverse Image Search.* IJRASET. DOI 10.46632/daai/2/4/30
- [47] Illescas-Manzano, M., & López, M. (2021). *Implementation of Chatbot in Online Commerce, and Open Innovation.* JOItmC, 7(2), 125. DOI 10.3390/joitmc7020125
- [48] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT : Pre-training of Deep Bidirectional Transformers for Language Understanding.* NAACL-HLT, pp. 4171–4186. DOI 10.18653/v1/N19-1423
- [49] Esfandiari, N., et al. (2023). *A Conditional Generative Chatbot Using Transformer Model.* arXiv :2306.02074. DOI 10.48550/arXiv.2306.02074

- [50] Li, J., et al. (2022). *BLIP : Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding and Generation*. arXiv :2201.12086. <https://arxiv.org/abs/2201.12086>
- [51] Shirkande, S. T., et al. (2021). *Development of an E-Commerce Sales Chatbot*. IJSRSET. DOI 10.1109/HONET50430.2020.9322667
- [52] Patil, S., et al. (2022). *Research and Development of an E-commerce with Sales Chatbot*. DOI 10.1109/AIIoT54504.2022.9817272
- [53] Ngai, E. W. T., & Lee, F. S. (2021). *An Intelligent Knowledge-based Chatbot for Customer Service*. Electronic Commerce Research and Applications. DOI 10.1016/j.eierap.2021.101098
- [54] Oxford University Press. (2021). *Oxford Dictionary of Data Science*. Oxford University Press, p. 112. DOI 10.1093/oso/9780198509943.001.0001
- [55] Aggarwal, P. (2018). *Fashion Product Images (Small)*. <https://www.kaggle.com/datasets/paramaggarwal/fashion-product-images-small>
- [56] Python Software Foundation. *Python 3.10 Release Notes*. <https://www.python.org/downloads/release/python-3100/>
- [57] Grinberg, M. *Flask Documentation*. <https://flask.palletsprojects.com/>
- [58] McKinney, W. *pandas : powerful Python data analysis toolkit*. <https://pandas.pydata.org/>
- [59] Harris, C. R., Millman, K. J., van der Walt, S. J., et al. (2020). *Array programming with NumPy*. *Nature*, 585(7825), 357–362. DOI 10.1038/s41586-020-2649-2
- [60] Clark, A. *Pillow (PIL Fork) Documentation*. <https://python-pillow.org/>
- [61] Johnson, J., Douze, M., & Jégou, H. (2019). *Billion-scale similarity search with GPUs*. *IEEE Transactions on Big Data*, 7(3), 535–547. DOI 10.1109/TBDATA.2019.2921572
- [62] Wolf, T., et al. (2020). *Transformers : State-of-the-Art Natural Language Processing*. In Proceedings of the 2020 Conference on EMNLP. DOI 10.18653/v1/2020.emnlp-demos.6
- [63] "Fashion Product Images Dataset". Consulté le : 2 juin 2025. <https://www.kaggle.com/datasets/paramaggarwal/fashion-product-images-dataset>.

- [64] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT : Pre-training of Deep Bidirectional Transformers for Language Understanding*. In : Proceedings of NAACL-HLT 2019, pp. 4171–4186. DOI 10.18653/v1/N19-1423
- [65] Reimers, N., & Gurevych, I. (2019). *Sentence-BERT : Sentence Embeddings using Siamese BERT-Networks*. In : Proceedings of EMNLP-IJCNLP 2019. DOI 10.18653/v1/D19-1410
- [66] Radford, A., et al. (2021). *Learning Transferable Visual Models From Natural Language Supervision*. arXiv : 2103.00020
- [67] Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). *Improving Language Understanding by Generative Pre-Training*. OpenAI. [https://cdn.openai.com/research-covers/language-unsupervised/language\\_understanding\\_paper.pdf](https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf)
- [68] Vaswani, A., et al. (2017). *Attention is All You Need*. arXiv : 1706.03762
- [69] Brown, T., et al. (2020). *Language Models are Few-Shot Learners*. arXiv : 2005.14165

## Abstract

Over the past decades, chatbots have taken on an important role in various fields, particularly in e-commerce, where they enrich the customer-product relationship by providing personalized assistance. In this thesis, an in-depth analysis of the literature on chatbots allowed us to identify various approaches, ranging from classical AI methods (LSTM, CNN) to modern Transformer-based models, known for their ability to understand context and produce relevant responses.

Based on this, we developed a hybrid architecture utilizing several complementary technologies : **NLTK** for simple interactions and linguistic preprocessing, **BERT** for fine-grained understanding of textual queries, **CLIP** for visual similarity search, and **GPT** for reformulating responses or generating original content. This combination enables the chatbot to offer effective semantic search and a smoother, more human-like interaction.

Despite these encouraging results, some limitations remain, particularly a rudimentary conversational memory and partial customer support. Future perspectives focus on long-term contextual tracking, integration with transactional systems, and the development of comprehensive assistance throughout the entire customer journey.

**Keywords** : Chatbots, E-commerce, AI, NLTK, BERT, CLIP, GPT, Multimodality, Semantic Search.

## Résumé

Au fil des dernières décennies, les chatbots ont pris une place importante dans de nombreux domaines, en particulier le e-commerce, où ils enrichissent la relation client-produit grâce à une assistance personnalisée. Dans ce mémoire, une analyse approfondie de la littérature sur les chatbots nous a permis d'identifier diverses approches, allant des méthodes classiques d'IA (LSTM, CNN) aux modèles modernes basés sur les Transformers, reconnus pour leur capacité à comprendre le contexte et produire des réponses pertinentes.

Sur cette base, nous avons conçu une architecture hybride exploitant plusieurs technologies complémentaires : **NLTK** pour les interactions simples et le prétraitement linguistique, **BERT** pour une compréhension fine des requêtes textuelles, **CLIP** pour la recherche par similarité visuelle, et **GPT** pour reformuler les réponses ou générer du contenu original. Cette combinaison permet au chatbot de proposer une recherche sémantique performante et une interaction plus fluide et humaine.

Malgré ces résultats encourageants, des limites subsistent, notamment une mémoire conversationnelle encore rudimentaire et un accompagnement client partiel. Les perspectives portent sur le suivi contextuel à long terme, l'intégration aux systèmes transactionnels, et le développement d'une assistance complète sur l'ensemble du parcours client.

**Mots-clés** : Chatbots, E-commerce, IA, NLTK, BERT, CLIP, GPT, Multimodalité, Recherche sémantique.