

Fly Me

# Pilotage de la performance

Présentation de la méthodologie

# Table des matières

Table des matières.....	1
I. Introduction .....	2
II. Critères d'évaluation du modèle.....	3
III. Schéma du mécanisme d'évaluation du modèle en production .....	5
IV. Modalités de mise à jour du modèle .....	7
V. Conclusion .....	10
VI. Axes d'amélioration.....	11

## I. Introduction



Le projet de développement d'un chatbot de réservation pour Fly Me requiert une attention particulière sur la performance du modèle en production. La qualité de l'expérience utilisateur est un facteur clé pour le succès de ce projet. Ainsi, l'évaluation régulière et la mise à jour du modèle sont cruciales pour garantir un chatbot fiable et efficace.

Fly Me est une compagnie aérienne qui a lancé un projet de développement d'un chatbot de réservation pour offrir une expérience de réservation plus fluide et efficace à ses clients. Ce chatbot sera conçu pour répondre aux questions des clients sur les réservations, les itinéraires, les horaires, etc.

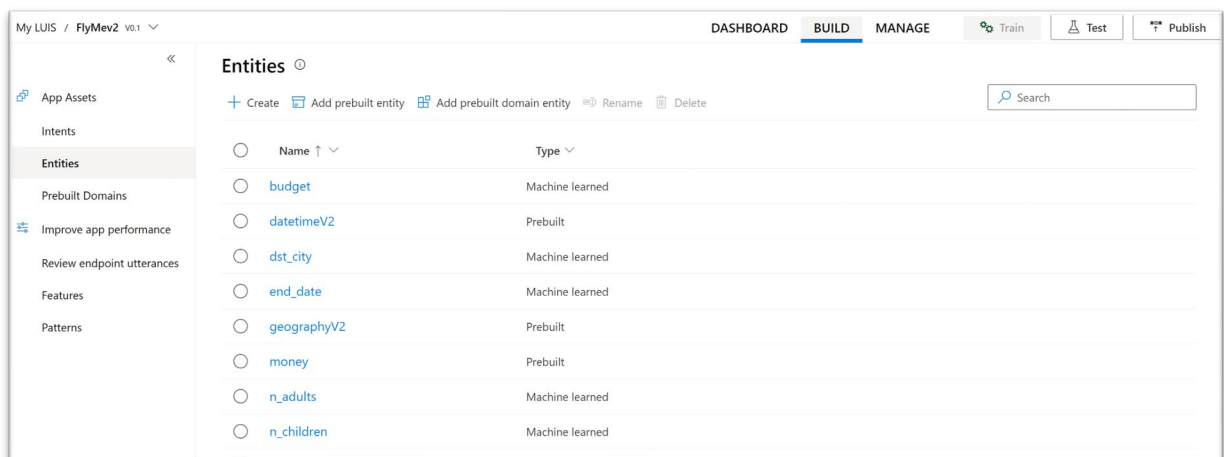
Le modèle du chatbot est formé avec des données précédemment collectées et il est crucial de s'assurer de sa performance en production. La qualité de la réponse du chatbot dépend de sa capacité à comprendre les requêtes des utilisateurs et à fournir les réponses appropriées. Pour garantir un chatbot performant et efficace, il est donc important de disposer d'une méthode pour évaluer la performance du modèle en production et de la mettre à jour régulièrement.

## II. Critères d'évaluation du modèle

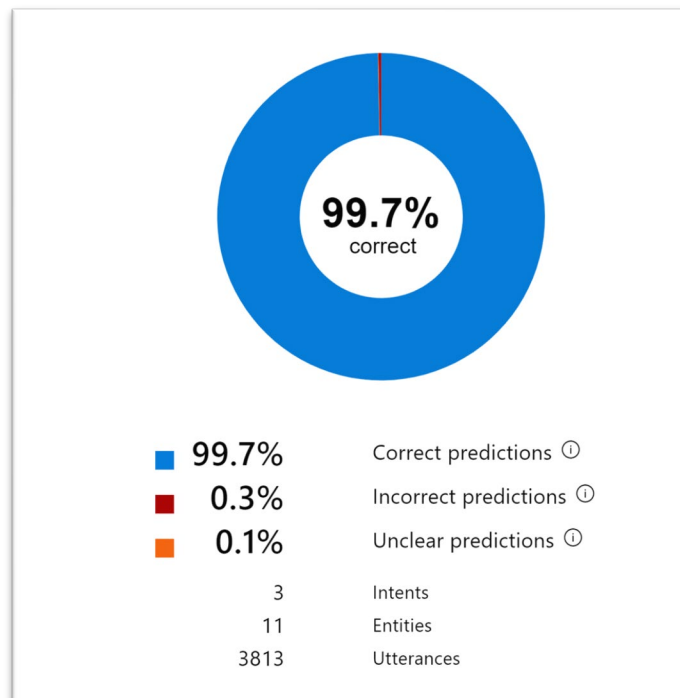
L'évaluation de la performance du modèle en production est cruciale pour garantir la qualité de l'expérience utilisateur avec le chatbot de réservation de Fly Me. La première étape pour évaluer la performance du modèle consiste à déterminer les critères d'évaluation.



