

Doctolib

Création d'un système *Agentic AI*

Mise en œuvre pratique de l'IA chez
Doctolib

Speakers of the day



Goulven LE DÜ

Staff engineer



Florian DUTRONC

Senior software engineer

Notre North Star

Together, let's build
the healthcare we all dream of.

Nos missions

#1



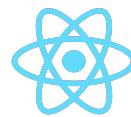
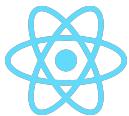
We improve the
daily lives of
health
professionals.

#2



We help people
to be **healthier**.

Notre stack technique



React Native



Ruby



Summary

1

Contexte

2

C'est quoi Agentic?

3

Alfred

4

Défis techniques &
décisions

5

Implémentation
technique

6

Et après ?

Contexte

Croissance de la plateforme ➔ croissance des demandes au support

Objectif : Avoir un coût de support soutenable tout en préservant la satisfaction

Gérer des demandes complexes qui nécessitent un raisonnement et la réalisation d'actions

C'est quoi Agentic ?

Un réseau d'agents IA spécialisés, travaillant ensemble comme une équipe d'experts bien coordonnée vers un objectif commun



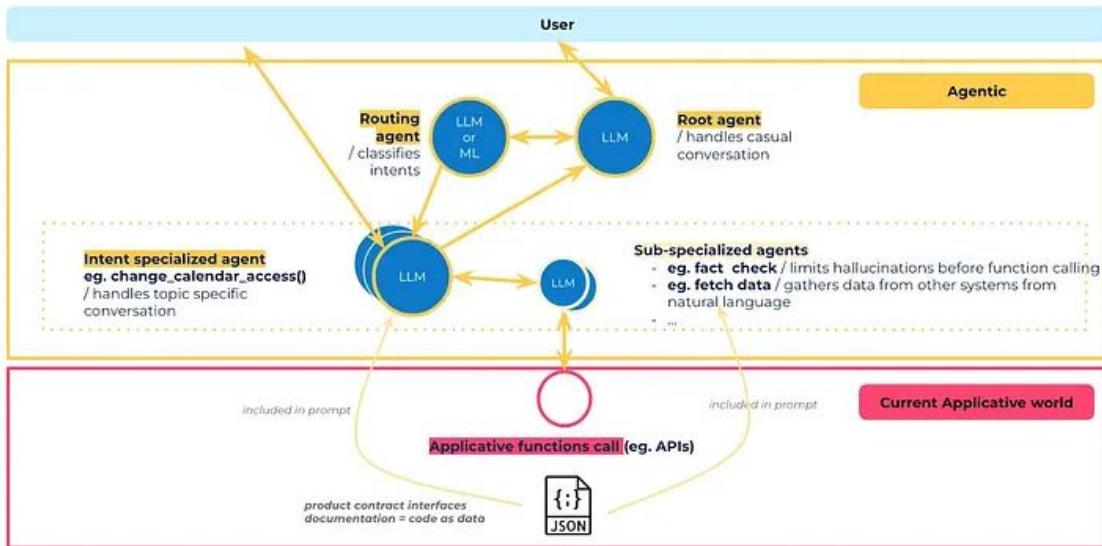
Agent IA

= un LLM ayant

- Un modèle (GPT4o, Llama, Claude, etc.)
- Un *prompt* spécifique qui définit son rôle, son contexte et son expertise
- Un ensemble spécifique d'outils qu'il peut utiliser

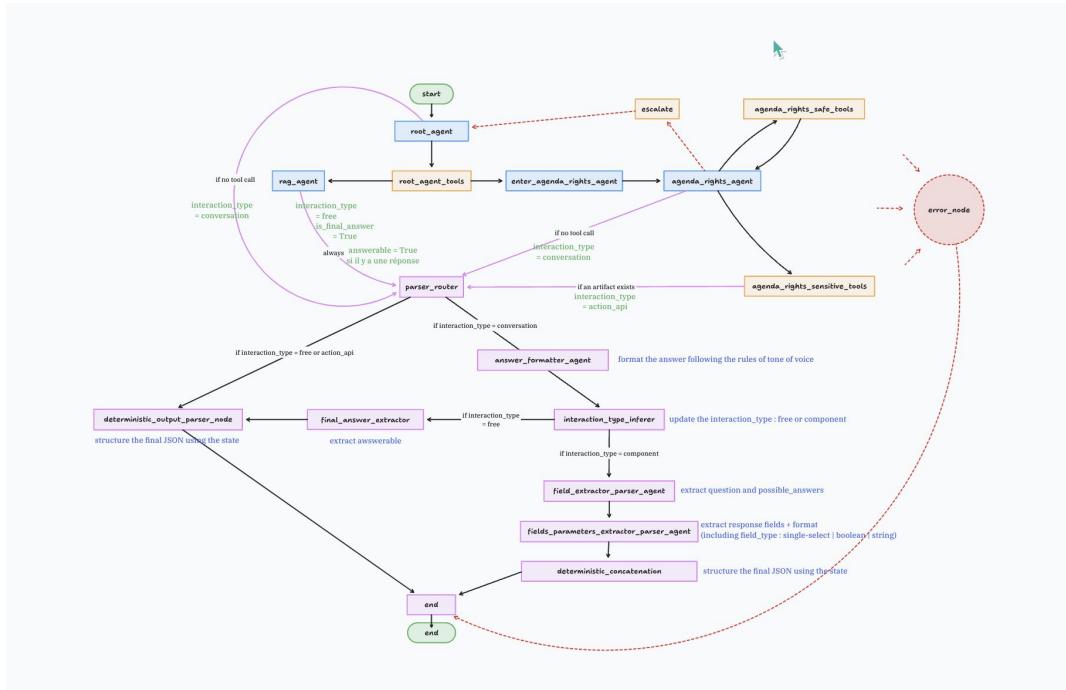
C'est quoi Agentic ?

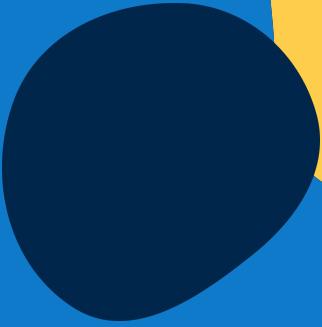
Une structure de graphe orienté



C'est quoi Agentic ?

LangGraph sous le capot





3

Alfred

L'expérience du chatbot débile



Please select an option above.

Let me see main menu



Please select an option above.

See original options



Please select an option above.

