Como construir chatbots inteligentes sin morir en el intento

@JordiCabot / @ingdesoftware / @xatkit– jordicabot.com / ingenieriadesoftware.es

1

Introduction

2

About

3

SOM Research Lab

Software runs the world. Models run the software

4

Nuestra misión

Interested in the broad area of systems and **software engineering**, especially promoting the rigorous use of software **models and engineering principles** in all software engineering tasks.

Flickr/clement127

5

6

Context

7

Once upon a time in 2018 (we were young and naïve)

Berger Levrault

Can you help us create a chatbot to help newcomers to write good bug reports on GitHub?

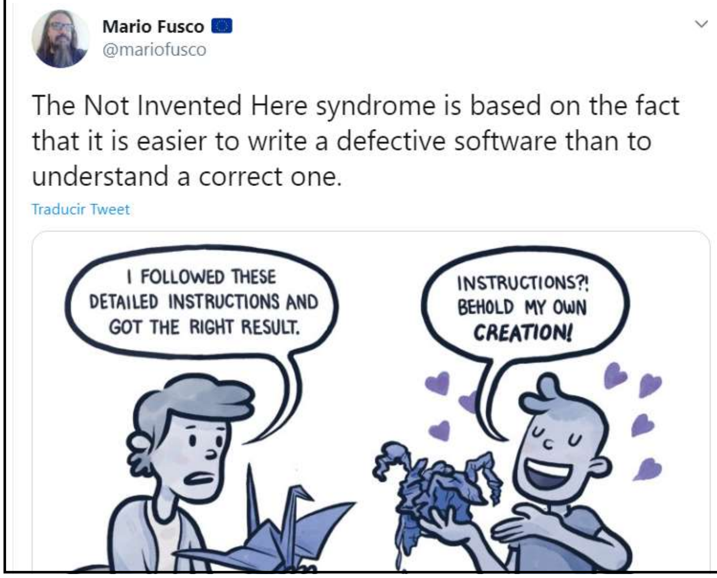
Sure! It's just a set of questions & answers, this will be pretty simple!

Narrator

It wasn't.

8

We were asked to build a bot. We ended up building a bot platform 3 times so far!



Mario Fusco
@mariofusco

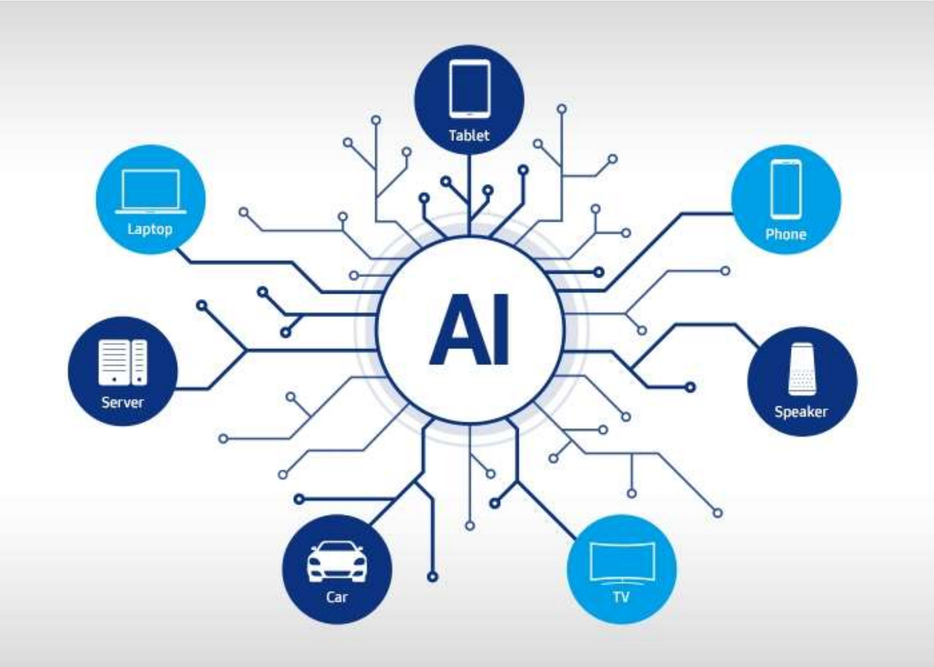
The Not Invented Here syndrome is based on the fact that it is easier to write a defective software than to understand a correct one.

