

Contexte

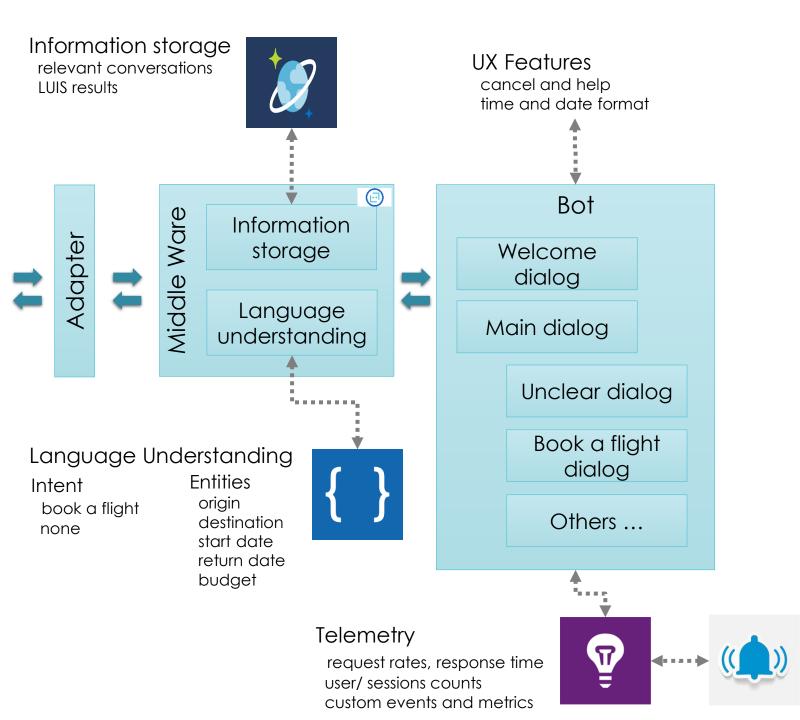
Fly Me, agence de voyage, souhaite développer son premier ChatBot de réservation.

Ce premier MVP inclut:

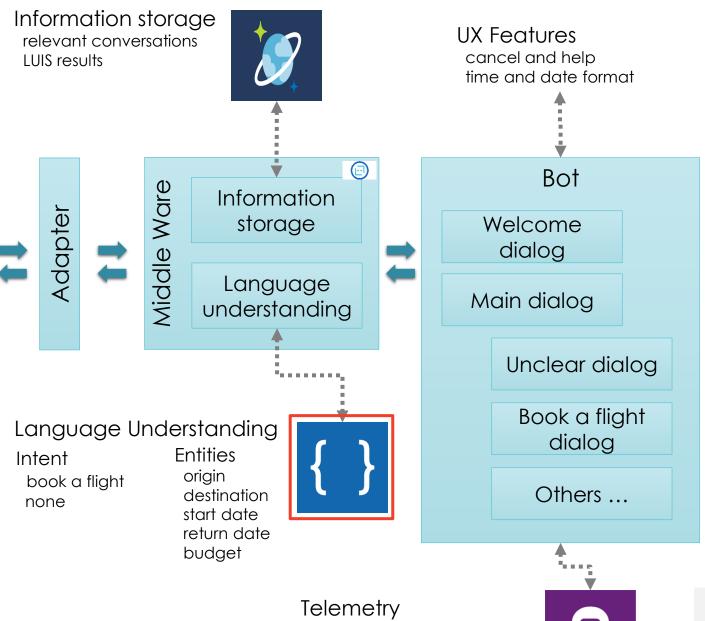
- l'identification des informations principales
- les fondations d'un monitoring et d'alertes
- des capacités d'entrainement et d'évaluation off-line

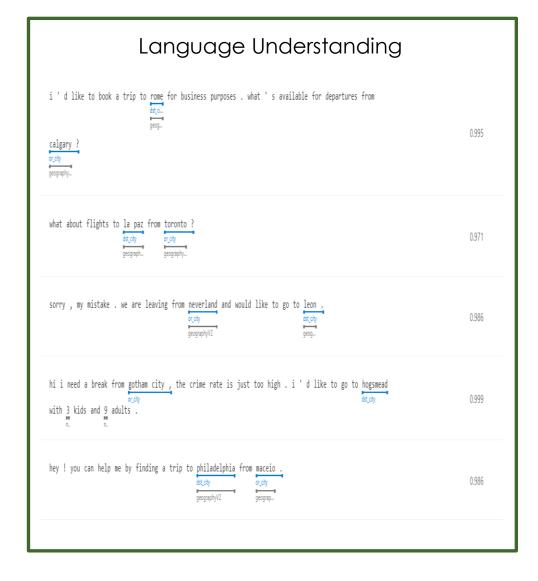
- Mise en forme du jeu de données MS 1369 dialogues
- Entrainement et déploiement du modèle LUIS
- Architecture et développement du Bot sur la base du Bot Framework SDK 4
- Intégration des ressources LUIS, Insights, CosmosDB
- Tests du bot en local
- Implémentation des tableaux de bord et alertes
- Déploiement du Bot sur Azure
- Tests sur le cloud

