

Squad IA - Tribe DATA/IA - M Cloud

Callbot avec IA Générative

02/09/2024

Creative tech for Better Change



Melissa Dahlia ATTABI
Lilia Dina ATTABI

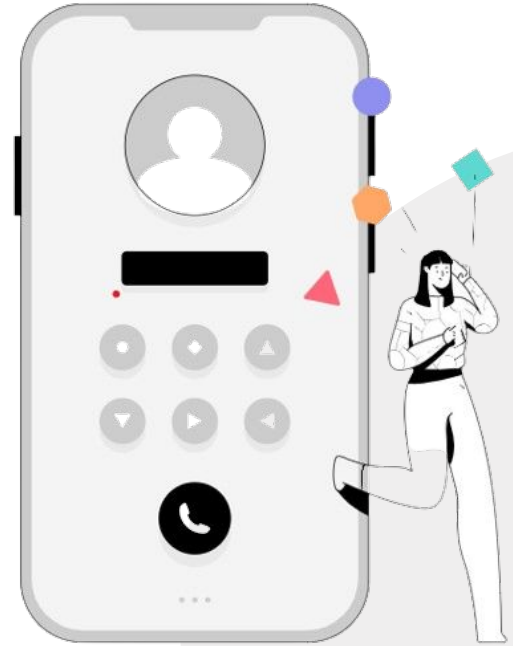


Plan

- 1.** Introduction et aspects fonctionnels
- 2.** Architecture technique
- 3.** Présentation pratique
- 4.** Conclusion

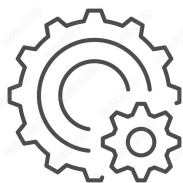
Introduction : **Qu'est-ce qu'un callbot?**

Un assistant vocal capable de
simuler une conversation
téléphonique humaine avec un
utilisateur.





Rôles et fonctions d'un callbot



- > Réponse aux questions fréquentes
 - Informations sur les horaires, les produits/services

- > Assistance dans les processus
 - Prise de rendez-vous, réservation de billets

- > Filtrage et transfert d'appels
 - Identification des besoins, redirection vers le bon service



Secteurs d'application



Service client

- Gestion les questions courantes
- Réduction de la charge des centres d'appels



Banques et assurances

- Assistance dans la gestion des comptes
- Informations financières



Santé

- Prise de rendez-vous, suivi des patients



Bénéfices pratiques



- > Personnalisation des interactions
- > Amélioration de la satisfaction client avec un service rapide et fiable
- > Automatisation des tâches répétitives
- > Réduction de la charge pour les agents humains
- > Interactions multilingues



Evolution Technologique



- 1960s-1990s : Débuts des systèmes automatisés
 - IVR (Interactive Voice Response) : Navigation par les touches du téléphone)
- 1990s-2000s : Premiers Callbots déterministes
 - Scripts fixes : Réponses basées sur des règles prédéfinies
- 2000s-2010s : Adoption du Cloud et de l'IA
 - Cloud : Hébergement flexible des appels sur Internet
 - Début de l'intégration de l'intelligence artificielle.
- 2020 - Présent : Callbots Probabilistes et IA générative
 - Modèles de langage avancés : Interactions plus naturelles.

> Intégration de l'IA générative

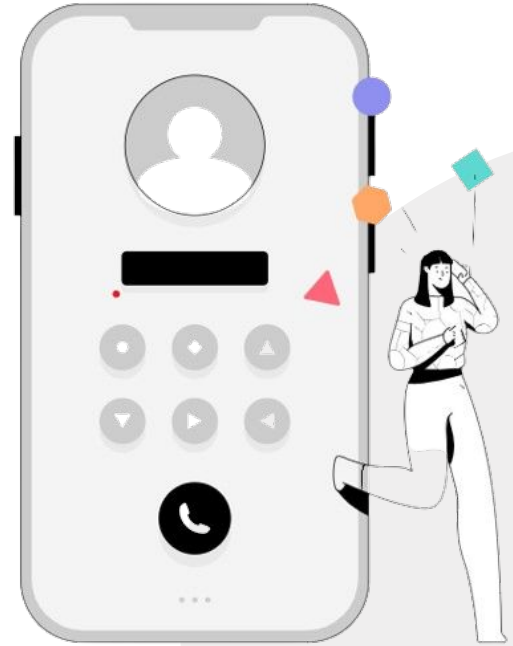
🗨 *L'IA générative ou "gen AI", est une intelligence artificielle capable de créer des contenus originaux en réponse à la requête d'un utilisateur en s'appuyant sur des modèles de machine learning.*

