Université Abou Bekr Belkaid
Tlemcen Algérie
جامعة أبي بكر بلقايد
تلمسان الجزائر

République Algérienne Démocratique et Populaire
Université Abou Bakr Belkaid - Tlemcen
Faculté des Sciences
Département d'Informatique

Mémoire de fin d'études dans le cadre du decrêt 1275

Pour l'obtention du diplôme de Master en Informatique

Option: Intelligence Artificielle (I.A)

Thème

PsyBot: Chatbot intelligent d'accompagnement psychologique

Réalisé par:

- Charef Fatma
- Dib Samira

Présenté le 2 juillet, 2025 devant le jury composé de:

- Mme. ILES Nawel (Président)
- M. SMAHI Mohammed Ismail (Encadrant)
- Mme. KAZI TANI Adila (Examinateur)
- M. MERZOUG Mohamed (Expert I2E)

Année Universitaire 2024/2025

Nous dédions ce modeste travail à :

À mon cher père, pour ses innombrables sacrifices. Qu'Allah t'accorde santé, paix du cœur et la plus haute des récompenses. Cette réussite est aussi la tienne.
À ma douce mère, dont l'amour, les prières et la sagesse m'ont portée. Que chaque fatigue vécue soit une source de bénédiction ici-bas et dans l'au-delà. Je t'aime infiniment.
À mes frères, Fouad et Mohamed, mes repères et protecteurs. Qu'Allah vous accorde réussite et préservation. Merci pour votre amour fraternel.
À ma petite sœur Karima, mon trésor. Que ta vie soit remplie de lumière, de savoir et de bonheur.
À Farah, mon amie et sœur de cœur. Merci pour ton écoute, ton soutien fidèle et ta loyauté. Que notre lien reste fort et sincère.
À mon oncle Rachid, pour son aide sincère, sa générosité et son soutien désintéressé. Que Dieu te récompense pour tout le bien que tu m'as apporté.
À ma binôme et amie Samira, pour ta rigueur, ton engagement et ton soutien tout au long de ce travail. Qu'Allah t'accorde réussite et épanouissement.
À ma famille et à tous ceux qui m'ont soutenue, merci pour votre amour, vos prières et votre bienveillance. Que Dieu vous protège et vous récompense.

Fatma

À ma mère, pilier de ma vie, pour son amour inépuisable, ses encouragements constants et sa foi indéfectible en moi. C'est grâce à toi que j'ai tenu bon. Que Dieu te comble de Ses bienfaits ici-bas et dans l'au-delà.
À mon père, ma sœur Sihem et mon frère Chakib, pour leur affection, leur soutien moral et les moments de joie partagés qui m'ont donné force et motivation.
À mes amis, pour leur écoute, leur encouragement, leur patience et leurs mots justes. Vous m'avez accompagnée même dans le silence.
À ma binôme Fatma, merci pour ton soutien et ta collaboration précieuse tout au long de ce travail.
À toute ma famille, pour l'amour, la prière et les valeurs que vous m'avez transmises.
À tout mes amis qui m'ont soutenue, conseillée, aidée, ou simplement cru en moi, de près ou de loin. Merci du fond du cœur. Cette réussite est aussi la vôtre.
Samira

Remerciement

Tout d'abord, nous remercions profondément ALLAH, qui nous a toujours aidés par Sa grande générosité, en nous donnant le courage, la volonté et la détermination nécessaires pour mener à bien tous nos projets d'études.

Nous remercions également Monsieur M. I. SMAHI, notre encadrant, pour son accompagnement, ses précieux conseils, sa disponibilité et son soutien tout au long de ce travail. Son expertise et sa bienveillance ont grandement contribué à la réalisation de ce projet.

Nous exprimons notre sincère gratitude aux professeurs Iles Nawel et Kazi Tani Adila pour avoir accepté de présider notre jury et d'évaluer ce travail avec bienveillance.

Nous exprimons notre sincère gratitude aux professeurs Benazzouz Mortada, pour tout le savoir qu'il nous a transmis et la qualité de son enseignement, ainsi qu'à Hadjila Fethallah, pour ses conseils précieux.

Et aussi Monsieur Meghebbar Ayoub pour son aide et ses presieux conseils.

Enfin, nous tenons à exprimer toute notre gratitude à nos chers parents pour leur amour inconditionnel, leur patience, leurs encouragements constants et les sacrifices qu'ils ont consentis pour nous permettre d'avancer. Un grand merci également à nos familles et à nos amis, dont le soutien moral, la compréhension et la présence bienveillante nous ont été d'un grand réconfort tout au long de ce parcours. Nous n'oublions pas non plus toutes les personnes, de près ou de loin, qui nous ont apporté leur aide, leurs conseils ou leur motivation à un moment ou un autre : à chacun, merci du fond du cœur.

Résumé

Ces dernières années, les enjeux liés à la santé mentale ont pris une place centrale dans nos sociétés. En parallèle, les technologies numériques, et en particulier l'intelligence artificielle, offrent de nouvelles pistes pour accompagner les individus de manière innovante. Ce mémoire explore la mise en place d'un chatbot capable d'interagir avec des utilisateurs en tenant compte de la nature de leurs demandes, en particulier dans un contexte psychologique. Le fonctionnement du système repose sur plusieurs étapes successives. Une première analyse, effectuée avec le modèle DistilBert avec un fine-tuning, permet de classer les requêtes entre celles qui relèvent d'une simple recherche d'information, et celles à caractère psychologique. Les requêtes informatives sont traitées, par la suite, à l'aide de la bibliothéque duckduckgo_search, qui fournit une réponse issue de sources web fiables. Lorsqu'il s'agit d'une demande plus sensible, une détection émotionnelle est lancée grâce à une combinaison de BERT et VADER, afin d'évaluer l'intensité du message. Si l'émotion détectée reste dans un seuil jugé acceptable, une réponse est générée par le modèle GPT. Dans le cas contraire, l'utilisateur reçoit un message de redirection ou de prudence. Ce travail met en lumière les apports, mais aussi les limites, d'un tel système dans un domaine aussi délicat. L'architecture proposée, à la fois simple et évolutive, constitue une base solide pour envisager de futurs outils d'assistance psychologique automatisée.

Mots-clés: Santé mentale, Chatbot psychologique, Intelligence artificielle, Traitement du langage naturel, Émotion, Intention, Interaction humain-IA.

ملخص

في السنوات الأخيرة، أصبحت قضايا الصحة النفسية أكثر حضورًا وأهمية في مجتمعاتنا. في المقابل، توفّر التقنيات الرقمية، وخاصة الذكاء الاصطناعي، آفاقًا جديدة لمرافقة الأفراد بطريقة مبتكرة. يتناول هذا البحث تطوير روبوت محادثة (Chatbot) قادر على التفاعل مع المستخدمين بناءً على طبيعة استفساراتهم، وخصوصًا في السياقات النفسية. يعتمد النظام على سلسلة من المراحل المتتالية. يتم أولاً تصنيف نوع الطلب باستخدام DistilBERT لتحديد ما إذا كان طلبًا لمعلومة عامة، أو استفسارًا ذا طابع نفسي. بعد ذلك تُعالج الطلبات المعلوماتية باستخدام أداة duckduckgo_search التي تُرجع إجابات من مصادر موثوقة على الإنترنت. بالنسبة للطلبات ذات الطابع النفسي، يتم تحليل المشاعر باستخدام مزيج من النموذجين BERT وVADER لتقييم شدة الانفعالات. أخيرا إذا كانت الحالة العاطفية ضمن الحدود المقبولة، يتم توليد الرد باستخدام GPT. وإذا كانت المشاعر قوية جدًا، يتم إرسال رسالة تحذيرية أو توجيه المستخدم إلى مساعدة أخرى. يسلّط هذا العمل الضوء على إمكانيات وحدود مثل هذا النظام في مجال حساس. ويقترح بنية بسيطة وقابلة للتطوير، تمثّل أساسًا لتطوير أدوات مستقبلية للدعم النفسي المعتمد على الذكاء الاصطناعي.

الكلمات الدالة :الصحة النفسية، روبوت محادثة نفسي، الذكاء الاصطناعي، معالجة اللغة الطبيعية، المشاعر، النية، التفاعل بين الإنسان والآلة.

Abstract

In recent years, mental health has become an increasingly important issue in our societies. At the same time, digital technologies-especially artificial intelligence—offer new ways to support individuals in innovative ways. This thesis explores the development of a chatbot designed to interact with users based on the nature of their requests, particularly in a psychological context. The system operates through several key steps. First, an initial classification using the model DistilBERT with fine-tuning identifies whether the user's request is a general information query or a psychologically oriented question. Next, an informational requests are handled using duckduckgo_search, which retrieves answers from reliable web sources. In contrast, psychological queries undergo an emotional analysis using a combination of BERT and VADER to determine the intensity of the user's emotions. Finaly, if the emotional state is within an acceptable threshold, a response is generated using GPT; otherwise, the user receives a cautionary message or is redirected accordingly. This work highlights both the potential and the limitations of such a system in a sensitive domain. The proposed architecture, which is simple yet scalable, provides a solid foundation for the development of future AI-based psychological support tools.

Keywords: Mental health, Psychological chatbot, Artificial intelligence, Natural language processing, Emotion, Intent, Human-AI interaction.

Liste des Acronymes

Abréviation	Signification
DistilBERT	Distilled Bidirectional Encoder Representations from Transformers
BERT	Bidirectional Encoder Representations from Transformers
VADER	Valence Aware Dictionary and sEntiment Reasoner
GPT	Generative Pre-trained Transformer
NLP	Natural Language Processing
IA	Intelligence Artificielle
LLMs	Large Language Models
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
GAN	Generative Adversarial Network
NER	Named Entity Recognition
LSTM	Long Short-Term Memory
GRU	Gated Recurrent Unit
Word2Vect	Word to Vector
GloVe	Global Vectors for Word Representation
HMM	Hidden Markov Model
RI	Recherche dInformation (Information Retrieval)
BM25	Best Matching 25
TF-IDF	Term Frequency Inverse Document Frequency
CBOW	Continuous Bag of Words
TCC	Thérapie Cognitive et Comportementale (Cognitive Behavioral Therapy - CBT)
TAG	Thérapie dAcceptation et dEngagement
TOG	Thérapie dOrientation Gestalt
BLEU	Bilingual Evaluation Understudy
ROUGE	Recall-Oriented Understudy for Gisting Evaluation
METEOR	Metric for Evaluation of Translation with Explicit ORdering
BMC	Business Model Canvas

Table 1 – Abréviations et leurs significations

Table des Matières

Remerciement v

Résumé vii

ملخص ix

Abstract xi

Liste des Acronymes xi

Table des Matières xvii

Liste des Figures xix

Liste des Tableaux xxi

1 Introduction Générale 1

2 Fondements théoriques et conceptuels 5
2.1 Introduction 6
2.2 Fondements de l'intelligence artificielle et du NLP 6
2.2.1 Fondements de l'intelligence artificielle 6
2.2.2 Concepts et tâches en traitement du langage naturel (NLP) 8
2.2.3 Architectures et technologies avancées pour le NLP 10
2.3 Fondements de la psychologie 16
2.3.1 Définition et importance de la psychologie 16
2.3.2 Historique de la psychologie 17
2.3.3 Courants majeurs de la psychologie 18
2.3.4 Psychologie clinique et psychothérapie 18
2.3.5 Les émotions en psychologie 19
2.3.6 Problèmes psychologiques fréquents (anxiété, dépression, etc.) . . 19
2.3.7 Barrières d'accès à la thérapie (coût, distance, stigmatisation) . . 20
2.4 La convergence entre l'intelligence artificielle et la psychologie 20
2.4.1 Applications NLP pour la santé mentale 20
2.4.2 Utilisation de modèles NLP pour les chatbots psychologiques . . . 21
2.4.3 Détection d'émotions et d'humeurs avec le NLP 21

2.4.4 Classification des émotions avec BERT . . . 22
2.4.5 Limites éthiques et défis d'utilisation de l'IA en santé mentale . . . 22
2.5 Conclusion . . . 23

3 Description de l'Architecture du Système PsyBot 25
3.1 Introduction . . . 26
3.2 Modèle proposé . . . 26
3.3 Modèles et algorithmes . . . 27
3.3.1 DistilBERT pour la classification des intentions . . . 28
3.3.2 MiniBERT et VADER pour la classification des émotions . . . 28
3.3.3 GPT pour la génération de réponses . . . 28
3.4 Modélisation de l'architecture du PsyBot . . . 29
3.4.1 Fonctionement globale . . . 30
3.4.2 Pseudo-algorithme . . . 31
3.5 Conclusion . . . 31

4 Expérimentations et interprétation des résultats 33
4.1 Introduction . . . 34
4.2 Classification des intetions DistilBERT (Phase 1) . . . 34
4.2.1 Présentation du dataset Utilisé . . . 34
4.2.2 Prétraitement des données . . . 35
4.2.3 Entraînement du modèle . . . 35
4.2.4 Résultats . . . 36
4.2.5 Évaluation et test . . . 37
4.2.6 Comparaison avec l'approche baseline . . . 38
4.3 Classification des emotions BERT et VADER (Phase 2) . . . 38
4.3.1 Présentation du dataset utilisé . . . 38
4.3.2 Prétraitement des données: . . . 39
4.3.3 Entraînement du modèle . . . 40
4.3.4 Résultats . . . 42
4.3.5 Évaluation et test . . . 43
4.3.6 Comparaison avec les autre travaux . . . 45
4.4 Génération de texte GPT (Phase 3) . . . 46
4.4.1 Présentation du dataset Utilisé . . . 46
4.4.2 Prétraitement des données . . . 47
4.4.3 Entraînement du modèle . . . 48
4.4.4 Résultats . . . 49
4.4.5 Évaluation et test . . . 50
4.4.6 Comparaison avec les autre travaux . . . 51
4.5 Conclusion . . . 52

5 Mise en œuvre pratique du psybot 55
5.1 Introduction . . . 56
5.2 Diagrammes . . . 56
5.2.1 Diagramme des cas d'utilisations . . . 56
5.3 Interfaces et explications . . . 57
5.3.1 Patients . . . 57
5.3.2 Psychologue . . . 62
5.3.3 Clinique . . . 62
5.4 Conclusion . . . 63

6 Conclusion générale 65

7 Business Modele Canva (BMC) 67
7.1 Introduction . . . 68
7.2 BMC . . . 68
7.2.1 Proposition de valeur . . . 68
7.2.2 Segments clients . . . 69
7.2.3 Canaux . . . 69
7.2.4 Relations clients . . . 69
7.2.5 Partenaires clés . . . 70
7.2.6 Activités clés . . . 70
7.2.7 Ressources clés . . . 71
7.2.8 Charges et Coûts . . . 72
7.2.9 Revenus . . . 73
7.3 Conclusion . . . 74

Références bibliographiques et webographiques 77

Liste des Figures

3.1 Architecture Globale du PsyBot 27
3.2 Architecture Détaillée du PsyBot 29

4.1 Distribution des classes dans le jeu de DistilBERT 35
4.2 Courbes de performance du modèle DistilBERT 37
4.3 Matrice de confusion apres le fine-tuning 38
4.4 Matrice de confusion avant le fine tuning 40
4.5 Comparaison des distributions d'émotions avant et après fusion 41
4.6 Distribution des émotions après prétraitement 42
4.7 Courbes de perte 42
4.8 Courbes d'accuracy 43
4.9 Matrice de confusion 45
4.10 Nuage de mots 47
4.11 Distribution de longueur des reponses 48
4.12 Courbes de perte 50
4.13 Mériques d'évaluation de GPT (BLEU,ROUGE-L,METEOR) 51
4.14 Métrique d'évaluation de GPT (Perplexity) 52

5.1 Diagramme de cas d'utilisation 57
5.2 Portail du site pour le Patient 58
5.3 Requête de recherche 58
5.4 Requête avec emotion non acceptable 59
5.5 Requête de soutien du psybot 59
5.6 Conversation en arabe 60
5.7 Communauté 60
5.8 Historique 61
5.9 Soutien psychologue 61
5.10 choix clinique 62
5.11 psychologue 62
5.12 Clinique 63

7.1 BMC 75

Liste des Tableaux

1 Abréviations et leurs significations xiii

2.1 Comparaison fonctionnelle entre les chatbots psychologues existants . . . 22

4.1 Résultats des metriques pour la classification d'intentions 38
4.2 Résultat de la classification d'intention 39
4.3 Résultats des métriques d'évaluation par émotion 44
4.4 Résultats de la classification des émotions avec score VADER et acceptabilité 46
4.5 tableau des f1-scores 46
4.6 Résultats pour chaque métrique 51
4.7 Évaluation des modèles sur le même dataset 52
4.8 Comparaison fonctionnelle entre PsyBot et des chatbots psychologues existants 53

1

Introduction Générale

Le traitement automatique du langage naturel (Natural Language Processing - NLP) est un domaine de l'intelligence artificielle (IA) qui permet aux machines de comprendre, d'évaluer et de générer intelligemment du contenu textuel. Grâce aux avancées en apprentissage automatique, les applications conversationnelles comme les chatbots ont connu une forte croissance ces dernières années. Leur intégration dans des domaines tels que la santé mentale et la psychologie a ouvert la voie à des compagnons virtuels capables de reconnaître les émotions des utilisateurs et de leur apporter un certain réconfort.

Dans ce contexte, les modèles de langage de grande taille (LLMs) tels que BERT (représentations encodées bidirectionnelles issues des Transformers) et GPT (transformers génératifs pré-entraînés) ont significativement amélioré l'efficacité des chatbots dans le domaine de la santé mentale. Ces architectures permettent une meilleure compréhension des requêtes des utilisateurs et la génération de réponses plus naturelles et empathiques [1], [2].

Problématique

Dans un contexte où les troubles psychologiques touchent une part croissante de la population, l'accès à un accompagnement psychologique reste difficile pour de nombreuses personnes. Deux obstacles majeurs se posent : d'une part, le coût élevé des consultations, qui rend les services de santé mentale inaccessibles à ceux qui disposent de moyens financiers limités ; d'autre part, la barrière géographique, notamment pour les personnes vivant dans des zones rurales ou isolées, où les professionnels de la santé mentale sont peu voire pas représentés.

En parallèle, les solutions numériques se multiplient, notamment les chatbots conversationnels. Cependant, la majorité de ces outils présentent encore des limites : compréhension superficielle du langage naturel, manque d'adaptabilité au contexte émotionnel, réponses génériques ou préprogrammées, et absence de sensibilité humaine.

Ainsi, une question essentielle émerge : Comment concevoir un agent conversationnel intelligent, accessible, multilingue et émotionnellement pertinent, capable d'apporter un

soutien psychologique crédible aux personnes en situation d'isolement ou de précarité, tout en garantissant une interaction éthique et fiable ?

Ce mémoire explore cette problématique à travers le développement de PsyBot, un chatbot psychologue basé sur les technologies de traitement du langage naturel (NLP), combinant des modèles de grande taille (LLM), la classification émotionnelle, et des approches hybrides de génération de réponse, dans le but d'offrir un soutien psychologique accessible à tous.

Objectifs

L'objectif principal de ce travail est de concevoir un agent conversationnel intelligent, nommé PsyBot, capable d'offrir un accompagnement psychologique de première intention, accessible, multilingue et émotionnellement pertinent, à des utilisateurs pouvant se trouver dans des situations d'isolement géographique ou de précarité économique.

Pour atteindre cet objectif général, plusieurs objectifs spécifiques ont été définis:

— Analyser les limites des chatbots existants en matière de prise en charge psychologique .

— Étudier les technologies NLP adaptées à la compréhension des émotions et des intentions dans un contexte conversationnel .

— Concevoir une architecture hybride intégrant la classification des requêtes, l'analyse émotionnelle, et la génération de réponses adaptées .

— Implémenter un pipeline technique combinant des outils comme BERT, VADER et GPT pour gérer les différents niveaux d'interaction .

— Tester et évaluer le chatbot en termes de pertinence, d'adaptabilité émotionnelle et de fluidité conversationnelle .

— Réfléchir aux enjeux éthiques liés à l'usage d'un chatbot psychologique, notamment en termes de limites d'intervention, de confidentialité et de sécurité émotionnelle.

Ce mémoire est structuré en six chapitres, permettant d'explorer de manière progressive la conception, le développement et la valorisation du chatbot PsyBot :

Le chapitre 1 introduit la problématique du projet, les objectifs visés, ainsi que le contexte scientifique et sociétal dans lequel s'inscrit ce travail.

Le chapitre 2 présente les fondements théoriques et conceptuels, en abordant les bases de l'intelligence artificielle, du traitement du langage naturel (NLP), et de la psychologie, tout en analysant les enjeux liés à leur convergence dans le domaine du soutien psychologique

Le chapitre 3 est dédié à la description de l'architecture du système PsyBot, en expliquant le fonctionnement global, les modules utilisés (BERT, VADER, GPT), et les choix techniques adoptés.

Le chapitre 4 concerne les expérimentations, avec une présentation des datasets, du prétraitement, des entraînements réalisés, ainsi qu'une évaluation des performances des différents modèles.

Le chapitre 5 est consacré à la mise en œuvre pratique du système, en décrivant les interfaces, les cas d'utilisation, et les aspects techniques liés au déploiement du prototype.

Le chapitre 6 présente le modèle économique du projet à travers le Business Model Canvas (BMC), mettant en évidence la valeur ajoutée, les segments cibles, les ressources clés, les partenaires potentiels et les sources de revenus.

2

Fondements théoriques et conceptuels

Sommaire

2.1 Introduction 6
2.2 Fondements de l'intelligence artificielle et du NLP 6
2.2.1 Fondements de l'intelligence artificielle 6
2.2.2 Concepts et tâches en traitement du langage naturel (NLP) 8
2.2.3 Architectures et technologies avancées pour le NLP 10
2.3 Fondements de la psychologie 16
2.3.1 Définition et importance de la psychologie 16
2.3.2 Historique de la psychologie 17
2.3.3 Courants majeurs de la psychologie 18
2.3.4 Psychologie clinique et psychothérapie 18
2.3.5 Les émotions en psychologie 19
2.3.6 Problèmes psychologiques fréquents (anxiété, dépression, etc.) . . 19
2.3.7 Barrières d'accès à la thérapie (coût, distance, stigmatisation) . . 20
2.4 La convergence entre l'intelligence artificielle et la psychologie 20
2.4.1 Applications NLP pour la santé mentale 20
2.4.2 Utilisation de modèles NLP pour les chatbots psychologiques . . . 21
2.4.3 Détection d'émotions et d'humeurs avec le NLP 21
2.4.4 Classification des émotions avec BERT 22
2.4.5 Limites éthiques et défis d'utilisation de l'IA en santé mentale . . . 22
2.5 Conclusion 23

2.1 Introduction

Ce chapitre constitue une revue théorique fondamentale qui vise à explorer les concepts clés mobilisés dans le cadre de notre projet. Il s'articule autour de trois axes majeurs : les bases de l'intelligence artificielle et du traitement automatique du langage naturel (NLP), les fondements de la psychologie clinique, et enfin, la convergence entre ces deux domaines dans le contexte de la santé mentale.

Dans un premier temps, nous abordons les principes fondamentaux de l'IA et du NLP, en présentant les approches classiques, les avancées modernes comme les modèles Transformers, et les techniques de classification utilisées dans le traitement du langage. Ensuite, nous dressons un panorama des fondements de la psychologie, notamment ses courants théoriques, les troubles émotionnels fréquents, et les obstacles à l'accès aux soins psychologiques.

Enfin, nous explorons comment les technologies NLP peuvent être appliquées à la santé mentale, à travers des cas d'usage concrets tels que les chatbots thérapeutiques, la détection des émotions et les défis éthiques associés. Cette partie pose ainsi les bases théoriques nécessaires à la conception et à la compréhension du système PsyBot.

2.2 Fondements de l'intelligence artificielle et du NLP

2.2.1 Fondements de l'intelligence artificielle

Définition et importance de l'intelligence artificielle

L'intelligence artificielle (IA) désigne un ensemble de méthodes informatiques visant à reproduire certaines capacités cognitives humaines telles que la perception, le raisonnement, ou encore la compréhension du langage [3]. À travers des techniques comme l'apprentissage automatique et les réseaux de neurones, elle permet aux machines d'apprendre à partir des données, sans être explicitement programmées pour chaque tâche spécifique.

L'IA occupe aujourd'hui une place centrale dans l'innovation technologique. Elle est au cœur de nombreuses avancées dans des domaines variés comme la médecine, la finance, les transports, et particulièrement le traitement du langage naturel (NLP), où elle facilite l'analyse sémantique, la traduction automatique, ou encore l'interaction homme-machine [4]. En analysant de grands volumes de textes, les systèmes d'IA sont capables d'identifier des tendances, de détecter des émotions, et de proposer des réponses contextuelles cohérentes.