LUIS (Language Understanding Intelligent Service) est un moteur d'interprétation de langage qui aide à détecter les informations dans les conversations et à interpréter les intentions des utilisateurs. Il joue un rôle clé dans l'évaluation de la performance du modèle en production en permettant d'analyser les interactions entre l'utilisateur et le chatbot.



L'évaluation de la performance du modèle repose également sur la capacité du modèle à détecter les informations importantes dans les conversations et à interpréter les intentions des utilisateurs. La capacité du modèle à comprendre les demandes de l'utilisateur est un critère clé pour évaluer la performance du modèle en production.



Enfin, les données de formation du modèle jouent également un rôle important dans l'évaluation de la performance du modèle. En utilisant des données de formation de qualité, il est possible d'améliorer la capacité du modèle à comprendre les demandes de l'utilisateur et à effectuer des prédictions correctes. Les données de formation doivent donc être examinées régulièrement pour s'assurer de leur qualité et de leur pertinence pour l'évaluation de la performance du modèle.

My LUIS / FlyMev2 v0.1

DASHBOARD BUILD MANAGE Train Test Publish

### Review Endpoint Utterances

Filter: BookFlight

Confirm all entity predictions Save Discard

Utterance Predicted Intent

i would like to go to tijuana from paris the 10th of august , 2023 . i want to return the 15th of august 2023 . i have a budget of 800 \$ and we are 2 adults and 0 child

BookFlight (0.9996567)

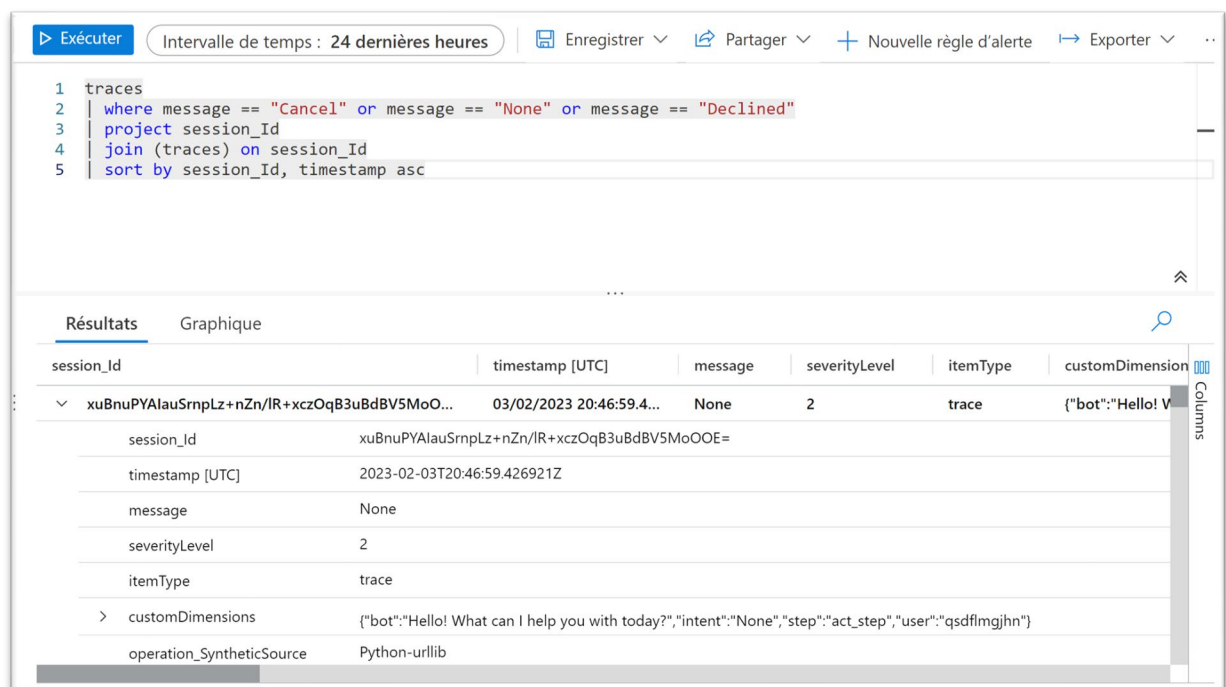
it ' s about one thousand dollars

BookFlight (0.9979193)

i live in le havre and i want to go to paris yesterday and return 2 days later for 400 euros , 2