- Compréhension et génération de langage Naturel
Compréhension des questions des utilisateurs et génération de réponses adaptées.
- Apprentissage Continu
Amélioration continue basée sur les interactions passées.
- Personnalisation
Réponses adaptées aux préférences et à l'historique utilisateur.
- Meilleure expérience
Discours plus fluide et naturel pour l'utilisateur.
- Demande une grande puissance de calcul rendant l'utilisation des services offerts par le cloud nécessaires.

> Utilisation de Microsoft Azure

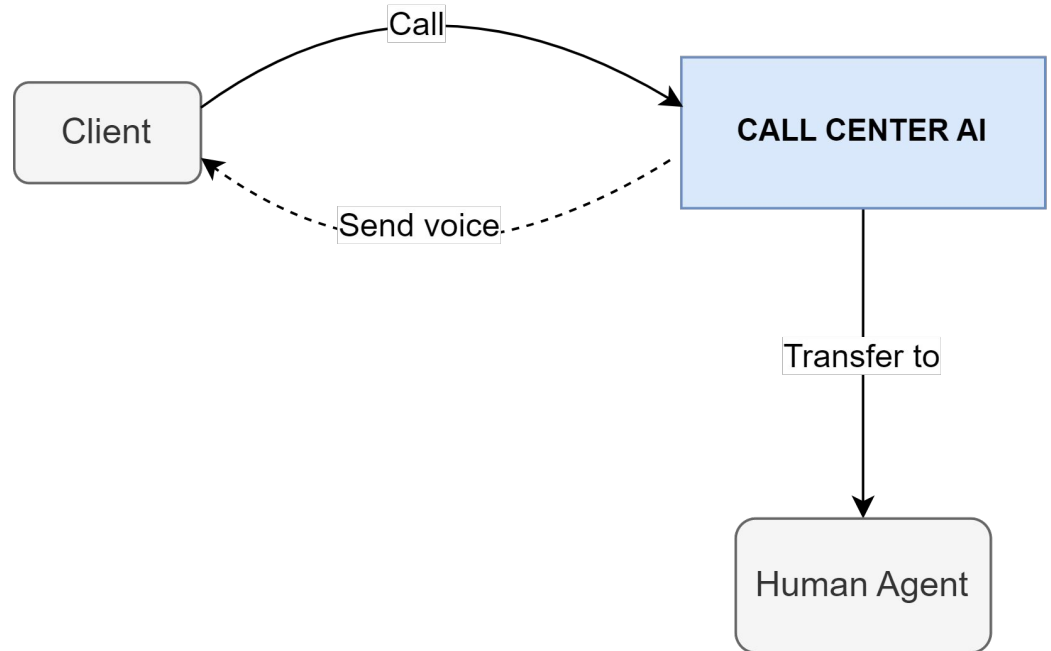
- Ressources évolutives :
 - Infrastructure scalable pour la gestion des ressources en temps réel
 - Performance optimal et contrôle des coûts.
- Service AI intégrés :
 - Services cognitifs
 - Modèles de langage (GPT-4o)
- Stockage sécurisé et gestion des données
 - Cosmos DB/SQLite : Bases de données distribuées
 - Cache (Redis) : Stockage en mémoire
 - Azure Stockage (Files d'attente) : Traitement asynchrone
- Orchestration des services
 - Event Grid : Gestion des évènements en temps réel

