

# Méthodologie de pilotage de la performance du modèle en production

Pierre Augereau

---



## Table des matières

<i>Présentation .....</i>	<i>2</i>
<i>Critères d'évaluation .....</i>	<i>2</i>
LUIS .....	2
Validation utilisateur .....	2
<i>Évaluation du modèle en production .....</i>	<i>3</i>
LUIS .....	3
Validation utilisateur .....	3
<i>Mise à jour du modèle .....</i>	<i>3</i>

# Présentation

Un chatbot nécessite une évaluation de sa performance au fil de l'eau. Des outils (comme Azure Application Insight) existent pour surveiller chaque étape du fonctionnement. Dans ce document nous allons définir les critères et les outils à mettre en place afin d'obtenir un chatbot qui répond au mieux aux demandes des utilisateurs.

Lorsque l'utilisateur saisit sa première demande celle-ci est évaluée par le service LUIS de Azure et dans un deuxième temps le chatbot pose des questions sur les éléments qui n'auraient pas été détectés. Les éléments sont : la ville de départ, la ville d'arrivée, la date de départ, la date de retour et le budget.

## Critères d'évaluation

Les critères d'évaluation sont basés sur deux principes ; le premier sur les performances de LUIS (le moteur d'interprétation de langage) et le deuxième sur la validation ou non par l'utilisateur de la réservation.

## LUIS

LUIS est conçu pour identifier des informations dans les conversations et interpréter les intentions des utilisateurs. C'est un service de Microsoft Azure. Nous avons utilisé ce service car il permet de faire un premier filtre sur les phrases saisies par l'utilisateur. Le modèle a été entraîné par nos soins avec des données de conversations issues de données fournies par [Microsoft](#).

Nous évaluerons le modèle par rapport aux données prédites qui ne sont pas conformes à la structure requise. C'est le cas par exemple d'une date qui ne serait pas saisie correctement ou d'un budget dont le montant n'aurait pas pu être déterminé bien qu'ayant été reconnu comme budget.

## Validation utilisateur

En fin de conversation l'utilisateur doit choisir entre valider ou non le voyage. Il a aussi la possibilité d'annuler la conversation à tout moment avec le mot clé « cancel ».

Bien qu'il ne soit pas possible de déterminer les raisons du refus de validation nous utiliserons ce critère pour évaluer la satisfaction client.

# Évaluation du modèle en production

## LUIS

Nous allons mettre en place des traces qui s'écriront lorsque LUIS aura prédit une donnée qui n'est pas conforme à ce qui est attendu. C'est le cas par exemple pour :



Où la date départ et d'arrivée ne sont pas correctement détectées. Au bout de trois erreurs dans un intervalle de 5 minutes un message est envoyé au gestionnaire de compte.

## Validation utilisateur

A l'instar des critères d'évaluation de LUIS si 3 refus sur une période de 5 minutes sont détectés, les gestionnaires du compte seront avertis.

## Mise à jour du modèle

Après une remontée d'erreurs une évaluation doit être faite et les mesures nécessaires doivent être prises. Cela peut être d'entraîner le modèle avec de nouvelles données ou d'avoir une meilleure interprétation des données saisies par l'utilisateur.

Nous avons mis en place un système de Workflow dans le processus de développement afin de garantir une qualité des programmes déployés.

1. Validation du code par rapport aux standards actuels ([Codacy](#))
2. Validation du code par rapport aux vulnérabilités connues (Security Policy GitHub)
3. Tests unitaires
4. Déploiement continu avec Azure App Service

La fréquence de mise à jour doit être variable dans le temps. En effet il convient dans un premier temps d'être attentif aux erreurs qui pourraient être remontées et agir au plus vite pour ne pas ternir l'expérience utilisateur.