Traducir Tweet

I FOLLOWED THESE DETAILED INSTRUCTIONS AND GOT THE RIGHT RESULT.
INSTRUCTIONS?! BEHOLD MY OWN CREATION!

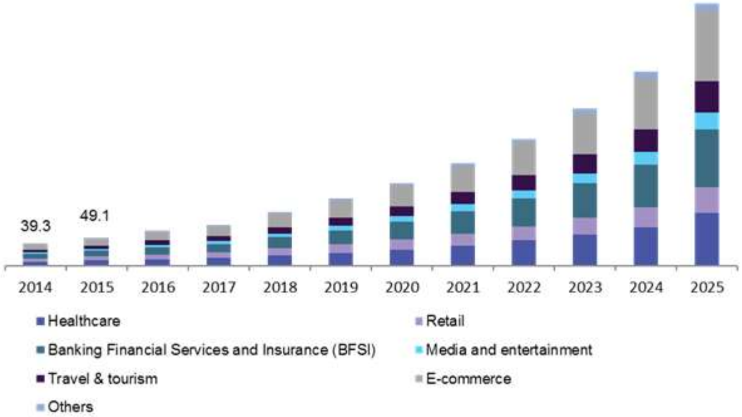
But at least we've learnt some things along the way that today we want to share with you

Why you all need to have some bot building notions



11

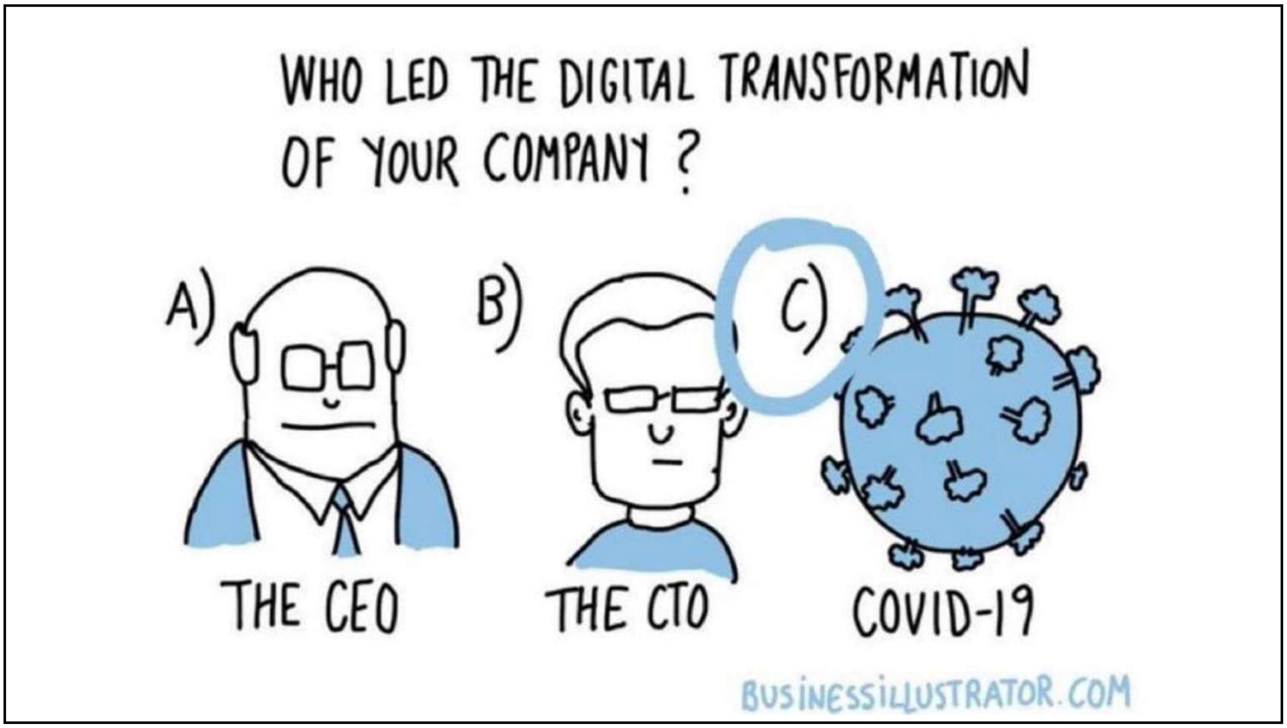
Chatbot market is worth over 1 Billion USD



And more predictions:

- By 2024, consumer retail spend via chatbots worldwide will reach $142 billion—up from just $2.8 billion in 2019.

