

Comparaison des solutions de création de chatbots

Sommaire

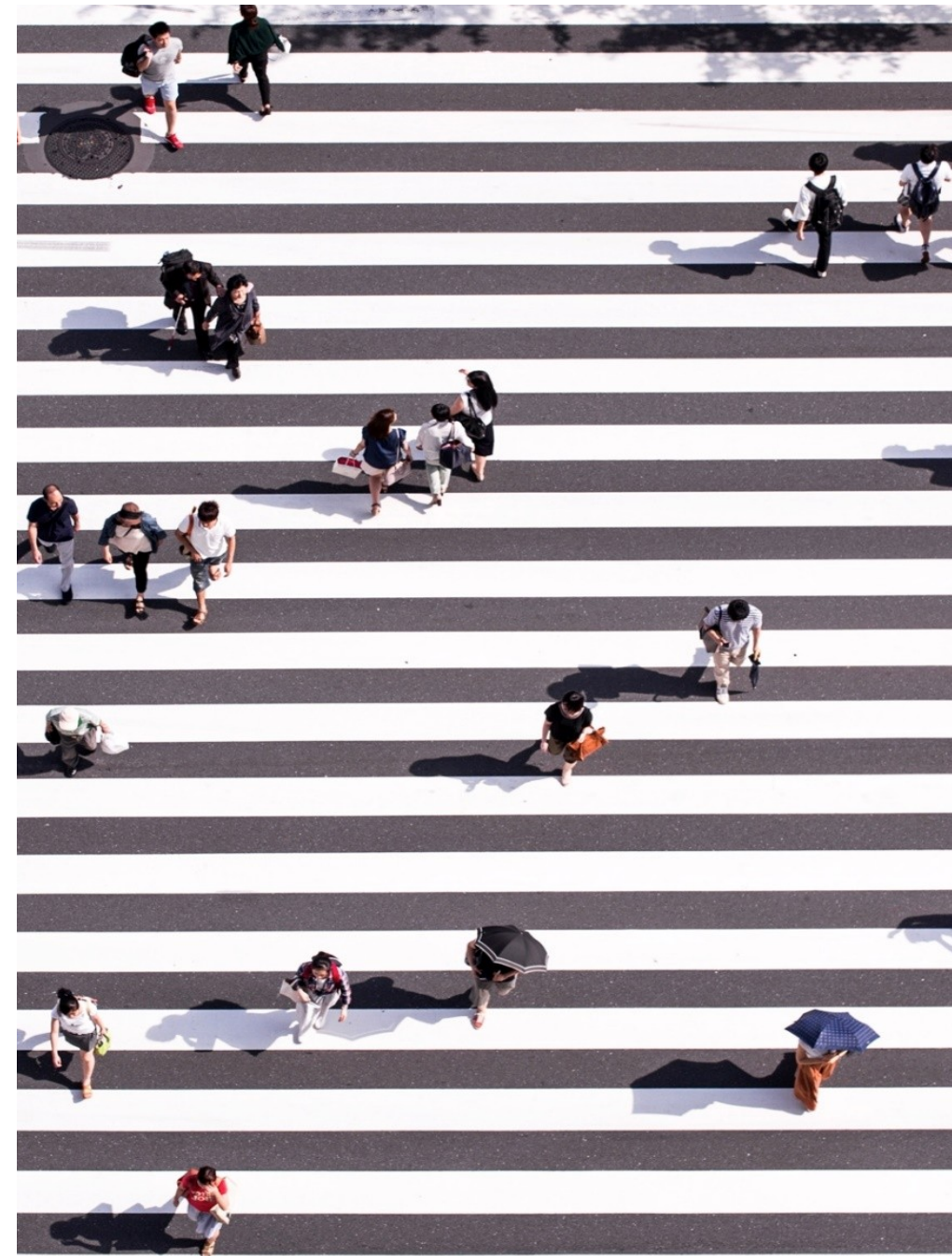
Introduction

Les deux catégories de chatbots

Fonctionnement d'un chatbot NLP

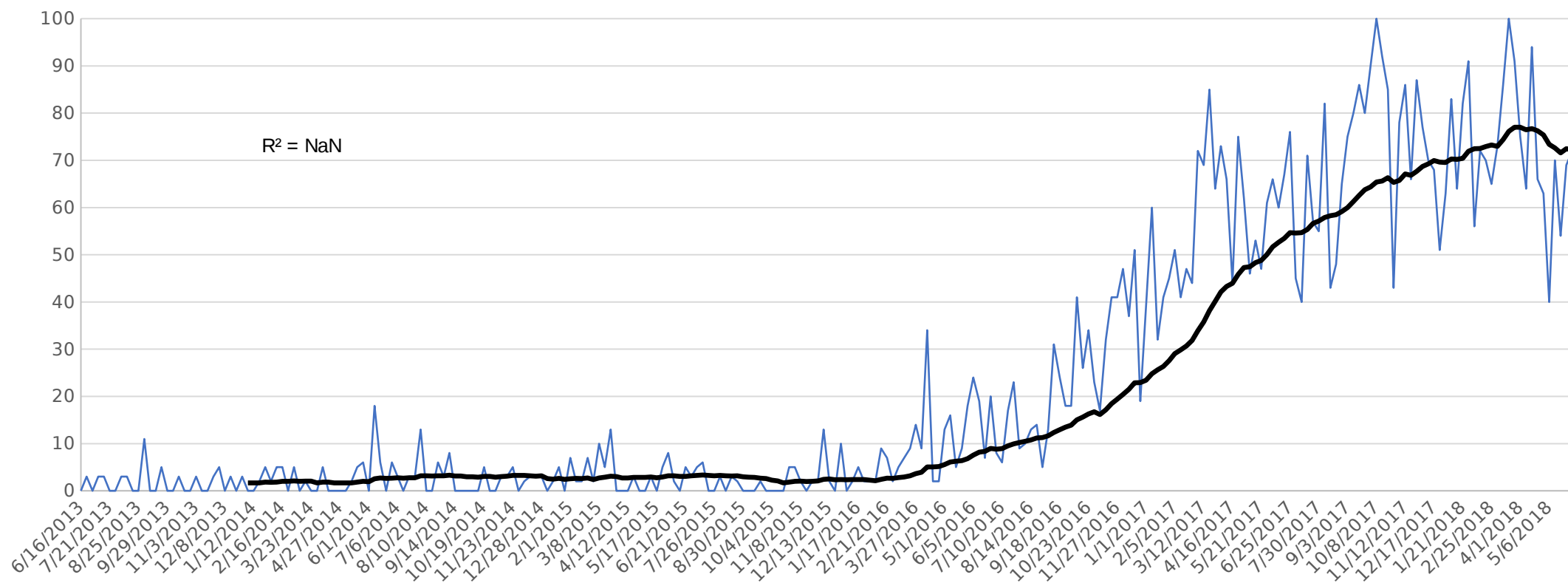
Comparaison des outils

Conclusions



Évolution de l'intérêt pour la recherche « chatbot » sur google

% des recherches (relatif au maximum du terme)



Les principaux chatbots



Historical timeline



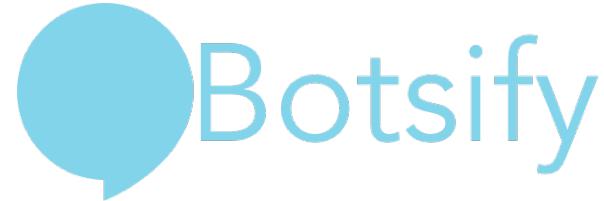
© Intento, Inc.

August 2017

D'autres outils

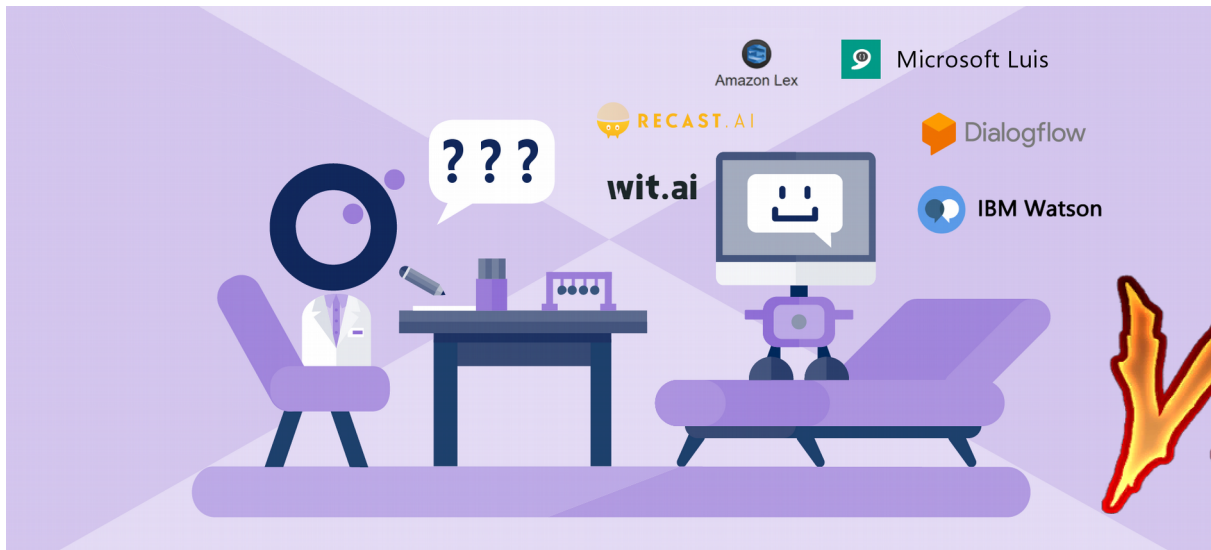
- Semantic Machines
- Chatfuel
- Rasa NLU
- ...

Il en existe des centaines

The logo for Chatfuel, featuring a dark blue speech bubble icon followed by the word "Chatfuel" in a sans-serif font, with "Chat" in dark blue and "fuel" in orange.The logo for Botsify, consisting of a light blue speech bubble icon followed by the word "Botsify" in a light blue sans-serif font.The logo for Semantic Machines, featuring a stylized, multi-colored wave or ribbon icon above the text "semanticmachines" in a lowercase, grey sans-serif font.The logo for Rasa NLU, featuring the word "RASA" in a bold, white, sans-serif font inside a purple speech bubble outline, followed by "NLU" in a purple sans-serif font.

Deux catégories de chatbots

Bots NLP



VS

Bots séquentiels



Les bots séquentiels

Le bot est comme sur des rails, et continue la discussion jusqu'à l'aboutissement.

Les humains ont tendance à changer souvent d'avis, ce modèle est donc très limité.