Dans le champ de la santé mentale, l'intelligence artificielle ouvre de nouvelles perspectives pour améliorer l'accessibilité aux soins. Elle permet notamment la création d'assistants virtuels capables de dialoguer avec les patients, de repérer des signes précoces de détresse psychologique ou de proposer un accompagnement personnalisé [5]. Ces avancées soulignent l'importance croissante de l'IA comme levier de transformation sociale et médicale.

Sous-ensembles de l'intelligence artificielle : Machine Learning et Deep Learning

Machine Learning (Apprentissage automatique)

Le Machine Learning est une branche de l'intelligence artificielle qui vise à concevoir des algorithmes permettant aux machines d'apprendre à partir des données et d'adapter leur comportement sans intervention humaine explicite. Ces systèmes sont capables d'identifier des patterns dans les données et d'améliorer leurs performances au fil du temps, ce qui les rend aptes à résoudre des problèmes complexes dans des domaines variés comme la reconnaissance d'images, la prédiction de tendances, ou encore la recommandation de contenus [6].

Les algorithmes de Machine Learning se divisent principalement en trois catégories :

— Apprentissage supervisé : L'algorithme apprend à partir de données étiquetées, c'est-à-dire où chaque entrée est associée à une sortie spécifique. L'objectif est d'établir une relation entre les deux pour effectuer des prédictions précises [7].

— Apprentissage non supervisé : Contrairement à l'apprentissage supervisé, les données utilisées ne sont pas étiquetées. L'algorithme cherche des structures ou des regroupements cachés dans les données, comme dans le cas du clustering [8].

— Apprentissage par renforcement : Ce type d'apprentissage s'inspire de la psychologie comportementale. Un agent apprend à effectuer des actions dans un environnement donné pour maximiser une récompense cumulative, en ajustant ses actions en fonction des résultats obtenus [9].

Le Machine Learning est au cœur de nombreuses technologies actuelles, de la conduite autonome à la détection de fraudes, en passant par les systèmes de personnalisation des services en ligne [10].

Deep Learning (Apprentissage Profond)

Le Deep Learning est une sous-catégorie du Machine Learning qui se concentre sur l'utilisation de réseaux neuronaux profonds, c'est-à-dire des architectures composées de multiples couches d'unités de traitement, appelées neurones. Ces réseaux sont capables d'apprendre des représentations complexes des données, et sont particulièrement efficaces pour des tâches comme la reconnaissance d'images, la traduction automatique, et la reconnaissance vocale [6].

Le Deep Learning repose sur des techniques d'apprentissage supervisé et non supervisé. Cependant, ce qui le distingue, c'est la capacité de ces modèles à extraire automatiquement des caractéristiques à partir de grandes quantités de données, sans nécessiter de prétraitement manuel des caractéristiques. Par exemple, dans les réseaux convolutifs (CNN), utilisés pour la reconnaissance d'images, les couches successives apprennent à détecter des motifs de plus en plus complexes à chaque niveau de profondeur [7].

Les réseaux neuronaux profonds peuvent être composés de plusieurs types d'architectures, parmi lesquelles :

— Réseaux de neurones convolutifs (CNN) : Utilisés principalement pour le traitement des images et des vidéos, ces réseaux apprennent des filtres hiérarchiques pour extraire des caractéristiques à différents niveaux [8].

— Réseaux de neurones récurrents (RNN) : Ces réseaux sont utilisés pour traiter des données séquentielles, comme les séries temporelles et le langage naturel, en mémorisant les informations précédentes pour influencer les décisions futures [9].

— Réseaux antagonistes génératifs (GAN) : Une architecture où deux réseaux neuronaux s'affrontent pour générer de nouvelles données réalistes, souvent utilisées dans la génération d'images ou de textes [10].

— Transformers : Utilisés principalement pour le traitement du langage naturel, les Transformers utilisent des mécanismes d'attention pour traiter l'ensemble de la séquence simultanément, ce qui les rend particulièrement efficaces pour capturer des dépendances à long terme et améliorer les performances dans des tâches telles que la traduction automatique et la compréhension du langage [7].

Le Deep Learning est devenu un pilier fondamental dans de nombreux domaines, notamment la vision par ordinateur, la reconnaissance vocale, et la création de contenu automatisée.

Classification en apprentissage automatique et apprentissage profond La classification est une tâche fondamentale en apprentissage automatique (Machine Learning - ML) et en apprentissage profond (Deep Learning - DL). Elle consiste à affecter une étiquette prédéfinie à une donnée en fonction de ses caractéristiques internes [10].

En ML, des algorithmes comme la régression logistique, les machines à vecteurs de support (SVM), ou les forêts aléatoires sont utilisés pour identifier des motifs à partir de données structurées [6]. Ils nécessitent généralement une bonne ingénierie des caractéristiques.

En DL, les réseaux de neurones profonds permettent d'apprendre automatiquement des représentations à partir de données brutes. Les CNN, RNN ou LSTM sont fréquemment utilisés dans des tâches de classification liées au texte ou aux images [11]. Dans le NLP, cette capacité à comprendre des séquences rend les modèles DL très performants pour des tâches telles que l'analyse des sentiments ou la classification d'émotions.

Le processus de classification comprend :

— Le prétraitement des données textuelles ;

— L'extraction ou l'apprentissage de caractéristiques ;

— L'entraînement supervisé sur des données étiquetées ;

— L'évaluation via des métriques comme la précision, le rappel ou la F-mesure [11].

2.2.2 Concepts et tâches en traitement du langage naturel (NLP)

Principales tâches en NLP Le NLP couvre plusieurs tâches fondamentales :

— Classification de texte : attribuer des catégories à un texte (ex. : spam/non-spam, analyse de sentiments).

— Génération de texte : produire automatiquement du texte cohérent (ex. : chatbot, résumé automatique).

— Analyse des sentiments : identifier les émotions et opinions exprimées dans un texte.

— Reconnaissance d'entités nommées (NER) : détecter et classer les entités telles que les noms de personnes, lieux, organisations, etc.

— Traduction automatique : convertir un texte d'une langue à une autre [12].

Évolution des modèles NLP: des approches classiques aux modèles neuronaux L'évolution des modèles NLP s'est faite en plusieurs étapes importantes :

— Approches basées sur les règles : utilisation des grammaires et des dictionnaires pour analyser la langue.

— Modèles statistiques : approches probabilistes telles que les modèles N-grammes et la régression logistique.

— Word Embeddings : représentation continue des mots dans un vecteur (ex: Word2Vec, GloVe).

— Modèles neuronaux : adoption des RNN, LSTM et GRU pour comprendre l'ordre et les détails des mots.

— Transformers : introduction des modèles BERT et GPT, basés sur l'auto-attention, qui ont révolutionné le domaine en offrant des performances inégalées [13].

Modèles de langage pour le NLP

Approches classiques du NLP L'analyse probabiliste N-gram prévoit les cotes d'occurrence en précédant les termes pour le traitement du langage. Une paire de mots adjacents est ce qu'un bigrrame recherche, tandis qu'un trimère considère trois en séquence.

— Formule générale du modèle N-grammes :

$$Pw_n|w_1, w_2, \ldots, w_{n-1} \approx Pw_n|w_{n-N1}, \ldots, w_{n-1}$$

où N est la taille de l'historique considéré.

Les modèles statistiques comme Hidden Markov Models (HMM) et Maximum Entropy Models (MEM) ont également été utilisés pour des tâches comme l'étiquetage grammatical et la reconnaissance d'entités nommées.

Cependant, ces approches souffrent de plusieurs limites, notamment :

— Explosion combinatoire : Les grands corpus nécessitent énormément de données pour estimer correctement les probabilités.

— Manque de généralisation : Les modèles N-grammes ne capturent pas bien la relation sémantique entre les mots éloignés.

Recherche d'information (RI) La recherche d'information (RI) est un domaine de l'informatique qui vise à retrouver, à partir d'une requête, l'information la plus pertinente dans un ensemble de documents. Dans le traitement automatique du langage naturel (TALN), elle permet d'améliorer l'accès aux données textuelles [11].

Elle repose sur plusieurs étapes clés :

— Tokenisation : découpe du texte en unités élémentaires (mots, phrases).

— Suppression des stop words : élimination des mots fréquents sans valeur sémantique forte.

— Lemmatisation : réduction des mots à leur forme de base.

— Stemming : réduction des mots à leur racine, souvent plus agressive que la lemmatisation.

— Indexation et pondération : transformation des textes en structures exploitables via des scores (comme TF-IDF ou BM25).

Ces traitements améliorent l'efficacité des systèmes de recherche documentaire [14].

TF-IDF et BM25 Le TF-IDF (Term Frequency - Inverse Document Frequency) mesure l'importance d'un mot dans un document, pondérée par sa fréquence dans l'ensemble du corpus :

$$\text{TF-IDF}t, d = \text{TF}t, d \times \text{IDF}t$$

$$\text{IDF}t = \log\left(\frac{N}{\text{df}t}\right)$$

où N est le nombre total de documents et df$t$ est le nombre de documents contenant le terme t.
Le BM25 est une version améliorée du TF-IDF qui prend en compte la longueur des documents pour pondérer les termes de manière plus précise.

2.2.3 Architectures et technologies avancées pour le NLP

Word Embeddings (Word2Vec, GloVe, FastText)

Les "word embeddings" sont des représentations vectorielles continues des mots, capturant leur signification contextuelle.

— Word2Vec [15] : repose sur deux architectures :
  — CBOW (Continuous Bag of Words) : prédit un mot à partir de son contexte.
  — Skip-gram : prédit le contexte à partir d'un mot cible.

— GloVe [16] : factorise une matrice de co-occurrence globale des mots.

— FastText [17] : représente les mots par des sous-unités, ce qui améliore la gestion des mots rares.

Réseaux de neurones pour le NLP

RNN, LSTM, GRU (architectures récurrentes)

Les réseaux de neurones récurrents (RNN) sont adaptés aux séquences comme les textes, car ils conservent une mémoire des éléments précédents.

Limites des RNN

— Vanishing gradient : perte du signal d'apprentissage dans les longues séquences.

— Difficulté à gérer des relations distantes dans le texte.

Pour résoudre ces problèmes, deux variantes ont été introduites :

LSTM (Long Short-Term Memory) Les LSTM utilisent une cellule mémoire et trois portes (entrée, oubli, sortie) pour gérer le flux d'information [18] :

$$h_t = o_t \times \tanh c_t$$

où o_t est la porte de sortie et c_t l'état mémoire.

GRU (Gated Recurrent Unit) Les GRU [19], plus simples, utilisent deux portes:

— Porte de mise à jour

— Porte de réinitialisation

Ils offrent des performances comparables aux LSTM, avec un entraînement plus rapide.

Limites des LSTM/GRU et apparition des Transformers Malgré leurs avantages, les LSTM/GRU ont plusieurs limitations :

— Traitement séquentiel limitant la parallélisation.

— Difficulté à capturer des dépendances très longues.

— Disparition du gradient avec le temps.

Ces contraintes ont motivé le développement des Transformers , qui utilisent une approche d'attention pour traiter toutes les positions d'une séquence simultanément.

Les modèles Transformers

Introduction aux Transformers Les Transformers sont une architecture de réseau de neurones introduite par Vaswani et al. en 2017 [20], conçue pour traiter les séquences de données (comme le texte) sans recourir à des mécanismes récurrents. Leur innovation clé est le mécanisme d'attention, qui permet au modèle de pondérer dynamiquement l'importance de chaque mot d'une séquence par rapport aux autres. Grâce à leur parallélisation efficace et à leur capacité à capturer des dépendances à longue distance, les Transformers sont devenus la base de nombreux modèles de pointe en traitement du langage naturel, comme BERT, GPT, T5 et autres.

Ils se composent de :

— Encodeur : convertit le texte en une représentation contextuelle.

— Décodeur : génère du texte à partir de l'encodeur.

— Attention multi-têtes : permet au modèle de se concentrer sur différentes parties du texte.

Ce traitement parallèle améliore la performance sur de grandes quantités de données.

Avantages des Transformers en NLP

Les Transformers ont apporté plusieurs avantages au domaine du NLP :

— Traitement efficace des grandes séquences sans perte de contexte.

— Capacité à généraliser sur de vastes corpus, améliorant ainsi la qualité de génération et de classification.

— Modélisation du contexte bidirectionnelle pour mieux comprendre le sens des mots en fonction de leur environnement.

Les grands modèles de langage (LLMs)

Les grands modèles de langage (LLMs), tels que BERT et GPT, représentent une avancée majeure dans le traitement automatique du langage naturel. Ces modèles sont basés sur l'architecture des Transformers, mais se distinguent par leur taille massive, leur capacité à généraliser sur de nombreux types de tâches, et leur pré-entraînement sur d'énormes corpus de texte [21].

Les LLMs peuvent être :

— Auto-encodeurs : comme BERT, qui excelle dans les tâches de compréhension de texte (classification, extraction d'information, etc.).

— Auto-régressifs : comme GPT, conçu pour la génération de texte cohérente, mot par mot, à partir d'un prompt donné.

— Hybrides : tels que T5 ou BART, combinant des aspects des deux pour des tâches de traduction, résumé, etc.

Les LLMs permettent aujourd'hui des applications avancées comme les agents conversationnels (chatbots), la rédaction automatique, ou encore l'analyse sémantique à grande échelle. Leur développement soulève toutefois des défis en termes de coût, d'éthique, et de biais [22].

Modèle BERT (Bidirectional Encoder Representations from Transformers)

Architecture et fonctionnement de BERT

BERT (Bidirectional Encoder Representations from Transformers) est un modèle de langue bidirectionnel et contextuel développé par Google en 2018 [21]. Contrairement aux modèles unidirectionnels, il analyse une séquence de mots dans les deux directions (gauche → droite et droite → gauche) simultanément, permettant une compréhension plus fine du contexte et des dépendances syntaxiques et sémantiques entre les mots.

BERT repose sur l'architecture Transformer, mais n'utilise que la partie encodeur du Transformer. Voici les principaux composants de cette architecture :

— Tokenisation (WordPiece tokenizer) :
Basée sur des sous-unités (subwords), utile pour gérer les mots rares [23].

— Embedding Layer :
Composé de trois types : token, segment et position embeddings.

— Encodeur Transformer (12 ou 24 couches) :
Composé de self-attention multi-tête et de réseaux feed-forward.

— Mécanisme d'attention (Self-attention) :
Basé sur les vecteurs Query, Key, Value :

$$\text{Attention}Q, K, V = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

— Normalisation et Résidus :
Chaque sous-couche est suivie d'une normalisation LayerNorm et d'une connexion résiduelle.

— Pré-entraînement de BERT :
Deux tâches : Masked Language Modeling (MLM) et Next Sentence Prediction (NSP).

— Fine-tuning :
BERT peut être affiné (fine-tuned) sur diverses tâches NLP en ajoutant une simple couche de sortie.

Variantes de BERT: De nombreux modèles ont été dérivés de BERT pour des langues ou des domaines spécifiques, parmi lesquels :

— BERT : (Bidirectional Encoder Representations from Transformers) est le modèle original développé par Google, capable de comprendre le contexte des mots dans les deux directions grâce à son architecture bidirectionnelle.

— RoBERTa : Une version optimisée de BERT développée par Facebook AI. Elle est entraînée plus longtemps, avec plus de données et sans la tâche NSP (Next Sentence Prediction), ce qui améliore sa performance.

— CamemBERT : Adaptation francophone de RoBERTa, préentraînée sur un corpus massif en langue française (OSCAR). Très performante sur les tâches en français.

— FlauBERT : Un autre modèle francophone, mais préentraîné sur un corpus plus varié, incluant du texte littéraire, journalistique, et scientifique. Il vise à maximiser la couverture linguistique du français.

— DistilBERT : Version allégée de BERT avec 40% de paramètres en moins tout en conservant environ 95% des performances. Il est plus rapide et mieux adapté aux appareils avec peu de ressources.

— MiniBERT : Une version encore plus compacte de BERT, conçue pour les appareils mobiles ou embarqués, avec une architecture réduite et une latence minimale.

— ALBERT (A Lite BERT) : Variante optimisée de BERT qui réduit le nombre de paramètres via le partage de poids et la factorisation des matrices, tout en conservant de bonnes performances.

— SpanBERT : Extension de BERT qui modélise directement des spans de texte (groupes de mots) plutôt que des mots isolés, utile pour l'extraction d'entités et la co-référence.

— SciBERT : Version spécialisée de BERT préentraînée sur des articles scientifiques (biomédecine et informatique principalement), utile pour le NLP dans la recherche.

— BioBERT : Variante entraînée sur des textes biomédicaux (PubMed, PMC), utilisée en NLP médical pour la reconnaissance d'entités, relations, etc.

Applications et limites de BERT :

Application : BERT est un modèle open source accessible sur des plateformes comme Hugging Face. Après son pré-entraînement, il peut être adapté (« fine-tuning ») à des tâches spécifiques, comme :

— L'analyse de sentiments.

— Les systèmes de questions-réponses.

— La classification des intentions.

Le fine-tuning consiste à réentraîner partiellement BERT sur un corpus ciblé pour réaliser une tâche précise.

Limites :

— Temps et ressources de calcul élevés.

— Moins adapté à la génération de texte.

— Difficultés d'interprétabilité

Modèle GPT (Generative Pre-trained Transformer)

Architecture et fonctionnement de GPT:

GPT, développé par OpenAI, est basé sur la partie décodeur des Transformers [24]. Contrairement à BERT, GPT est un modèle unidirectionnel, ce qui signifie qu'il prédit le prochain mot en fonction du contexte précédent. Il est donc particulièrement adapté aux tâches de génération de texte.

GPT est entraîné en deux phases :

— Pré-entraînement : le modèle apprend à prédire le mot suivant dans de grandes quantités de texte (modélisation du langage auto-régressive).

— Fine-tuning (facultatif) : pour certaines tâches spécifiques, GPT peut être ajusté avec un jeu de données supervisé.

Principaux composants de l'architecture GPT :

— Tokenisation (Byte Pair Encoding - BPE) :
Les textes sont découpés en sous-unités (subwords) via BPE, facilitant le traitement des mots rares ou inconnus.

— Embedding Layer :
Chaque token est représenté par la somme de deux vecteurs :

  — Token embeddings : vecteur associé au mot ou sous-mot.

  — Position embeddings : vecteur encodant la position du mot dans la séquence.

— Blocs Décodeurs du Transformer :
GPT empile plusieurs blocs identiques (ex : 12 pour GPT-1) comprenant chacun :

  — Masked Multi-head Self-Attention : empêche l'accès aux tokens futurs.

  — Feed-forward Network : réseau dense appliqué à chaque position.

  - Normalization et connexions résiduelles : stabilisent et accélèrent l'apprentissage.
- Tête de sortie linéaire (Output head) :
  Une couche linéaire suivie d'un softmax prédit le token suivant dans le vocabulaire.

Évolution (GPT-1, GPT-2, GPT-3, GPT-4):

- GPT-1 (2018): Premier modèle basé sur un décodeur de Transformer.
- GPT-2 (2019) : Modèle beaucoup plus grand (1,5 milliard de paramètres) capable de générer du texte fluide.
- GPT-3 (2020): 175 milliards de paramètres, permettant des applications plus complexes comme la création de contenu et les chatbots avancés.
- GPT-4 (2023): Amélioration en termes de précision, de compréhension et de génération de réponses pertinentes.

Applications et limites de GPT :

Applications :

- Génération de texte et création de contenu.
- Assistance conversationnelle.
- Traduction automatique.
- generation (ex: ChatGPT).

Limites:

- Production de réponses biaisées ou erronées.
- Coût élevé d'entraînement.
- Manque de contrôle sur la génération des réponses.

Analyse de sentiment avec VADER

En parallèle des approches neuronales complexes, il existe des méthodes légères et efficaces pour l'analyse de sentiment. VADER (Valence Aware Dictionary and sEntiment Reasoner) est un outil basé sur des règles lexicales, spécialement conçu pour analyser le ton émotionnel de textes courts, comme ceux présents sur les réseaux sociaux ou dans les messages de chat. Il repose sur un lexique enrichi d'expressions et de pondérations émotionnelles, ce qui le rend particulièrement performant sans nécessiter d'entraînement préalable [25]. VADER est donc souvent utilisé en complément ou en alternative aux réseaux neuronaux lorsqu'une solution rapide et légère est souhaitée.

Le fonctionnement de VADER repose sur plusieurs éléments clés :

- Lexique et pondérations : VADER utilise un lexique spécifique qui contient des mots et des expressions associés à des valeurs de valence (positives, négatives ou neutres). Chaque mot du lexique est également associé à une pondération qui reflète l'intensité du sentiment (par exemple, "excellent" pourrait avoir une pondération positive élevée, tandis que "horrible" aurait une pondération négative élevée).

— Prise en compte des intensificateurs et atténuateurs : VADER ajuste les scores de sentiment en fonction de certains mots qui intensifient ou atténuent l'émotion exprimée. Par exemple, des mots comme "très" ou "extrêmement" augmentent l'intensité d'un sentiment, tandis que des mots comme "un peu" ou "modérément" le diminuent.

— Gestion de la ponctuation et des majuscules : VADER prend également en compte des éléments non lexicaux comme la ponctuation. L'usage excessif de points d'exclamation ou de majuscules (ex : "FABULEUX!!!") augmente l'intensité émotionnelle du texte.

— Analyse des négations : VADER détecte les mots de négation comme "pas" ou "aucun", et ajuste le sentiment en conséquence. Par exemple, "pas génial" aura une valence moins positive que "génial".

— Calcul du score global : Une fois que chaque mot est analysé et pondéré en fonction de ces facteurs, VADER calcule un score global pour le texte. Ce score est divisé en trois catégories principales :

  — Positif : score indiquant le degré de sentiment positif.

  — Négatif : score indiquant le degré de sentiment négatif.

  — Neutre : score indiquant l'absence de sentiment marqué.

  VADER génère également un score de composé, qui résume le sentiment global du texte en une valeur comprise entre -1 (très négatif) et +1 (très positif).

2.3 Fondements de la psychologie

La psychologie est l'étude scientifique du comportement et des processus mentaux. Elle cherche à comprendre pourquoi nous pensons, ressentons et agissons comme nous le faisons, offrant des aperçus fascinants sur la complexité de l'expérience humaine.

2.3.1 Définition et importance de la psychologie

La psychologie est la science qui étudie les comportements, les processus mentaux, les émotions et les mécanismes sous-jacents à la pensée humaine. Elle cherche à comprendre comment les individus perçoivent le monde, comment ils réagissent face à des situations, et comment ces processus sont influencés par des facteurs internes (comme la biologie, la génétique, ou la cognition) et externes (comme l'environnement social et culturel).

Les psychologues s'intéressent à un large éventail de phénomènes, notamment les émotions, la perception, la mémoire, l'apprentissage, les motivations, la personnalité et la santé mentale. Ils étudient aussi les dysfonctionnements psychologiques, comme les troubles anxieux, les dépressions, les troubles de la personnalité, ou les psychoses.

La psychologie est une discipline multidimensionnelle et comprend plusieurs branches, dont les plus importantes sont :

— Psychologie clinique : Elle se concentre sur le diagnostic et le traitement des troubles mentaux, des dysfonctionnements comportementaux, et des problèmes émotionnels. Les psychologues cliniciens utilisent diverses approches thérapeutiques,

telles que la thérapie cognitivo-comportementale (TCC), la psychanalyse, ou la thérapie humaniste.

— Psychologie du développement : Cette branche étudie l'évolution des processus mentaux et comportementaux à travers les différentes étapes de la vie humaine, de la naissance à la vieillesse. Elle explore des sujets comme le développement de la langue, la cognition, les relations sociales, et l'adaptation au changement au fil du temps.

— Psychologie sociale : Elle analyse les effets des interactions sociales sur les comportements, les attitudes et les croyances des individus. Les psychologues sociaux examinent comment les individus sont influencés par leur environnement social, incluant des phénomènes comme l'obéissance, la conformité, la persécution, et l'agression.

— Psychologie cognitive : Cette branche se concentre sur les processus mentaux internes, tels que la perception, l'attention, la mémoire, le raisonnement et la prise de décision. Elle cherche à comprendre comment les individus traitent l'information, comment ils résolvent des problèmes, et comment les connaissances sont acquises et utilisées.

La psychologie est aussi un domaine appliqué, en ce sens qu'elle est utilisée pour améliorer la qualité de vie des individus à travers des interventions dans divers domaines, tels que l'éducation, la gestion des entreprises, la santé mentale, le sport, et la politique.