Etapes du projet



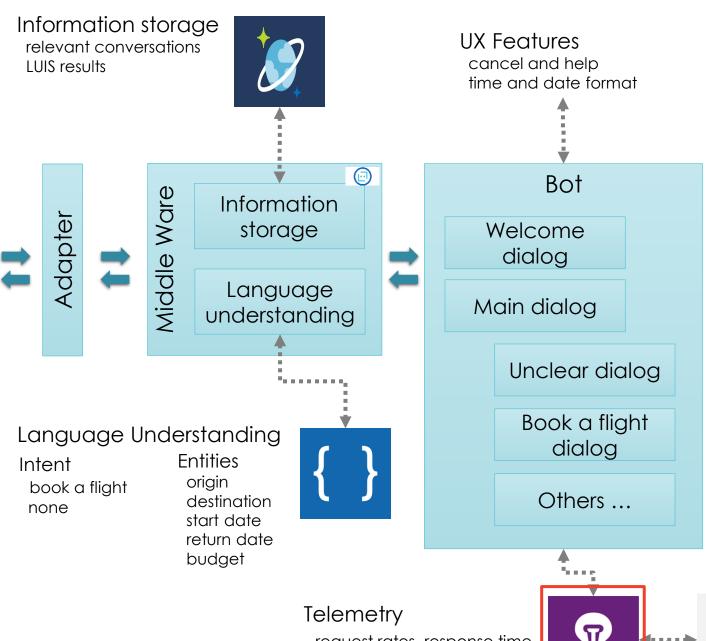
MVP – Features & Architecture

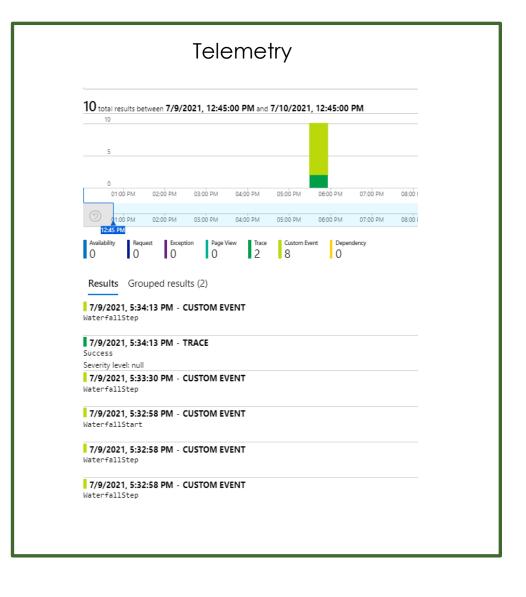




request rates, response time user/ sessions counts custom events and metrics

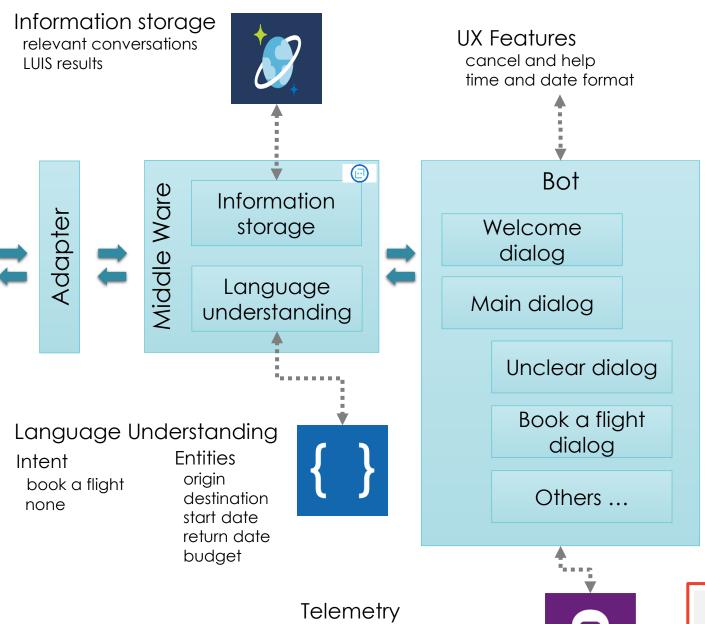


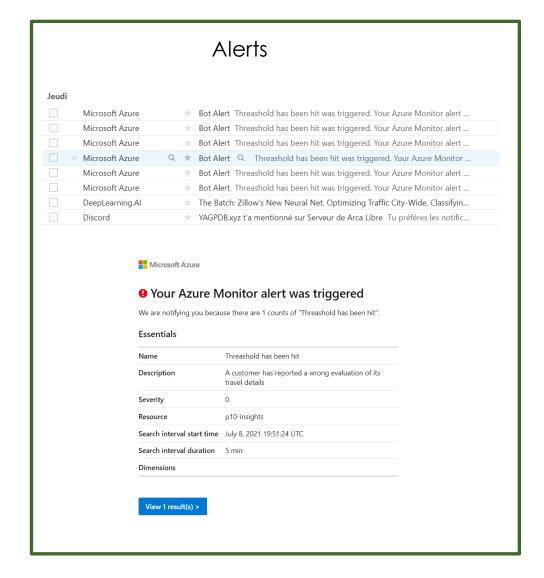




request rates, response time user/ sessions counts custom events and metrics

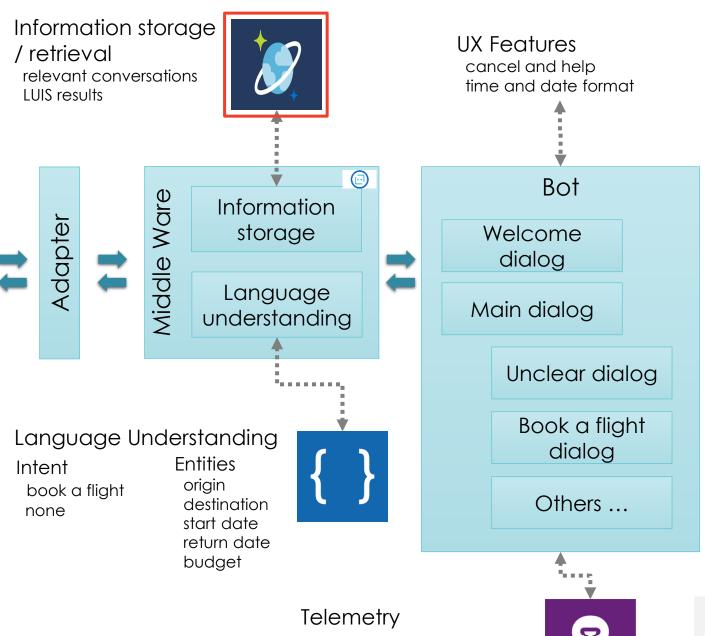






request rates, response time user/sessions counts custom events and metrics

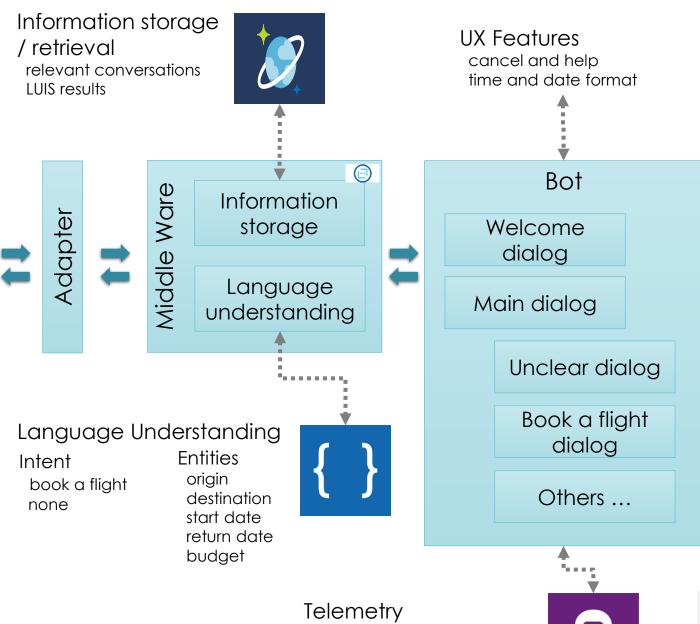


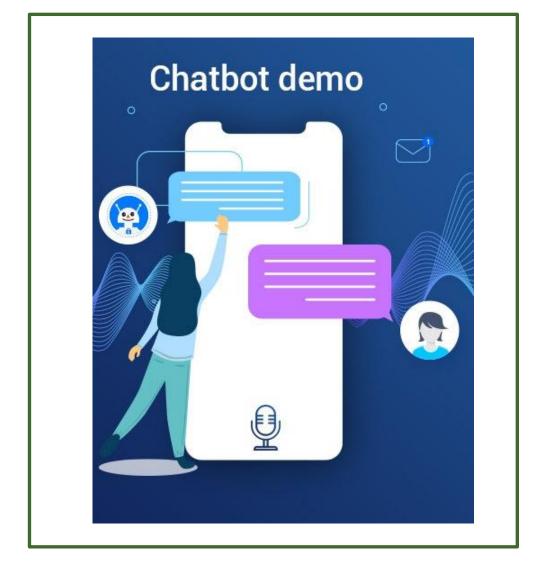


```
Information Storage / Retrieval
"id": "962e65c0-e0ce-11eb-b2c8-63296613ad40",
"realId": "962e65c0-e0ce-11eb-b2c8-63296613ad40",
"document": [
   "[User]: i want to book a flight",
   "[Bot]: Where would you like to travel to?",
   "[User]: to paris",
   "[Bot]: From what city will you be travelling?",