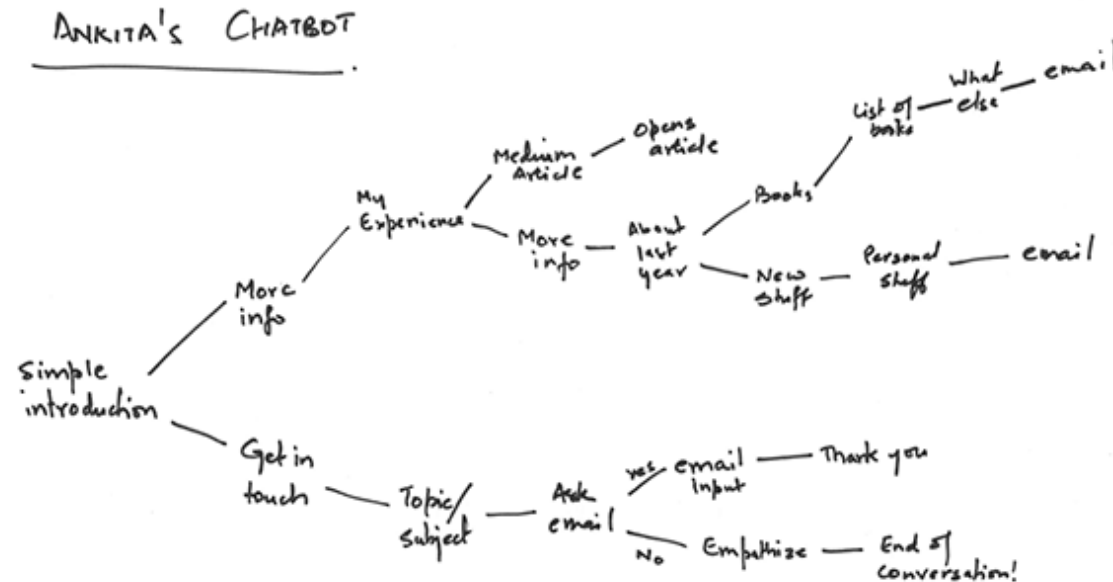


Schéma séquentiel

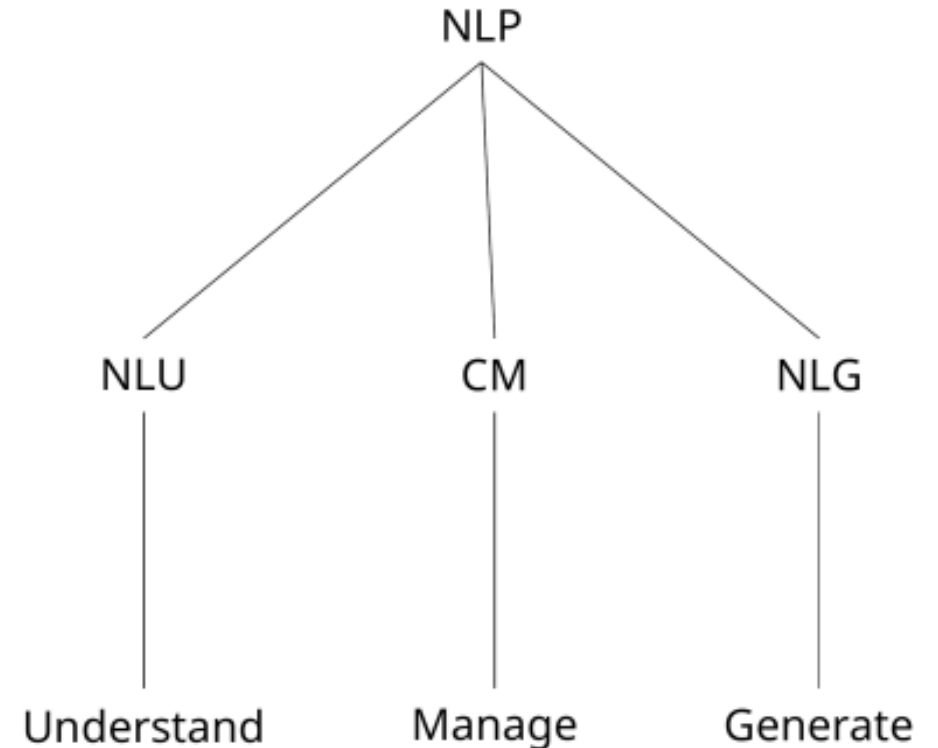
Les bots basés sur la technologie NLP

Les bots intelligents utilisent le modèle **NLP (Natural Language Processing)**, et sont composés de trois parties :

NLU (Natural Language Understanding) : Ce module utilise la syntaxe, l'étude sémantique pour reconnaître des modèles et extraire les entités.

CM (Conversation Manager) : Cette partie s'occupe de gérer la conversation. Elle permet la notion de contexte et la notion d'intention. Est aussi appelé **DM (Dialog Manager)**

NLG (Natural Language Generation) : Ce module permet de transformer les informations de la réponse en une phrase cohérente. C'est l'opposé du NLU