Les psychologues utilisent une large gamme de méthodes de recherche, incluant des observations, des expériences contrôlées, des entretiens, des questionnaires, et des tests psychométriques. Ces méthodes permettent de recueillir des données et de tester des hypothèses sur les comportements humains [26], [27].

2.3.2 Historique de la psychologie

La psychologie, en tant que discipline scientifique, a émergé à la croisée de la philosophie, de la biologie et de la médecine. Bien que les interrogations sur l'âme, la pensée ou les émotions remontent à l'Antiquité avec des penseurs tels que Platon ou Aristote, ce n'est qu'à la fin du XIXe siècle que la psychologie s'est structurée comme une science autonome [28].

Le tournant majeur survient en 1879, lorsque Wilhelm Wundt fonde le premier laboratoire de psychologie expérimentale à Leipzig. Ce moment marque la naissance officielle de la psychologie scientifique, centrée sur l'observation rigoureuse et la méthode expérimentale [29]. Par la suite, plusieurs courants théoriques se sont développés : le structuralisme, le fonctionnalisme, le béhaviorisme, la psychanalyse, la psychologie humaniste, puis les approches cognitives et neurobiologiques.

Au cours du XXe siècle, la psychologie s'est progressivement diversifiée, intégrant des méthodes quantitatives, des recherches empiriques, ainsi que des outils issus des sciences sociales et naturelles. L'apparition des neurosciences et, plus récemment, l'intégration de l'intelligence artificielle dans l'analyse des comportements humains témoignent de l'évolution continue de cette discipline [30].

2.3.3 Courants majeurs de la psychologie

Au fil de son développement, la psychologie a donné naissance à plusieurs courants théoriques qui reflètent différentes conceptions du fonctionnement mental et du comportement humain. Ces approches se sont souvent succédé ou opposées, tout en contribuant collectivement à l'enrichissement de la discipline.

Le premier courant structuré, le structuralisme, initié par Wilhelm Wundt et Edward Titchener, visait à analyser les composants de la conscience à travers l'introspection [29]. En réaction à cette approche introspective, le fonctionnalisme, influencé par William James, s'intéressait plutôt aux fonctions adaptatives de l'esprit et du comportement.

Au début du XXe siècle, le béhaviorisme, porté par John B. Watson puis B.F. Skinner, a recentré la psychologie sur l'observation des comportements mesurables, excluant les processus mentaux internes jugés non scientifiques [28]. Cette orientation a dominé la psychologie américaine pendant plusieurs décennies.

En parallèle, Sigmund Freud introduit la psychanalyse, qui met l'accent sur les conflits inconscients, les pulsions et les expériences précoces dans la formation de la personnalité [31]. Ce courant a fortement influencé la psychologie clinique.

À partir des années 1950, la psychologie humaniste, avec des figures comme Carl Rogers et Abraham Maslow, propose une vision plus positive de l'être humain, centrée sur l'épanouissement personnel et l'expérience subjective.

Enfin, la révolution cognitive marque un tournant majeur : elle réhabilite l'étude des processus mentaux tels que la mémoire, le langage, ou la perception, à l'aide de modèles inspirés des sciences de l'information [32].

Ces différents courants, bien que parfois divergents, ont tous contribué à forger une compréhension plus riche et nuancée de l'esprit humain.

2.3.4 Psychologie clinique et psychothérapie

La psychologie clinique est une branche de la psychologie qui se concentre sur l'évaluation, le diagnostic et le traitement des troubles mentaux, émotionnels et comportementaux. Elle repose sur l'observation directe, les entretiens cliniques et l'utilisation de tests psychométriques pour comprendre la souffrance psychique et accompagner les patients dans leur rétablissement [33].

Historiquement influencée par les travaux de Freud et les modèles psychanalytiques, la psychologie clinique s'est progressivement ouverte à des approches variées, telles que la thérapie cognitivo-comportementale (TCC), les thérapies humanistes ou systémiques. Chacune de ces approches propose une compréhension spécifique des mécanismes psychiques et des stratégies thérapeutiques adaptées [34].

La psychothérapie, au cœur de cette pratique clinique, désigne l'ensemble des méthodes utilisées pour aider une personne à surmonter ses difficultés psychologiques. Elle peut être individuelle, de groupe, de couple ou familiale, et s'adapte aux besoins spécifiques du patient. Selon l'orientation du praticien, l'intervention peut viser la modification des pensées dysfonctionnelles, la gestion des émotions, la reconstruction du récit personnel ou encore l'amélioration des habiletés sociales [35].

Aujourd'hui, la psychologie clinique s'appuie sur des données empiriques issues de la recherche pour valider ses interventions. Cette approche dite « fondée sur les preuves » (evidence-based) permet d'optimiser l'efficacité des traitements et d'assurer une prise en charge éthique et rigoureuse [36].

2.3.5 Les émotions en psychologie

Les émotions occupent une place centrale dans l'étude du comportement humain, car elles influencent nos décisions, nos relations sociales, ainsi que notre bien-être psychologique. En psychologie, une émotion est généralement définie comme un état affectif complexe, déclenché par un stimulus interne ou externe, et accompagné de modifications physiologiques, cognitives et comportementales [37].

De nombreuses théories ont été proposées pour expliquer la nature et la fonction des émotions. Les approches classiques, comme la théorie de James-Lange, soutiennent que l'émotion résulte de la perception des réactions physiologiques du corps face à un événement. À l'inverse, la théorie de Cannon-Bard postule que les émotions et les réponses corporelles surviennent simultanément [38], [39].

Les travaux de Paul Ekman ont identifié des émotions dites « universelles » — telles que la joie, la colère, la peur, le dégoût, la surprise et la tristesse — présentes dans toutes les cultures et associées à des expressions faciales spécifiques [37]. Ces recherches ont profondément influencé la compréhension contemporaine des émotions, notamment dans les domaines de la reconnaissance automatique des expressions ou du traitement affectif par les machines.

D'autres modèles contemporains, comme celui de Scherer ou de Lazarus, insistent sur l'évaluation cognitive de la situation par l'individu. Selon eux, l'émotion ne dépend pas seulement du stimulus, mais aussi de la manière dont il est interprété sur le plan personnel, social ou culturel [40].

Ainsi, la psychologie moderne considère les émotions comme des processus dynamiques et multidimensionnels, essentiels à la régulation du comportement, à la prise de décision et à la construction de la vie affective.

2.3.6 Problèmes psychologiques fréquents (anxiété, dépression, etc.)

Parmi les troubles psychologiques les plus répandus dans la population générale, l'anxiété et la dépression occupent une place prépondérante. Ces affections touchent des millions de personnes à travers le monde et constituent un enjeu majeur de santé publique [41].

Les troubles anxieux englobent un ensemble de conditions caractérisées par une inquiétude excessive, une tension constante et des réactions de peur inappropriées face à des situations perçues comme menaçantes. On distingue, entre autres, le trouble anxieux généralisé (TAG), les phobies spécifiques, le trouble panique ou encore le trouble obsessionnel-compulsif (TOC) [42].

La dépression, quant à elle, se manifeste par une humeur dépressive persistante, une perte d'intérêt pour les activités habituelles, une diminution de l'énergie, ainsi que des troubles du sommeil, de l'appétit et de la concentration. Le trouble dépressif majeur

(TDM) est la forme la plus étudiée et peut avoir des conséquences graves sur la qualité de vie et le fonctionnement quotidien [43].

Ces deux catégories de troubles partagent plusieurs symptômes et peuvent coexister chez un même individu, rendant leur prise en charge plus complexe. Les approches thérapeutiques incluent des interventions psychologiques telles que la thérapie cognitivo-comportementale, ainsi que des traitements pharmacologiques (antidépresseurs, anxiolytiques) lorsque nécessaire [44].

La précocité du diagnostic et l'accès à un soutien psychologique adapté sont des facteurs essentiels pour améliorer le pronostic de ces troubles. Malgré cela, de nombreuses personnes ne reçoivent pas de soins adéquats, notamment en raison de la stigmatisation ou du manque de ressources disponibles [45].

2.3.7 Barrières d'accès à la thérapie (coût, distance, stigmatisation)

Malgré l'augmentation de la sensibilisation aux questions de santé mentale, de nombreux individus confrontés à des troubles psychologiques n'ont pas accès à des services thérapeutiques adéquats. Plusieurs obstacles structurels, économiques et socioculturels limitent l'accès aux soins, aggravant ainsi la souffrance psychique non prise en charge [45].

Le coût élevé des consultations psychologiques ou psychiatriques constitue l'un des freins majeurs, en particulier dans les pays où la couverture santé est partielle ou inexistante. Les séances de psychothérapie, souvent non remboursées, représentent une charge financière importante pour les personnes à faibles revenus.

La distance géographique est un autre facteur déterminant. Dans de nombreuses zones rurales ou mal desservies, l'offre en professionnels de santé mentale est faible, voire absente. Les personnes concernées doivent parfois parcourir de longues distances pour consulter, ce qui décourage la démarche thérapeutique, notamment pour des populations vulnérables ou à mobilité réduite [46].

À ces contraintes pratiques s'ajoute la stigmatisation persistante liée aux troubles psychologiques. Dans certaines cultures ou contextes sociaux, consulter un psychologue est perçu comme un signe de faiblesse ou de folie, ce qui pousse les individus à dissimuler leur détresse et à éviter toute forme d'aide formelle [47]. Cette stigmatisation peut être intériorisée, réduisant la propension des personnes à rechercher un accompagnement, même en présence de symptômes sévères.

Ces barrières soulignent l'importance de promouvoir des alternatives d'accès, telles que la télépsychologie, les interventions numériques ou les campagnes de sensibilisation, afin de rendre la thérapie plus inclusive, accessible et socialement acceptée.

2.4 La convergence entre l'intelligence artificielle et la psychologie

2.4.1 Applications NLP pour la santé mentale

Le traitement automatique du langage naturel (NLP) connaît une adoption croissante dans le domaine de la santé mentale, offrant des outils capables d'analyser des textes ou discours pour détecter des signaux de détresse psychologique. L'analyse linguistique

permet notamment d'identifier des signes précoces de troubles comme la dépression, l'anxiété ou le stress post-traumatique à partir de journaux personnels, de publications sur les réseaux sociaux ou d'échanges verbaux [5].

Ces applications se fondent sur l'hypothèse que la manière dont une personne utilise le langage reflète son état émotionnel. Par exemple, une fréquence accrue de mots à connotation négative, un usage excessif de la première personne du singulier ou une focalisation sur des thèmes sombres peuvent être indicateurs d'un état dépressif [48].

Le NLP est également utilisé pour le suivi de la thérapie, en analysant l'évolution du discours des patients au fil des séances. Cela permet aux thérapeutes de mieux comprendre l'impact des interventions et de personnaliser les traitements [49].

2.4.2 Utilisation de modèles NLP pour les chatbots psychologiques

Les avancées en intelligence artificielle ont permis le développement de chatbots thérapeutiques, capables de simuler des conversations empathiques avec des utilisateurs confrontés à des difficultés psychologiques. Ces agents virtuels reposent sur des modèles NLP avancés pour comprendre le langage humain, adapter leur réponse et guider l'utilisateur dans un processus d'auto-assistance [50].

Ils peuvent offrir un soutien émotionnel, des techniques de gestion du stress, ou orienter vers des ressources adaptées. L'avantage de ces chatbots réside dans leur accessibilité permanente, leur anonymat, et leur faible coût, ce qui les rend particulièrement utiles dans les contextes à ressources limitées.

Exemples de chatbots psychologiques : Woebot, Wysa, Replika

Dans cette partie, nous explorons diverses applications concrètes des chatbots dans le domaine de la psychologie, offrant un aperçu de leurs fonctionnalités et de leur potentiel. Parmi les exemples les plus notables, Woebot utilise la thérapie cognitivo-comportementale (TCC) pour proposer un soutien quotidien basé sur l'analyse des émotions [50]. Wysa adopte une approche plus personnalisée, intégrant la pleine conscience et des exercices psychologiques variés. Replika, quant à lui, se concentre davantage sur la conversation empathique et la compagnie émotionnelle. La table 2.1 résume les principales caractéristiques de ces chatbots.

Bien que prometteurs, ces outils ne remplacent pas une thérapie professionnelle, mais peuvent agir comme compléments, notamment dans les phases de prévention ou de suivi.

2.4.3 Détection d'émotions et d'humeurs avec le NLP

La détection des émotions à partir du langage écrit ou parlé représente un domaine clé du NLP affectif. Les modèles modernes sont capables d'identifier l'émotion sous-jacente à une phrase, comme la tristesse, la colère ou la joie, à partir d'indices lexicaux, syntaxiques et contextuels [51].

Cette capacité est essentielle pour des applications en santé mentale, car elle permet de suivre les fluctuations émotionnelles de l'utilisateur, de repérer les tendances négatives persistantes, et d'alerter les professionnels en cas de risque accru.

Critère	Woebot	Wysa	Replika	PsyBot
Thérapie de base	TCC	TCC + écoute active	Compagnon émotionnel	...
Détection des émotions	Non	Partielle	Générale	...
Génération des réponses	Scriptées	Scriptées	Semi-personnalisé	...
Multilingue	Non	Partiel	Non	...
Adaptation culturelle	Non	Non	Non	...
Filtrage émotionnel	Non	Non	Non	...
Compréhension d'intention	Non	Non	Non	...
Confidentialité / vie privée	Moyenne	Moyenne	Faible	...
Accompagnement par exercices	Non	**Oui (relaxation, respiration)**	Non	...
Relation affective long terme	Non	Non	**Oui (mémoire émotionnelle)**	...

Table 2.1 – Comparaison fonctionnelle entre les chatbots psychologues existants

2.4.4 Classification des émotions avec BERT

Les modèles de type BERT (Bidirectional Encoder Representations from Transformers) ont considérablement amélioré la performance des systèmes de classification émotionnelle. Grâce à leur capacité à prendre en compte le contexte bilatéral des mots, ils offrent une compréhension fine du contenu émotionnel des textes [21].

Dans le cadre de la santé mentale, ces modèles permettent de construire des systèmes de détection automatique de détresse émotionnelle à partir de journaux intimes, d'e-mails, ou de messages sur les réseaux sociaux, facilitant ainsi l'identification de patients à risque [52].

2.4.5 Limites éthiques et défis d'utilisation de l'IA en santé mentale

Bien que les technologies basées sur l'IA offrent des perspectives prometteuses pour la santé mentale, leur utilisation soulève d'importantes questions éthiques. Le respect de la vie privée, la protection des données sensibles et le risque de dépendance aux outils numériques figurent parmi les principales préoccupations [53].

En outre, le manque de transparence des modèles d'IA, notamment les réseaux neu-

ronaux profonds, rend difficile l'explication des décisions prises, ce qui pose problème en contexte clinique. Il est donc crucial de développer des systèmes responsables, auditables et encadrés juridiquement pour garantir leur usage éthique [54].

2.5 Conclusion

Ce chapitre a permis de poser les bases théoriques essentielles à la compréhension du projet PsyBot. D'une part, nous avons exploré les principes fondamentaux de l'intelligence artificielle et du traitement automatique du langage naturel (NLP), en insistant sur les architectures modernes comme BERT et GPT, largement utilisées dans les systèmes conversationnels.

D'autre part, nous avons présenté les fondements psychologiques nécessaires à la conception d'un chatbot adapté au soutien mental : définitions clés, approches thérapeutiques, et émotions humaines. Cette double approche a montré que la convergence entre IA et psychologie ouvre des perspectives prometteuses pour créer des outils numériques d'accompagnement émotionnel, à condition de prendre en compte les limites éthiques, culturelles et technologiques.

PsyBot s'inscrit dans cette vision hybride, en intégrant des modèles NLP capables d'analyser les émotions et les intentions, tout en respectant les principes de bienveillance, de confidentialité, et d'adaptation au contexte algérien.

3

Description de l'Architecture du Système PsyBot

Sommaire

3.1 Introduction 26
3.2 Modèle proposé 26
3.3 Modèles et algorithmes 27
3.3.1 DistilBERT pour la classification des intentions 28
3.3.2 MiniBERT et VADER pour la classification des émotions 28
3.3.3 GPT pour la génération de réponses 28
3.4 Modélisation de l'architecture du PsyBot 29
3.4.1 Fonctionement globale 30
3.4.2 Pseudo-algorithme 31
3.5 Conclusion 31

3.1 Introduction

Ce chapitre est consacré à la description de l'architecture du système PsyBot, conçu pour analyser les intentions et les émotions des utilisateurs dans un contexte conversationnel. L'objectif est de fournir une vision claire et structurée des différentes composantes qui permettent à PsyBot de traiter les requêtes en langage naturel de manière intelligente et adaptée.

L'architecture du système repose sur une combinaison de modules spécialisés, chacun jouant un rôle bien défini dans la chaîne de traitement de l'information : classification de l'intention, détection de l'émotion, génération de réponses et, si nécessaire, une recherche professionnel. Ces modules interagissent selon une logique de flux de données pensée pour maximiser la pertinence des réponses tout en garantissant une adaptation au contexte émotionnel de l'utilisateur.

La Figure 3.1 présente une vue d'ensemble de cette architecture modulaire, où les modèles préentraînés comme DistilBERT, MiniBERT et GPT sont intégrés de manière cohérente pour assurer un fonctionnement fluide et réactif du système. Cette organisation permet non seulement d'optimiser les performances du système, mais aussi de favoriser une évolutivité facilitée pour d'éventuelles améliorations futures.

3.2 Modèle proposé

Le système PsyBot est un chatbot intelligent conçu pour simuler un échange conversationnel à visée psychologique. Son objectif principal est de comprendre l'état émotionnel et les intentions des utilisateurs afin de générer des réponses appropriées, cohérentes et empathiques. Pour atteindre cet objectif, le modèle proposé repose sur une architecture modulaire combinant plusieurs composants de traitement du langage naturel. Le modèle global est composé de trois modules principaux :

1. Module de classification des intentions : Ce module permet d'identifier le type de requête formulée par l'utilisateur (par exemple : demande de conseil, expression d'un mal-être, simple salutation, etc.). Il utilise le modèle DistilBERT, un transformer optimisé pour la classification de texte, capable de traiter des entrées brèves et informelles avec efficacité.
2. Module de détection des émotions : Ce composant analyse le contenu émotionnel du message. Il s'appuie sur une double approche combinant BERT, pour la classification supervisée d'émotions spécifiques (comme la tristesse, la colère, ou la joie), et VADER, pour une estimation rapide de la polarité (positive, négative, neutre) du texte.
3. Module de génération de texte : Ce module est responsable de la réponse générée par PsyBot. Il intègre les sorties des deux premiers modules (intention + émotion) pour produire une réponse textuelle contextuelle. Ce module repose sur un modèle de type GPT, pré-entraîné puis ajusté pour produire des réponses adaptées au ton de la conversation.

L'originalité du modèle réside dans la synergie entre la classification des intentions,

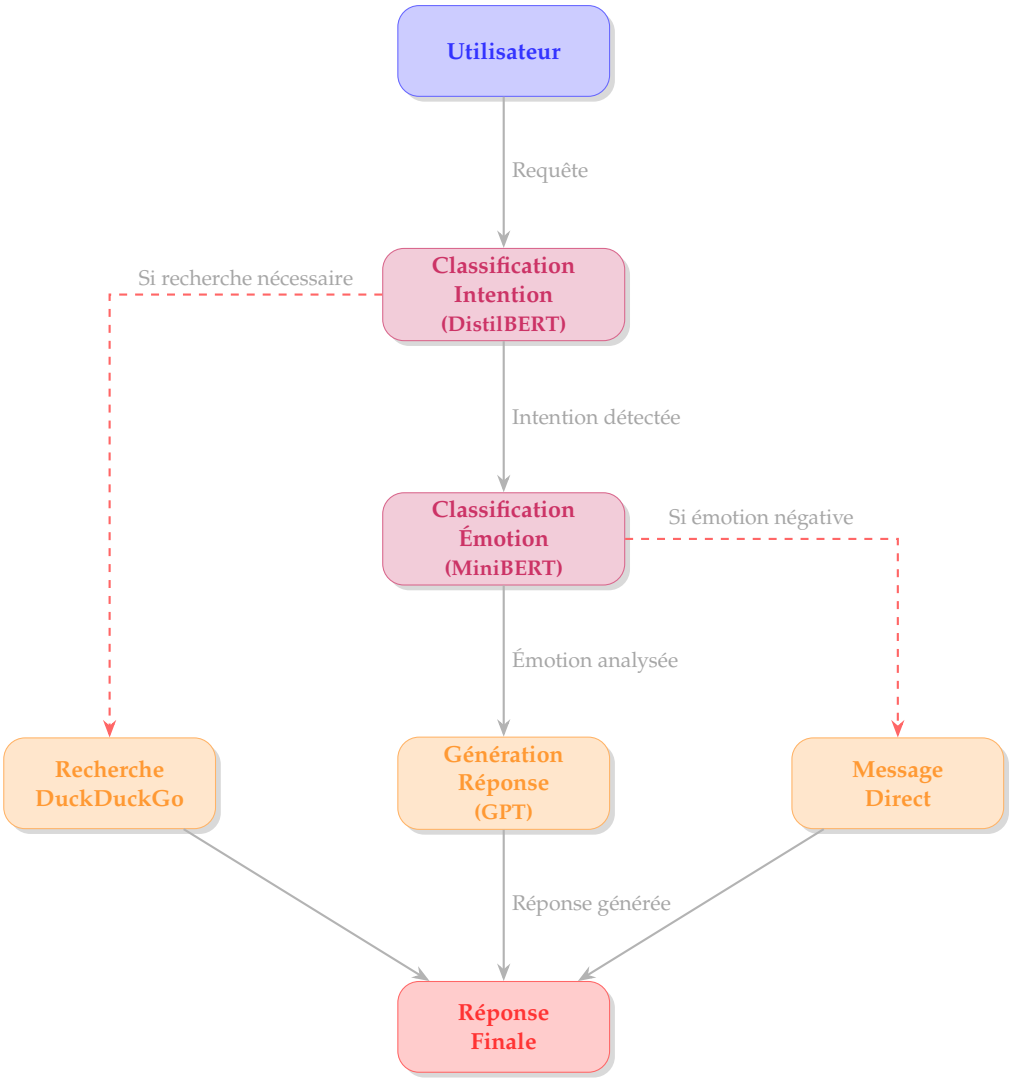


Figure 3.1 – Architecture Globale du PsyBot

l'analyse émotionnelle et la génération de texte, permettant au chatbot de proposer un échange naturel, sensible et personnalisé. Cette approche modulaire assure également une certaine flexibilité : chaque composant peut être amélioré ou remplacé indépendamment, sans modifier l'ensemble de l'architecture.

3.3 Modèles et algorithmes

Dans cette section, nous présentons les différents modèles d'intelligence artificielle utilisés dans la conception de PsyBot. Chaque module repose sur un algorithme ou une combinaison d'algorithmes spécialisés dans une tâche précise : la compréhension de l'intention, la détection des émotions et la génération de réponses naturelles. Le choix de ces modèles repose sur un compromis entre performance, coût computationnel et disponibilité des ressources.

3.3.1 DistilBERT pour la classification des intentions

DistilBERT est une version allégée et optimisée de BERT(Bidirectional Encoder Representations from Transformers), conçue pour offrir des performances comparables à celles de BERT tout en étant plus rapide et moins coûteuse en ressources. Il conserve 97% de la performance de BERT tout en réduisant de 40% le nombre de paramètres.

Dans le cadre du projet PsyBot, DistilBERT est utilisé pour la classification des intentions des utilisateurs, c'est-à-dire pour identifier automatiquement l'objectif ou le besoin exprimé dans une requête. Cela permet au système de comprendre le contexte global d'une interaction, qu'il s'agisse d'une demande de conseil, d'un besoin d'écoute, ou d'une question personnel ou émotionnel. Le modèle est pré-entraîné sur un corpus généraliste, puis affiné (fine-tuning) sur un jeu de données spécifique à la détection d'intentions psychologiques.

3.3.2 MiniBERT et VADER pour la classification des émotions

MiniBERT est une version simplifiée de BERT, conçue pour être plus légère et plus rapide, tout en conservant une bonne capacité de compréhension contextuelle. Cette compacité en fait un candidat idéal pour des environnements à ressources limitées ou des applications en temps réel, comme un chatbot. Dans PsyBot, MiniBERT est utilisé pour la classification supervisée des émotions, en se basant sur un jeu de données étiqueté avec des émotions comme la joie, la tristesse, la peur, ou la colère. Il permet une interprétation fine des intentions émotionnelles dans des messages souvent courts et ambigus.