   "[User]: I will take off from london",
   "[Bot]: On what date would you like to travel in?",
   "[User]: mar 23 21",
   "[Bot]: On what date would you like to travel back?",
   "[User]: 2/4/21",
   "[Bot]: What is your maximum budget for this trip ?",
   "[User]: 500",
   "[Bot]: Please confirm: I have you traveling to paris\nfrom london on 2021-03-23.\nYour flight back is sche
   "[User]: No",
   "[Bot]: What else can I do for you?"
" rid": "VXRuAKbnxNoBAAAAAAAAAA==",
" self": "dbs/vXRuAA==/colls/vXRuAKbnxNo=/docs/vXRuAKbnxNoBAAAAAAA==/",
" etag": "\"0f000a10-0000-0d00-0000-60e872480000\"",
"_attachments": "attachments/",
" ts": 1625846344
```

request rates, response time user/ sessions counts custom events and metrics







request rates, response time user/sessions counts custom events and metrics

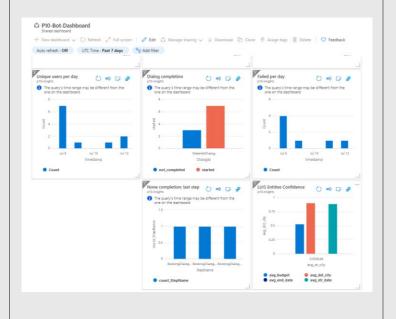


« Cahier des charges » du pilotage de la performance

- Monitorer la qualité de l'UX: la conversation avec le chat bot doit être fluide et permettre au client d'obtenir aisément le service souhaité.
- Monitorer la qualité des réponses du chat bot: le chat bot étant en contact direct avec le client, il contribue à l'image de marque. Il est essentiel de s'assurer de la qualité des informations fournies par le chat bot.
- Synthétiser l'information dans des tableaux de bord, Alerter en cas de dépassement de seuil: la collecte des données de monitoring n'est pertinente que si l'information est accessible facilement, au niveau de granularité adapté
- Intégrer la collecte des données pour le ré-entrainement du modèle dans l'UX: les conversations des clients avec le chat bot sont une source de données précieuses pour ré-entrainer le modèle.
- Mesurer la performance du modèle, Ré-entrainer et Re-déployer le modèle: pour améliorer les prédictions du modèle IA intégré à la solution.