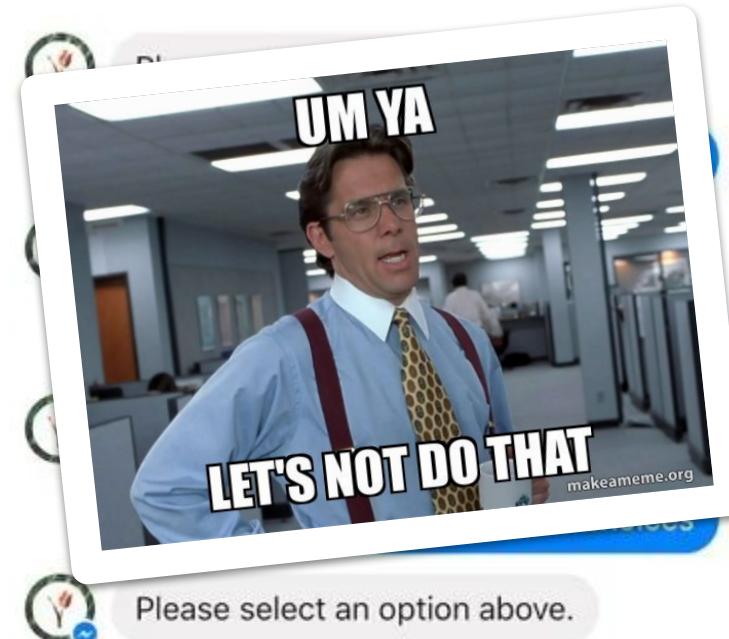
Go back to all choices



Please select an option above.

This bot is stupid

L'expérience du chatbot débile



Please select an option above.

This bot is stupid

Alfred ! Le bouton rouge, s'il te plaît



- Comprend l'intention de la demande
- Sait quelles questions poser et quand les poser
- Collecte des informations depuis nos systèmes (API)
- Présente une solution claire
- Tout ça via une UI dynamique

Cas réel

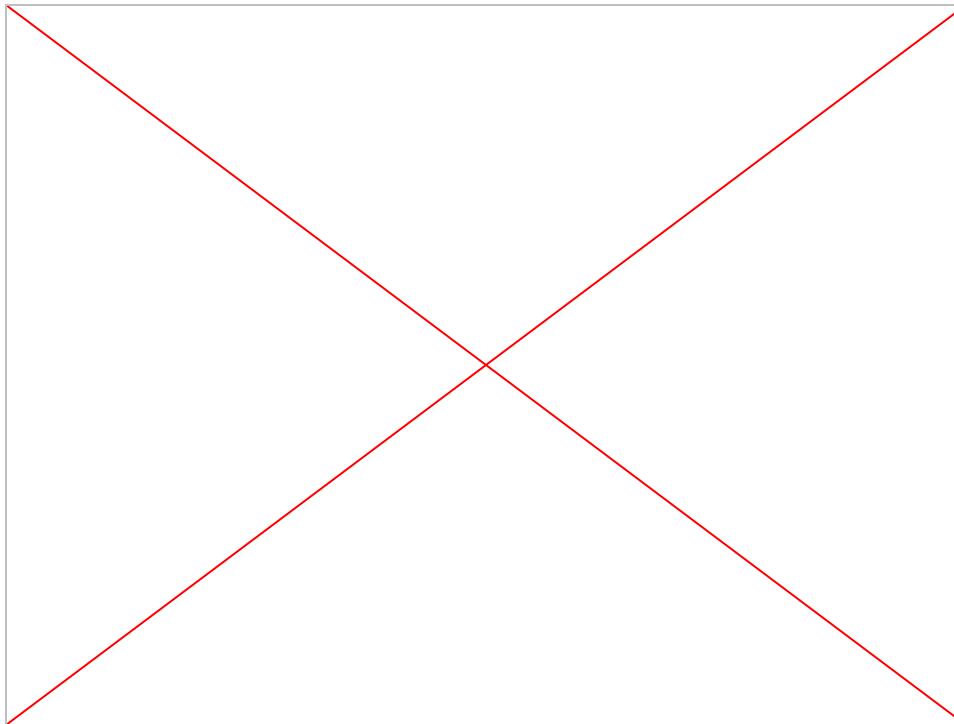
**“ Eh Alfred,
je voudrais donner à Maria Smith un accès en lecture seule
à mon agenda de consultations à domicile ”**

je voudrais donne
à mon age

en lecture seule
domicile ”



La vision



Démo

L'agent spécialisé est identifié mais tout le contexte ne lui est pas donné

The screenshot shows a web-based support interface for Doctolib Silo. At the top, there's a navigation bar with a search bar ('Rechercher un patient'), account switcher ('Switch Account'), help link ('Aide et contact'), and various system icons. On the left, a sidebar lists navigation items: Agenda, Tâches, Doctolib Silo (selected), Messagerie patients, and Gestion des patients. The main content area has a title 'Comment pouvons-nous vous aider ?'. A yellow callout box at the top right says: 'Nous résolvons actuellement 1 incident. Avant de nous contacter, vérifiez si vous êtes impacté.' and 'Vous pouvez demander à être averti par e-mail lorsque l'incident sera résolu.' Below this, there are two radio button options for the subject: 'L'utilisation de Doctolib' (selected) and 'Votre contrat ou vos prélevements'. A dropdown menu for 'Quel est le produit Doctolib concerné?' shows 'Tous les produits'. The 'Votre numéro de téléphone' field contains '+33676767672'. The 'Quel est le titre de votre demande?' section has a note: 'Le sujet doit faire moins de 150 caractères.' and a text input field with the placeholder 'donner accès à mon agenda' and character count '25/150'. A 'Détails de votre demande' section at the bottom has a note: 'Votre message ne doit contenir aucune donnée sensible, personnelle ou confidentielle (nom, prénom, données bancaires, etc.).' and a text input field with placeholder 'Donnez l'adresse de la page où vous avez rencontré ce problème'.

Démo

L'agent spécialisé est identifié et une partie du contexte lui est donnée

The screenshot shows a weekly calendar view from January 20 to January 25, 2025. The calendar grid is color-coded by day of the week: Monday (light blue), Tuesday (light green), Wednesday (light orange), Thursday (light purple), Friday (light pink), Saturday (light yellow), and Sunday (light grey). Specific days are highlighted with a red dot and labeled 'Absence'. For example, January 23rd has a red dot and a tooltip 'Absence 9:00'. January 24th has a red dot and a tooltip 'Absence 10:20'. The sidebar on the left contains links for 'Agenda', 'Tâches', 'Doctolib Site', 'Messagerie patients', and 'Gestion des patients'. Below the calendar, there are sections for 'Agendas (2/2)', 'Status patients (7/7)', and 'Motifs de consultation (4/4)'. The bottom status bar shows the URL <https://pro-staging.doctolib.fr/calendar/today/opening>.