VADER, de son côté, est utilisé pour compléter cette analyse par une évaluation plus globale du ton émotionnel (positif, négatif, neutre). Son approche lexicon-based, combinée à des règles grammaticales, est particulièrement bien adaptée aux textes issus de la communication en ligne, comme les messages de chat.

L'intégration de MiniBERT et VADER permet donc d'obtenir une classification émotionnelle à la fois fine et robuste. MiniBERT apporte la précision contextuelle grâce à l'apprentissage profond, tandis que VADER offre une validation rapide et explicite à travers des scores de polarité. Cette combinaison renforce la capacité de PsyBot à détecter l'état émotionnel de l'utilisateur de manière fiable, ce qui est essentiel dans une application à visée psychologique.

3.3.3 GPT pour la génération de réponses

La génération de texte constitue une étape cruciale dans le fonctionnement du PsyBot, puisqu'elle permet de produire des réponses naturelles, cohérentes et adaptées au contexte émotionnel et intentionnel de l'utilisateur. Pour cela, le modèle GPT (Generative Pre-trained Transformer) est utilisé.

GPT est un modèle autoregressif de génération de langage naturel basé sur l'architecture Transformer. Il a été pré-entraîné sur une vaste quantité de données textuelles et peut être affiné pour s'adapter à des domaines spécifiques. Dans le cadre de PsyBot, GPT est utilisé pour générer des réponses empathiques, pertinentes et contextualisées, en s'appuyant sur les sorties des modules de détection d'intention et d'émotion.

Le modèle est configuré pour intégrer à la fois :

L'intention détectée (ex. : recherche de soutien, demande de conseil)

L'émotion identifiée (ex. : anxiété, tristesse)

Le contenu des messages précédents

Cela permet à PsyBot de produire des réponses non seulement correctes sur le plan linguistique, mais également sensibles à l'état émotionnel de l'utilisateur.

3.4 Modélisation de l'architecture du PsyBot

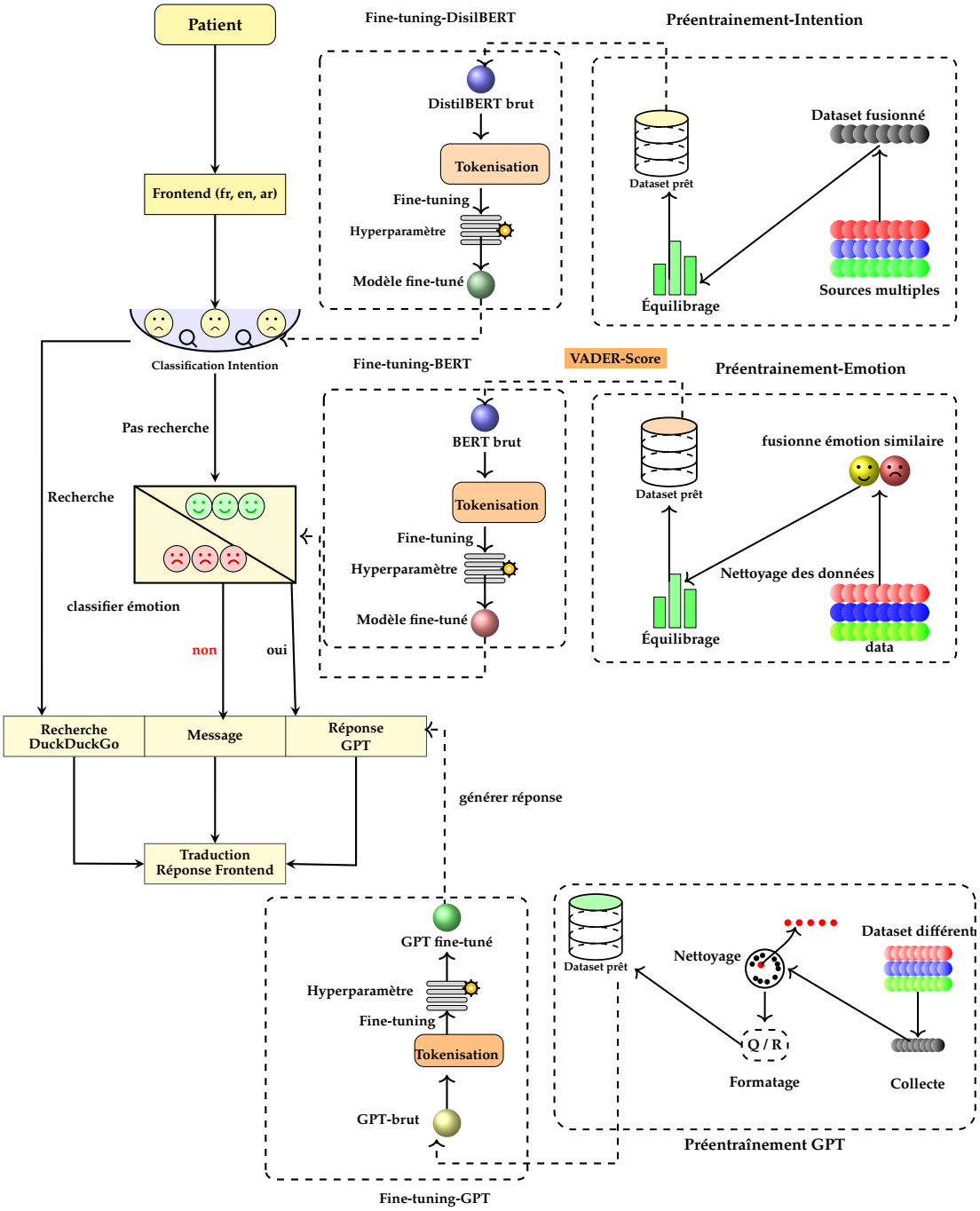


Figure 3.2 – Architecture Détaillée du PsyBot

3.4.1 Fonctionement globale

La Figure 3.2 représente l'architecture détaillée de notre PsyBot. Elle est composée de trois parties principales: la détection de l'intention, la classification émotionnelle, puis la génération de la réponse.

Avant toute analyse, la langue de la requête est détectée automatiquement afin de garantir que les données textuelles soient traitées dans une langue compatible avec les modèles préentraînés. Si la langue détectée est différente de l'anglais, une traduction automatique est effectuée. Cela permet d'assurer la cohérence des entrées et d'améliorer la performance globale du système.

Détection de l'intention:

La première étape consiste à analyser la requête de l'utilisateur afin de déterminer son intention. Deux cas sont distingués : soit l'utilisateur cherche une information générale (intention de recherche), soit il s'adresse directement au psybot pour une interaction d'ordre psychologique. Pour accomplir cette tâche, nous avons utiliséle modèle (DistilBERT fine-tuné) pour une classification binaire en deux classes : Recherche et Classification émotions. Le modèle apprend ainsi à différencier les requêtes techniques (ex: "Comment traiter l'anxiété ?") de celles exprimant une émotion ou une situation personnelle (ex: "Je me sens constamment stressé").

Si la requête est classée comme une intention de recherche, elle est automatiquement redirigée vers une bibliothèque intégrée, DuckDuckGo, qui permet de répondre avec des résultats web pertinents, sans nécessiter de clé API. Dans le cas contraire, la requête est transmise au module suivant pour une analyse émotionnelle.

Classification émotionnelle:

La deuxième étape repose sur une analyse émotionnelle de la requête, permettant de déterminer l'état affectif exprimé par l'utilisateur. Cette classification est réalisée à l'aide d'un modèle MiniBERT, lui aussi affiné sur un corpus annoté d'émotions, pour assurer une reconnaissance fine des sentiments. Afin de renforcer la fiabilité de cette étape, une double validation est effectuée en intégrant l'outil VADER, qui attribue un score dit compound mesurant l'intensité de la polarité émotionnelle de la phrase.

Une règle empirique est alors appliquée : si l'émotion détectée est de type anger (colère) et que le score compound est inférieur à -50 (cette valeur étant obtenue après avoir multiplié le score original par 100), la requête est considérée comme révélant une colère excessive ou anormale. Dans ce cas, le système interrompt le dialogue en envoyant un message d'avertissement à l'utilisateur, l'invitant à se reconnecter plus tard, dans un état émotionnel plus stable. Cette classification permet de préserver l'éthique du système en évitant d'aggraver une situation émotionnelle instable par une réponse automatique non adaptée.

Génération de réponse:

Lorsque la requête est jugée émotionnellement stable et d'ordre psychologique, elle est alors transmise au troisième et dernier module du PsyBot : la génération de réponse.

Ce module repose sur un modèle de langage de type GPT (Generative Pretrained Transformer), qui a été affiné (fine-tuning) sur un corpus de dialogues psychologiques soigneusement prétraité, nettoyé et structuré selon des critères qualitatifs.

L'objectif est de produire une réponse cohérente, empathique et adaptée à la situation émotionnelle de l'utilisateur. Le modèle génère ainsi des phrases qui imitent le langage d'un conseiller ou thérapeute, tout en respectant une neutralité émotionnelle et une bienveillance constante. L'intégration séquentielle de ces trois modules (intention, émotion, génération) permet d'assurer un traitement intelligent et humanisé des requêtes utilisateurs, conférant au PsyBot une capacité d'adaptation contextuelle avancée.

3.4.2 Pseudo-algorithme

L'algorithme 1 décrit le fonctionnement de notre chatbot conçu pour interagir avec les utilisateurs. Lorsque le bot reçoit une requête, sa première étape est de classifier l'intention de cette requête. Si l'intention est de type "Recherche", PsyBot utilise DuckDuckGo pour trouver et renvoyer le résultat pertinent à l'utilisateur. En revanche, si la requête n'est pas une recherche, le bot procède à une analyse émotionnelle. Il identifie l'émotion principale présente dans la requête et utilise l'outil VADER pour une analyse de sentiment plus approfondie, obtenant un score "compound". Si l'émotion détectée est la colère et que le score "compound" est inférieur à -50 (indiquant une intensité négative forte), PsyBot renvoie un message prédéfini concernant l'état émotionnel de l'utilisateur. Dans tous les autres cas, le bot génère une réponse en utilisant un modèle comme GPT et la retourne à l'utilisateur.

Algorithm 1 Pseudo-Algorithme de PsyBot

```
1:  function PsyBot(requete)
2:      intention ← classifierIntentions(requete)
3:      if intention = Recherche then
4:          resultat ← rechercherAvecDuckDuckGo(requete)
5:          return resultat
6:      else
7:          emotion ← classifierEmotion(requete)
8:          compound ← analyserAvecVADER(requete)
9:          if emotion = colere and compound < -50 then
10:             return "Votre état émotionnel est instable..."
11:         else
12:             reponse ← genererAvecGPT(requete)
13:             return reponse
14:         end if
15:     end if
16: end function
```

3.5 Conclusion

En résumé, l'architecture de PsyBot repose sur une intégration efficace de modules de classification et de génération, permettant une compréhension fine des intentions et émotions des utilisateurs. Cette structure modulaire garantit à la fois flexibilité,

performance et adaptabilité pour des interactions plus naturelles et pertinentes.

4

Expérimentations et interprétation des résultats

Sommaire

4.1 Introduction . . . 34
4.2 Classification des intetions DistilBERT (Phase 1) . . . 34
4.2.1 Présentation du dataset Utilisé . . . 34
4.2.2 Prétraitement des données . . . 35
4.2.3 Entraînement du modèle . . . 35
4.2.4 Résultats . . . 36
4.2.5 Évaluation et test . . . 37
4.2.6 Comparaison avec l'approche baseline . . . 38
4.3 Classification des emotions BERT et VADER (Phase 2) . . . 38
4.3.1 Présentation du dataset utilisé . . . 38
4.3.2 Prétraitement des données: . . . 39
4.3.3 Entraînement du modèle . . . 40
4.3.4 Résultats . . . 42
4.3.5 Évaluation et test . . . 43
4.3.6 Comparaison avec les autre travaux . . . 45
4.4 Génération de texte GPT (Phase 3) . . . 46
4.4.1 Présentation du dataset Utilisé . . . 46
4.4.2 Prétraitement des données . . . 47
4.4.3 Entraînement du modèle . . . 48
4.4.4 Résultats . . . 49
4.4.5 Évaluation et test . . . 50
4.4.6 Comparaison avec les autre travaux . . . 51
4.5 Conclusion . . . 52

4.1 Introduction

Dans ce chapitre, nous présentons la phase expérimentale de notre travail, qui vise à démontrer l'efficacité de l'approche que nous avons proposée pour la mise en œuvre du système PsyBot. Cette phase est essentielle pour évaluer la performance des différents modèles sélectionnés à chaque étape clé du traitement: classification de l'intention, détection de l'émotion et génération de texte.

La première partie est consacrée à la tâche de classification des intentions. Elle a pour but de déterminer si une requête utilisateur relève d'une simple recherche d'information ou d'une intention de dialogue psychologique. Pour cela, nous avons fine-tuné le modèle DistilBERT sur un dataset annoté, en appliquant un prétraitement rigoureux et en mettant en place des métriques d'évaluation précises.

Dans la deuxième partie, nous détaillons la classification émotionnelle en combinant deux approches: un modèle MiniBERT pour la reconnaissance des émotions textuelles, et VADER pour mesurer l'intensité affective à travers un score de polarité. Cette étape est cruciale pour filtrer les situations à forte charge émotionnelle négative.

La troisième partie concerne la génération de texte via un modèle de type GPT, entraîné pour produire des réponses adaptées au contexte émotionnel et psychologique de l'utilisateur. Cette génération repose sur un corpus structuré de dialogues, permettant de simuler l'intervention d'un conseiller empathique.

Chaque section expérimentale est structurée autour de la présentation du dataset utilisé, du prétraitement appliqué, de la configuration du modèle, de l'entraînement, de l'évaluation, des résultats obtenus, et d'une comparaison avec les approches existantes. Cette organisation nous permet de valider de manière rigoureuse les performances et les limites de chaque composant du système PsyBot.

4.2 Classification des intetions DistilBERT (Phase 1)

4.2.1 Présentation du dataset Utilisé

Il existe très peu de jeux de données psychologiques publiquement disponibles et adaptés au contexte algérien ou maghrébin, notamment en ce qui concerne les langues utilisées (arabe, français). C'est pourquoi nous avons eu recours à la solution suivante:

- Utilisation de jeux de données internationaux open-source comme GoEmotions (Google) et ESConv, qui comportent des annotations émotionnelles fiables.
- Création d'un dataset personnalisé pour la classification d'intention (Recherche vs GPT), annoté manuellement à partir d'exemples collectés via des forums et des discussions simulées, avec en plus la fusion de certaines parties de jeux de données existants (qui seront détaillés dans la section Génération avec le modèle GPT).
- Ce dataset personnalisé a été construit à partir de plusieurs sources:
  - Une partie du dataset GoEmotions, annotée comme appartenant à la classe "GPT".
  - Une sélection du dataset Amod/mental_health_counseling_conversations,

utilisée pour enrichir davantage la classe "GPT" avec des questions à caractère personnel.

— Un bloc de questions réelles extraites du web (source: blooq), annotées comme appartenant à la classe "Recherche".

— Des ajouts manuels de phrases synthétiques afin d'équilibrer les deux classes (GPT et Recherche).

- Un effort de data augmentation a également été réalisé pour introduire des variantes linguistiques et améliorer la robustesse du modèle.

Pour le problème lié à la langue (arabe et français), nous avons opté pour une solution basée sur la traduction, en suivant les étapes suivantes:

1. Dès l'entrée, une détection de langue est effectuée grâce à la bibliothèque langdetect, qui identifie la langue de la requête.
2. Ensuite, une traduction vers l'anglais est réalisée pour permettre un traitement en anglais, comme on vas l'expliqué dans la suite du chapitre.
3. Enfin, après traitement, une traduction inverse (de l'anglais vers la langue détectée initialement) est effectuée afin d'afficher le résultat à l'utilisateur dans la langue appropriée.

4.2.2 Prétraitement des données

La Figure 4.1 illustre la distribution des exemples étiquetés pour la classification d'intention par DistilBERT. On y constate que la classe GPT compte environ 5 772 exemples, tandis que la classe Recherche en compte 5 002, soit un déséquilibre modéré de l'ordre de 53% 47%. Cette répartition relativement proche des effectifs est globalement satisfaisante pour l'apprentissage d'un classifieur robuste.

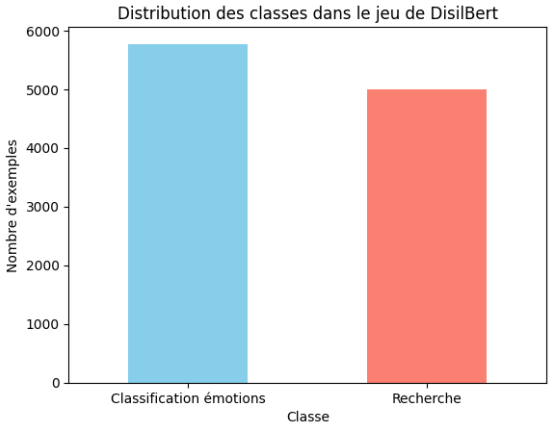


Figure 4.1 – Distribution des classes dans le jeu de DistilBERT

4.2.3 Entraînement du modèle

Chargement du modèle et configuration: Le modèle a été configuré pour une tâche de classification binaire, avec deux classes de sortie (gpt et recherche). Bien qu'il repose sur l'architecture de BERT, sa taille réduite permet un déploiement plus rapide, notamment sur des machines disposant de ressources limitées.

Préparation des ensembles d'apprentissage: Les textes ont été tokenisés avec (BertTokenizer), avec (padding) et (truncation) à une longueur maximale de 256 tokens. Les ensembles d'apprentissage ont ensuite été encapsulés sous forme de Dataset PyTorch personnalisé, puis chargés via des (DataLoader) avec un (batch_size) de 16.

Hyperparamètres et stratégie d'optimisation: En utilisant un taux d'apprentissage initial de (1e-6) et une pénalisation L2 (weight_decay) de (0.05). Une stratégie de planification du taux d'apprentissage par (cosine scheduling avec warmup) a été adoptée, avec 10% des étapes totales consacrées au warmup. Cette approche permet une montée progressive du taux d'apprentissage, suivie d'une décroissance douce, favorisant une meilleure stabilité durant l'entraînement.

Entraînement et stratégie d'arrêt: L'entraînement a été effectué sur un maximum de 3 époques. À chaque époque, le modèle est évalué sur l'ensemble de validation.

4.2.4 Résultats

Les métriques de perte (loss) et de précision (accuracy) sont visualisées pour les ensembles d'entraînement et de validation tout au long des époques d'apprentissage. Le modèle affichant la meilleure performance, c'est-à-dire la perte la plus faible sur l'ensemble de validation, est sauvegardé. Il sera ensuite rechargé pour être utilisé lors de la phase de test finale.

Interprétation des Résultats: Les courbes de précision et de perte peuvent être interprétées comme des indicateurs clés de la progression de l'apprentissage du modèle, révélant si ce dernier généralise efficacement ou s'il rencontre des difficultés comme le surapprentissage.

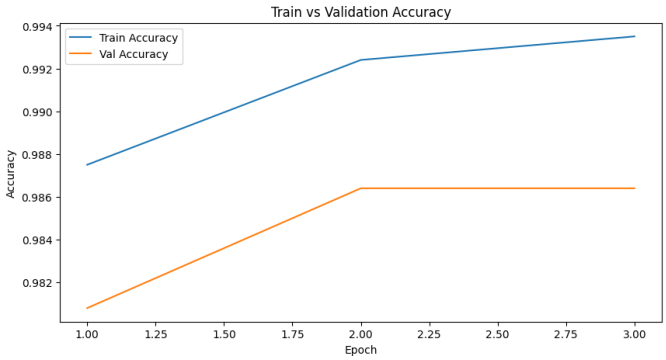
1. Accuracy (Train et Validation): Le premier graphique (Figure 4.2a) présente l'évolution de l'accuracy (exactitude) sur les données d'entraînement et de validation au fil des époques. On observe une amélioration progressive de la précision pour les deux courbes:

   — L'accuracy d'entraînement augmente de manière constante, atteignant environ 0.993 à la fin de la troisième époque.

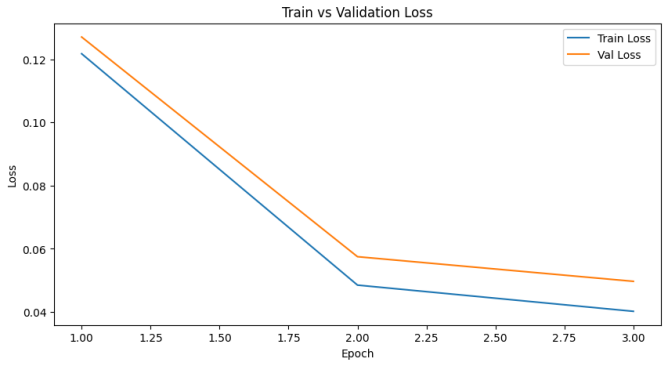
   — L'accuracy de validation s'améliore également entre la première et la deuxième époque, puis se stabilise autour de 0.986, suggérant une performance généralisable mais avec un léger écart vis-à-vis de l'entraînement.

   Cette stabilité après l'époque 2 peut indiquer que le modèle commence à atteindre une performance optimale sur les données de validation, sans pour autant surapprendre.

2. Loss (Train et Validation) Le deuxième graphique (Figure 4.2b) montre la diminution de la perte (loss) au cours des époques. Les deux courbes suivent une tendance décroissante, ce qui est un bon indicateur de convergence:



(a) Évolution de l'accuracy



(b) Diminution de la perte

Figure 4.2 – Courbes de performance du modèle DistilBERT

— La loss d'entraînement diminue de manière régulière, passant d'environ 0.12 à 0.04, indiquant que le modèle apprend efficacement.

— La loss de validation baisse également jusqu'à l'époque 2, puis décroît lentement, suggérant que le modèle généralise correctement sans surajustement significatif.

L'écart modéré entre les courbes de perte d'entraînement et de validation reste maîtrisé, ce qui témoigne d'un bon équilibre entre biais et variance.

3. Conclusion: Le modèle DistilBERT semble converger correctement avec des performances élevées et une stabilité dès la deuxième époque. Aucun signe flagrant d'overfitting n'est observé, ce qui valide la qualité du fine-tuning réalisé.

4.2.5 Évaluation et test

Interprétation de la matrice de confusion: La matrice de confusion (voir Figure 4.3) présente les performances du modèle sur la tâche de classification binaire entre deux catégories: émotions et Recherche. Les résultats montrent une très bonne séparation entre les deux classes.

Le modèle a correctement prédit 854 cas sur 866 pour la classe émotions, et 744 cas sur 751 pour la classe Recherche. Les erreurs sont limitées à seulement 19 échantillons au total, réparties de manière équilibrée (12 faux positifs et 7 faux négatifs). Ces chiffres témoignent d'une précision et d'un rappel élevés pour les deux catégories, ce qui confirme l'efficacité du modèle dans la différenciation entre les intentions émotionnelles et les

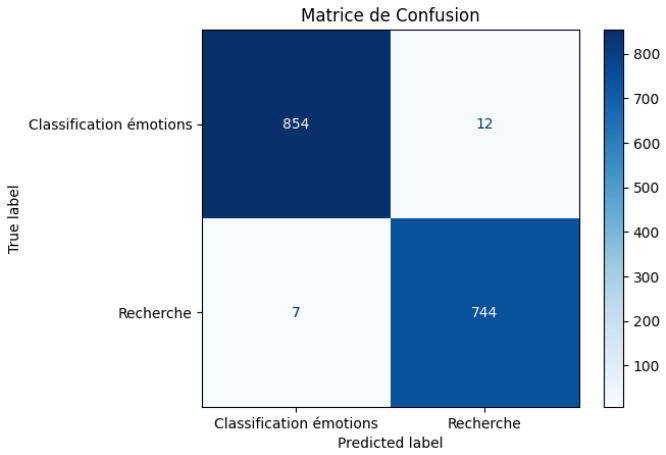


Figure 4.3 – Matrice de confusion apres le fine-tuning

requêtes de recherche. Pour illustrer davantage ces performances, nous vous présentons des exemples de classification pour chacune des deux classes dans le Tableau 4.2.

