

Un système d'aide au dialogue en santé intime des femmes

Xingyu Liu^{1,2}, François Portet¹, Didier Schwab¹, Juliette Mauro²

¹ Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, 38000 Grenoble, France
² Shesmet

PFIA - Journée Santé & IA
6 juillet 2023



Plan

Introduction

Système d'aide au dialogue envisagé

Un module du système : module de recherche de réponse

ChatGPT en dialogue sur la santé intime des femmes

Travaux futurs

Contexte

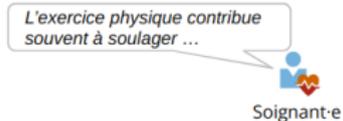
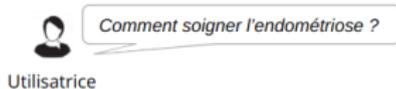
- ▶ Santé intime des femmes :
 - ▶ sujet encore peu abordé (cf. rapport sénat santé femmes¹)
 - ▶ sujet encore tabou alors qu'il impacte le bien-être mental et physique des femmes (cf. dossier fondation pour la recherche sur le cancer)²
- ▶ MySLife³ a été créé par la société Shesmet pour informer les femmes sur leur santé intime : service d'échange avec un·e soignant·e dans l'application
- ▶ Pour le soutien à l'échange entre utilisatrice et soignant·e : besoin d'un système d'aide au dialogue
⇒ Thèse CIFRE (collaboration entre le Laboratoire d'Informatique de Grenoble (LIG) et Shesmet)

1. Rapport du Sénat : santé des femmes

2. Santé sexuelle : briser les tabous

3. <https://myslife.co/mobile/>

Apports du système d'aide au dialogue



Apports du système d'aide au dialogue



1. Diminuer le temps de saisie de réponse
2. Proposer un cadre de réponse commun
3. Gérer plusieurs threads
4. Laisser de la liberté aux soignant.e-s.

Défis scientifiques de la thèse

1. Beaucoup de système de dialogue dans la santé (santé mentale [Fitzpatrick et al., 2017, Miner et al., 2016, Hudlicka, 2013]; psychiatrie [Vaidyam et al., 2019]) mais pas sur la santé intime des femmes.
Système de recommandation de réponses (aide au dialogue) beaucoup moins populaire en recherche.
2. Une tâche peu dotée : peu de ressources dans le domaine en français
3. Système industriel nécessitant une maintenance et une adaptation aux évolutions des informations dans le domaine
4. Adaptation des réponses aux utilisatrices