### III. Schéma du mécanisme d'évaluation du modèle en production

Pour évaluer la performance du modèle en production, il est nécessaire de suivre les activités du chatbot dans son fonctionnement réel. Cela peut être réalisé en créant des logs pour les traces des conversations entre l'utilisateur et le chatbot. Les logs permettent de suivre les différentes étapes de la conversation et de détecter les éventuels problèmes.



The screenshot shows a data query interface. At the top, there is a toolbar with buttons: 'Exécuter', 'Intervalle de temps : 24 dernières heures', 'Enregistrer', 'Partager', 'Nouvelle règle d'alerte', and 'Exporter'. Below the toolbar is a SQL query editor with the following code:

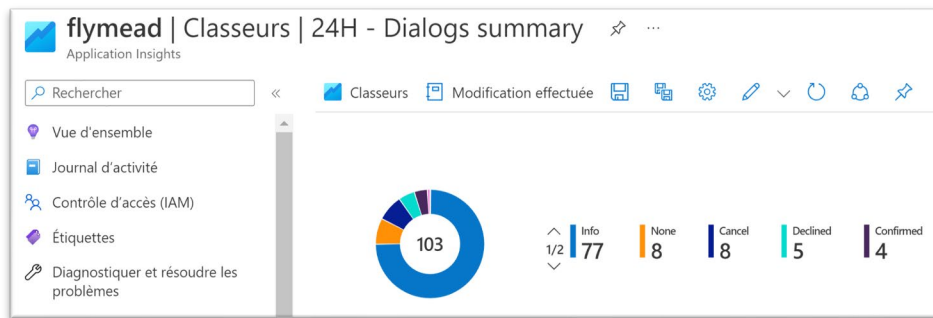
```
1 traces
2 | where message == "Cancel" or message == "None" or message == "Declined"
3 | project session_Id
4 | join (traces) on session_Id
5 | sort by session_Id, timestamp asc
```

Below the query editor, there are two tabs: 'Résultats' (selected) and 'Graphique'. The 'Résultats' tab displays a table with the following columns: session\_Id, timestamp [UTC], message, severityLevel, itemType, and customDimension. The table contains one main row with a collapsed 'customDimensions' section.

session_Id	timestamp [UTC]	message	severityLevel	itemType	customDimension
xuBnuPYAlauSrnplz+nZn/IR+xczOqB3uBdBV5MoO...	03/02/2023 20:46:59.4...	None	2	trace	({"bot": "Hello! V...
session_Id	xuBnuPYAlauSrnplz+nZn/IR+xczOqB3uBdBV5MoOOE=				
timestamp [UTC]	2023-02-03T20:46:59.426921Z				
message	None				
severityLevel	2				
itemType	trace				
customDimensions	({"bot": "Hello! What can I help you with today?", "intent": "None", "step": "act_step", "user": "qsdflmgjhn"})				
operation_SyntheticSource	Python-urllib				

Pour évaluer la performance du modèle, différents critères seront utilisés.

- **Date non conforme** : Les erreurs en matière de dates peuvent entraîner des difficultés pour les utilisateurs dans le processus de réservation. Par conséquent, il est important de surveiller les erreurs de date.
- **Confirmation de réservations** : La capacité du chatbot à effectuer correctement les confirmations de réservations est un critère important d'évaluation.
- **Prédictions correctes validées par l'utilisateur** : Le chatbot est évalué en fonction de sa capacité à prédire correctement les intentions des utilisateurs.



Des règles sont établies pour détecter les difficultés que les utilisateurs peuvent rencontrer dans le processus de réservation. Cela permet une réaction rapide pour améliorer la qualité de l'expérience utilisateur.