Interprétation: Le modèle atteint une excellente performance globale avec une précision, un rappel et un F1-score de 0.99 pour les deux classes. Cela indique une classification très fiable et équilibrée entre ”Classification émotions” et ”Recherche”. Le résultat est présenté dans la tableau suivant (Tableau 4.1):

Type	Precision	Recall	F1-score	Support
Classification émotions	0.99	0.99	0.99	866
Recherche	0.98	0.99	0.99	751

Table 4.1 – Résultats des metriques pour la classification d’intentions

4.2.6 Comparaison avec l’approche baseline

Avant le fine-tuning, le modèle DistilBERT appliqué sur le jeu de test affichait des performances très faibles, avec une accuracy de seulement 47%, une précision de 8%, un rappel de 1,3%, et un F1-score de 2,3%. La matrice de confusion montre une mauvaise capacité de discrimination entre les deux classes: le modèle classait presque toutes les requêtes comme GPT , avec 755 requêtes GPT correctement identifiées mais aussi 741 requêtes Recherche mal classées, ce qui démontre un fort déséquilibre de la prédiction (voir la figure 4.4).

4.3 Classification des emotions BERT et VADER (Phase 2)

4.3.1 Présentation du dataset utilisé

Le dataset GoEmotions, développé par Google Research, est une ressource riche pour la classification des émotions dans le langage naturel. Il contient plus de 54 000 phrases extraites de Reddit, annotées manuellement avec 27 émotions distinctes, ainsi qu’une

Requête	Prédiction
I'm scared if being left alone	Émotions
No one cares about me	Émotions
I keep failing at everything I do	Émotions
Life feels so heavy lately	Émotions
What is the role of serotonin in depression?	Recherche
What is impostor syndrome?	Recherche
What is social anxiety?	Recherche
How does childhood trauma affect adults?	Recherche
What are grounding techniques	Recherche
I cry almost every day	Émotions
I can't talk to anyone about how I feel	Émotions
Can meditation reduce stress?	Recherche
I feel broken inside	Émotions
Why is emotional regulation important?	Recherche
Are panic attacks dangerous	Recherche

Table 4.2 – Résultat de la classification d'intention

étiquette neutre. [55]

Chaque phrase peut être associée à une ou plusieurs émotions, rendant ce dataset initialement adapté à des tâches de classification multi-label. Les émotions couvertes incluent des états comme la joie, la colère, la tristesse, la peur, ou encore la fierté. La diversité des émotions et la nature informelle des textes en font un corpus particulièrement utile pour entraîner des modèles sur des données textuelles réalistes et variées.

Ce dataset est structuré en trois ensembles: train, validation et test, avec un déséquilibre notable entre les classes émotionnelles, justifiant un traitement spécifique avant l'entraînement des modèles.

4.3.2 Prétraitement des données:

Pour effectuer le fine-tuning sur le dataset GoEmotions en vue de la classification des émotions, nous avons utilisé le modèle MiniBERT, une variante allégée de BERT permettant un entraînement plus rapide. Le prétraitement a consisté en plusieurs étapes essentielles:

1. Fusion des ensembles de données: Les trois parties du dataset original (train, validation, test) ont été combinées en un seul ensemble afin de faciliter les étapes de traitement suivantes.
2. Suppression des annotations multi-labels: Environ 20% des entrées du dataset

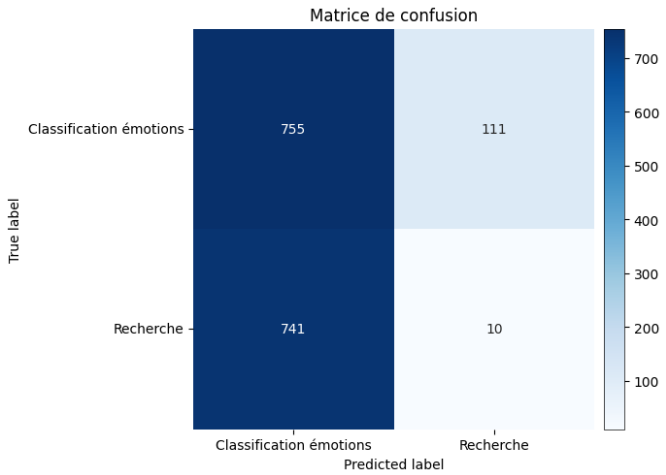


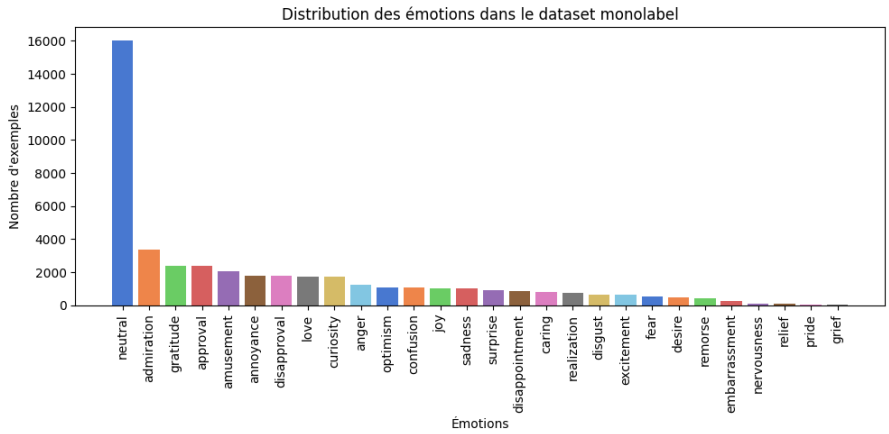
Figure 4.4 – Matrice de confusion avant le fine tuning

contiennent plusieurs étiquettes émotionnelles pour une seule phrase. Étant donné notre objectif de classification mono-label, ces lignes ont été éliminées.(voir figure:4.5a)

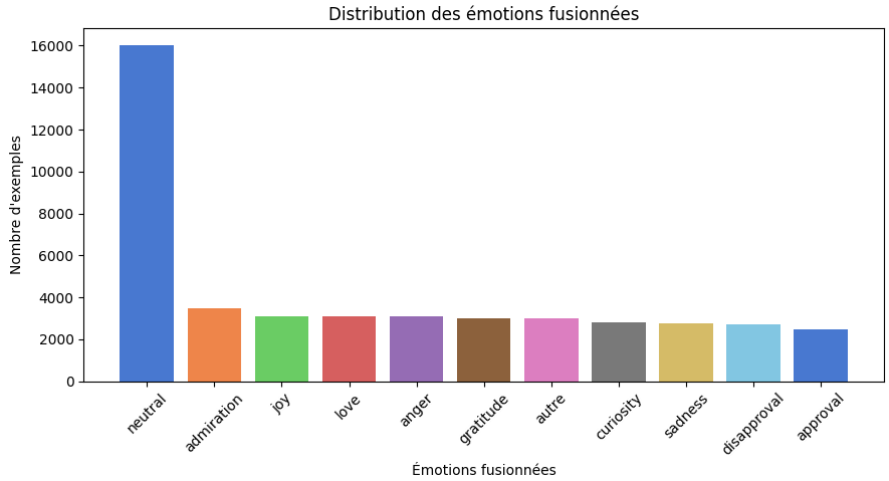
3. Fusion des émotions rares: Le dataset présente un déséquilibre important entre les émotions, certaines classes contiennent plus de 16 000 exemples, tandis que d'autres en comptent moins de 50. Pour atténuer ce déséquilibre, nous avons procédé à une fusion des émotions proches sémantiquement. Cette similarité a été mesurée à l'aide de la similarité cosinus (cosine similarity) entre les embeddings des étiquettes émotionnelles, suivie d'une réduction de dimension par ACP (Analyse en Composantes Principales) pour la visualisation. [56] (voir figure:4.5b)
4. Sous-échantillonnage équilibré: Un sous-échantillonnage (undersampling) a ensuite été appliqué afin de conserver un total de 3 000 exemples par émotion, garantissant ainsi un dataset final relativement équilibré entre les différentes classes émotionnelles.Au total, cette opération a permis de constituer un corpus homogène de plus de 31,000 exemples, prêt à être utilisé pour l'entraînement du modèle. (voir figure:4.6)
5. Redistribution des ensembles: Une fois les données prétraitées et équilibrées, l'ensemble a été redistribué selon un ratio de 80% pour l'entraînement, 10% pour le test et 10% pour la validation. Cela a abouti à une répartition d'environ 25 000 exemples pour l'entraînement, 3 000 pour le test, et 3 000 pour la validation, assurant une base solide pour le fine-tuning du modèle.
6. Tokenisation: Chaque phrase a été tokenisée à l'aide du tokenizer associé à MiniBERT (BertTokenizer), garantissant une compatibilité optimale avec le modèle utilisé. Cette étape inclut la mise en minuscules, le découpage en sous-mots (wordpieces), et le padding/truncation des séquences

4.3.3 Entraînement du modèle

Chargement du modèle et configuration: Le modèle a été initialisé avec une configuration adaptée à la classification multi-classe, en spécifiant 11 sorties correspondant



(a) Distribution des émotions avant fusions



(b) Distribution des émotions après fusions

Figure 4.5 – Comparaison des distributions d'émotions avant et après fusion

aux émotions fusionnées. L'architecture sous-jacente reste celle de BERT, mais avec une taille réduite, ce qui permet un déploiement plus rapide sur des machines à ressources limitées.

Préparation des ensembles d'apprentissage: Les textes ont été tokenisés avec (BertTokenizer), avec (padding) et (truncation) à une longueur maximale de 256 tokens. Les ensembles d'apprentissage ont ensuite été encapsulés sous forme de Dataset PyTorch personnalisé, puis chargés via des (DataLoader) avec un (batch_size) de 16.

Hyperparamètres et stratégie d'optimisation: L'optimisation a été réalisée avec l'algorithme (AdamW), en utilisant un taux d'apprentissage initial de (2e-5) et une pénalisation L2 (weight_decay) de (0.01). Une stratégie de planification du taux d'apprentissage par (cosine scheduling avec warmup) a été adoptée, avec 10% des étapes totales consacrées au warmup. Cette approche permet une montée progressive du taux d'apprentissage, suivie d'une décroissance douce, favorisant une meilleure stabilité durant l'entraînement.

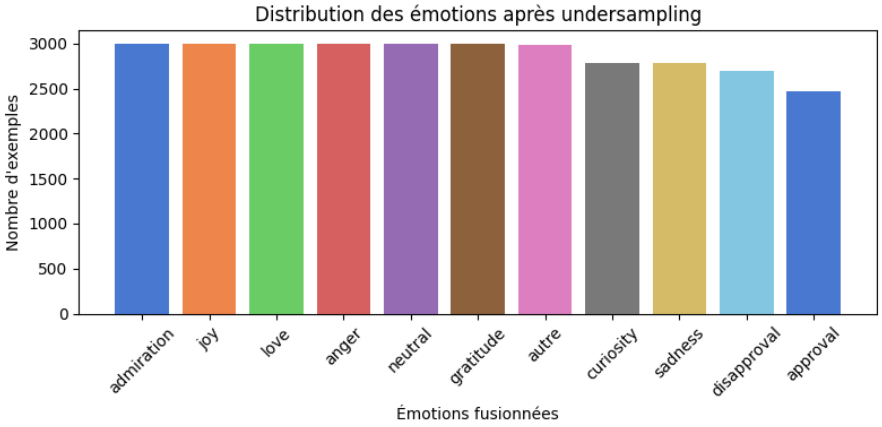


Figure 4.6 – Distribution des émotions après prétraitement

Entraînement et stratégie d'arrêt: L'entraînement a été effectué sur un maximum de 20 époques. À chaque époque, le modèle est évalué sur l'ensemble de validation. Un mécanisme d'early stopping a été mis en place: si aucune amélioration de la perte de validation n'est observée pendant trois époques consécutives, l'entraînement est arrêté pour éviter le surapprentissage.

Les métriques de loss et accuracy sont suivies et visualisées pour les ensembles d'entraînement et de validation tout au long des époques. Le modèle affichant la meilleure performance (plus faible perte sur la validation) est sauvegardé et rechargé à la fin pour être utilisé dans la phase de test finale.

4.3.4 Résultats

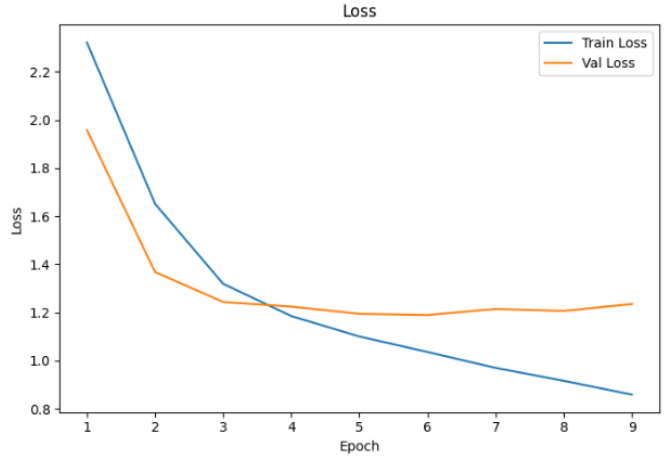


Figure 4.7 – Courbes de perte

Interprétation des courbes de perte: La perte d'entraînement a diminué régulièrement, passant de 2.32 à 0.85, ce qui montre que le modèle s'ajuste efficacement aux données d'entraînement. Concernant la perte de validation, une forte baisse est observée durant les premières époques, atteignant un minimum à l'époque 6 (1.19). Cependant, cette perte cesse de diminuer par la suite, voire augmente légèrement, ce qui indique un début

de surapprentissage sur les données d'entraînement (overfitting). L'early stopping a donc été déclenché à temps a la 9éme époque, ce qui confirme son efficacité pour maintenir un bon compromis entre biais et variance (voir figure:4.7).

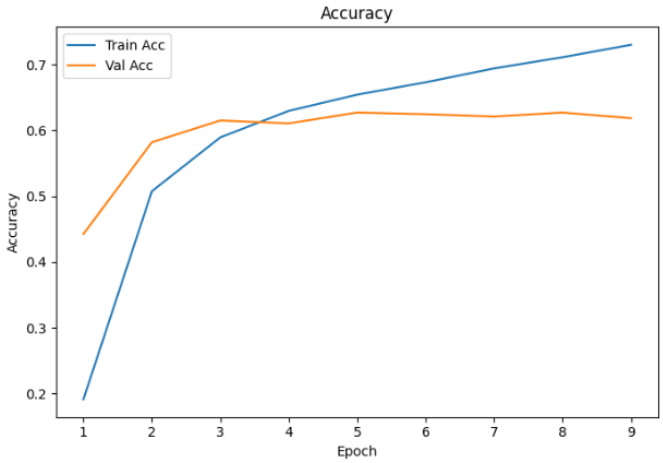


Figure 4.8 – Courbes d'accuracy

Interprétation des courbes d'Accuracy: L'accuracy sur l'ensemble d'entraînement est passée de 19.14% à 72.98%, montrant une amélioration constante au fil des époques. Cela indique que le modèle a appris à mieux classer les données au fil de l'entraînement. Sur l'ensemble de validation, l'accuracy a également progressé de manière significative pour atteindre un pic à 62.69% à l'époque 5, avant de se stabiliser. Cette stabilité, suivie d'une légère baisse, suggère que le modèle commence à atteindre ses limites de généralisation à partir de la 6éme époque. L'arrêt anticipé (early stopping) à la 9éme époque a permis d'éviter un surapprentissage (voir figure:4.8.

4.3.5 Évaluation et test

Après l'entraînement, le modèle a été évalué sur l'ensemble de test afin de mesurer sa capacité de généralisation sur des données jamais vues. Les textes ont été tokenisés à l'aide du tokenizer MiniBERT, puis transformés en DataLoader pour l'évaluation.

Le modèle a généré les prédictions sans mise à jour des poids (mode évaluation), et ces résultats ont été comparés aux étiquettes réelles pour produire un rapport de classification détaillé. Ce rapport inclut plusieurs métriques par classe émotionnelle: précision (precision), rappel (recall), F1-score ainsi que l'exactitude globale (accuracy) (voir Tableau 4.3).

Interprétation: L'analyse des métriques de précision et de rappel montre une trés bonne performances selon les émotions. On observe une très bonne précision pour des émotions comme gratitude (0.89) et joy (0.78), ce qui indique que lorsqu'elles sont prédites, elles le sont rarement à tort. En revanche, des classes comme neutral (0.41), approval (0.46) ou disapproval (0.47) présentent une précision plus faible, traduisant une plus grande confusion avec d'autres émotions.

En ce qui concerne le rappel, les émotions curiosity (0.78), joy (0.83) et sadness (0.70)

Émotion	Précision	Rappel	F1-Score	Support
admiration	0.68	0.72	0.70	300
anger	0.52	0.58	0.55	300
approval	0.46	0.44	0.45	247
autre	0.61	0.52	0.56	299
curiosity	0.63	0.78	0.70	279
disapproval	0.47	0.49	0.48	269
gratitude	0.89	0.75	0.81	308
joy	0.78	0.83	0.80	300
love	0.69	0.72	0.70	300
neutral	0.41	0.29	0.34	300
sadness	0.66	0.70	0.68	278
Accuracy		0.62 (3172 exemples)		
moyenne	0.62	0.62	0.62	3172

Table 4.3 – Résultats des métriques d'évaluation par émotion

sont bien détectées dans la majorité des cas où elles apparaissent réellement. À l'inverse, des émotions comme neutral (0.29) et approval (0.44) affichent un rappel bas, ce qui signifie que le modèle échoue souvent à les reconnaître quand elles sont présentes.

Globalement, la moyenne (0.62) montre que le modèle est globalement cohérent, mais pas optimal pour toutes les classes.

Interprétation de la matrice de confusion: La matrice de confusion montre que le modèle parvient globalement à bien différencier les émotions, comme en témoignent les fortes valeurs présentes sur la diagonale principale. Les émotions telles que joy, gratitude, curiosity ou encore admiration sont particulièrement bien reconnues.

Cependant, plusieurs confusions subsistent:

— L'étiquette neutral est fréquemment confondue avec approval, autre et disapproval, ce qui reflète une certaine ambiguïté sémantique dans ces catégories.

— De nombreuses confusions sont également observées entre approval et disapproval, deux émotions proches dans certaines formulations linguistiques.

On constate aussi que certaines émotions sont bien distinctes et donc plus faciles à prédire correctement. Cela indique que le modèle est capable de capter des signaux lexicaux ou contextuels forts pour des émotions bien marquées, mais peut avoir plus de mal avec des émotions plus subtiles ou ambiguës (voir figure:4.9).

Enfin, pour enrichir l'interprétation des résultats, l'analyse a été complétée par l'utilisation de l'outil VADER grâce à son compound score, qui est une valeur comprise entre

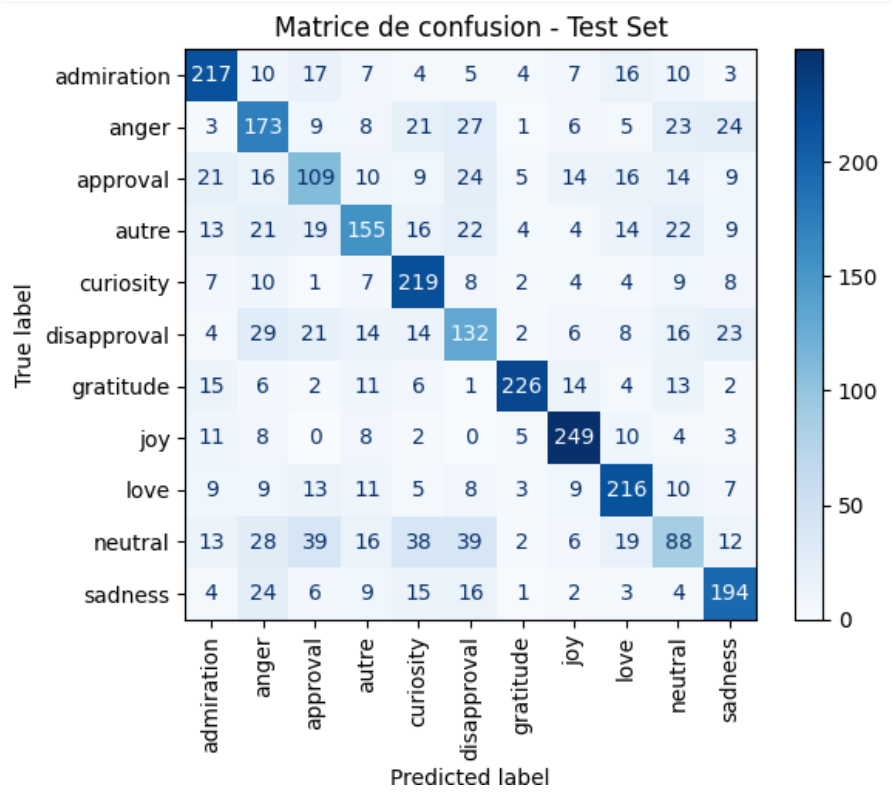


Figure 4.9 – Matrice de confusion

-1 et 1, calculant ainsi l'intensité de l'émotion. Nous avons utilisé une condition du type: si l'émotion détectée est anger et que le compound est inférieur à -50 (car nous l'avons multiplié par 100), alors l'émotion est jugée inacceptable ; sinon, elle est considérée comme acceptable (voir tableau 4.4).

4.3.6 Comparaison avec les autre travaux

Afin de situer la performance de notre approche, nous avons comparé nos résultats avec ceux obtenus par d'autres modèles standards sur dataset GoEmotion , notamment BERT, CNN, ROBERTA, en utilisant les F1-scores par émotion disponibles dans la littérature. [57]

comparaison: entre les modèles CNN, BERT, ROBERTA et notre modèle MiniBERT pour la classification des émotions met clairement en évidence la supériorité de ce dernier. En effet, MiniBERT obtient les meilleurs scores F1 sur 10 des 11 émotions évaluées, surpassant systématiquement les autres architectures. Les émotions telles que gratitude (0.81), joy (0.80), curiosity (0.70) et admiration (0.70) présentent des performances particulièrement élevées, ce qui témoigne de la capacité du modèle à bien capter les nuances émotionnelles dans les textes. Par rapport aux modèles de référence, les écarts de performance sont significatifs, atteignant parfois plus de 0.30 points sur certaines émotions comme love ou curiosity. Seule l'émotion neutral constitue une exception, où BERT atteint un score légèrement supérieur (0.50 contre 0.34 pour MiniBERT), ce qui pourrait être lié à une confusion avec d'autres émotions ou à un déséquilibre dans les données. Globalement, ces résultats démontrent que MiniBERT est particulièrement

Phrase	Émotion (confiance)	Score VADER	Acceptabilité
"I feel like my temper is getting worse every day."	Sadness (0.71)	-52.67	Acceptable
"The anxiety of this situation is unbearable."	Sadness (0.86)	-17.79	Acceptable
"I can't stop shaking, I'm too nervous to think straight."	Sadness (0.86)	29.73	Acceptable
"I'm furious that nobody is listening to me!"	Anger (0.78)	-61.14	Not Acceptable
"I was so mad I almost lost control!"	Anger (0.62)	-75.18	Not Acceptable

Table 4.4 – Résultats de la classification des émotions avec score VADER et acceptabilité

adapté à la tâche de classification fine des émotions, offrant un compromis optimal entre légèreté du modèle et efficacité, ce qui le rend pertinent pour des applications en environnement à ressources limitées (voir tableau 4.5).

Emotions	BERT	CNN	ROBERTA	MiniBERT
admiration	0.51	0.33	0.60	0.70
anger	0.31	0.31	0.33	0.55
approval	0.32	0.19	0.31	0.45
autre	0.35	0.31	0.30	0.56
curiosity	0.31	0.36	0.36	0.70
disaproval	0.18	0.21	0.22	0.48
gratitude	0.52	0.52	0.59	0.81
joy	0.56	0.45	0.65	0.80
love	0.29	0.43	0.37	0.70
neutral	0.50	0.52	0.51	0.34
sadness	0.38	0.30	0.43	0.68

Table 4.5 – tableau des f1-scores

4.4 Génération de texte GPT (Phase 3)

4.4.1 Présentation du dataset Utilisé

Un corpus personnalisé sur la santé mentale a été construit en fusionnant quatre ensembles de données open source, spécifiquement orientés vers les conversations psy-

chologiques: (voir figure:4.10)

- `Amod/mental_health_counseling_conversations`
- `jerryjalapeno/nart-100k-synthetic`
- `CalebE/new_mental_health_conversations_all1`
- `CalebE/new_mental_health_conversations_all`

Ces jeux de données contiennent des conversations structurées entre questions/réponse.