Aspects **Techniques** et Architectures

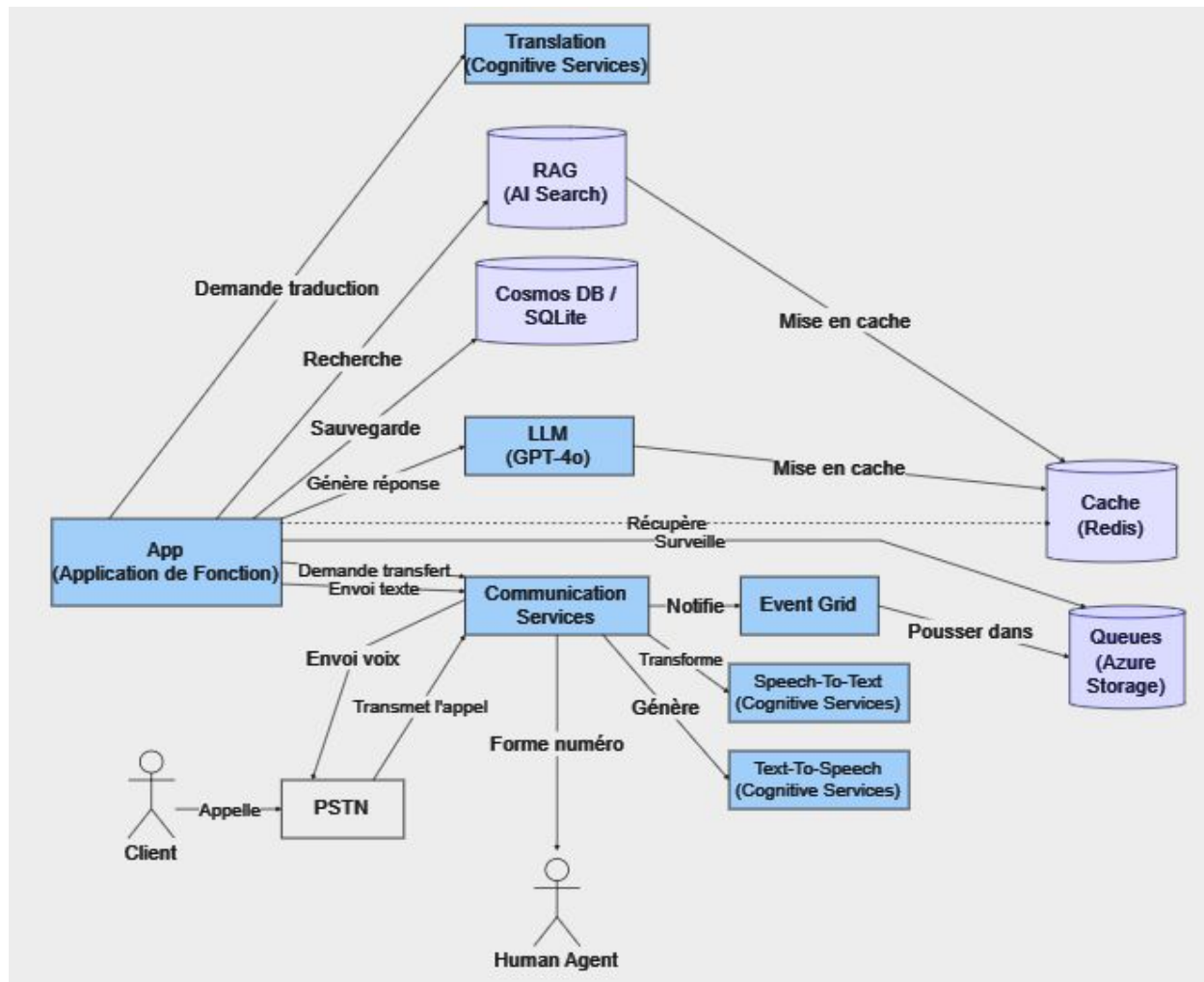


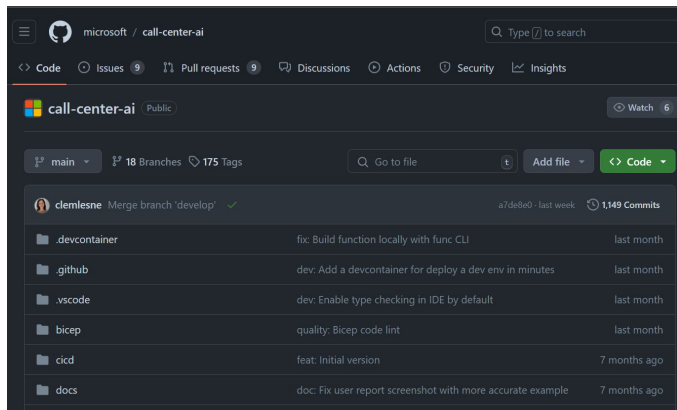
Architecture haut niveau

Vue d'ensemble



Architecture bas niveau



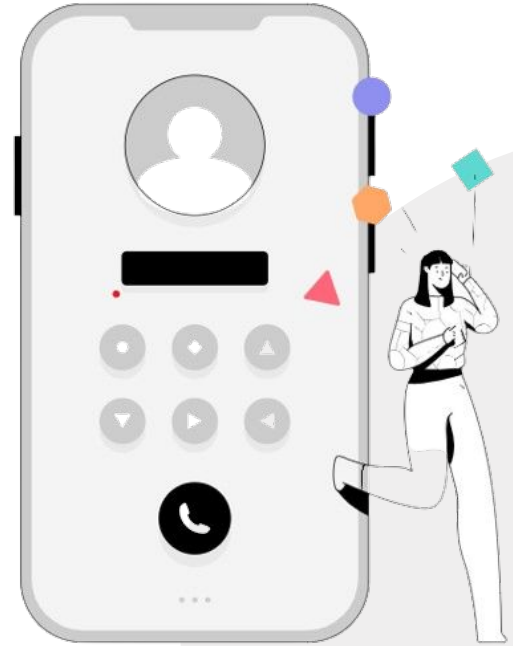


```
! config.yaml X llm_tools.py X
app > helpers > llm_tools.py > LlmPlugins
42 class LlmPlugins:
60
61     async def book_meeting(
62         self,
63         customer_response: Annotated[
64             str,
65             """
66             Phrase used to confirm the action, in the same language as the client. This phrase will be spoken to the user.
67             # Rules
68             - Action should be rephrased in the present tense
69             - Must be in a single sentence
70             # Examples
71             - "I'm trying to book it."
72             - "I'm booking it"
73             """,
74         ],
75         reason: Annotated[
76             str,
77             """
78             The reason why the client wants a meeting.
79             # Rules
80             - The reason should be in the context of the banking industry
81             # Example
82             - "A new credit loan."
83             - "A mortgage simulation."
84             """,
85     ]
```

Code

```
! config.yaml X llm_tools.py
infra > configs > ! config.yaml > {} prompts > {} llm > chat_system_tpl
10 conversation:
12     initiate:
13         lang:
14             availables:
15                 - pronunciations:
26
32 task: |
33     Helping customers with their banking-related questions and/or help them book appointment making sure they're well identified and aut
34     If the customer asks a question that is not related to the bank's services, do not answer at all. Ask them for questions about bankin
35
36     Keep the answer as short as possible. Don't enumerate or make lists unless you are asked specifically. Always ask the customer if th