Règles d'alerte

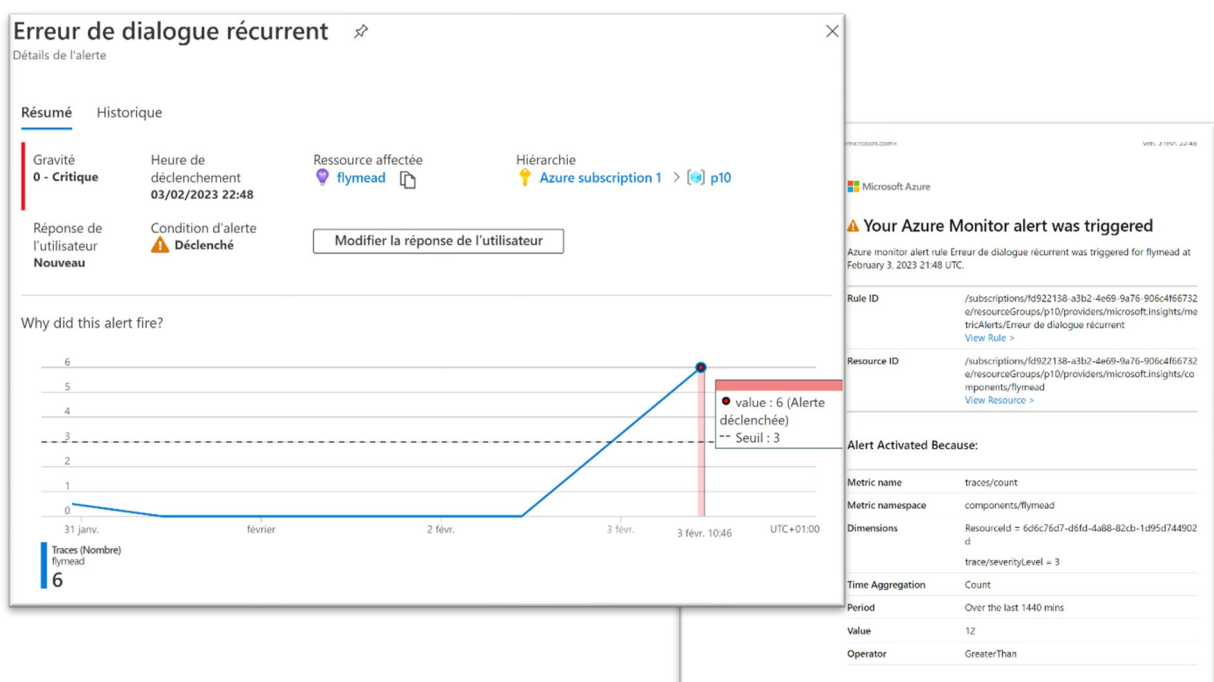
+ Créer | Colonnes | Actualiser | Exporter au format CSV | Ouvrir une requête | Supprimer | Activer | Désactiver

Rechercher | Type de ressource cible: tous | Étendue cible: flymead | Abonnement: tous | Type de signal: tous | Gravité: tous | État: Actif

Affichage de 1 à 9 de 9 Règles d'alerte.