Démo

L'utilisateur pose une question annexe

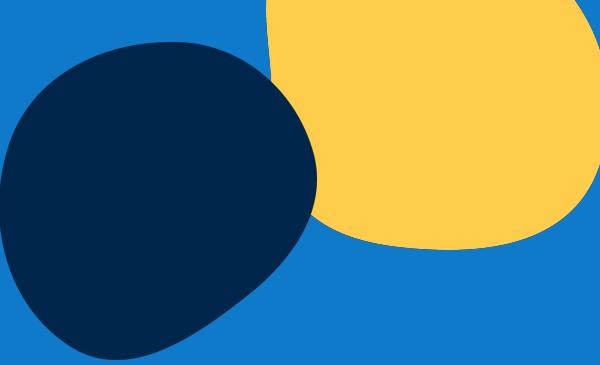
The screenshot shows the Doctolib customer support interface. At the top, there is a search bar labeled "Rechercher un patient" and a "Switch Account" button. Below the search bar, a sidebar on the left contains icons for Agenda, Tâches, Doctolib Sillo, Messagerie patients, and Gestion des patients. The main content area has a title "Comment pouvons-nous vous aider ?". A yellow box contains a warning message: "⚠ Nous résolvons actuellement 1 incident. Avant de nous contacter, vérifiez si vous êtes impacté. Vous pouvez demander à être averti par e-mail lorsque l'incident sera résolu." Below this, there are two radio buttons: one selected for "L'utilisation de Doctolib" and another for "Votre contrat ou vos prélèvements". A dropdown menu for "Quel est le produit Doctolib concerné ?" shows "Tous les produits". The "Votre numéro de téléphone" field contains "+33676767672". The "Quel est le titre de votre demande ?" field has the placeholder "Le sujet doit faire moins de 150 caractères." and contains the text "donner accès à mon agenda". A character counter "25/150" is shown next to the title field. At the bottom, there is a "Détails de votre demande" section with a note about not including sensitive information, and a large text input area for the message. A blue circular icon with a white cross is located in the bottom right corner of the message area.

Démo

Graceful degradation

The screenshot shows a contact form titled "Comment pouvons-nous vous aider ?" (How can we help you?). The interface includes a sidebar with navigation links: Agenda, Tâches, Doctolib Silo, Messagerie patients, and Gestion des patients. The main form fields are:

- A message box stating: "Nous résolvons actuellement 1 incident. Avant de nous contacter, vérifiez si vous êtes impacté." (We are currently resolving 1 incident. Before contacting us, check if you are impacted.)
- A question "À quel sujet avez-vous besoin d'aide ?" (What subject do you need help with?) with two options:
 - L'utilisation de Doctolib
 - Votre contrat ou vos prélevements
- A dropdown menu "Quel est le produit Doctolib concerné ?" (Which Doctolib product is concerned?) set to "Tous les produits".
- A text input "Votre numéro de téléphone" (Your phone number) containing "+33676767672".
- A text area "Quel est le titre de votre demande ?" (What is the title of your request?) with placeholder text "J'aimeerais obtenir l'assistant de consultation" and character count "46/150".
- A section "Détails de votre demande" (Details of your request) with a note: "Votre message ne doit contenir aucune donnée sensible, personnelle ou confidentielle (nom, prénom, données bancaires, etc.)."



4

Défis techniques

Défis techniques, décisions & conséquences

Gérer les hallucinations	Sécurité & droits d'accès	Scaling for production	Fiabilité
<p><u>Décision</u> LLM n'exécutera jamais directement d'actions sensibles.</p> <p><u>Conséquences</u> Comment garantir que ce que nous montrons aux utilisateurs correspond réellement à ce qui sera réalisé ?</p>	<p><u>Décision</u> Pas d'accès 🤝 admin 🤝 . Les agents agissent au nom des utilisateurs.</p> <p><u>Conséquences</u> Propagation du contexte utilisateur.</p>	<p><i>Helphub</i> ouvert ~7k/jour ~10 msg/conversation → ~70k msg/jour. 10x → ~700k msg/jour.</p> <p><u>Conséquences</u> Maintenir un contexte de conversation à travers plusieurs échanges.</p> <p>Assurez un temps de réponse constant.</p>	<p>Les erreurs peuvent survenir à n'importe quelle étape.</p> <p><u>Conséquences</u> Gérer les erreurs, <i>gracefull degradation</i>. Observabilité.</p>

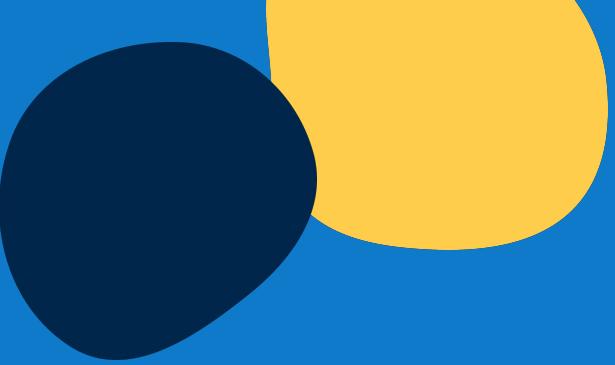
Défis techniques, décisions & conséquences