Plan

Introduction

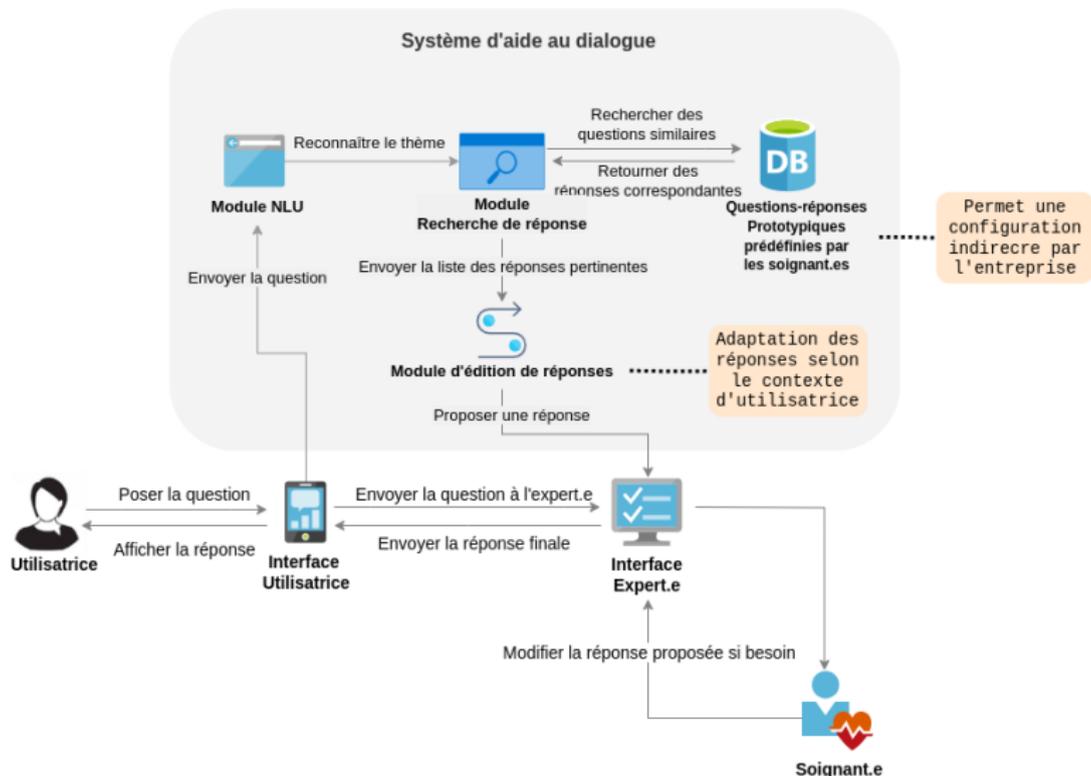
Système d'aide au dialogue envisagé

Un module du système : module de recherche de réponse

ChatGPT en dialogue sur la santé intime des femmes

Travaux futurs

Système d'aide au dialogue envisagé



Plan

Introduction

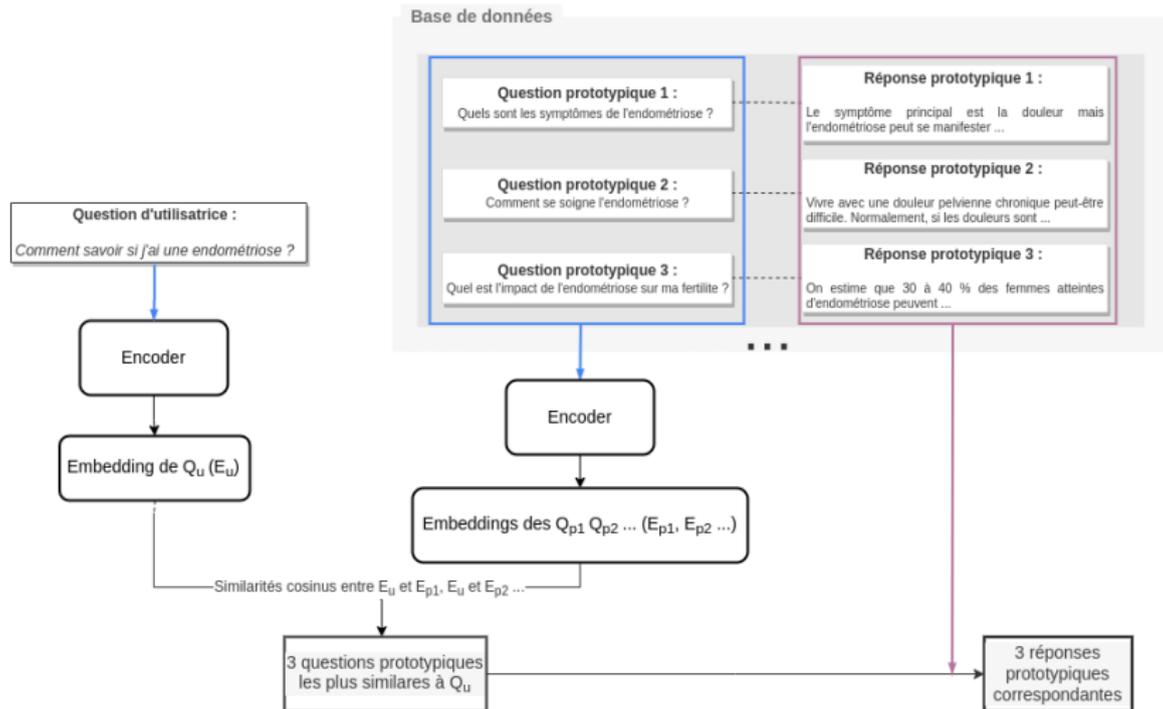
Système d'aide au dialogue envisagé

Un module du système : module de recherche de réponse

ChatGPT en dialogue sur la santé intime des femmes

Travaux futurs

Approche de recherche de réponse par la similarité sémantique



Expériences

- ▶ Méthodes de plongement de phrases : moyenne pondérée des vecteurs des mots avec FastText, Sentence-BERT (SBERT) et Universal Sentence Encoder (USE)
- ▶ Données d'entraînement (pour FastText) : corpus Doctissimo-Gynécologie (491,738 paires de questions-réponses) [Nivon, 2021]
- ▶ Données d'évaluation : 100 paires de questions-réponses (28 paires de questions-réponses prototypiques écrites par les soignant·e·s + augmentation de données par traduction aller-retour)

Résultats

- ▶ Métriques d'évaluation : *accuracy* et *Mean Reciprocal Rank* (MRR) (cf. équation (1)). Pour décider si une réponse retournée est correcte, la réponse doit respecter deux contraintes :
 1. la réponse correspondante dans le corpus d'évaluation se trouve dans la liste de réponses retournées par le système ;
 2. le score de similarité entre la réponse retournée et la réponse d'évaluation dépasse 0,5.

$$MRR = \frac{1}{Q} \sum_{i=1}^Q \frac{1}{rang_i} \quad (1)$$

Méthode	Données de test	Accuracy	MRR
FastText entraîné sur Doctissimo	eval-100	0.56	0.82
Universal Sentence Encoder	eval-100	0.37	0.83
Sentence-Bert	eval-100	0.50	0.79