Pilotage de la performance

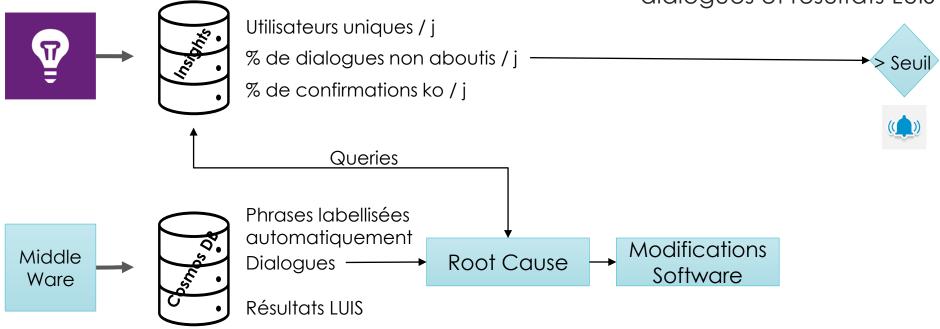
Garantir la satisfaction et donc l'engagement client



Objectif	Métrique
Garantir la satisfaction client	% Utilisateurs uniques / j
Garantir la qualité de l'UX	% de dialogues non aboutis / j
Garantir la qualité des réponses	% de confirmation ko / j

<u>Pilotage de la performance: qualité de l'expérience utilisateur</u>

Lorsque le % de dialogues non-aboutis dépasse un seuil, on recherche la cause racine du problème pour apporter les modifications pertinentes au software. Le diagnostique est facilité par le requêtage évènements enregistrés par Insights et les dialogues et résultats LUIS dans Cosmos DB.

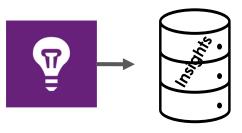


Objectif	Métrique
Garantir la satisfaction client	% Utilisateurs uniques / j
Garantir la qualité de l'UX	% de dialogues non aboutis / j
Garantir la qualité des réponses	% de confirmation ko / j

<u>Pilotage de la performance: qualité des réponses</u>

Lorsque le % de confirmations de l'utilisateur dépasse un seuil, on ré-entraine le modèle sur des données générées automatiquement à partir des interactions clients et / ou sur des données labellisées à la main suite à diagnostique des erreurs.

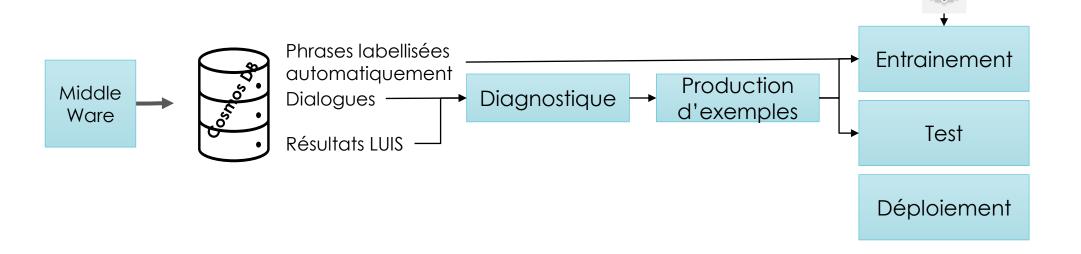
> Seuil



Utilisateurs uniques / j

% de dialogues non aboutis / j

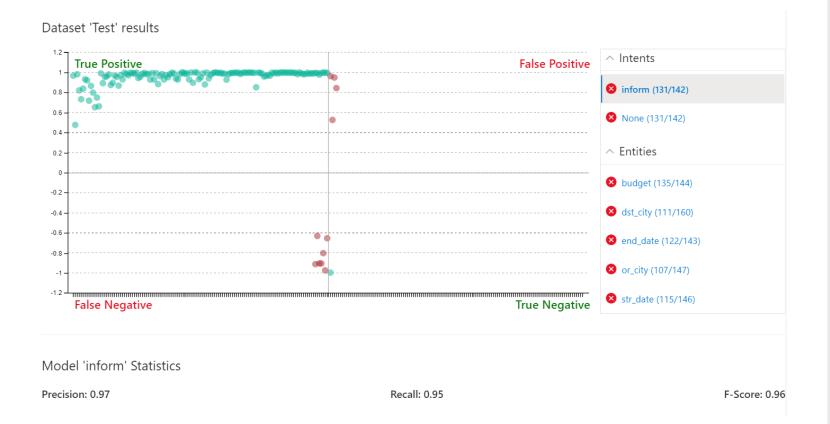
% de confirmations ko / j -





L'utilisateur pourra identifier les informations inexactes sur une adaptative card.

Démonstration d'un test de LUIS par batch



Pilotage de la performance

- Gestion de versions sous GitHub
- Classes de tests
- Tests lancés automatiquement avant chaque commit
- Déploiement automatique sur Azure à chaque push

Best Practices mises en œuvre