Développer les API

API REST suivant notre *inter-service standard*

Documenter les API

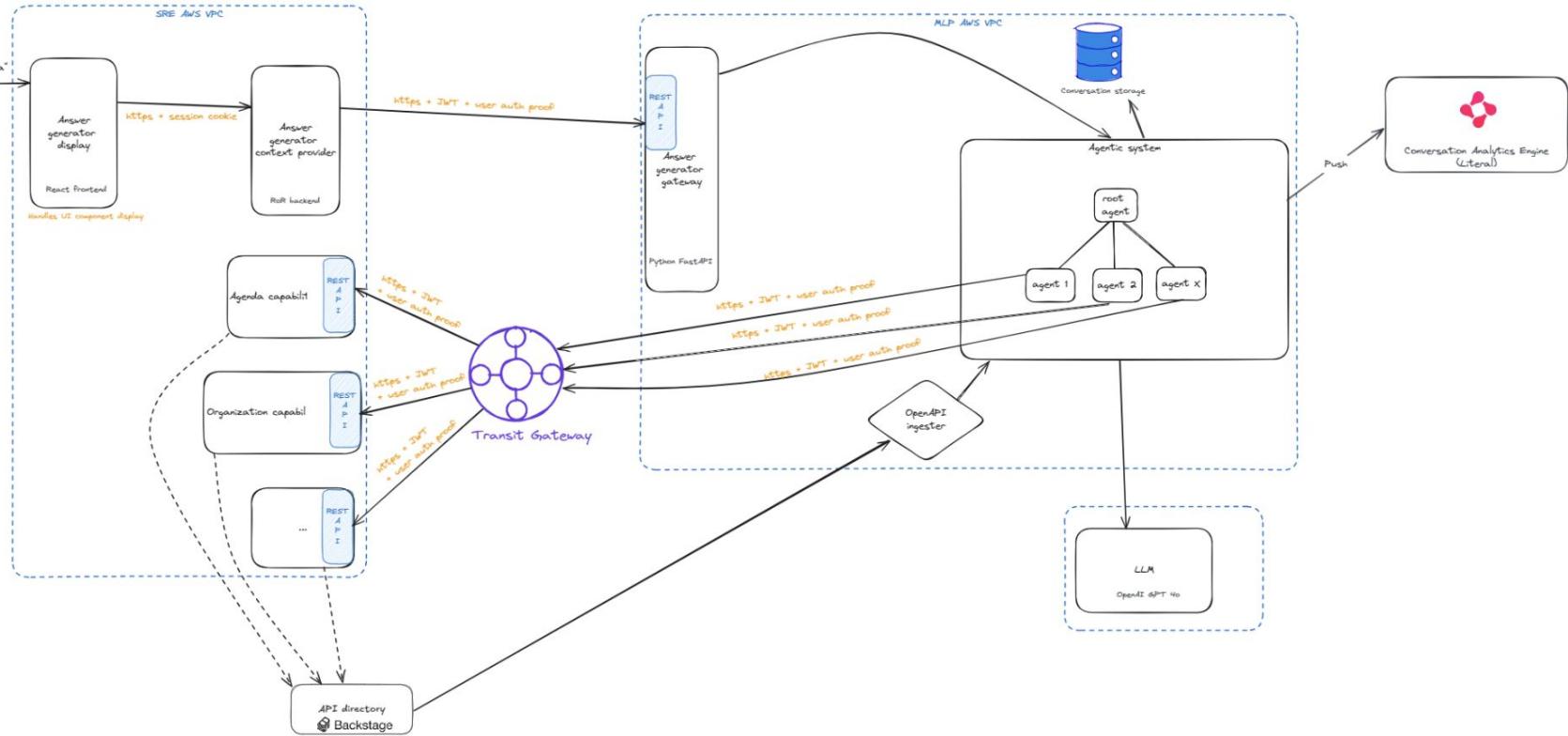
Spec OpenAPI pour des agents IA



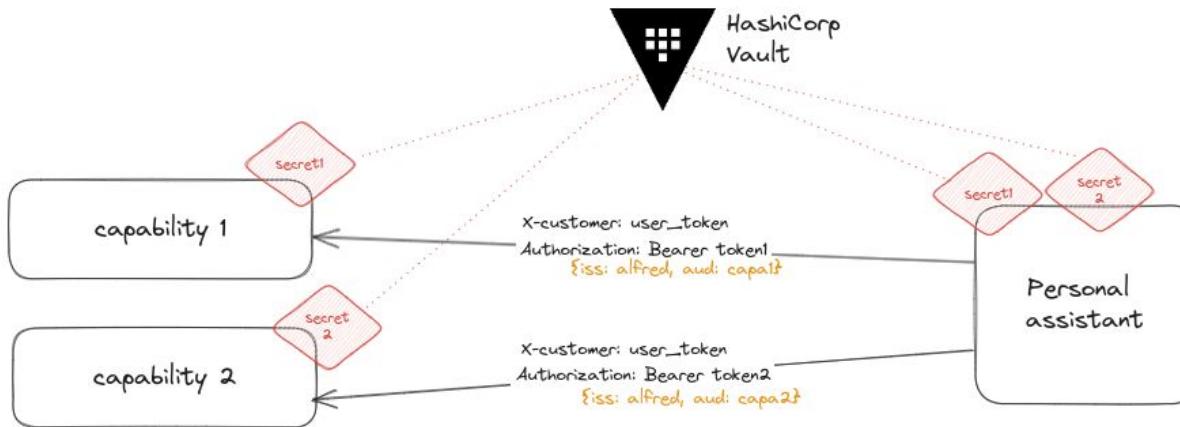
5

Implémentation technique

Vue d'ensemble



Communication inter-services

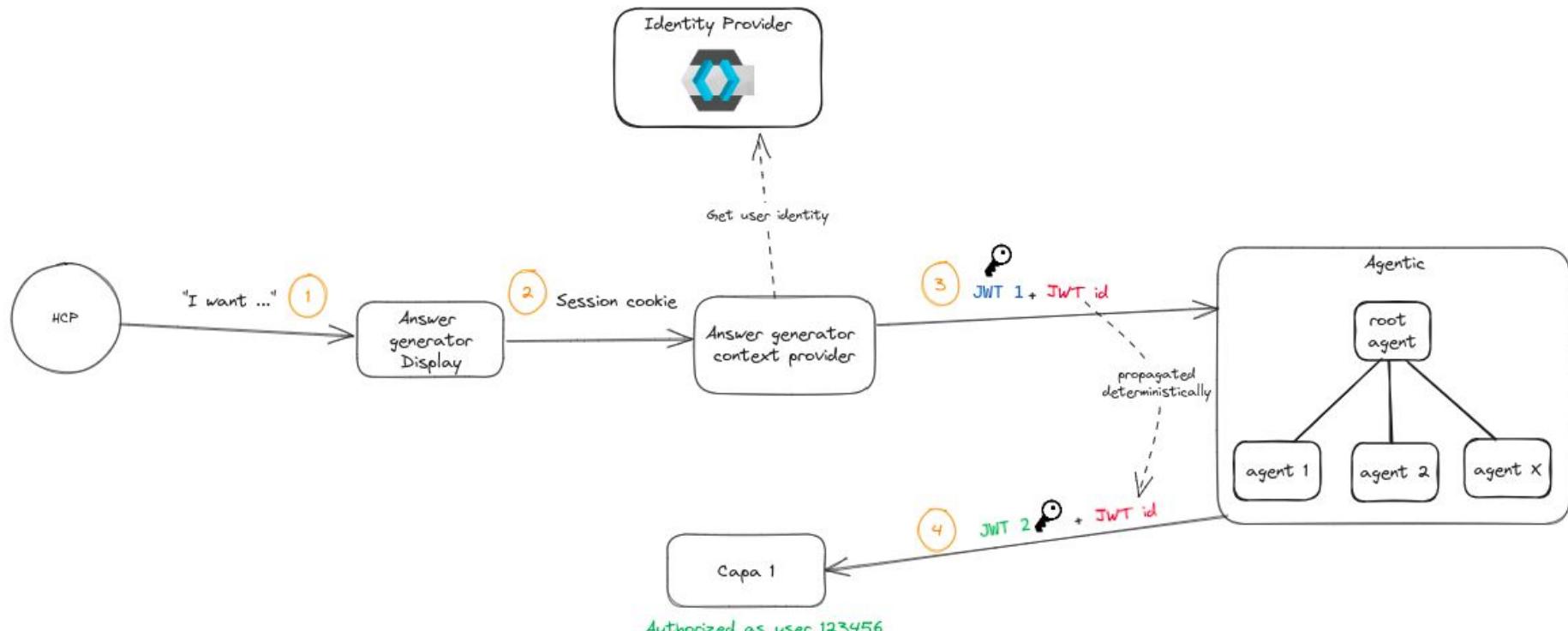


```
INTERSERVICE_SERVER_CAPA1_CONFIG:  
{  
    "service_name": "capa1",  
    "algorithm": "HS256",  
    "local_name": "monolith",  
    "clients": {  
        "personal-assistant": "my_secret"  
    }  
}
```