Discussion sur les résultats

- ▶ Scores MRR sont relativement bons pour toutes les trois méthodes : si la bonne réponse a été trouvée, elle est souvent au premier ou deuxième rang
- ▶ Scores d'accuracy sont faibles pour les trois méthodes :
USE et SBERT non adaptés à nos données contenant une part importante de termes concernant la santé intime des femmes
- ▶ Une piste d'amélioration : créer un jeu de données spécifique à la santé intime des femmes avec lequel nous pouvons ajuster USE et SBERT

Plan

Introduction

Système d'aide au dialogue envisagé

Un module du système : module de recherche de réponse

ChatGPT en dialogue sur la santé intime des femmes

Travaux futurs

ChatGPT en dialogue sur la santé intime des femmes

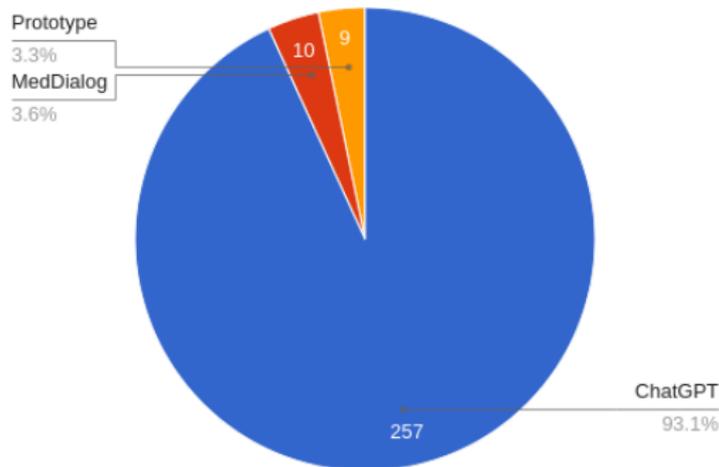
- ▶ 30 novembre 2022, l'arrivée de ChatGPT provoque à la fois de la panique et de l'emballement dans divers domaines
- ▶ ChatGPT sur les sujets de la santé intime des femmes : réponses riches en informations et empathiques ; variation de réponses selon le contexte (par exemple, réponse sur la conception adaptée à l'âge de femme)

Évaluation des réponses générées par ChatGPT

- ▶ Objectif : tester la pertinence des réponses ChatGPT sur la santé intime des femmes et le niveau d'acceptabilité des soignant-e-s
- ▶ Source de questions : 150 questions de la santé intime des femmes depuis le corpus anglais MedDialog [Zeng et al., 2020], corpus scrappé depuis un site de consultation en ligne, traduites en français
- ▶ Source de réponses : réponse prototypique, réponse ChatGPT (version 13 février 2023), réponse MedDialog
- ▶ Soignantes sollicitées : 1 médecin gynécologue, 2 sages-femmes