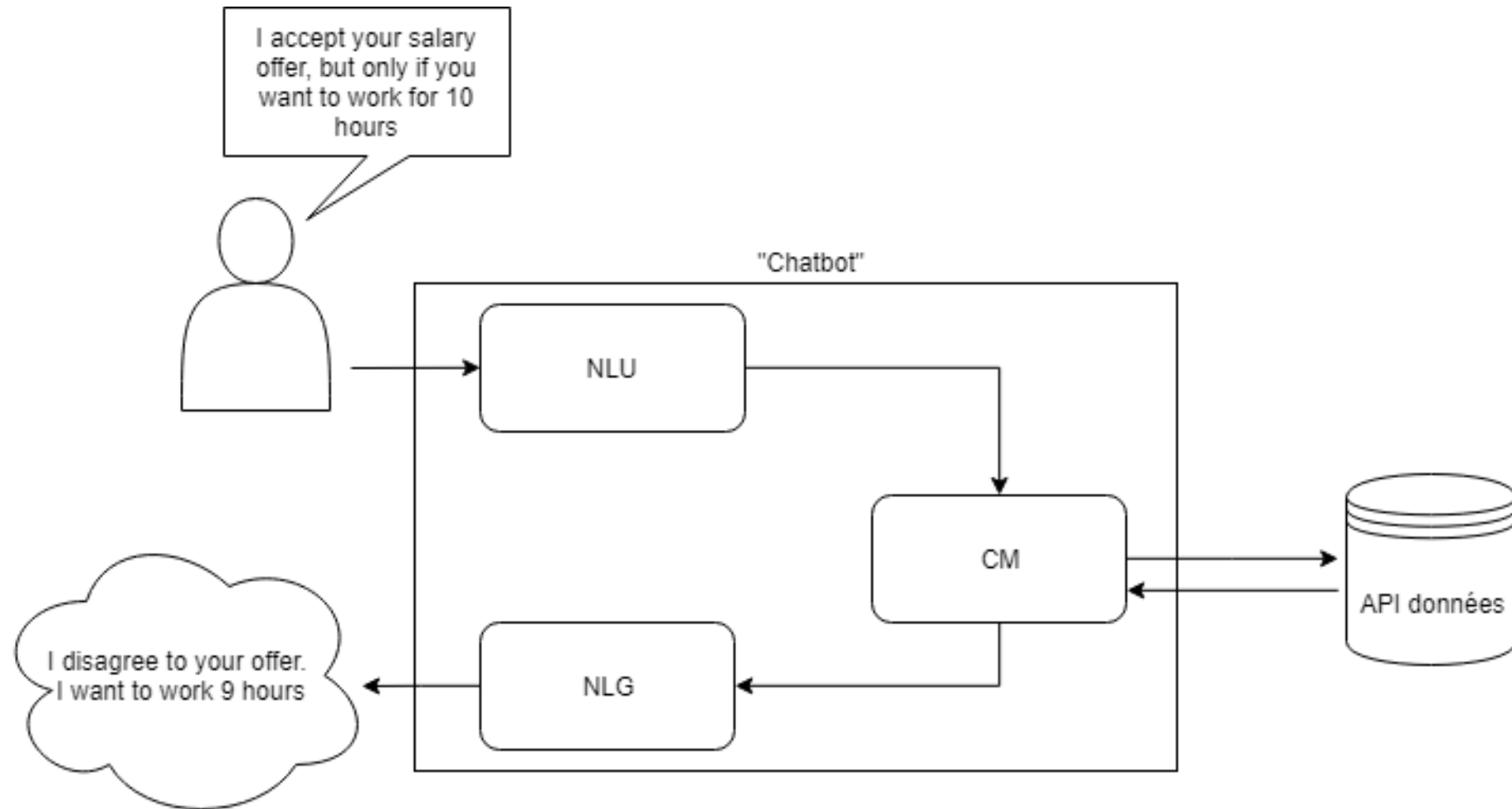


Les limites du NLP

- Sujet nouveau
- Conception nécessitant beaucoup de compétences (UX, Dev, Archi, créatif...)
- Phase de test complexe
- MCO / MCE

Le fonctionnement d'un NLP

Schéma de fonctionnement d'un NLP

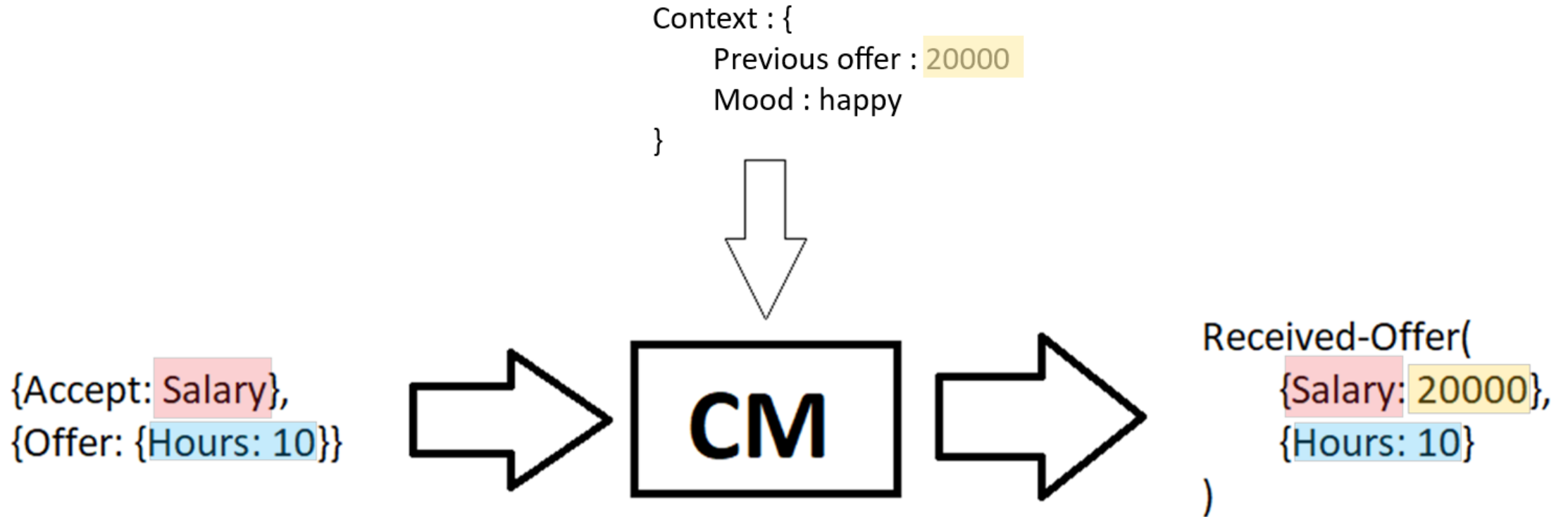


Natural Language Understanding



Le NLU est un moteur permettant de découper le langage naturel en éléments exploitables. Le NLU est généralement une black box, difficile de savoir exactement ce qu'il s'y passe.