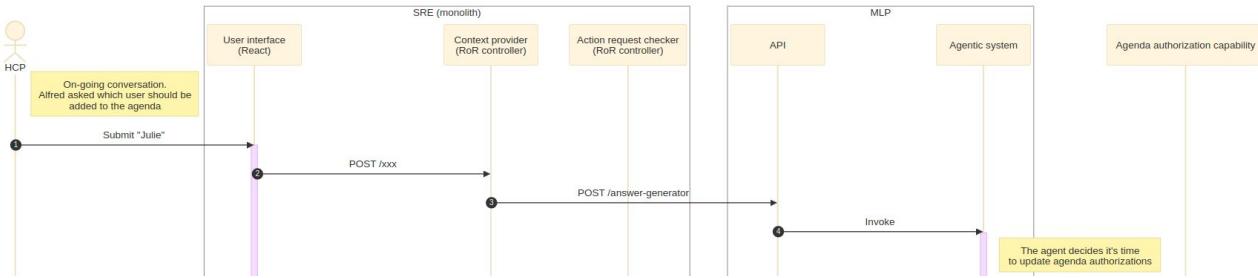
```
JWT claims:  
{  
    "iat": 12345789,  
    "exp": 12457890,  
    "aud": "capa1",  
    "iss": "personal-assistant",  
    "sub": 123  
}
```

Propagation de contexte utilisateur

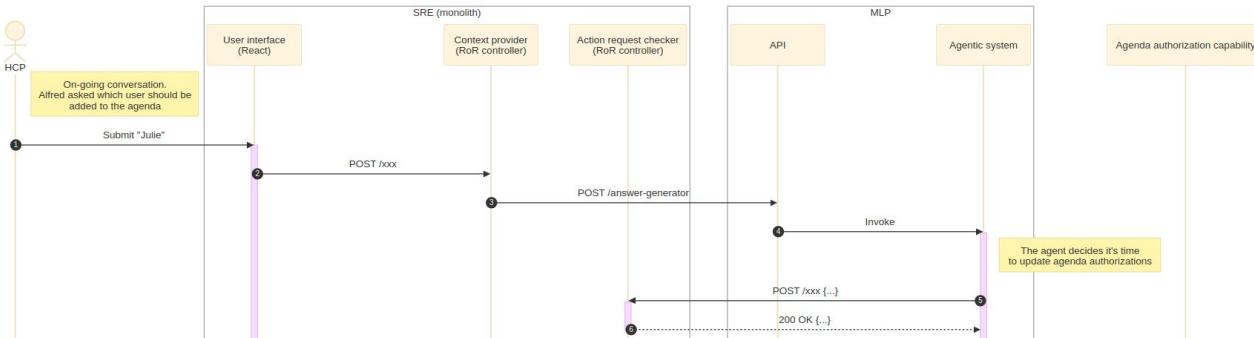
User context propagation



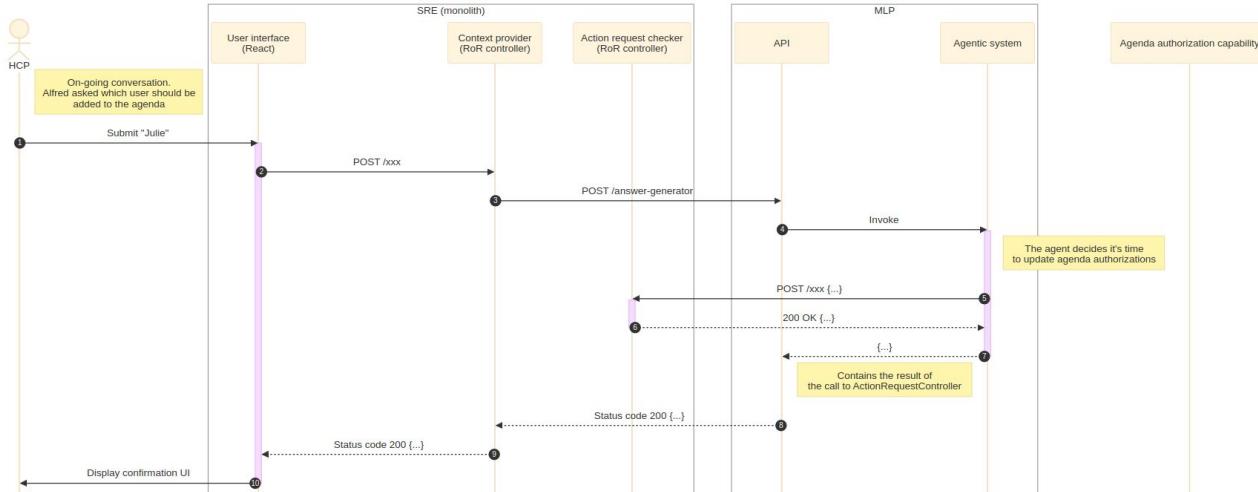
Effectuer une action “sensible”