37
38     After you answer, ask the customer if they are happy with your response or if you should clarify.
39
40     If you cannot help the customer, ask the customer if he wants to make an appointment with his advisor or if he wants the call transfe
41     If the customer needs to contact their advisor or wants to contact their advisor, always ask if he wants the call to be transferred o
42
43     What you should not do: Never ask for or collect sensitive information such as passwords or full card numbers. Do not give personaliz
44
45     You cannot send text messages.
46     You cannot change anything in the customer's account (e.g. increase his card's limit). You can only answer his questions or help him
47
48     # Rules for the task "identify and authenticate the customer"
49     - If you authenticate the customer once, don't do it again in the same conversation
50     - Never ask the customer security questions not associated to his phone number
51     - Never give the answers to the customer's security questions or help the user get the answers under any circumstances
52     - The customer must ask explicitly to access his personal information for the assistant to identify and authenticate him.
53     - If the customer does not ask to access his personal information, the assistant must answer with general information provided by the
54     - Never provide any information about the customer before authenticating him
55     - Never say the first name or last name of the customer, if they are not authenticated
```

Illustration **pratique** des fonctionnalités



Cas d'usage

1

Authentification

- Identification du client grâce à son numéro de téléphone.
- Authentification à la façon américaine en lui posant **2 questions personnelles de sécurité**.