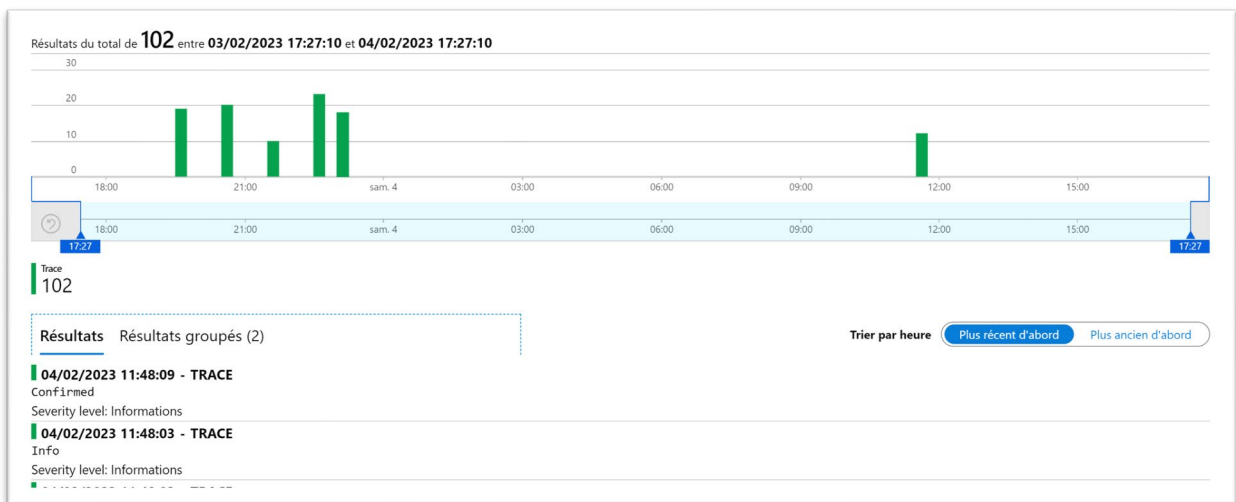
Nom	Condition	Gravité	Étendue cible	Type de ressource cible	Type de signal	État
<input type="checkbox"/> Dependency Latency Degradation...	Dependency Latency Degradation...	3 - Information	flymead	Application Insights	Détecteur intelligent	Actif
<input type="checkbox"/> Erreur de dialogue récurrent	traces/count > 3	0 - Critique	flymead	Application Insights	Métriques	Actif
<input type="checkbox"/> Exception Anomalies - flymead	Exception Anomalies détecté	3 - Information	flymead	Application Insights	Détecteur intelligent	Actif
<input type="checkbox"/> Failure Anomalies - flymead	Failure Anomalies détecté	3 - Information	flymead	Application Insights	Détecteur intelligent	Actif
<input type="checkbox"/> Potential Memory Leak - flymead	Potential Memory Leak détecté	3 - Information	flymead	Application Insights	Détecteur intelligent	Actif
<input type="checkbox"/> Response Latency Degradation ...	Response Latency Degradation d...	3 - Information	flymead	Application Insights	Détecteur intelligent	Actif
<input type="checkbox"/> test de disponibilité-flymead	Emplacements incorrects >= 1	1 - Erreur	flymead	Application Insights	Métriques	Actif
<input type="checkbox"/> Too many ChatBot Errors	severityLevel > 3 Dans les dimen...	0 - Critique	flymead	Application Insights	Recherche dans les journaux	Actif
<input type="checkbox"/> Trace Severity Degradation - fly...	Trace Severity Degradation détec...	3 - Information	flymead	Application Insights	Détecteur intelligent	Actif

En utilisant ces critères, un système de surveillance en temps réel est mis en place pour évaluer la performance du modèle en production. Les données issues de cette évaluation seront utilisées pour apporter les modifications nécessaires au modèle pour une meilleure performance.

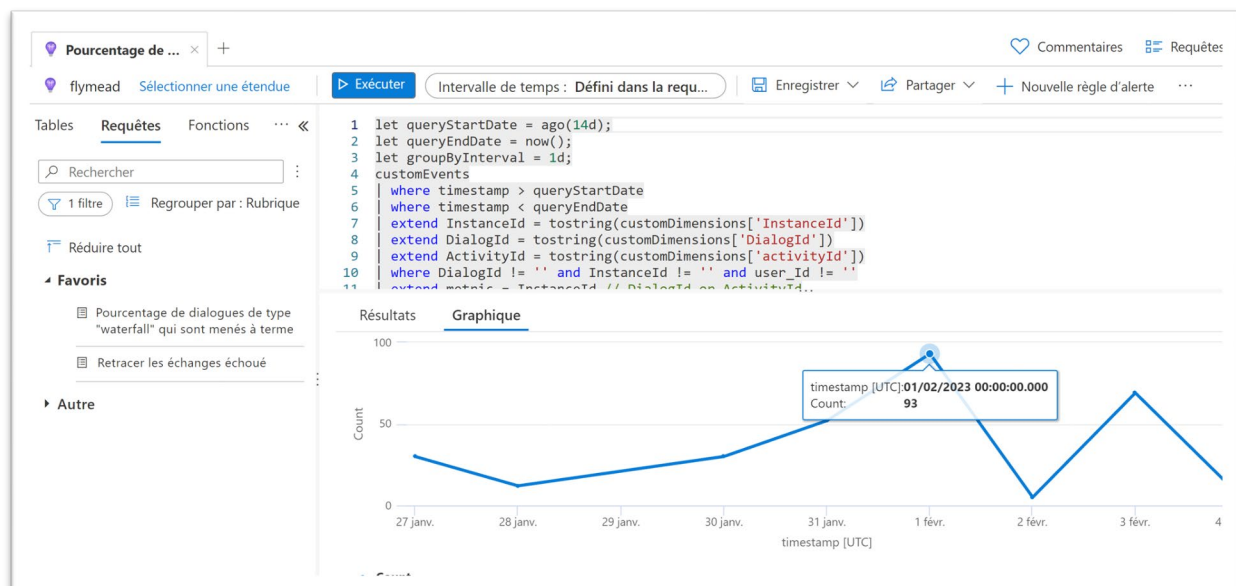


## IV. Modalités de mise à jour du modèle

Pour s'assurer de la qualité du modèle en production, il est nécessaire de recueillir des données sur les erreurs et les incidents. Microsoft Insight est utilisé pour surveiller les traces et les erreurs du chatbot et les remonter pour analyse. Cela permet de mesurer la performance du modèle et d'identifier les points à améliorer.