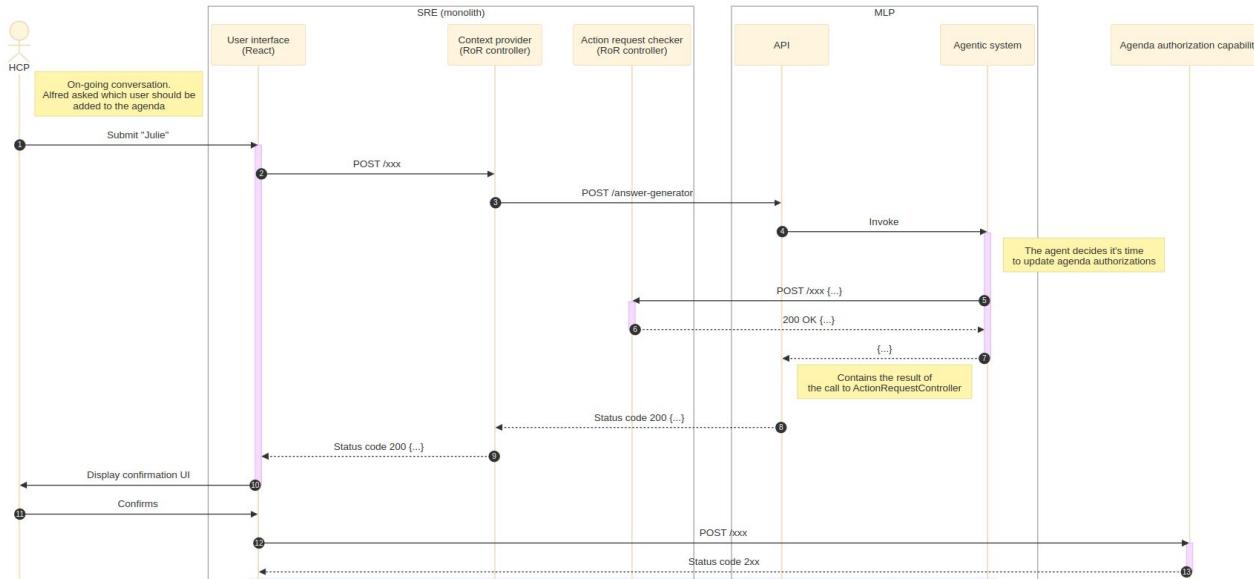
Effectuer une action “sensible”



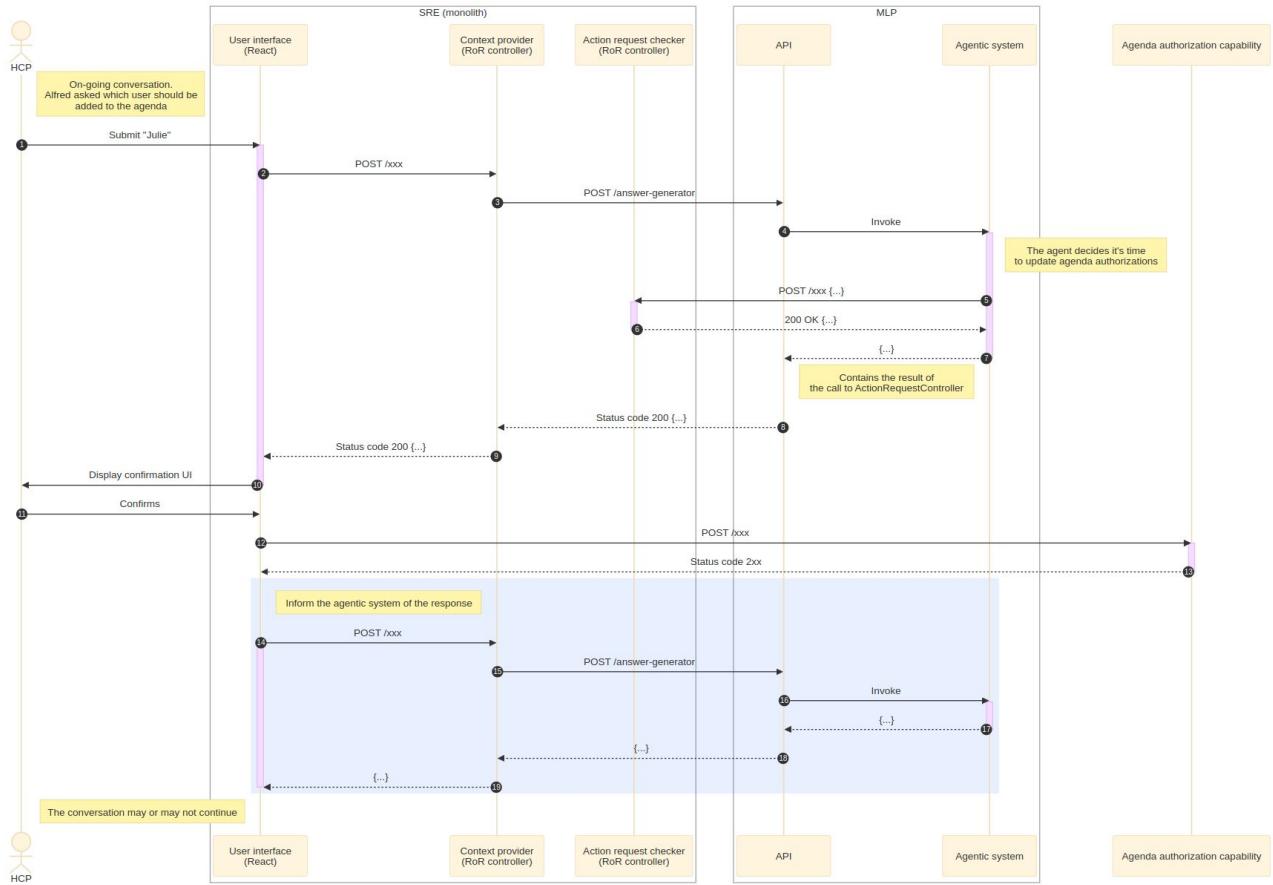
Effectuer une action “sensible”



Effectuer une action “sensible”



Effectuer une action “sensible”



Des composants pour une meilleure expérience utilisateur

I can't grant my secretary access to my calendar

Hello Marie Duchamp, thanks for contacting us. We're going to help you solve your problem right now. Who do you want to grant access to your calendar?

Accounts

- Sophie Boiseau 
- Martin Gouse 
- Rayna Vetrovs 
- Jaob Montesquieu Botosh 

Write your message

Answers are automatically generated. Always double check information before confirming any action.

↑

Alright. Now, select the access level you want to grant to **Sophie Boiseau**.

Account rights to the calendar

CONFIRM

↑

Write your message

Answers are automatically generated. Always double check information before confirming any action.

↑

Implémentation technique

Système Agentic

Alfred ne décide pas quel composant UX afficher

Pour converser avec le client, le système Agentic doit :

- Déterminer quel type de réponse il attend. Par exemple, il attend une valeur dans une liste de valeurs.