12

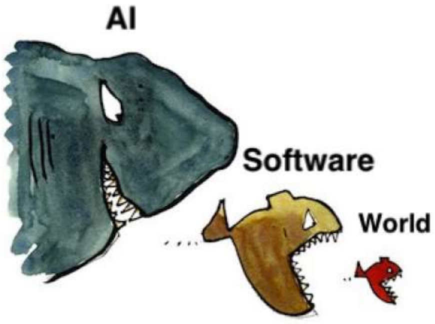


13

Chatbots <-> Software engineers

- Software Engineers -> Chatbot
  - Chatbots are a new type of software artefact that we'll need to consider in our dev projects
- Chatbot -> Software Engineering
  - Bots are more and more playing a supportive role in many SE tasks (e.g. in GH) so better to know how they work and their limitations

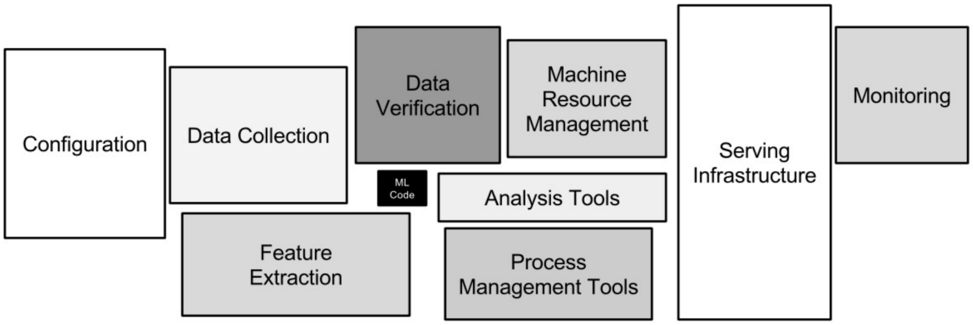
14



“Software is eating the world”
Andreessen Horowitz, HP (2011)

“Software is eating the world, but AI is going to eat software”
Jensen Huang, Nvidia CEO (2017)

15



Configuration

Data Collection

Data Verification

Machine Resource Management

ML Code

Analysis Tools

Feature Extraction

Process Management Tools

Serving Infrastructure

Monitoring

Figure 1: Only a small fraction of real-world ML systems is composed of the ML code, as shown by the small black box in the middle. The required surrounding infrastructure is vast and complex.

Hidden Technical debt in Machine Learning Systems - Google - https://proceedings.neurips.cc/paper/2015/file/86df7dcfd896fcaf2674f757a2463eba-Paper.pdf

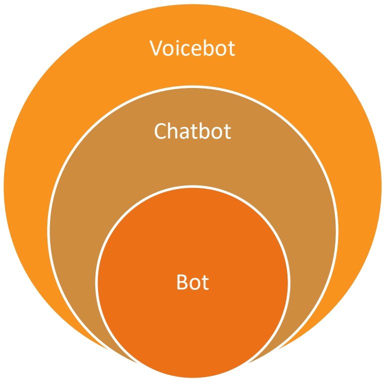
16

Core chatbot concepts

17

What is exactly a Chatbot?

- Bot are software components
- Chatbot = Bot + Chat interface
  - Tb llamado asistente digital, asistente virtual,…
- Voicebot = Chatbot + Voice Interface