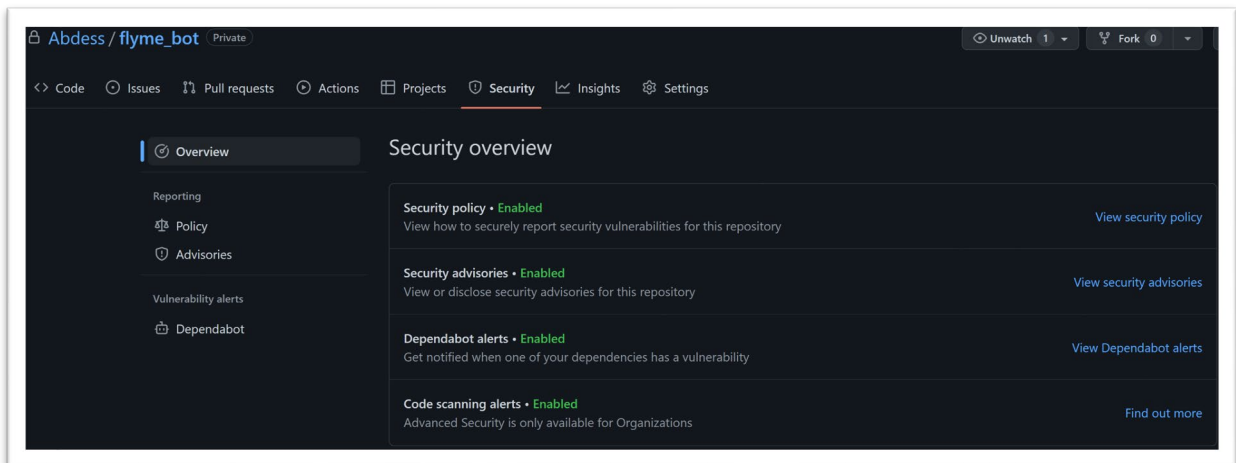
Une évaluation régulière de la performance du modèle est effectuée en utilisant les données recueillies avec **Microsoft Insight**. Les mesures de mise à jour sont déterminées en fonction de ces évaluations pour améliorer les résultats du modèle.