- Formater sa réponse selon un format strict.

```
● ● ●

{
  "thread_id": "71442511-d83d-55a7-94d1-9a7ddd05afb7",
  "action_payload": {},
  "component_title": "Préciser à quel collègue donner accès à l'agenda.",
  "answer": "Voici la liste des utilisateurs associés à votre organisation :\n\n1. Laurence ADDI\n2. Patrick LE ROY\n3. Paul PAUL\n4. Louis LOUIS\nPourriez-vous me préciser à quel collègue vous souhaitez donner accès à votre agenda ?",
  "interaction_type": "component",
  "fields": [
    {
      "label": "Choix du collègue",
      "field_type": "single-select",
      "parameters": [
        {
          "type": "str",
          "value": "Laurence ADDI"
        },
        {
          "type": "str",
          "value": "Patrick LE ROY"
        },
        {
          "type": "str",
          "value": "Paul PAUL"
        },
        {
          "type": "str",
          "value": "Louis LOUIS"
        }
      ]
    },
    "is_final_answer": false,
    "answerable": true
  }
}
```

Implémentation technique

Notre Backend

```
{  
    "personal_assistant_response": {  
        "action_payload": {  
            "confirmation_text": null,  
            "modification_api_url": null,  
            "modification_api_params": null,  
            "thread_id": null  
        },  
        "answerable": true,  
        "content": "Voici la liste des utilisateurs associés à votre organisation  
:\n\n1. Laurence ADDI\n2. Patrick LE ROY\n3. Paul PAUL\n4. Louis LOUIS\nPourriez-vous  
me préciser à quel collègue vous souhaitez donner accès à votre agenda ?",  
        "is_final_answer": true,  
        "thread_id": "71442511-d83d-55a7-94d1-9a7ddd05afb7",  
        "type": "component",  
        "form_schema": {  
            "title": "Préciser à quel collègue donner accès à l'agenda.",  
            "fields": [  
                {  
                    "label": "Choix du collègue",  
                    "name": "choixducollègue",  
                    "type": "string",  
                    "parameters": [],  
                    "validation": []  
                }  
            ]  
        }  
}
```

Le rôle du backend dans ce système est de :

- **Maintenir son contrat API avec Alfred et avec l'application Frontend.**



- **Vérifie que les données que nous recevons peuvent être traitées.**
- **Gère les erreurs avec le concept de graceful degradation.**



Implémentation technique

Notre Frontend

Notre code Frontend **décidera** quel composant UX afficher.

Notre Frontend sera responsable de :

- **Décider quel composant offrira la meilleure expérience utilisateur selon des règles métier spécifiques.**
- **Construire des formulaires à la volée**
- **Mettre en œuvre la graceful degradation lorsqu'une payload invalide est envoyée.**

Hello Marie Duchamp, thanks for contacting us. We're going to help you solve your problem right now. Who do you want to grant access to your calendar?

Sophie Boiseau

skylar.vaccaro@doctolib.fr
Secretary



Martin Gouse

martin.gouse@doctolib.fr
Secretary



Rayna Vetrovs

rayna.vetrovs@doctolib.fr
Secretary



Jakob Montesquieu Botosh

jakob.botosh@doctolib.fr
Secretary



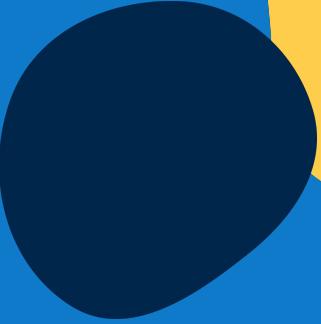
Hello Marie, thanks for contacting us. We're going to help you solve your problem right now.

Here are all the users in your practice:

1. Sophie Boiseau
2. Martin Gouse
3. Rayna Vetrovs
4. Jakob Montesquieu Botosh
5. Ahmad Franci
6. Maren Pressi

Who do you want to grant access to your calendar?



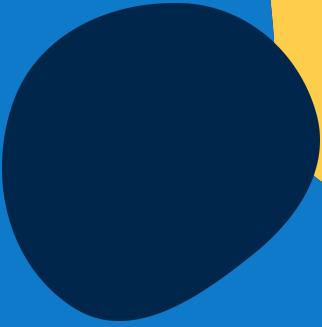


6

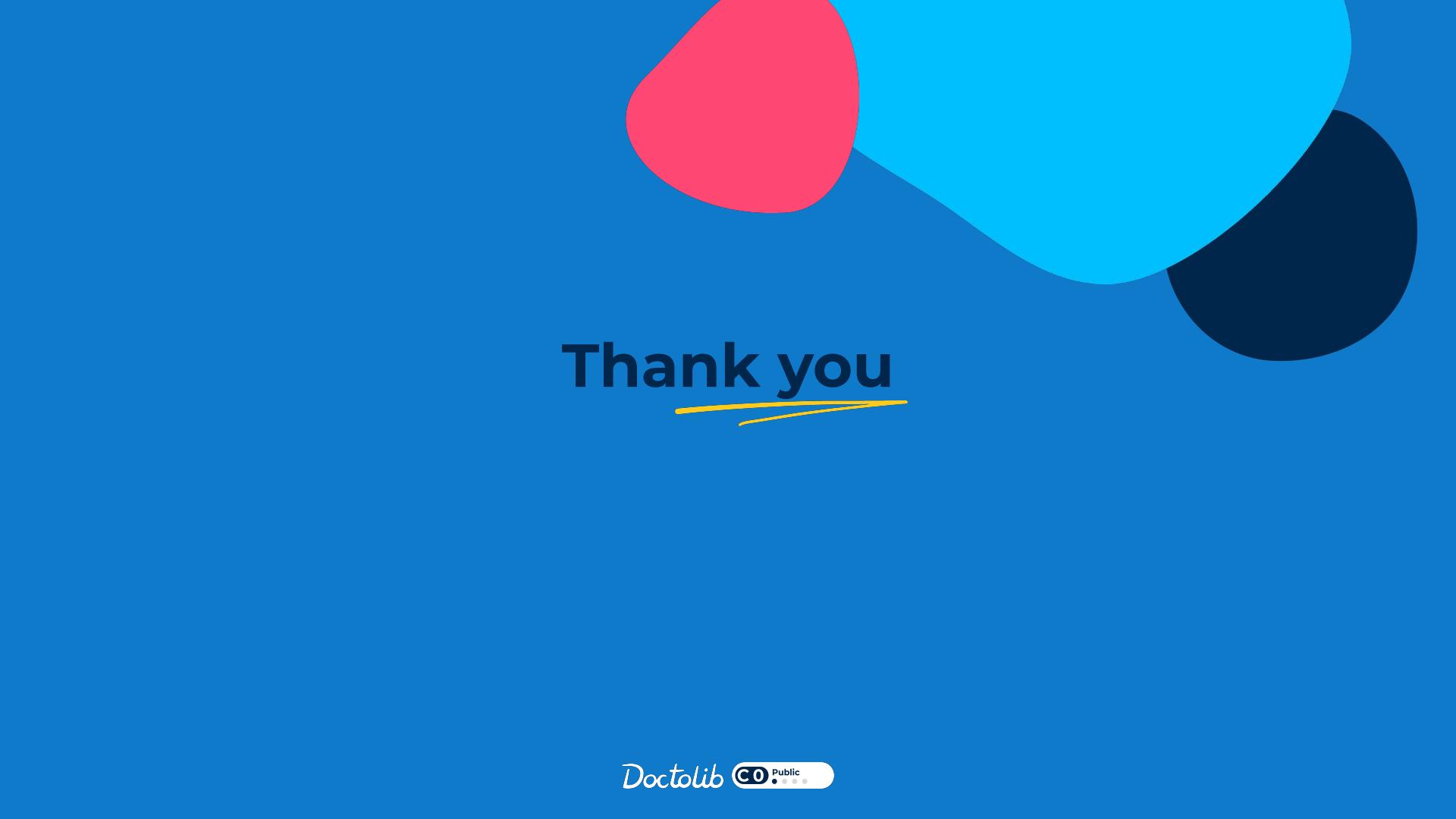
Et après ?