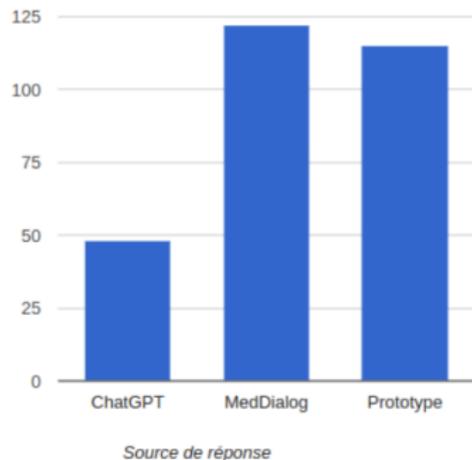
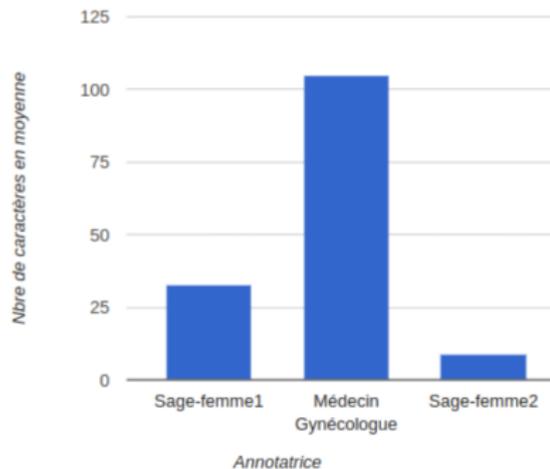
Évaluation des réponses générées par ChatGPT (analyse quantitative)

- ▶ ChatGPT : 257/276
- ▶ MedDialog : 10/276
- ▶ Prototype : 9/276



Proportion des sources de réponses choisies par soignant

Évaluation des réponses générées par ChatGPT (analyse quantitative)



Distance de modification

Limites à l'utilisation de ChatGPT en entreprise

Pour une tâche spécifique sur le terrain (i.e. QA en santé intime des femmes) :

1. Commentaires des soignantes sur les réponses ChatGPT :
 - + riches en informations pertinentes
 - + empathiques
 - parfois contenant des redondances
2. Confidentialité des données : serveurs aux États-Unis
3. Origine morale et éthique : par exemple, demande relative à l'IVG

Plan

Introduction

Système d'aide au dialogue envisagé

Un module du système : module de recherche de réponse

ChatGPT en dialogue sur la santé intime des femmes

Travaux futurs

Travaux futurs

- ▶ Approches par apprentissage automatique nécessitant de grands corpus équilibrés vs. Accès difficile aux données de la santé en français : traduction du corpus anglais MedDialog
- ▶ Dialogue à tour unique (question-réponse) vers dialogue contextuel : utilisation des connaissances des soignant-e dans le domaine cible pour guider le dialogue
- ▶ Évaluation du système d'aide au dialogue : mesure du gain temporel et de l'effet d'un cadre commun

Références I



Fitzpatrick, K. K., Darcy, A., and Vierhile, M. (2017).

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, 4(2) :e19.



Hudlicka, E. (2013).

Virtual training and coaching of health behavior : Example from mindfulness meditation training.
Patient education and counseling, 92(2) :160–166.



Miner, A. S., Milstein, A., Schueller, S., Hegde, R., Mangurian, C., and Linos, E. (2016).

Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health.
JAMA internal medicine, 176(5) :619–625.



Nivon, Q. (2021).

Study of the adaptation of generalist chatbots to a gendered application domain : Application to the women's health's dialogues.
Research project report, MOSIG.



Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S., and Torous, J. B. (2019).

Chatbots and conversational agents in mental health : a review of the psychiatric landscape.
The Canadian Journal of Psychiatry, 64(7) :456–464.



Zeng, G., Yang, W., Ju, Z., Yang, Y., Wang, S., Zhang, R., Zhou, M., Zeng, J., Dong, X., Zhang, R., et al. (2020).

Meddialog : Large-scale medical dialogue dataset.
In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.