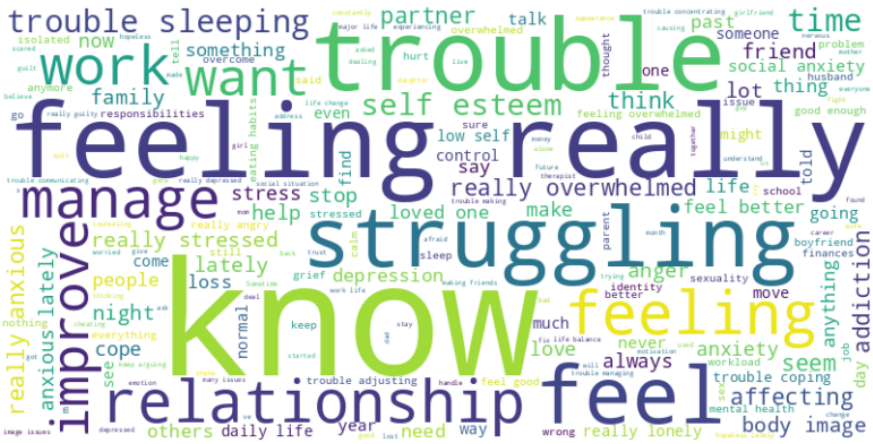


Figure 4.10 – Nuage de mots

4.4.2 Prétraitement des données

Avant de procéder à l'entraînement du modèle pour la génération de réponses, plusieurs étapes de prétraitement ont été appliquées afin d'assurer la cohérence, la qualité et l'homogénéité des données utilisées:

1. Réorganisation du format des données: Certains jeux de données présentaient la question et la réponse dans une seule cellule. Afin de faciliter l'entraînement, ces cas ont été restructurés sous la forme d'une paire explicite (question, réponse).
2. Fusion des sources de données: Après restructuration, les différents jeux de données (les quatre dataset) ont été fusionnés en un seul corpus unifié. Cette étape a permis d'enrichir la diversité des interactions tout en augmentant le volume global.
3. Suppression des informations sensibles: Toutes les lignes contenant des noms propres ou des identifiants personnels de patients ont été supprimées, afin de garantir l'anonymat et la conformité éthique des données.
4. Élimination des doublons question-réponse: Les paires où la question et la réponse étaient identiques ou dupliquées ont été supprimées intégralement pour éviter l'introduction de biais ou d'ambiguïtés dans les réponses générées.
5. Nettoyage des textes: Les entrées contenant des champs vides ou des caractères indésirables (emojis, balises HTML, parenthèses erronées, etc.) ont été filtrées pour garantir la qualité linguistique du corpus.
6. Réécriture des réponses redondantes: Lorsque plusieurs réponses étaient très similaires ou identiques, elles ont été reformulées manuellement tout en conservant le

même sens. Cette étape visait à renforcer la diversité lexicale tout en maintenant la cohérence sémantique (ex. : remplacer “joy” par une reformulation équivalente exprimant la même émotion).

7. Filtrage selon la longueur des réponses: Afin d’éviter les réponses trop courtes ou excessivement longues, un seuil minimal et maximal de longueur a été défini. Les paires ne respectant pas ces critères ont été supprimées pour maintenir une qualité constante.

À l’issue de ces étapes, le dataset obtenu était nettoyé, structuré et prêt à être utilisé pour le fine-tuning du modèle de génération.

voici la distribution de longueur des reponses (figure:4.11):

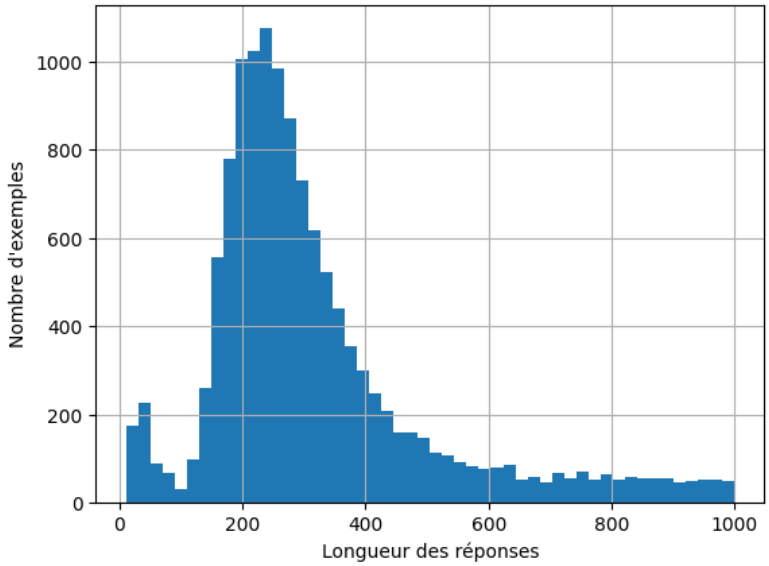


Figure 4.11 – Distribution de longueur des reponses

4.4.3 Entraînement du modèle

Paramètres de génération

Les paramètres suivants ont été utilisés pour contrôler la génération des réponses du modèle:

— Top-k: Limite les choix au top k mots les plus probables pour éviter les sorties trop aléatoires.

— Top-p: Sélectionne dynamiquement les mots les plus probables dont la somme cumulée atteint p (nucleus sampling).

— Temperature: Contrôle la créativité du modèle, plus la valeur est élevée, plus la génération est aléatoire.

— Répetition penalty: Applique une pénalité aux mots déjà générés pour éviter les répétitions excessives.

Le modèle utilisé a été soumis à un entraînement supervisé à l’aide d’une configuration optimisée, spécifiquement adaptée à la génération de texte dans un cadre de

dialogue émotionnel. L'objectif était d'obtenir une réponse cohérente et empathique en contexte, tout en maximisant l'efficacité computationnelle.

Chargement du modèle et configuration: Le modèle a été initialisé avec une architecture de type causal LM, permettant la génération de texte conditionnée. Il intègre le mécanisme de gradient_checkpointing, permettant de réduire l'usage mémoire durant l'entraînement. La configuration prévoit une gestion fine des gradients et un suivi précis des métriques, avec sauvegarde automatique du meilleur modèle en fonction de la perte de validation (eval_loss).

Préparation des ensembles d'apprentissage: Les données textuelles ont été traitées à l'aide du tokenizer approprié, avec ajout de prompts de dialogue structurés. Les jeux d'apprentissage et de validation ont été encodés, puis traités sous forme de lots (batch_size de 8) avec une accumulation de gradients sur deux étapes pour simuler un batch effectif de 16.

Hyperparamètres et stratégie d'optimisation: L'optimisation a été réalisée avec une stratégie de warmup dynamique, correspondant à 10% des étapes totales, suivie d'un ajustement progressif du taux d'apprentissage. Ce dernier a été initialisé à 5e-6, avec un facteur de pénalisation L2 (weight_decay) de 0.01. L'entraînement a été effectué sur 6 époques, avec évaluation et sauvegarde du modèle à chaque époque.

Entraînement et génération contrôlée: Le processus d'entraînement a intégré une logique de génération contrôlée à la fin de chaque époque via une TrainerCallback personnalisée. Cette dernière permettait d'évaluer qualitativement le comportement du modèle sur un exemple fixe, à l'aide de paramètres de génération (top-k=50, top-p=0.9, température=1.2, pénalité de répétition de 1.2). L'objectif était de suivre l'évolution du langage généré en parallèle des métriques classiques.

L'ensemble des métriques (perte, performance de validation) a été loggé toutes les 20 étapes. Le modèle présentant la meilleure perte de validation a été automatiquement rechargé à la fin pour les phases de test ou d'inférence.

4.4.4 Résultats

Interprétation:

On observe une diminution régulière de la perte d'entraînement, passant d'environ 1.7 à 1.27, ce qui indique que le modèle apprend progressivement à minimiser l'erreur sur les données d'apprentissage. De manière similaire, la perte de validation diminue également, atteignant une valeur approximative de 1.22 à la dernière époque. Fait intéressant, la courbe de validation reste constamment inférieure à celle de l'entraînement. Cette situation, bien que peu fréquente, peut s'expliquer par un jeu de validation plus homogène ou moins bruité, ou encore par la désactivation de certaines techniques de régularisation comme le dropout durant la phase d'évaluation. L'absence de remontée de la courbe de validation et la convergence progressive des deux courbes suggèrent que

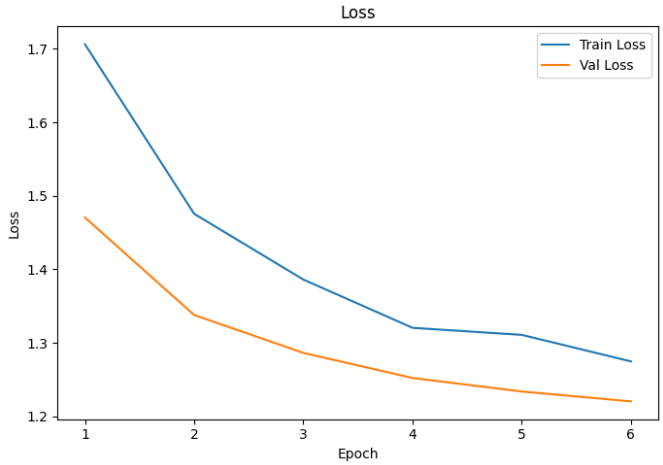


Figure 4.12 – Courbes de perte

le modèle ne souffre pas de surapprentissage (overfitting) et que l'entraînement est stable et efficace (voir figure:4.12).

4.4.5 Évaluation et test

Pour évaluer la performance de notre modèle de génération de réponses, nous avons utilisé quatre métriques standard largement adoptées dans le domaine du traitement automatique du langage naturel: Perplexity, BLEU, ROUGE-L et METEOR. Chacune de ces métriques apporte une perspective complémentaire sur la qualité des réponses générées.

— Perplexity: Il s'agit d'une mesure de la capacité du modèle à prédire une séquence de texte. Une perplexité faible indique que le modèle est plus confiant dans ses prédictions, ce qui signifie que les réponses générées sont grammaticalement et syntaxiquement cohérentes [58].

— BLEU (Bilingual Evaluation Understudy): Cette métrique évalue la correspondance entre la réponse générée et une ou plusieurs réponses de référence en se basant sur des n-grammes. Elle est couramment utilisée pour mesurer la précision dans les tâches de traduction ou de génération de texte [59].

— ROUGE-L (Recall-Oriented Understudy for Gisting Evaluation): Elle se concentre sur le rappel et compare la plus longue sous-séquence commune (Longest Common Subsequence) entre la sortie du modèle et la référence, ce qui permet de juger de la couverture du contenu [60].

— METEOR (Metric for Evaluation of Translation with Explicit ORdering): Contrairement à BLEU, cette métrique prend en compte les correspondances synonymiques, la racinisation des mots (stemming) et l'ordre des mots. Elle est plus sensible au contenu sémantique réel [61].

Les résultats obtenus à l'aide de ces métriques sont résumés dans le tableau **??** suivant:

L'analyse des distributions des scores de génération révèle plusieurs éléments importants:

Métrique	Score
Perplexity	*3.4*
BLEU	*0.31*
ROUGE-L	*0.47*
METEOR	*0.81*

Table 4.6 – Résultats pour chaque métrique

— BLEU (Moyenne: 0.31): Bien que le score BLEU soit relativement modéré, cela est cohérent avec le fait que cette métrique pénalise fortement toute divergence lexicale, même si le sens est conservé. La majorité des scores se situent entre 0.2 et 0.4, indiquant que les réponses générées partagent plusieurs n-grammes avec les réponses de référence.

— METEOR (Moyenne: 0.81): Ce score élevé montre que les réponses générées sont très proches des réponses attendues sur le plan sémantique. METEOR, prenant en compte les synonymes et la racinisation, est plus tolérant aux variations lexicales, ce qui le rend pertinent pour évaluer des tâches de génération libre comme ici.

— ROUGE-L (Moyenne: 0.47): Ce score intermédiaire indique une bonne capacité du modèle à capturer des séquences significatives de mots présentes dans les réponses de référence. Il souligne un bon niveau de couverture de contenu, même si certaines différences de formulation existent. (voir figure:4.13)

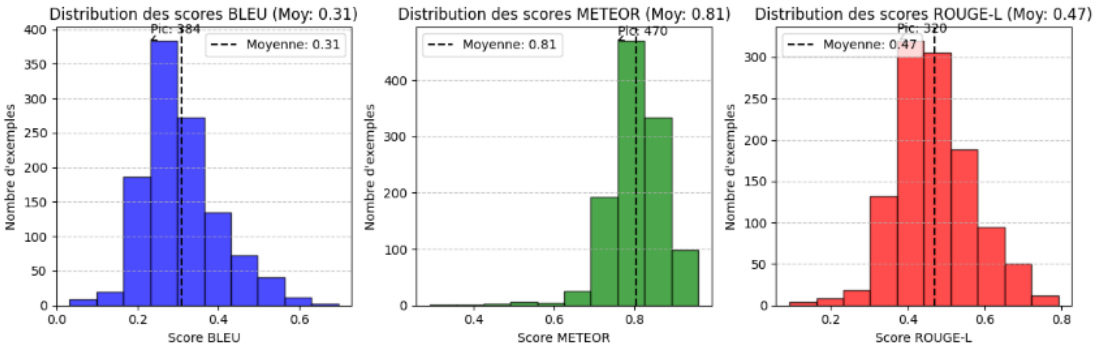


Figure 4.13 – Mériques d'évaluation de GPT (BLEU,ROUGE-L,METEOR)

Le graphique de l'évolution de la perplexité en validation montre une diminution progressive au fil des epochs, passant de 4.32 à 3.4 après 6 itérations. Cette baisse régulière indique que le modèle devient de plus en plus confiant dans ses prédictions et qu'il apprend efficacement à générer des séquences linguistiques cohérentes. Une perplexité plus faible correspond à une meilleure capacité du modèle à prédire la suite d'un texte, ce qui est essentiel dans les tâches de génération de réponses (voir figure:4.14).

4.4.6 Comparaison avec les autre travaux

Dans l'article PsyChat: A Client-Centric Dialogue System for Mental Health Support, leur modèle READER a obtenu un score ROUGE-L de 27.6, un METEOR de 22.9, un BLEU de 24.1 et une perplexité (PPL) de 1.26. Dans notre cas, bien que la PPL soit

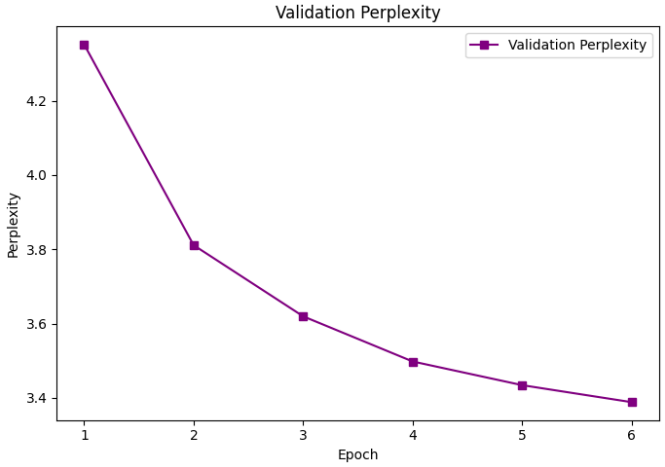


Figure 4.14 – Métrique d'évaluation de GPT (Perplexity)

légèrement inférieure, les scores METEOR, BLEU et ROUGE-L sont meilleurs, ce qui reflète une meilleure qualité sémantique dans les réponses générées [62].

Modèle	Perplexity	BLEU	ROUGE-L	METEOR
PsyChat	1.26	0.24	0.27	0.22
PsyBot	3.4	0.31	0.46	0.81

Table 4.7 – Évaluation des modèles sur le même dataset

Aussi, dans l'article TherapyBot: a chatbot for mental well-being using transformers, un score PPL de 1.26 a été rapporté [63]. Dans notre cas, bien que la PPL soit légèrement inférieure, il est important de noter que notre modèle GPT-2 a été fine-tuné sur un corpus sensiblement plus riche que ceux employés dans les travaux de référence, lesquels n'exploitent qu'une fraction des données que nous avons fusionnées (Amod, Nart 100k Synthetic, CalebE, etc.) (voir tableau:4.7)

Cette augmentation significative du volume et de la diversité des exemples d'entraînement se traduit par une meilleure couverture lexicale et une plus grande robustesse face à des formulations variées, ce qui peut expliquer en grande partie les valeurs supérieures que nous obtenons sur les métriques automatiques (BLEU, ROUGE, METEOR).

À l'inverse, les modèles comparés, entraînés sur un sous-ensemble plus restreint, présentent deux limites:

— une capacité moindre à généraliser à des contextes inédits ;

— une couverture thématique plus étroite.

De ce point de vue, l'avantage de notre approche réside dans la fusion de plusieurs sources de données psychologiques, offrant ainsi un apprentissage plus complet du registre émotionnel et conversationnel attendu pour PsyBot. (voir tableau:4.8)

4.5 Conclusion

Ce chapitre a permis de valider expérimentalement les différents modules qui composent l'architecture fonctionnelle de PsyBot. À travers une série d'expériences dis-

Critère	Woebot	Wysa	Replika	PsyBot
Thérapie de base	TCC	TCC + écoute active	Compagnon émotionnel	TCC + détection IA
Détection des émotions	Non	Partielle	Générale	**Oui (BERT + VADER)**
Génération des réponses	Scriptées	Scriptées	Semi-personnalisées	**GPT-2 fine-tuné**
Multilingue	Non	Partiel	Non	**Oui (français, arabe, anglais)**
Adaptation culturelle	Non	Non	Non	**Oui (contexte algérien)**
Filtrage émotionnel	Non	Non	Non	**Oui (alerte en cas de colère)**
Compréhension d'intention	Non	Non	Non	**Oui (DistilBERT)**
Confidentialité / vie privée	Moyenne	Moyenne	Faible	**Forte (anonymat, DuckDuckGo)**
Accompagnement par exercices	Non	**Oui (relaxation, respiration)**	Non	**Non (non intégré)**
Relation affective long terme	Non	Non	**Oui (mémoire émotionnelle)**	**Non**

Table 4.8 – Comparaison fonctionnelle entre PsyBot et des chatbots psychologues existants

tinctes, nous avons pu évaluer la performance de trois modèles principaux, chacun jouant un rôle clé dans le processus d'interaction avec l'utilisateur.

Le modèle DistilBERT a montré une bonne capacité à distinguer entre les requêtes de type « recherche » et celles destinées à une interaction psychologique, avec un score d'exactitude globalement satisfaisant malgré quelques erreurs liées à des formulations ambigües. De son côté, le module de classification émotionnelle, basé sur MiniBERT et renforcé par VADER, a permis de filtrer les requêtes à forte charge émotionnelle, offrant ainsi une première couche de sécurité éthique au système.

Enfin, la phase de génération, assurée par un modèle GPT fine-tuné, a démontré une aptitude prometteuse à produire des réponses pertinentes et contextuellement adaptées. Bien que les scores de qualité de génération (BLEU, ROUGE-L, METEOR, Perplexity) restent perfectibles, ils confirment la faisabilité de notre approche.

Dans l'ensemble, les résultats obtenus valident les choix méthodologiques effectués et fournissent une base solide pour envisager l'intégration complète du système dans une plateforme d'assistance virtuelle.

5

Mise en œuvre pratique du psybot

Sommaire

5.1 Introduction 56
5.2 Diagrammes 56
5.2.1 Diagramme des cas d'utilisations 56
5.3 Interfaces et explications 57
5.3.1 Patients 57
5.3.2 Psychologue 62
5.3.3 Clinique 62
5.4 Conclusion 63

5.1 Introduction

Après avoir décrit l'architecture technique et les modèles d'intelligence artificielle qui constituent le coeur du système PsyBot, ce chapitre est consacré à la présentation de l'interface web développée pour permettre l'interaction entre l'utilisateur et le chatbot, ainsi que l'intégration de deux autres type de client essentielles : les psychologues et les cliniques.

L'interface a été conçue de manière simple, intuitive et fonctionnelle afin de faciliter non seulement la communication avec le système, mais aussi la gestion des rôles liés à l'environnement thérapeutique. Elle permet aux utilisateurs de soumettre des requêtes, de recevoir des réponses via le chatbot, d'interagir avec des psychologues, et offre aux cliniques la possibilité de gérer à la fois les praticiens et les patients.

Dans ce chapitre, nous présentons d'abord les cas d'utilisation à travers des diagrammes UML, illustrant les différentes interactions possibles entre les acteurs et le système. Nous décrivons ensuite les principales interfaces de l'application web à l'aide de captures d'écran accompagnées d'explications fonctionnelles. Enfin, une conclusion revient sur les apports de cette interface dans le cadre du fonctionnement global de PsyBot.

5.2 Diagrammes

5.2.1 Diagramme des cas d'utilisations

Le diagramme de cas d'utilisation présenté ci-dessus (voir figure 5.1) illustre les principales interactions possibles entre les différents utilisateurs du site web PsyBot. On y distingue plusieurs types d'acteurs : le visiteur, le patient, le psychologue, l'administrateur de clinique, ainsi qu'un psychologue clinique.

Le visiteur peut uniquement consulter les services offerts, s'inscrire ou se connecter, et choisir la langue d'affichage. Une fois connecté, l'utilisateur devient un patient et accède à plusieurs options : initier une conversation avec le chatbot PsyBot, consulter ou modifier son profil, dialoguer avec un psychologue, prendre rendez-vous avec un praticien ou une clinique, et régler les paiements associés. Le patient peut également consulter son historique d'interaction, ou supprimer son compte ou certaines informations personnelles.

Les psychologues disposent d'une interface leur permettant de visualiser la liste de leurs patients, de gérer les rendez-vous, de répondre aux messages, de prendre des notes, et de suivre les paiements. Quant à l'administrateur clinique, il accède à une interface dédiée depuis laquelle il peut superviser les profils des utilisateurs (patients et psychologues), consulter les statistiques générales de la plateforme, ainsi que visualiser les courbes émotionnelles générées par le système.

Enfin, le diagramme inclut des relations de type «include» et «extend» pour montrer les dépendances fonctionnelles entre certaines actions. Par exemple, la suppression du compte est une extension du profil utilisateur, tandis que les interactions cliniques reposent sur des fonctionnalités de base accessibles au patient. Ce schéma permet ainsi de mieux comprendre l'architecture fonctionnelle de PsyBot et les parcours possibles des

utilisateurs en fonction de leur rôle. (voir figure:5.1)

Figure 5.1 – Diagramme de cas d'utilisation

5.3 Interfaces et explications

5.3.1 Patients

Après inscription/connexion le patient se retrouve devant 3 choix comme le montre la figure 5.2.

Figure 5.2 – Portail du site pour le Patient

Psybot (Discutez avec un modèle intelligent pour un soutien immédiat)

Si l'utilisateur choisit PsyBot (premier choix), une interface de chat s'ouvre, lui permettant de discuter avec notre modèle comme s'il parlait avec un psychologue. Contrairement à un psychologue humain, il peut échanger avec PsyBot à tout moment, sans avoir besoin de prendre rendez-vous. Cependant, des limites subsistent, car il n'est pas encore aussi performant qu'un véritable psychologue.

Comme le montrent les figures suivantes, plusieurs exemples de requêtes sont présentés.

requête de recherche (DuckDuckGo): Cette requête est dédiée à la recherche scientifique en psychologie. Grâce à l'intégration de la bibliothèque DuckDuckGo dans notre modèle, une réponse adéquate est retournée à l'utilisateur (voir figure5.3).

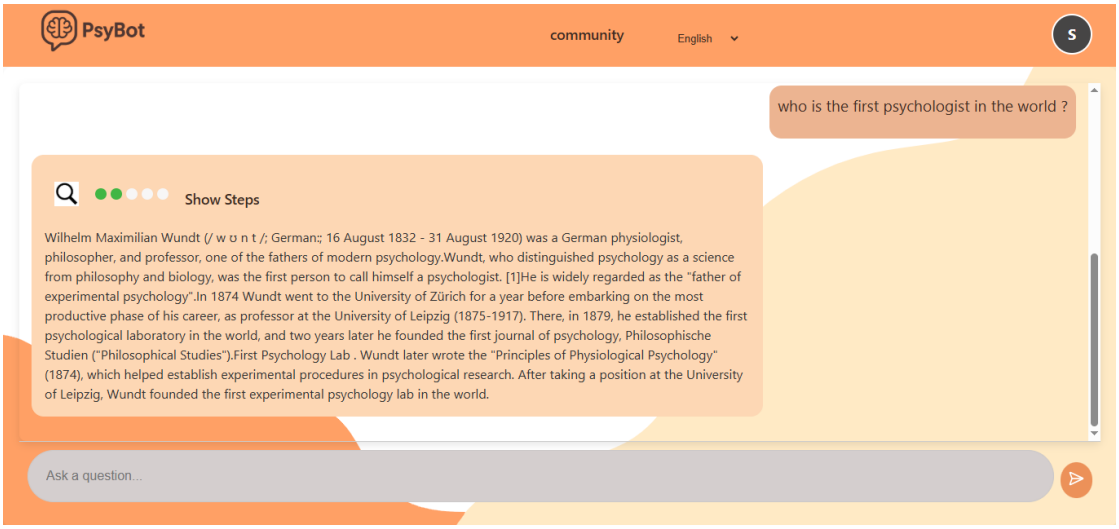


Figure 5.3 – Requête de recherche

requête avec émotions inacceptable: Dans cette requête délicate, exprimant une émotion de colère intense, le système retourne un message d'erreur et demande à l'utilisateur de se calmer (voir figure5.4).