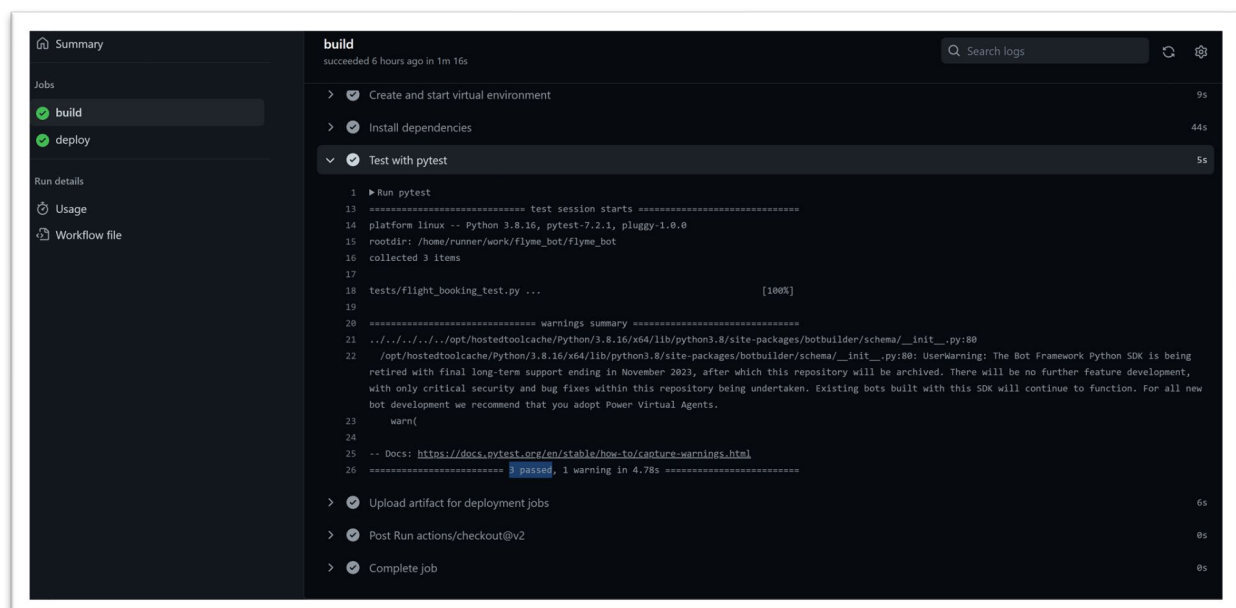


Des systèmes de **Workflow** sont en place pour garantir la qualité du développement du modèle. Cela permet d'optimiser les processus de mise à jour pour s'assurer que le modèle est toujours opérationnel et performant.

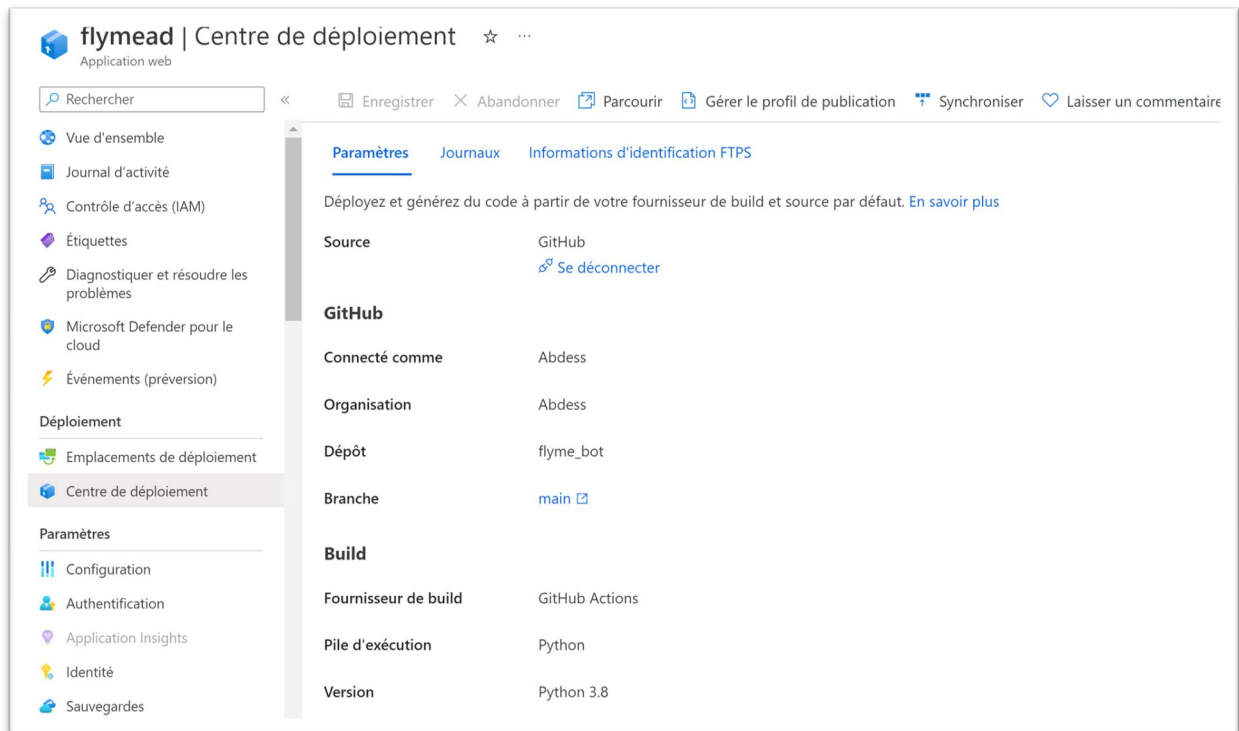
Pour s'assurer que le code est de qualité, des outils de validation de code tels que **SonarLint** et **GitHub Security** sont utilisés. Cela garantit que le code est sécurisé et conforme aux bonnes pratiques de développement.



Des **tests unitaires** sont effectués pour valider le fonctionnement correct du modèle avant sa mise en production.



Le déploiement continu est également mis en place avec **Azure App Service** pour s'assurer que le modèle est toujours opérationnel et performant.