Et après ?

- 18% de déflexion avec du RAG uniquement
- Objectif de 30% avec Agentic
- En production (bêta) depuis 1 semaine 🎉
- Ajout de nouveaux use-cases
- Alfred partout !



Des questions ?



Thank you

A large, bold, dark blue sans-serif font displays the words "Thank you". The word "you" is underlined with a thick, yellow, curved line.

Évaluation



Stratégie

End-to-end

- Plus simple de contrôler les entrées/sorties uniquement
- Pas d'évaluation noeud par noeud



North star

Niveau de réussite

Mesure la similarité entre la sortie du système et un *ground truth*

1 : niveau le plus bas
3 : niveau le plus élevé - parfaitement aligné avec *ground truth*

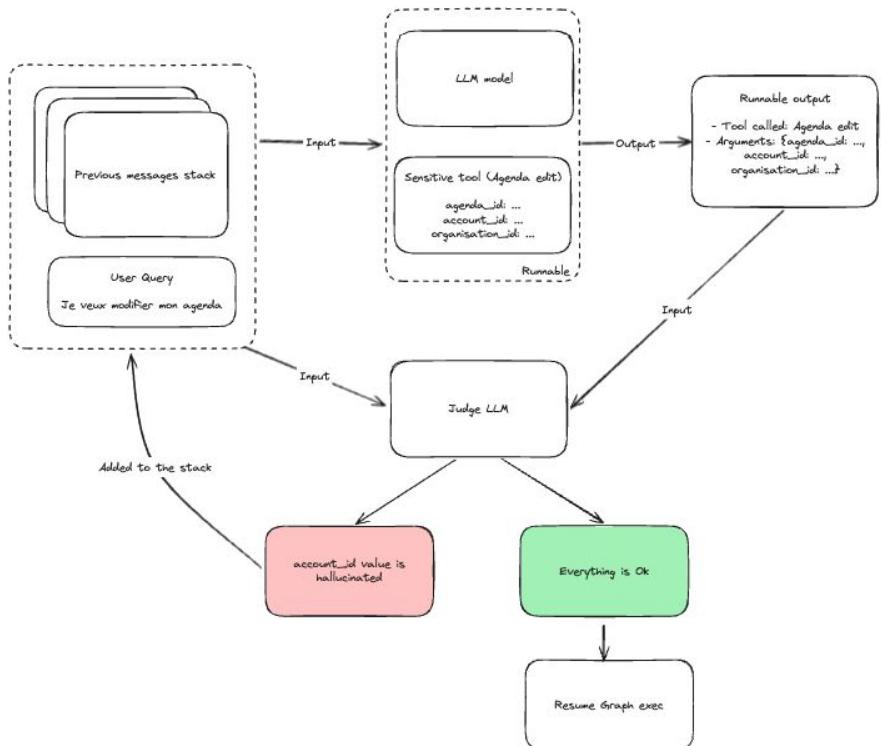


Métriques

Latence (en secondes) : rapidité avec laquelle le système accomplit la tâche.

Δ nb nœuds : le système a-t-il accompli la tâche avec un nombre minimal d'étapes par rapport à la réalité

Sensitive tool invocation is validated by a judge

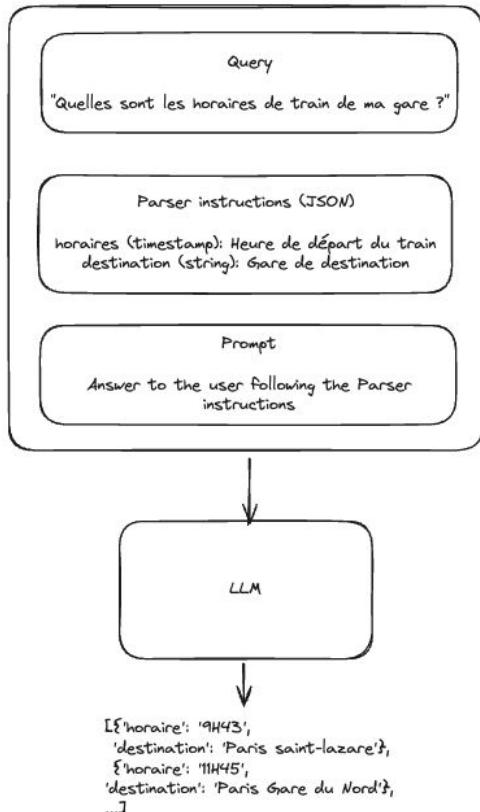


Using another LLM as a judge

- Sensitive tools invocation (such as agenda modifications) require arguments grounded in context
 - In the **messages** from the user
 - Within calls to capabilities (**Monolith data**)
- The arguments values filled by the Specialised agent, the tool definition and the stack of messages are passed to an LLM to judge the correctness of the tool invocation
- Any hallucination captured by the judge is passed back to the Specialised agent to correct the tool call

The judge LLM can also suffer hallucinations, but a second safety net is present at the controller lever

Other types of Guardrails in the code



Outputs of Parsers can be validated in the code

1. Make sure that **all keys** specified in the Parser instructions **are present**
2. Validate the **type of data** expected in each key-value pair
3. If a definitive set of values are expected for a key, validate that **only these possibilities are in the answer**

If one of these criteria is not met, the LLM is invoked once more until satisfaction or until a limit is reached