Figure 5.4 – Requête avec emotion non acceptable

requête qui demande le soutien du psybot: C'est une requête qui n'est pas dédiée à la recherche scientifique et qui exprime une émotion normale. À travers celle-ci, l'utilisateur cherche un soutien psychologique, des conseils, ou même des encouragements personnalisés (voir figure 5.5)

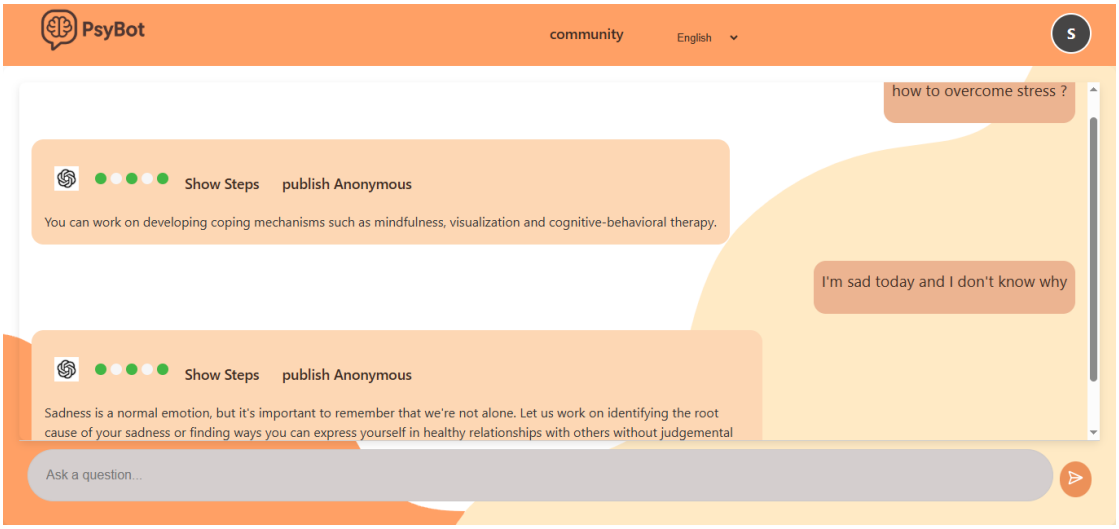


Figure 5.5 – Requête de soutien du psybot

Tout cela peut être effectué aussi bien en français qu'en arabe (voir figure5.6).

Figure 5.6 – Conversation en arabe

Dans cette section de PsyBot, l'utilisateur a également la possibilité de publier sa question ainsi que la réponse de PsyBot dans la section "Communauté", afin que les autres utilisateurs puissent les voir, commenter et aimer sa publication (voir figure5.7).

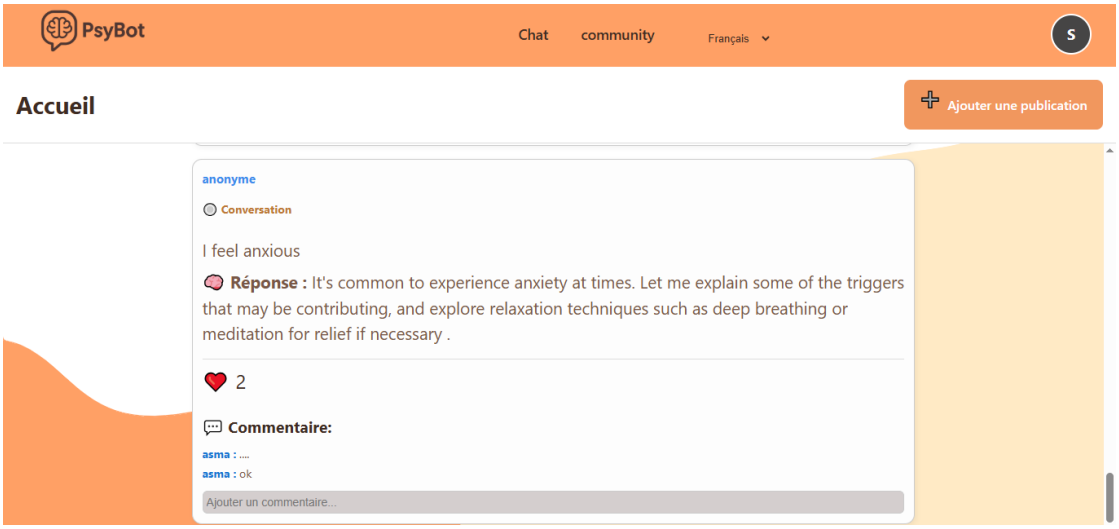


Figure 5.7 – Communauté

il peut aussi voir l'historique de ses conversations(voir figure5.8).

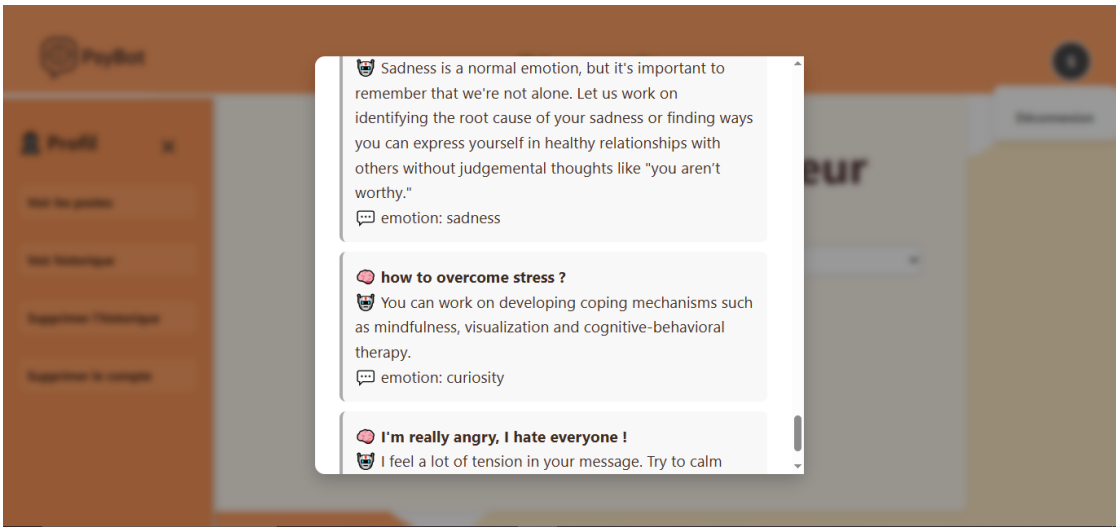


Figure 5.8 – Historique

Psychologue(Prenez contact avec un psychologue humain pour un suivi personnalisé)

Comme deuxième option, le patient peut choisir de parler avec un véritable psychologue humain. Il doit d'abord payer la séance, puis prendre un rendez-vous avec le psychologue de son choix. Une fois la séance commencée, il aura accès au chat pour bénéficier d'un suivi psychologique (voir figure 5.9).

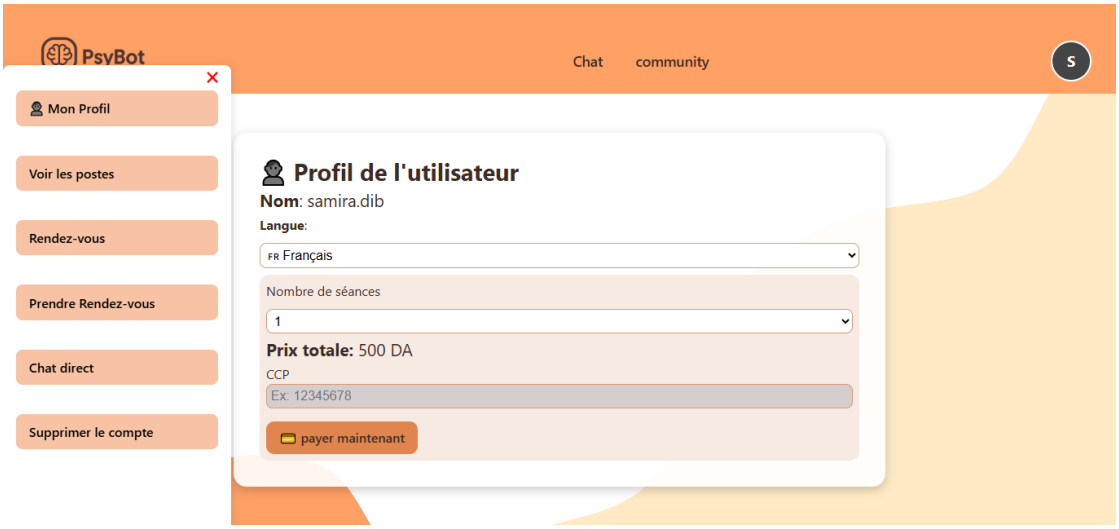


Figure 5.9 – Soutien psychologue

clinique(Trouvez une clinique spécialisée selon vos besoins)

Enfin, le troisième choix proposé au patient est de consulter une clinique spécialisée en psychologie, disposant de ses propres psychologues.(voir figure 5.10).

Le patient peut soit choisir un psychologue affilié à la clinique, soit opter pour un PsyBot avancé, entraîné à partir de nombreuses conversations précédentes enregistrées (avec l'accord des utilisateurs, bien entendu) et accessible via un modèle payant. Cette fonctionnalité n'est pas encore active, car le site web n'est pas encore utilisé par un nombre suffisant d'utilisateurs.

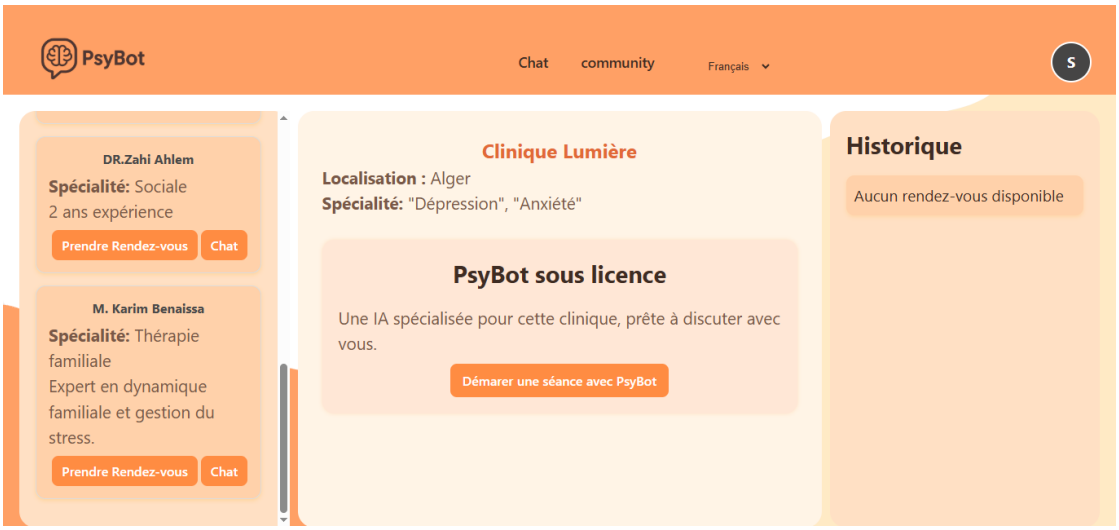


Figure 5.10 – choix clinique

5.3.2 Psychologue

Le deuxième type d'utilisateur est le psychologue. Il s'agit d'un praticien indépendant, non affilié à une clinique. Il bénéficie du privilège de travailler à domicile, sans avoir à se déplacer, en échangeant directement avec ses patients via la plateforme.

Le psychologue a également la possibilité de rédiger des remarques ou des notes privées sur ses patients. Il peut gérer ses rendez-vous librement (supprimer ou modifier les rendez-vous), selon ses.(voir figure 5.11)

Figure 5.11 – psychologue

5.3.3 Clinique

Enfin, les cliniques représentent le troisième type d'utilisateur (ou administrateur) dans notre système. Leur rôle est de gérer presque l'ensemble des données statistiques, notamment celles liées aux émotions des utilisateurs, ainsi que de s'occuper de la gestion des paiements.

Pour les psychologues affiliés à ces cliniques, c'est la clinique qui prend en charge

leur rémunération. Un pourcentage sur chaque séance est prélevé au profit de notre plateforme, contrairement aux psychologues indépendants qui gèrent directement leurs paiements.(voir figure 5.12)

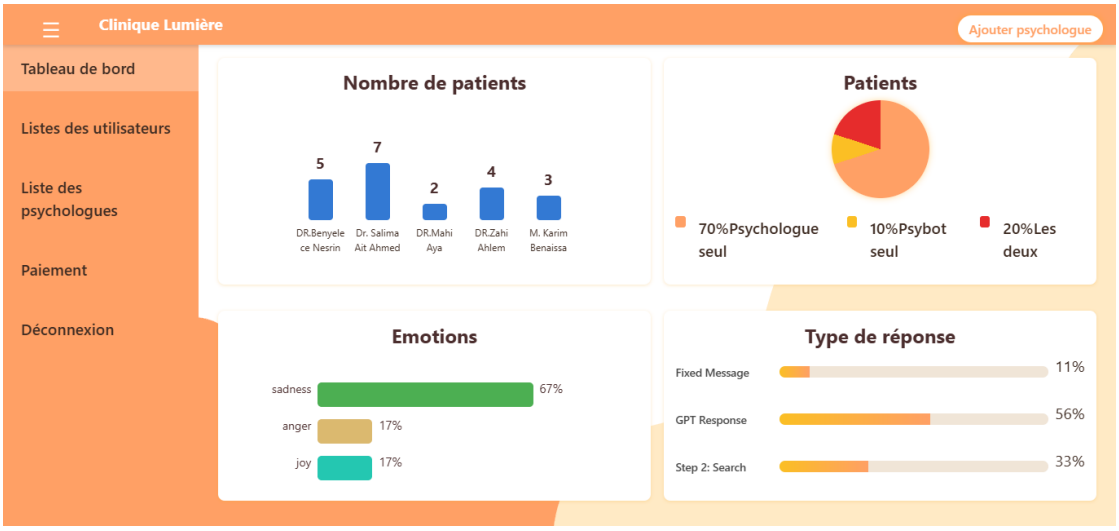


Figure 5.12 – Clinique

5.4 Conclusion

Ce chapitre a permis de mettre en lumière la dimension applicative et concrète de notre projet à travers l'interface web de PsyBot. En complément des modules intelligents développés précédemment, cette interface joue un rôle central dans l'accessibilité et l'exploitation du système par les utilisateurs finaux.

Nous avons présenté les différents cas d'usage à l'aide de diagrammes UML, permettant de visualiser les interactions entre les patients, les psychologues, les cliniques et le chatbot. Ensuite, les interfaces de l'application ont été décrites en détail à travers des captures d'écran, illustrant les fonctionnalités clés offertes à chaque type d'utilisateur.

L'interface assure une passerelle fluide entre l'utilisateur et les modules d'intelligence artificielle, tout en intégrant une logique de gestion adaptée au contexte thérapeutique. Elle incarne la finalité du projet : offrir une solution complète, intuitive et éthique, au service du soutien psychologique assisté par l'IA.

6

Conclusion générale

Ce mémoire a présenté la conception, le développement et l'évaluation de PsyBot, un assistant virtuel intelligent dédié au soutien psychologique. Dans un contexte où les besoins en santé mentale deviennent de plus en plus pressants, mais où l'accès aux professionnels reste limité par des contraintes géographiques, financières ou sociales, notre solution vise à proposer un outil d'accompagnement émotionnel à la fois accessible, éthique et techniquement robuste.

Le système repose sur une architecture modulaire articulée autour de trois modules principaux : la classification de l'intention, la détection d'émotions, et la génération de réponses. Chaque composant a été développé à partir de modèles de traitement automatique du langage naturel (NLP) modernes, entraînés sur des jeux de données adaptés, et complétés par des étapes de prétraitement, d'évaluation et de filtrage émotionnel. Bien que les résultats obtenus soient perfectibles, notamment en ce qui concerne la finesse des émotions détectées ou la fluidité des réponses générées, ils témoignent d'un fonctionnement globalement cohérent et prometteur.

Nous avons également intégré cette intelligence dans une plateforme web, pensée pour être intuitive et inclusive. Elle offre un accès simplifié au chatbot, mais aussi des fonctionnalités supplémentaires dédiées aux psychologues et aux cliniques, permettant d'envisager des cas d'usage hybrides mêlant assistance automatique et suivi humain.

Ce travail ouvre ainsi plusieurs perspectives. Il serait pertinent d'enrichir les jeux de données avec des sources multilingues et spécifiques au contexte culturel maghrébin, d'améliorer la sensibilité du modèle aux émotions subtiles, ou encore d'introduire une mémoire conversationnelle à long terme pour mieux contextualiser les échanges. Par ailleurs, des expérimentations à plus grande échelle pourraient permettre d'évaluer l'impact réel de PsyBot sur le bien-être des utilisateurs.

En somme, ce projet démontre qu'il est possible de combiner intelligence artificielle, sensibilité psychologique et accessibilité numérique dans une solution unifiée. PsyBot constitue une première étape vers une nouvelle génération d'agents conversationnels plus empathiques, plus responsables, et potentiellement utiles dans des contextes thérapeutiques encadrés.

7

Business Modele Canva (BMC)

Sommaire

7.1 Introduction 68
7.2 BMC 68
7.2.1 Proposition de valeur 68
7.2.2 Segments clients 69
7.2.3 Canaux 69
7.2.4 Relations clients 69
7.2.5 Partenaires clés 70
7.2.6 Activités clés 70
7.2.7 Ressources clés 71
7.2.8 Charges et Coûts 72
7.2.9 Revenus 73
7.3 Conclusion 74

7.1 Introduction

Ce chapitre est dédié à la présentation du modèle économique de notre projet à travers le Business Model Canvas (BMC), un outil stratégique permettant de structurer de manière claire et synthétique la proposition de valeur, les clients cibles, les canaux de distribution, les sources de revenus et les ressources clés du projet.

Dans le cadre de PsyBot, nous avons adapté le BMC pour refléter à la fois les aspects techniques liés à l'intelligence artificielle et les spécificités du domaine de la santé mentale en Algérie. Ce modèle permet de formaliser la stratégie de développement du produit, tout en assurant sa viabilité financière et son adéquation aux besoins des utilisateurs finaux : patients, psychologues et cliniques partenaires.

Chaque bloc du canevas a été analysé de manière détaillée afin de mettre en évidence les atouts différenciateurs de PsyBot, les opportunités de croissance, et les défis opérationnels à prendre en compte pour un déploiement à grande échelle.

7.2 BMC

7.2.1 Proposition de valeur

a) Quel problème principal résolvons-nous ?

- Accès difficile à un accompagnement psychologique fiable, rapide et abordable.
- Stigmatisation persistante autour de la santé mentale, freinant la prise d'initiative.
- Temps d'attente et indisponibilité des professionnels.
- Manque d'outils numériques adaptés au contexte algérien.

b) Quelle valeur offrons-nous au client ?

- Accompagnement psychologique confidentiel, rapide et accessible 24h/24 via PsyBot.
- Mise en relation directe avec des psychologues certifiés pour des consultations en ligne ou en présentiel.
- Suivi émotionnel personnalisé, conseils adaptés, et accès à des ressources.
- Plateforme intuitive, multilingue (ar/fr), respectueuse de la confidentialité.

c) Quels besoins ou désirs spécifiques comblons-nous ?

- Besoin d'écoute immédiate sans jugement.
- Besoin d'un accès fluide à un professionnel qualifié.
- Désir de discrétion, de flexibilité et de confort.
- Volonté de mieux gérer son état mental au quotidien.

d) Qu'est-ce qui rend notre solution unique ?

- PsyBot basé sur un modèle IA multilingue entraîné localement sur des données émotionnelles pertinentes.
- Intégration fluide entre chatbot, consultations humaines et cliniques partenaires.
- Matching intelligent entre l'utilisateur et le professionnel.
- Tarification accessible et adaptée au marché local.

7.2.2 Segments clients

a) Pour qui créons-nous de la valeur ?

- Grand public recherchant du soutien psychologique discret et abordable.
- Étudiants, jeunes adultes, travailleurs stressés, parents.
- Psychologues souhaitant digitaliser leurs services.
- Cliniques cherchant à toucher une clientèle plus large.

b) Quels sont nos segments de clientèle prioritaires ?

- Utilisateurs francophones et arabophones entre 18 et 45 ans.
- Psychologues agréés souhaitant proposer des services en ligne.
- Cliniques partenaires dans les grandes villes (Alger, Oran, Constantine…).

7.2.3 Canaux

a) Comment atteignons-nous nos clients ?

- Plateforme web (application PWA) accessible sur mobile et desktop.
- Réseaux sociaux (Instagram, TikTok, Facebook).
- Partenariats avec établissements universitaires, cliniques, influenceurs.

b) Par quels canaux livrons-nous la valeur ?

- Chatbot IA (PsyBot) accessible en ligne 24h/24.
- Interface de prise de rendez-vous avec professionnels humains.
- Tableau de bord utilisateur avec suivi, messages, historique, rappels.

7.2.4 Relations clients

a) Quel type de relation chaque segment attend-il ?

- Relation automatisée avec une touche humaine si besoin.
- Confidentialité totale, disponibilité rapide, et interface conviviale.
- Assistance ponctuelle (chat, email) pour les utilisateurs bloqués.

b) Comment intégrons-nous les utilisateurs ?

- Parcours onboarding fluide dès l'inscription.
- Gratuité d'utilisation de PsyBot pour découverte.
- Notification et rappels réguliers.
- Gamification : badges, progression, historique personnalisé.

c) Comment assurons-nous la fidélisation ?

- Suivi personnalisé avec conseils adaptés.
- Contenu exclusif (webinaires, vidéos, fiches pratiques).
- Accès prioritaire à certains professionnels.

7.2.5 Partenaires clés

a) Qui sont nos partenaires stratégiques ?

- Psychologues agréés partenaires.
- Cliniques spécialisées en santé mentale.
- Universités et laboratoires de recherche en psychologie / IA.
- Plateformes techniques (Render, Hugging Face, Firebase).

b) Quels sont les partenariats nécessaires pour réussir ?

- Partenariats de validation clinique et éthique.
- Accords avec prestataires de paiement (CIB, Edahabia).
- Partenaires de distribution (influenceurs, médias).

7.2.6 Activités clés

a) Quelles actions principales devons-nous entreprendre pour livrer notre proposition de valeur ?

- Développement et amélioration continue du PsyBot via fine-tuning, maintien des performances, rapidité et stabilité de l'API.
- Intégration fluide des psychologues et des cliniques : vérification des diplômes, création de profils, formation à la plateforme.
- Protection des données et conformité réglementaire : anonymisation, hébergement sécurisé conforme.
- Support client multicanal : assistance via chat et mail, base de connaissances, FAQ et tutoriels vidéo.

b) Quelles sont les opérations essentielles pour notre entreprise ?

- Développement et maintenance technique : mises à jour IA, monitoring API, gestion des bugs.
- Onboarding professionnel : vérification, création de comptes, accompagnement initial.
- Cybersécurité et protection des données : anonymisation, cryptage, hébergement conforme.
- Service client et accompagnement utilisateur : gestion des requêtes, amélioration continue de la satisfaction.

c) Quelles activités créent le plus de valeur pour nos clients ?

- Réponses instantanées et pertinentes de PsyBot, 24h/24.
- Matching intelligent entre utilisateur et professionnel selon critères linguistiques, émotionnels, etc.
- Suivi émotionnel personnalisé : détection de l'état émotionnel, conseils, rappels, alertes.
- Prise de rendez-vous intuitive avec rappels automatiques.
- Tableau de bord professionnel pour psychologues et cliniques : statistiques, revenus, consultations.

7.2.7 Ressources clés

a) Quels sont nos actifs matériels, immatériels et humains essentiels ?