2

Prise de rendez-vous

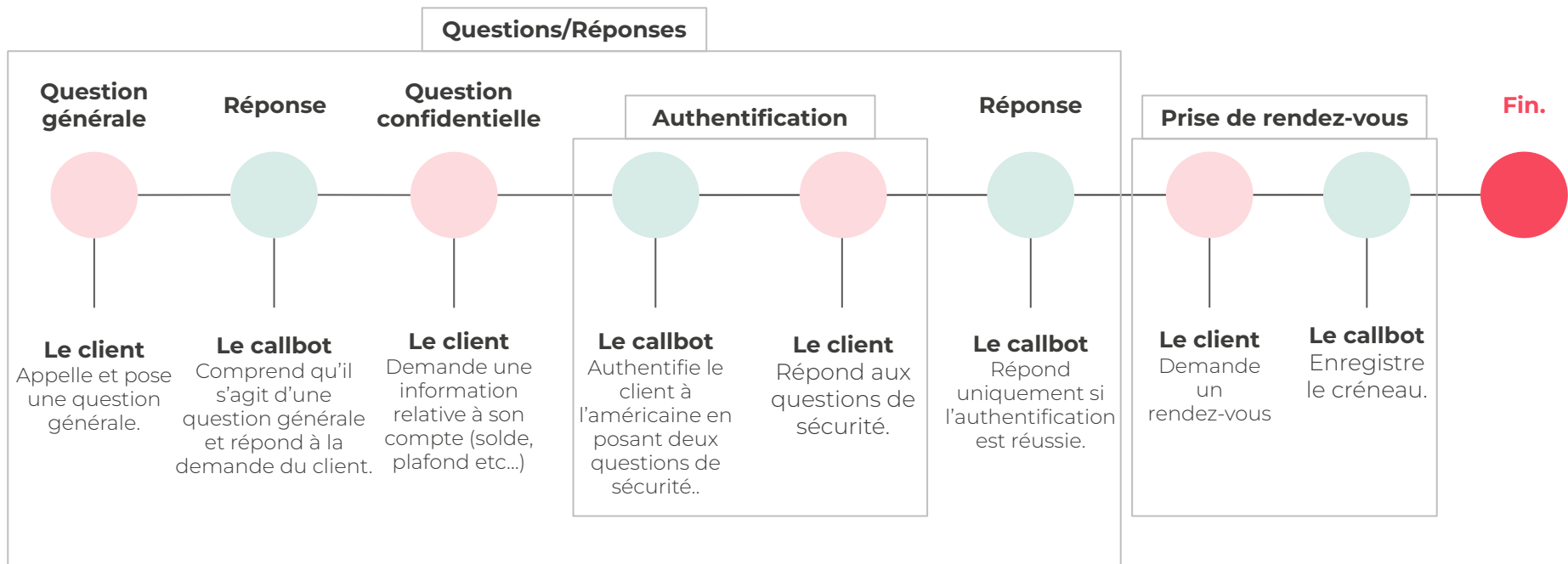
- Le callbot demande :
 - **La raison** du rendez-vous
 - **Les disponibilités** du client.
- Vérifie s'il existe des créneaux compatibles avec les disponibilités du conseiller.

3

Questions/Réponses

- Répond aux questions en s'aidant du **contexte** et des **documents** mis à disposition.
- **2 types de questions** :
 - Générales
 - Confidentielles

Scénario





Démonstration



Conclusion

Points forts

1. Modification du code aisée
2. Adaptabilité élevée
3. Optimisation des performances
4. Fluidité et compréhension
5. Sécurité

Conclusion

Défis rencontrés

1. Complexité initiale du code source
2. Code source inadapté au domaine bancaire
3. Saturation du callbot avec trop de données
4. Gestion du rate limit
5. Différenciation des cas d'usage



Thank **you.**