Conversation Manager



Le CM permet de gérer la conversation. Dans la plupart des cas il gère les contextes, les variables de session. Il peut être inclus dans une solution toute prête, ou à développer soit même. Les CM possèdent des approches et des philosophies différentes. Les fonctionnalités et les possibilités diffèrent elles aussi.

Le CM appelle ensuite nos fonctions (ici, Received-Offer) qui communiqueront par exemple avec une base de donnée

Natural Language Generation



Le NLG convertit les résultats de notre action en langage naturel.
Cette partie prend souvent la forme de templates et n'est que partiellement présente dans les solutions du marché

Les outils

		NLU	CM	NLG
Google	DialogFlow	+	++	+
SEP	Recast.ai	+	+	+
Amazon	LEX	+		
IBM	Watson Assistant	+		
Facebook	Wit.ai	+	+	+
Microsoft	LUIS	+		
Microsoft	Azure bot service		+++	+
RASA*	Rasa NLU	+		
RASA*	Rasa Core		++	+
SNIPS*	Snips NLU	+		

L'Open source, une possibilité ?

Avantages :

- contrôle total des données,
- prix,
- contrôle du déploiement

Inconvénients :

- développé par une startup et la communauté
- documentation parfois limitée
- certaines fonctionnalités expérimentales

Migration entre outils

- Il est relativement facile de changer de NLU, et beaucoup plus difficile de changer de CM
- Il existe des tutoriels d'import des données de DialogFlow / wit.AI / LUIS vers RASA-nlu
- RASA-nlu permet aussi d'émuler le fonctionnement des autres NLU (qui est une simple conversion du format JSON de sortie)

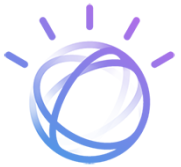
Les meilleurs outils de création de chatbots



- Dialogflow :
 - Bonnes performances
 - Conversation Manager plutôt complet et rapide à prendre en main



- LUIS :
 - Rapidité (pour une API)
 - La galaxie des outils Microsoft



- Watson :
 - Performances



- Snips NLU :
 - Open source
 - Simplicité



- Rasa :
 - Open source
 - Conversation Manager (Rasa-core)

Les outils éliminés

- Wit.AI :
 - Pas de version professionnelle
- LEX :
 - Pas de support du français. (MAJ 2018 : le français est en beta)

Documentations des chatbots

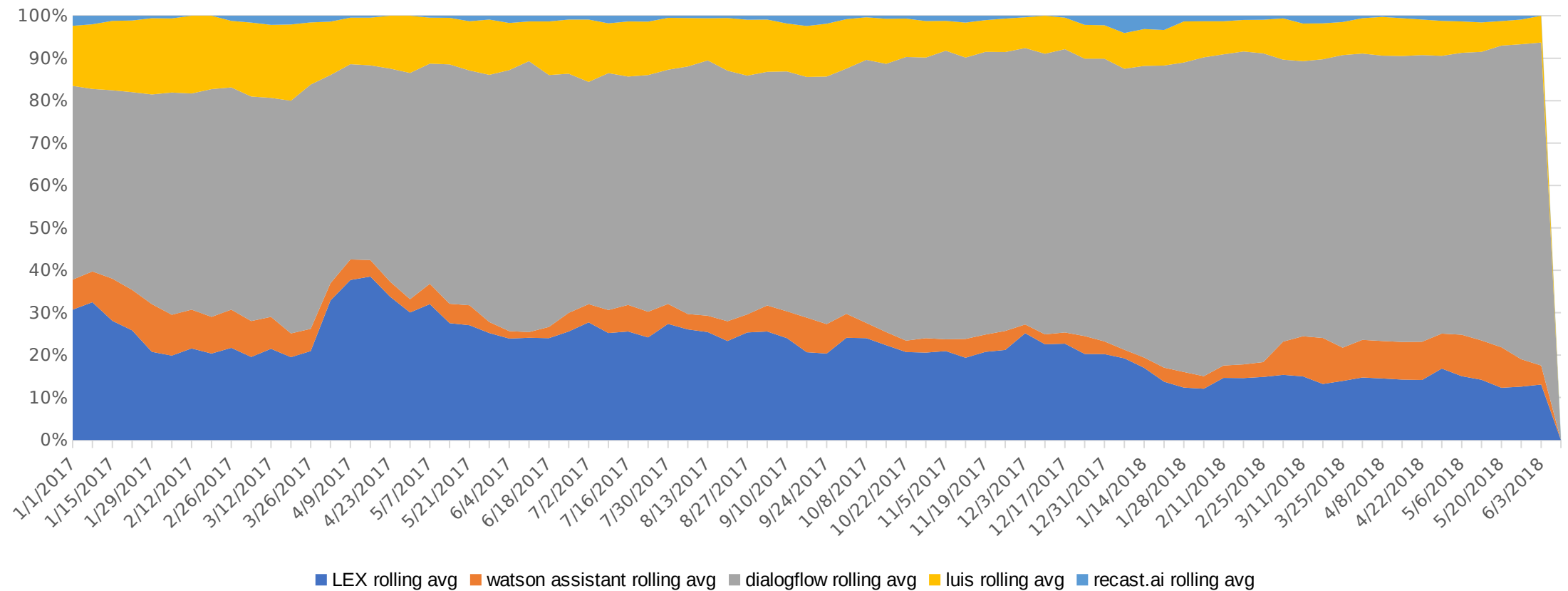
- <https://docs.microsoft.com/fr-fr/azure/>
- <https://aws.amazon.com/fr/lex/>
- <https://www.ibm.com/watson/services/conversation/>
- <https://recast.ai/docs/create-your-bot>
- <https://dialogflow.com/docs/getting-started/basics>
- <https://wit.ai/docs>