Ressources matérielles :

- Serveurs cloud (Render, Hugging Face, Firebase).
- Systèmes de sauvegarde locale sécurisée.

Ressources immatérielles :

- Modèle IA "PsyBot" pré-entraîné et affiné, multilingue.
- Base de données anonymisée pour amélioration continue.
- Propriété intellectuelle : marque, domaine, design UI/UX.
- Algorithme de matching émotionnel et linguistique.

Ressources humaines :

- Équipe produit (développeurs IA, frontend/backend, designers).
- Psychologues agréés pour supervision clinique.
- Chercheurs en psychologie et NLP pour validation scientifique.

b) Quels outils, technologies ou partenariats sont nécessaires pour réussir ?

Technologies essentielles :

— Hugging Face, Google Colab, Kaggle pour l'entraînement IA.

— FastAPI, FirebaseAuth, Firestore pour backend.

— React, Vite, StackBlitz pour frontend PWA.

— Intégration CIB et Edahabia pour paiements.

Partenariats stratégiques :

— Ordre des psychologues, universités, laboratoires pour validation scientifique.

c) Quels sont les principaux avantages concurrentiels de nos ressources ?

— Modèle IA localisé et multilingue (arabe, français, anglais).

— Base de données émotionnelle propriétaire pour apprentissage continu.

— Écosystème intégré : chatbot + psychologues + cliniques.

— Partenariats locaux de confiance renforçant la crédibilité.

— Équipe tech + psy : rare synergie assurant innovation et conformité.

7.2.8 Charges et Coûts

1. Coûts d'investissement initiaux (une fois, DZD)

Poste de dépense	**Description**	**Estimation (DZD)**
Développement de PsyBot	IA de base + fine-tuning	400 000 – 600 000
Développement application web	Front-end + back-end simple (inscriptions, RDV, chat)	300 000 – 500 000
Hébergement & cloud (1 an)	Serveurs, API, base de données (Hugging Face, Render...)	80 000 – 120 000
Design UX/UI	Interface simple et intuitive	50 000 – 100 000
Marketing de lancement	Réseaux sociaux, publicité locale (Facebook, Insta)	70 000 – 100 000
Total estimé		**900 000 – 1 420 000**

2. Revenus mensuels estimés (scénario modéré)

Source de revenu	**Hypothèse**	**Estimation mensuelle**
Abonnements psy (50 psys à 2 500 DA)	Forfait pour présence sur la plateforme	125 000 DA
Commissions consultations	300 consultations/mois Œ 100 DA de commission	30 000 DA
Licences cliniques (5 Œ 15 000 DA)	Accès à leur dashboard complet	75 000 DA
Publicité ou partenariats	Visibilité produits (ex. livres, apps santé)	25 000 – 50 000 DA
Total estimé		**255 000 – 350 000 DA**

3. Bénéfices nets estimés (par mois)

Mois	**Revenus estimés**	**Dépenses mensuelles**	**Bénéfices nets**
Mois 1	255 000 DA	150 000 DA	~105 000 DA
Mois 2	280 000 DA	150 000 DA	~130 000 DA
Mois 3	350 000 DA	160 000 DA	~190 000 DA
Année 1	~5 000 000 DA	~2 000 000 DA	~3 000 000 DA

b) Quels sont les coûts les plus importants ?

- Infrastructure cloud et IA (GPU, sécurité, scalabilité).
- Salaires techniques (développeurs backend / frontend).
- Marketing et acquisition (publicité, influenceurs psy).
- Paiements et commissions (passerelles, reversements).

c) Comment réduire les coûts ou améliorer l'efficacité ?

- Utiliser les offres gratuites et crédits startup (Render, Google Cloud...).
- Activer l'API IA à la demande (autoscaling intelligent).
- Prioriser le marketing organique (réseaux sociaux, lives, collaborations).
- Centraliser les outils (ex : Firebase tout-en-un).
- Travailler à distance pour éviter les frais fixes.

7.2.9 Revenus

a) Quels produits ou services nos clients sont-ils prêts à payer ?

- Consultations en ligne via chat sécurisé.
- Séances en clinique partenaire.
- Formations et webinaires thématiques animés par des professionnels.

b) Quels sont les moyens par lesquels nous générons des revenus ?

Source de revenu	**Description**
Commission par séance	∼10% prélevés sur chaque paiement effectué par le patient.
Abonnement psy Pro	2 500 DA/mois pour les psychologues : visibilité + outils (agenda, Dashboard, support).
Licence clinique annuelle	750 000 DA/an pour un accès multi-comptes + Dashboard + *chatbot* intégré.

c) Quel est notre modèle de tarification ?

Patients :

Service	**Tarif**
Utilisation de *PsyBot*	Gratuit (0 DA)
Consultation en ligne	500 DA/séance d1h (*Nous prenons 10%*)
Séance en clinique	Tarif librement fixé par la clinique (*Nous prenons 10%*)

Psychologues :

Offre	**Tarif**
Abonnement psy Pro	2 500 DA/mois

Cliniques :

Offre	**Tarif**
Licence annuelle	750 000 DA (inclut 5 comptes psy, support, *chatbot*)
Compte psy supplémentaire	+ 20 000 DA / an / psy

7.3 Conclusion

Le Business Model Canvas présenté dans ce chapitre a permis de structurer et de clarifier l'approche stratégique adoptée pour le déploiement du projet PsyBot. En définissant avec précision les segments de clientèle visés, les canaux de distribution, les sources de revenus, ainsi que les partenariats et ressources nécessaires, nous avons pu évaluer la faisabilité économique et l'impact potentiel de notre solution.

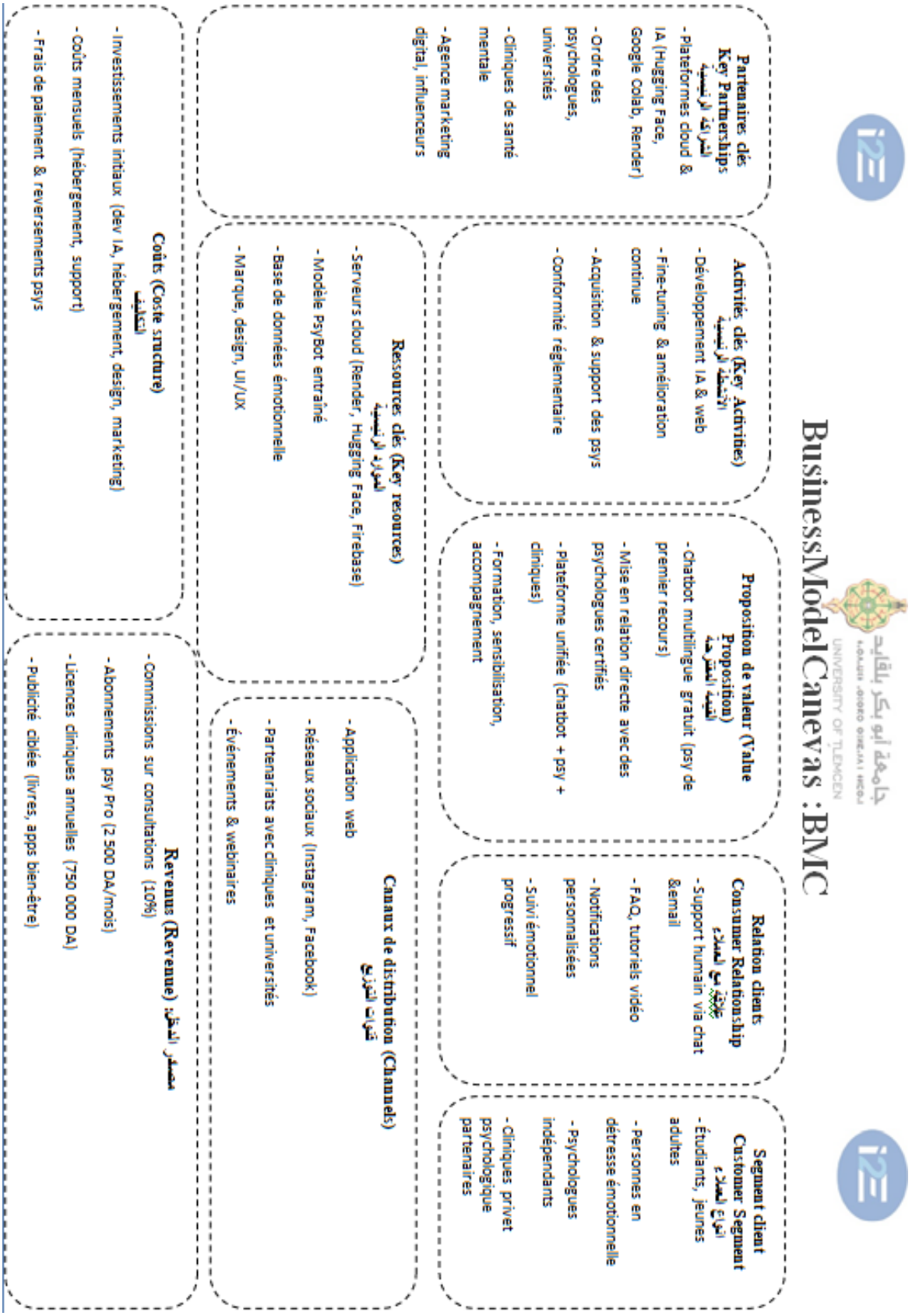


Figure 7.1 – BMC

Ce modèle met en évidence l'originalité de PsyBot dans le domaine de la e-santé mentale, en combinant intelligence artificielle, accessibilité numérique et accompagnement humain. Il constitue une base solide pour envisager des extensions futures du projet, qu'il s'agisse d'une version mobile, d'une internationalisation ou de l'ajout de nouvelles fonctionnalités cliniques.

Enfin, le BMC s'inscrit comme un support essentiel pour les démarches de financement, de partenariat et de communication autour du projet, tout en garantissant la cohérence entre les objectifs technologiques et les réalités du marché.

Références bibliographiques et webographiques

[1] J. Devlin, M.-W. Chang, K. Lee **and** K. Toutanova, **?**BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding**?**, *arXiv preprint arXiv:1810.04805*, 2019.

[2] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal **andothers**, **?**Language models are few-shot learners**?**, *Advances in Neural Information Processing Systems*, **jourvol** 33, **pages** 1877–1901, 2020.

[3] S. Russell **and** P. Norvig, *Artificial Intelligence: A Modern Approach*, 3rd. Prentice Hall, 2010.

[4] D. Jurafsky **and** J. H. Martin, *Speech and Language Processing*, 3rd. 2021, draft. **url**: `https://web.stanford.edu/~jurafsky/slp3/`.

[5] R. A. Calvo, D. N. Milne, M. S. Hussain **and** H. Christensen, **?**Natural language processing in mental health applications using non-clinical texts**?**, *Natural Language Engineering*, **jourvol** 23, **number** 5, **pages** 649–685, 2017.

[6] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[7] T. Hastie, R. Tibshirani **and** J. Friedman, *The Elements of Statistical Learning*. Springer, 2009.

[8] J. K-means, **?**Clustering Algorithms in Machine Learning**?**, *Journal of Machine Learning*, 2015.

[9] R. S. Sutton **and** A. G. Barto, *Reinforcement Learning: An Introduction*. MIT Press, 2018.

[10] I. Goodfellow, Y. Bengio **and** A. Courville, *Deep Learning*. MIT Press, 2016.

[11] C. D. Manning, P. Raghavan **and** H. Schütze, *Introduction to Information Retrieval*. Cambridge University Press, 2008.

[12] T. Young, D. Hazarika, S. Poria **and** E. Cambria, *Neural Network Methods for Natural Language Processing*. 2018.

[13] C. D. Manning **and** H. Schütze, *Foundations of Statistical Natural Language Processing*. MIT Press, 1999.

[14] I. H. Witten, A. Moffat **and** T. C. Bell, *Managing Gigabytes: Compressing and Indexing Documents and Images*. Morgan Kaufmann, 1999.

[15] T. Mikolov, I. Sutskever, K. Chen, G. Corrado **and** J. Dean, **?**Distributed Representations of Words and Phrases and their Compositionality**?**, **in** *Advances in Neural Information Processing Systems (NeurIPS)*, 2013.

[16] J. Pennington, R. Socher **and** C. D. Manning, **?**GloVe: Global Vectors for Word Representation**?**, **in** *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014.

[17] P. Bojanowski, E. Grave, A. Joulin **and** T. Mikolov, **?**Enriching Word Vectors with Subword Information**?**, *Transactions of the Association for Computational Linguistics*, **jourvol** 5, **pages** 135–146, 2017.

[18] S. Hochreiter **and** J. Schmidhuber, **?**Long Short-Term Memory**?**, *Neural Computation*, **jourvol** 9, **number** 8, **pages** 1735–1780, 1997.

[19] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk **and** Y. Bengio, **?**Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation**?**, **in** *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014.

[20] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser **and** I. Polosukhin, **?**Attention Is All You Need**?**, **in** *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

[21] J. Devlin, M.-W. Chang, K. Lee **and** K. Toutanova, *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*, 2018. eprint: `arXiv:1810.04805`.

[22] A. Radford, K. Narasimhan, T. Salimans **and** I. Sutskever, *Improving Language Understanding by Generative Pre-Training*, 2018.

[23] Y. Wu **andothers**, **?**Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation**?**, *arXiv preprint arXiv:1609.08144*, 2016.

[24] A. Radford, K. Narasimhan, T. Salimans **and** I. Sutskever, **?**Improving Language Understanding by Generative Pre-Training**?**, 2018.

[25] D. G. Myers, *Psychology*, 10th. Worth Publishers, 2013.

[26] U. Neisser, *Cognitive Psychology*. Appleton-Century-Crofts, 1967.

[27] C. J. Hutto, *vaderSentiment*, GitHub repository, 2014. **url**: `https://github.com/cjhutto/vaderSentiment`.

[28] D. P. Schultz **and** S. E. Schultz, *A History of Modern Psychology*, 11th. Cengage Learning, 2015.

[29] C. J. Goodwin, *A History of Modern Psychology*, 5th. Wiley, 2016.

[30] K. E. Stanovich, *How to Think Straight About Psychology*, 10th. Pearson, 2012.

[31] S. Freud, *The Ego and the Id*. International Psycho-Analytical Press, 1923.

[32] G. A. Miller, **?**The Cognitive Revolution: A Historical Perspective**?**, *Trends in Cognitive Sciences*, **jourvol** 7, **number** 3, **pages** 141–144, 2003.

[33] D. R. Shaffer **and** K. Kipp, *Developmental Psychology: Childhood and Adolescence*, 8th. Cengage Learning, 2008.

[34] J. C. Norcross, *Psychotherapy Relationships That Work: Evidence-Based Responsiveness*. Oxford University Press, 2011.

[35] J. O. Prochaska **and** J. C. Norcross, *Systems of Psychotherapy: A Transtheoretical Analysis*, 8th. Cengage Learning, 2014.

[36] American Psychological Association, **?**Evidence-based practice in psychology**?**, *American Psychologist*, **jourvol** 61, **number** 4, **pages** 271–285, 2006.

[37] P. Ekman, **?**An argument for basic emotions**?**, *Cognition and Emotion*, **jourvol** 6, **number** 3-4, **pages** 169–200, 1992.

[38] W. James, **?**What is an emotion?**?**, *Mind*, **jourvol** 9, **number** 34, **pages** 188–205, 1884.

[39] W. B. Cannon, **?**The James-Lange theory of emotions: A critical examination and an alternative theory**?**, *The American Journal of Psychology*, **jourvol** 39, **number** 1/4, **pages** 106–124, 1927.

[40] K. R. Scherer, **?**What are emotions? And how can they be measured?**?**, *Social Science Information*, **jourvol** 44, **number** 4, **pages** 695–729, 2005.

[41] W. H. Organization, **?**Mental health and COVID-19: Early evidence of the pandemics impact**?**, World Health Organization, 2021. **url**: `https://www.who.int/publications/i/item/9789240026618`.

[42] American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*. American Psychiatric Publishing, 2013.

[43] A. T. Beck **and** B. A. Alford, *Depression: Causes and Treatment*, 2nd. University of Pennsylvania Press, 2008.

[44] P. Cuijpers, E. Karyotaki, M. Reijnders **and** M. Purgato, **?**Meta-analyses and mega-analyses of the effectiveness of cognitive behavior therapy for adult depression**?**, *World Psychiatry*, **jourvol** 15, **number** 3, **pages** 245–258, 2016.

[45] R. Kohn, S. Saxena, I. Levav **and** B. Saraceno, **?**The treatment gap in mental health care**?**, *Bulletin of the World Health Organization*, **jourvol** 82, **number** 11, **pages** 858–866, 2004.

[46] A. Gulliver, K. M. Griffiths **and** H. Christensen, **?**Perceived barriers and facilitators to mental health help-seeking in young people: A systematic review**?**, *BMC Psychiatry*, **jourvol** 10, **number** 1, **page** 113, 2010.

[47] P. W. Corrigan, **?**How stigma interferes with mental health care**?**, *American Psychologist*, **jourvol** 59, **number** 7, **pages** 614–625, 2004.

[48] N. Ramirez-Esparza **andothers**, **?**Depressed and non-depressed individuals' word use on the Internet: A content analysis**?**, *Journal of Language and Social Psychology*, **jourvol** 27, **number** 4, **pages** 387–394, 2008.

[49] A. S. Miner **andothers**, **?**Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health**?**, *JAMA Internal Medicine*, **journvol** 176, **number** 5, **pages** 619–625, 2016.

[50] K. K. Fitzpatrick, A. Darcy **and** M. Vierhile, **?**Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial**?**, *JMIR Mental Health*, **journvol** 4, **number** 2, e19, 2017.

[51] S. M. Mohammad **and** P. D. Turney, **?**Crowdsourcing a word–emotion association lexicon**?**, *Computational Intelligence*, **journvol** 29, **number** 3, **pages** 436–465, 2013.

[52] R. Xu **and** T.-L. Wong, **?**Emotion recognition from text using deep learning: A survey**?**, *IEEE Access*, **journvol** 8, **pages** 48 674–48 688, 2020.

[53] B. D. Mittelstadt **andothers**, **?**The ethics of algorithms: Mapping the debate**?**, *Big Data & Society*, **journvol** 4, **number** 2, **pages** 1–20, 2017.

[54] A. Jobin, M. Ienca **and** E. Vayena, **?**The global landscape of AI ethics guidelines**?**, *Nature Machine Intelligence*, **journvol** 1, **number** 9, **pages** 389–399, 2019.

[55] D. Demszky, A. Makharia, B. Hermes, J. McAuley **and** D. Jurafsky, **?**GoEmotions: A dataset of fine-grained emotions**?**, **in** *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL 2020)*, 2020, **pages** 4040–4054.

[56] NomAuteur, *Titre du dépôt GitHub*, GitHub, 2024. **url**: `https://github.com/jsong336/emotion-bert`.

[57] R. S. Sitoula, M. Pramanik **and** R. Panigrahi, **?**Fine-Grained Classification for Emotion Detection Using Advanced Neural Models and GoEmotions Dataset**?**, techreport, 2024.

[58] Hugging Face, *Perplexity — Hugging Face Documentation*, 2023. **url**: `https://huggingface.co/docs/transformers/en/perplexity`.

[59] K. Papineni, S. Roukos, T. Ward **and** W.-J. Zhu, **?**BLEU: a method for automatic evaluation of machine translation**?**, **in** *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, 2002, **pages** 311–318.

[60] C.-Y. Lin, **?**ROUGE: A package for automatic evaluation of summaries**?**, **in** *Text Summarization Branches Out: Proceedings of the ACL-04 Workshop*, 2004, **pages** 74–81.

[61] S. Banerjee **and** A. Lavie, **?**METEOR: An automatic metric for MT evaluation with improved correlation with human judgments**?**, **in** *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for MT and/or Summarization*, 2005, **pages** 65–72.

[62] H. Qiu, A. Li, L. Ma **and** Z. Lan, **?**PsyChat: A Client-Centric Dialogue System for Mental Health Support**?**, techreport, 2023.

[63] D. Dharrao **and** S. Gite, **?**TherapyBot: a chatbot for mental well-being using transformers**?**, techreport, 2024.

Résumé

Ces dernières années, les enjeux liés à la santé mentale ont pris une place centrale dans nos sociétés. En parallèle, les technologies numériques, et en particulier l'intelligence artificielle, offrent de nouvelles pistes pour accompagner les individus de manière innovante. Ce mémoire explore la mise en place d'un chatbot capable d'interagir avec des utilisateurs en tenant compte de la nature de leurs demandes, en particulier dans un contexte psychologique. Le fonctionnement du système repose sur plusieurs étapes successives. Une première analyse, effectuée avec le modèle DistilBert avec un fine-tuning, permet de classer les requêtes entre celles qui relèvent d'une simple recherche d'information, et celles à caractère psychologique. Les requêtes informatives sont traitées, par la suite, à l'aide de la bibliothéque duckduckgo_search, qui fournit une réponse issue de sources web fiables. Lorsqu'il s'agit d'une demande plus sensible, une détection émotionnelle est lancée grâce à une combinaison de BERT et VADER, afin d'évaluer l'intensité du message. Si l'émotion détectée reste dans un seuil jugé acceptable, une réponse est générée par le modèle GPT. Dans le cas contraire, l'utilisateur reçoit un message de redirection ou de prudence. Ce travail met en lumière les apports, mais aussi les limites, d'un tel système dans un domaine aussi délicat. L'architecture proposée, à la fois simple et évolutive, constitue une base solide pour envisager de futurs outils d'assistance psychologique automatisée.

Mots-clés: Santé mentale, Chatbot psychologique, Intelligence artificielle, Traitement du langage naturel, Émotion, Intention, Interaction humain-IA.

ملخص

في السنوات الأخيرة، أصبحت قضايا الصحة النفسية أكثر حضورًا وأهمية في مجتمعاتنا. في المقابل، توفّر التقنيات الرقمية، وخاصة الذكاء الاصطناعي، آفاقًا جديدة لمرافقة الأفراد بطريقة مبتكرة. يتناول هذا البحث تطوير روبوت محادثة (Chatbot) قادر على التفاعل مع المستخدمين بناءً على طبيعة استفساراتهم، وخصوصًا في السياقات النفسية. يعتمد النظام على سلسلة من المراحل المتتالية. يتم أولاً تصنيف نوع الطلب باستخدام DistilBERT لتحديد ما إذا كان طلبًا لمعلومة عامة، أو استفسارًا ذا طابع نفسي. بعد ذلك تُعالج الطلبات المعلوماتية باستخدام أداة duckduckgo_search التي تُرجع إجابات من مصادر موثوقة على الإنترنت. بالنسبة للطلبات ذات الطابع النفسي، يتم تحليل المشاعر باستخدام مزيج من النموذجين BERT وVADER لتقييم شدة الانفعالات. أخيرا إذا كانت الحالة العاطفية ضمن الحدود المقبولة، يتم توليد الرد باستخدام GPT. وإذا كانت المشاعر قوية جدًا، يتم إرسال رسالة تحذيرية أو توجيه المستخدم إلى مساعدة أخرى. يسلّط هذا العمل الضوء على إمكانيات وحدود مثل هذا النظام في مجال حساس. ويقترح بنية بسيطة وقابلة للتطوير، تمثّل أساسًا لتطوير أدوات مستقبلية للدعم النفسي المعتمد على الذكاء الاصطناعي.

الكلمات الدالة :الصحة النفسية، روبوت محادثة نفسي، الذكاء الاصطناعي، معالجة اللغة الطبيعية، المشاعر، النية، التفاعل بين الإنسان والآلة.

Abstract

In recent years, mental health has become an increasingly important issue in our societies. At the same time, digital technologies-especially artificial intelligence—offer new ways to support individuals in innovative ways. This thesis explores the development of a chatbot designed to interact with users based on the nature of their requests, particularly in a psychological context. The system operates through several key steps. First, an initial classification using the model DistilBERT with fine-tuning identifies whether the user's request is a general information query or a psychologically oriented question. Next, an informational requests are handled using duckduckgo_search, which retrieves answers from reliable web sources. In contrast, psychological queries undergo an emotional analysis using a combination of BERT and VADER to determine the intensity of the user's emotions. Finaly, if the emotional state is within an acceptable threshold, a response is generated using GPT; otherwise, the user receives a cautionary message or is redirected accordingly. This work highlights both the potential and the limitations of such a system in a sensitive domain. The proposed architecture, which is simple yet scalable, provides a solid foundation for the development of future AI-based psychological support tools.

Keywords: Mental health, Psychological chatbot, Artificial intelligence, Natural language processing, Emotion, Intent, Human-AI interaction.