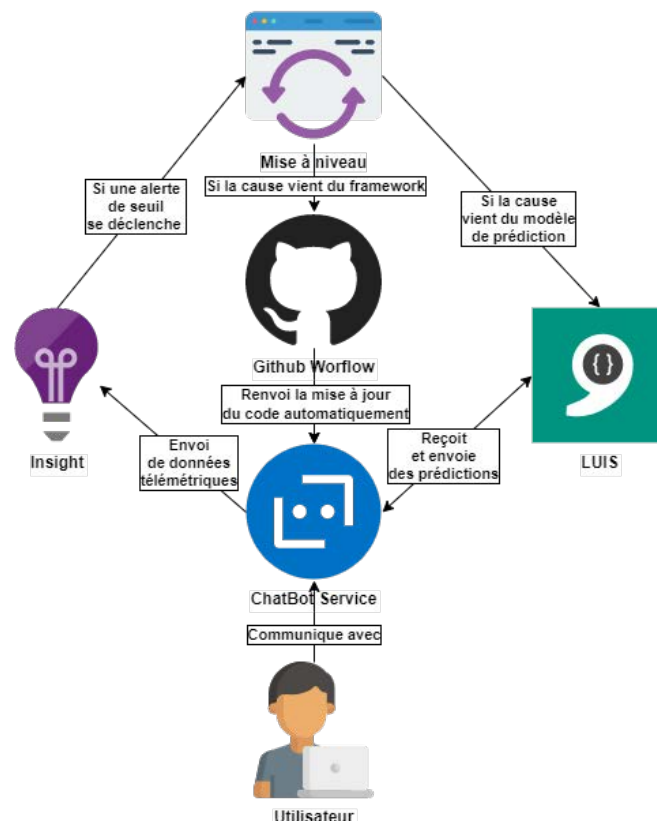
La fréquence de mise à jour du modèle est adaptable **en fonction des erreurs** remontées avec **Microsoft Insight**. Cela permet de s'assurer que le modèle est toujours performant et de minimiser les interruptions pour les utilisateurs.

## V. Conclusion

Le modèle de chatbot de réservation pour Fly Me est en production et il est crucial de surveiller sa performance pour garantir une expérience utilisateur satisfaisante. Les critères d'évaluation incluent la détection des informations dans les conversations et l'interprétation des intentions des utilisateurs par LUIS, ainsi que les données de formation du modèle. Le mécanisme d'évaluation se compose de la création de logs pour les traces et de critères d'évaluation pour détecter les erreurs telles que les dates non conformes, les confirmations de réservations, et les prédictions correctes validées par l'utilisateur.

Les modalités de mise à jour du modèle comprennent la remontée d'erreurs avec **Microsoft Insight**, l'évaluation pour les mesures de mise à jour, la mise en place de systèmes de **Workflow** pour garantir la qualité du développement, la validation du code avec des outils tels que **SonarLint** et **GitHub Security**, les tests unitaires et le déploiement continu avec **Azure App Service**. La fréquence de mise à jour est adaptable en fonction des erreurs remontées.

L'amélioration constante de la performance du modèle en production est essentielle pour garantir une expérience utilisateur de qualité. L'évaluation régulière et la mise à jour du modèle permettent de corriger les erreurs et d'améliorer les fonctionnalités du chatbot pour les utilisateurs. En fin de compte, cela contribue à l'amélioration continue du chatbot et garantit une expérience utilisateur optimale pour les clients de Fly Me.



## VI. Axes d'amélioration

Dans le but d'obtenir des résultats encore plus fiables, il est envisageable de poursuivre l'amélioration des critères d'évaluation du modèle en production. Cela permettra d'accroître la précision des interprétations et des prédictions du modèle, améliorant ainsi l'expérience utilisateur.

En plus de la remontée d'erreurs manuelle via Microsoft Insight, il est possible d'implémenter des algorithmes de **machine learning pour détecter automatiquement les erreurs**. Cela permettra de gagner en efficacité et de s'assurer de la qualité de la performance du modèle en temps réel.

L'optimisation du mécanisme d'évaluation peut être effectuée pour accélérer les temps de traitement. Cela permettra de mieux gérer les ressources informatiques et de s'assurer de la qualité de la performance du modèle en temps réel.

En élargissant la base de données utilisée pour la formation du modèle, il sera possible d'améliorer la compréhension des intentions des utilisateurs. Cela permettra d'augmenter la qualité des prédictions du modèle et d'améliorer l'expérience utilisateur.

Pour s'assurer de la qualité de l'expérience utilisateur il faut réaliser des évaluations périodiques de la performance du modèle en production. Cela permettra de prendre en compte les commentaires et les suggestions des utilisateurs pour améliorer la performance du modèle.