Autres sources

- <https://tech.fpt.com.vn/language/en/natural-language-processing-problems-chatbot-system-development/>
- <https://chatbotsmagazine.com/>
- https://www.researchgate.net/figure/Dialog-System-Architecture-Example-starts-at-the-top-left-corner_fig1_273518960
- <https://www.slideshare.net/KonstantinSavenkov/nlu-intent-detection-benchmark-by-intento-august-2017>
(benchmark)
- <https://blog.octo.com/comparatif-moteurs-chatbots/>

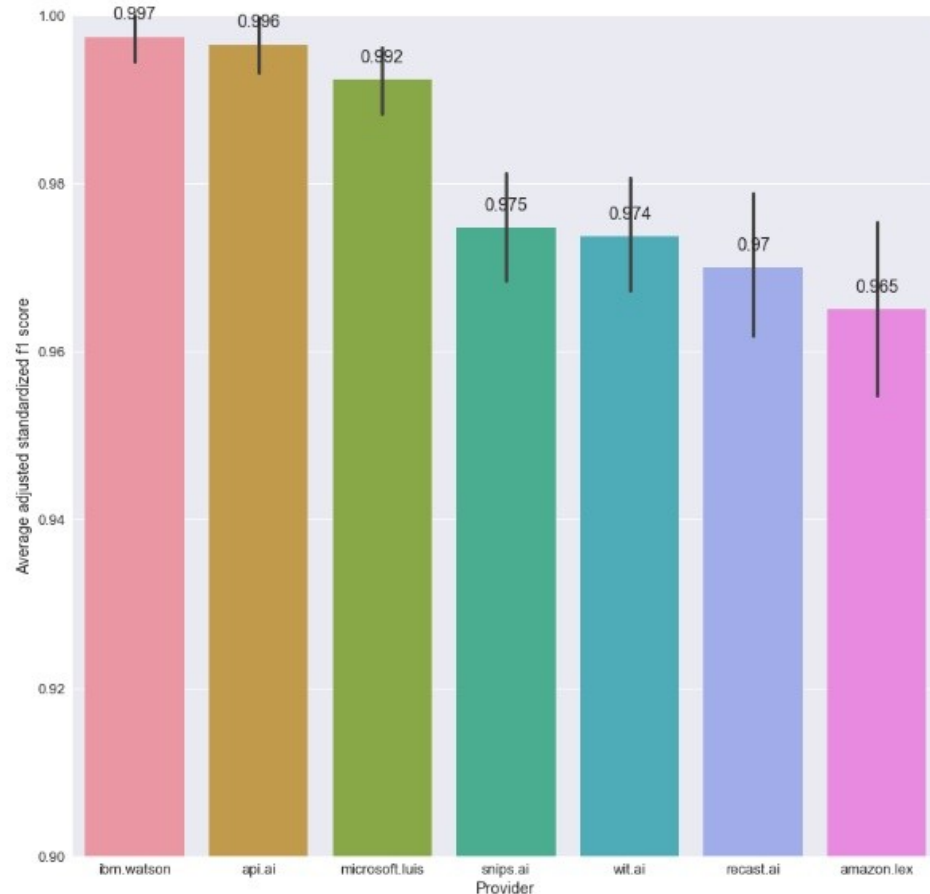
Popularité des outils propriétaires

popularité relative sur google





Detection Performance*



- Black bars indicate confidence intervals
- Amazon Lex is trained on a smaller dataset due to its API limits

Watson, Dialogflow, LUIS devant

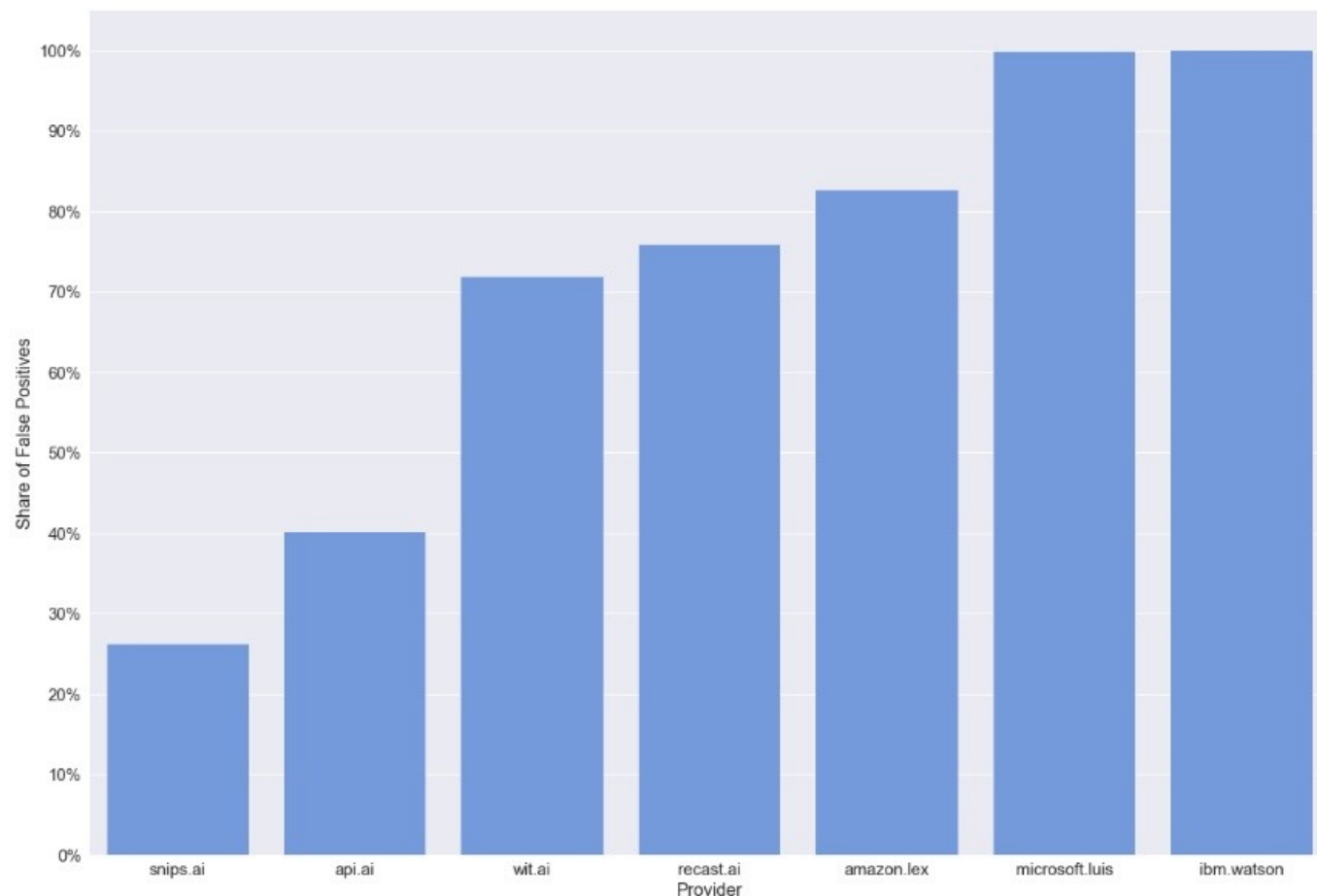
A noter aussi que même si il n'est pas représenté, les performances sont comparables entre SNIPS et Rasa

* Mean standardized F1 scores, adjusted to the initial scale using global mean and std
© Intento, Inc.

August 2017



Out-of-the-Domain Samples



© Intento, Inc.

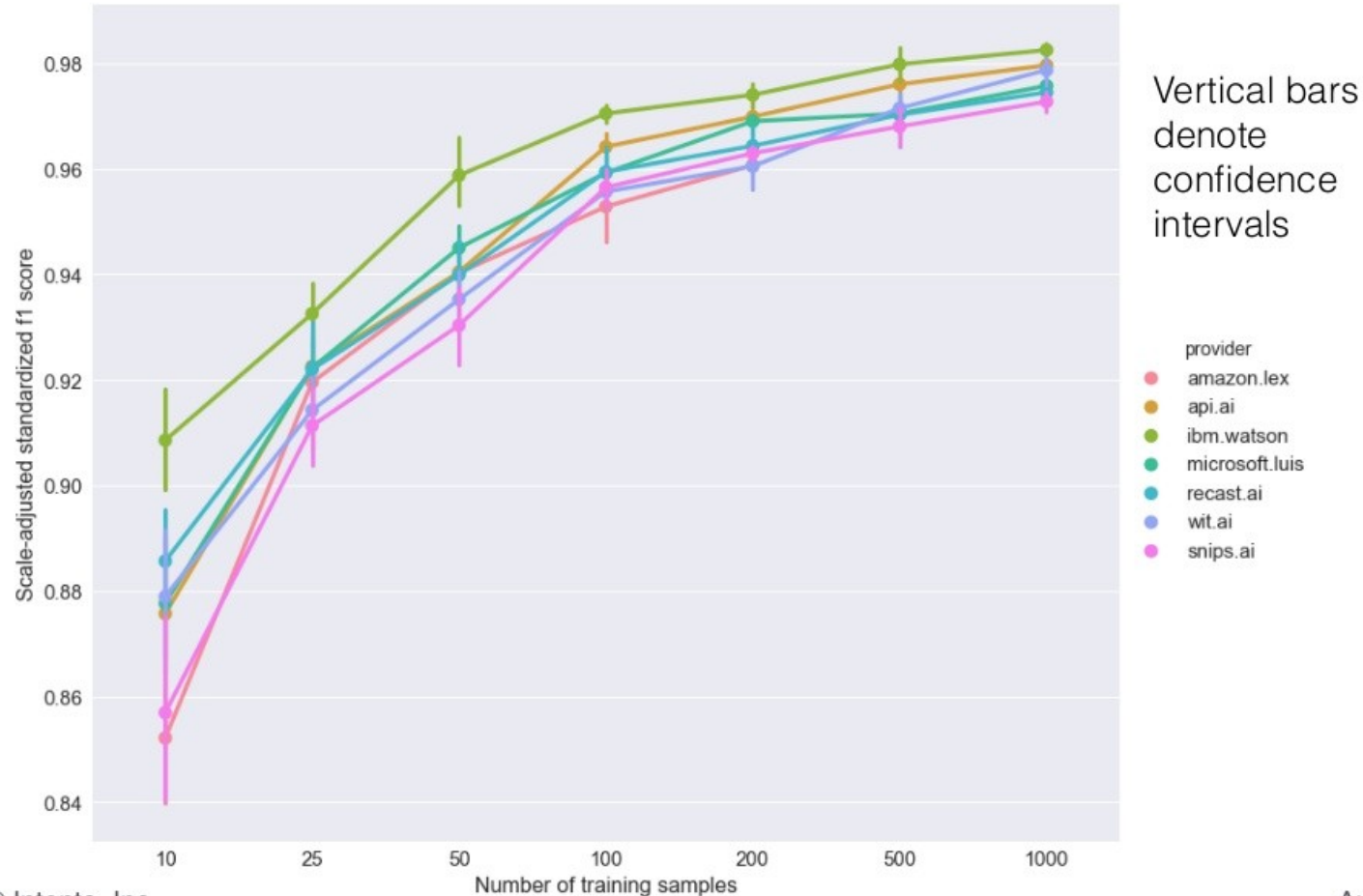
August 2017

Les Faux positifs :
Dialogflow s'en sort très bien. Snips s'en sort encore mieux.

Watson et LUIS tentent de renvoyer une intent à chaque fois, donc ils renvoient énormément de faux positifs (ici 100% !)



Leaning curve by provider

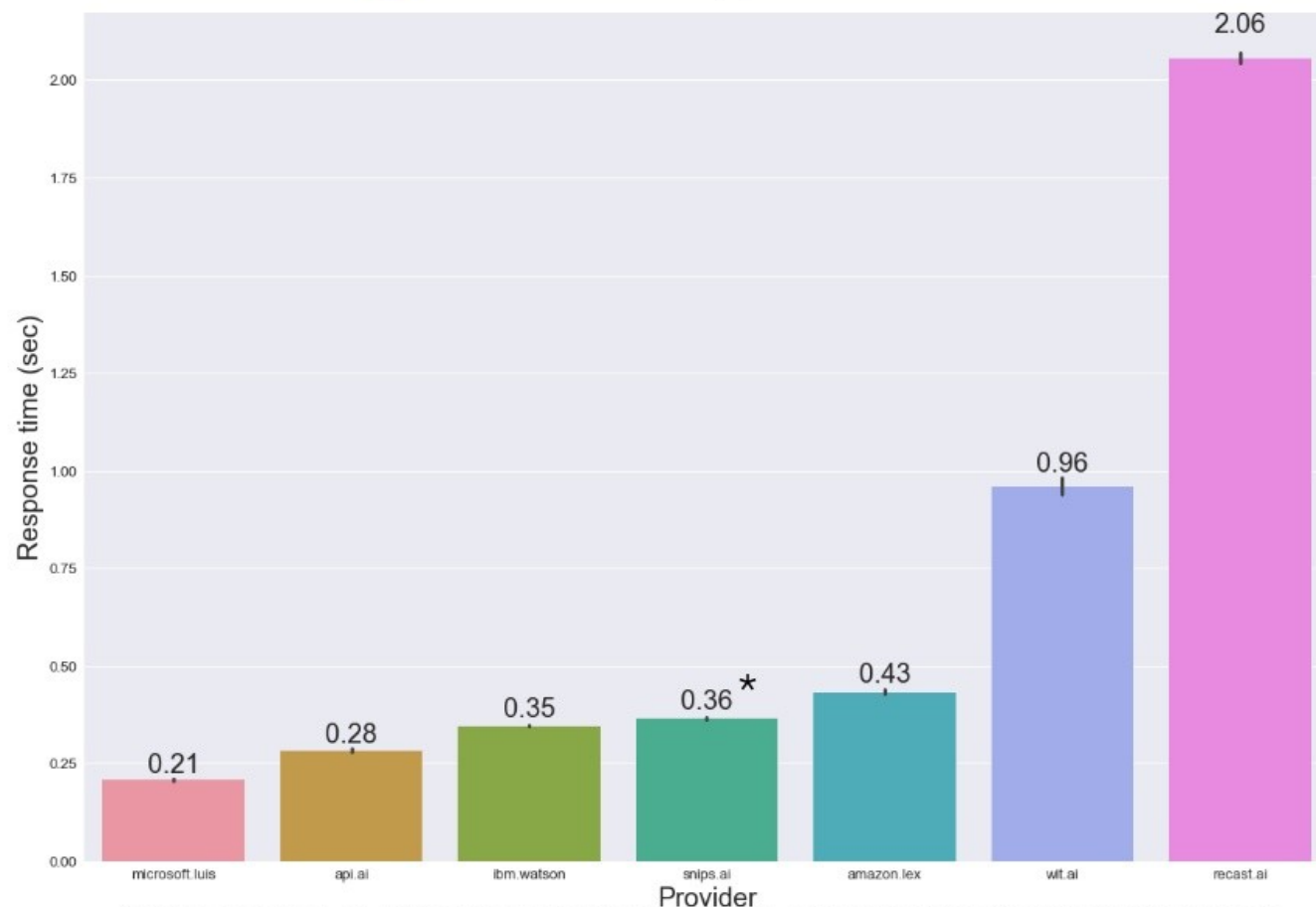


Courbe d'apprentissage :
Watson devant, en particulier
pour les faibles nombres de
données. Pour les grands
nombres de données
Dialogflow le rattrape.

La plupart des intents par
défaut sont construites avec
10 - 50 uttérances (phrases
d'entraînement)



Average response time



LUIS, Dialogflow,
Watson en tête

A noter que SNIPS NLU
(tout comme Rasa) est
aussi disponible en
déploiement sur une
machine, rendant la
latence proche de 0

© Intento, Inc. * Snips assumes on-device deployment, we put here a response time for a hosted test bench August 2017