Voicebot

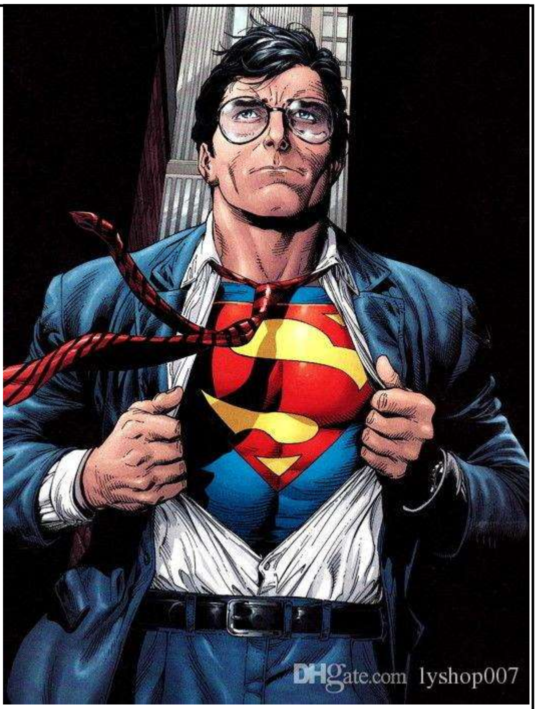
Chatbot

Bot

18

VoiceBot

- Alexa, Google Assistant ,…
- They all hide a chatbot inside
  1. Speech-to-text
  2. Process the text as a chatbot
  3. Text-to-speech to reply



19

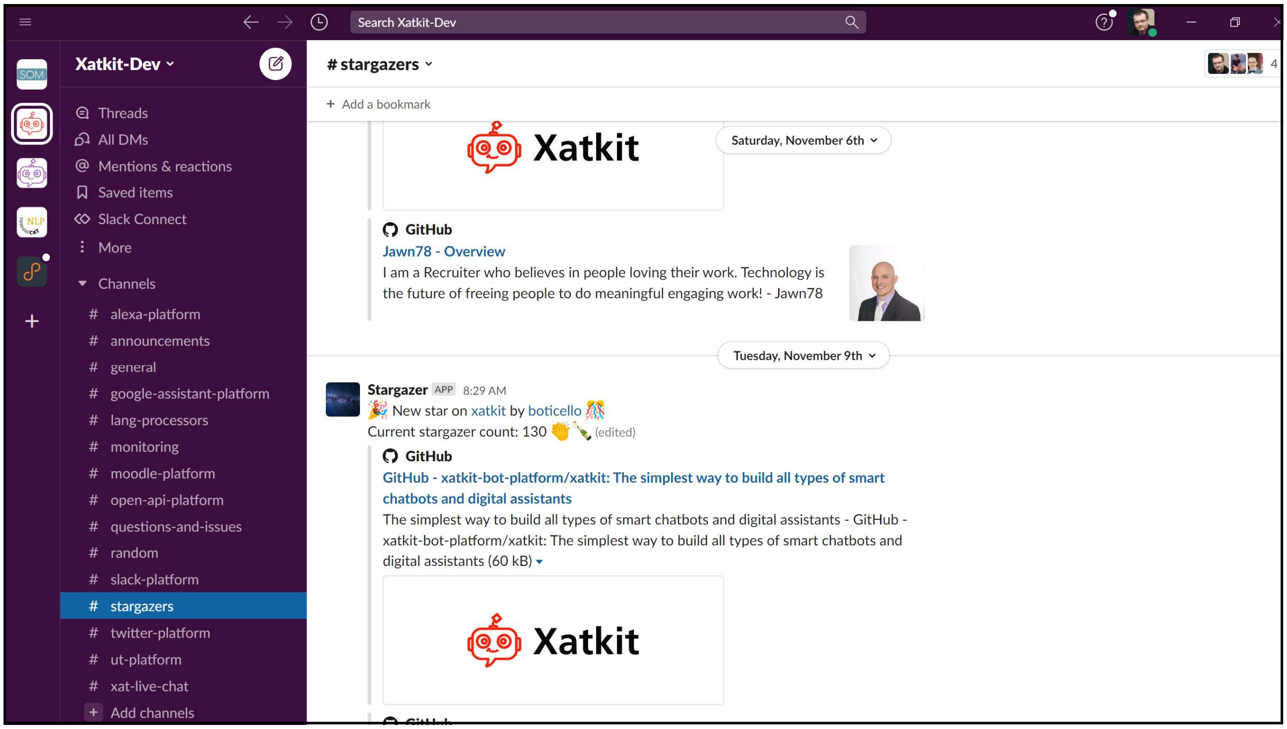
Chatbot vs LiveChat

- LiveChat: a person is behind
- They can be combined
  - Bot → Person: when the bot doesn't know how to answer (or the visitor is specially "interesting")
  - Person → Bot: when no support person is available or the client is not worth

20

Examples of bots

21



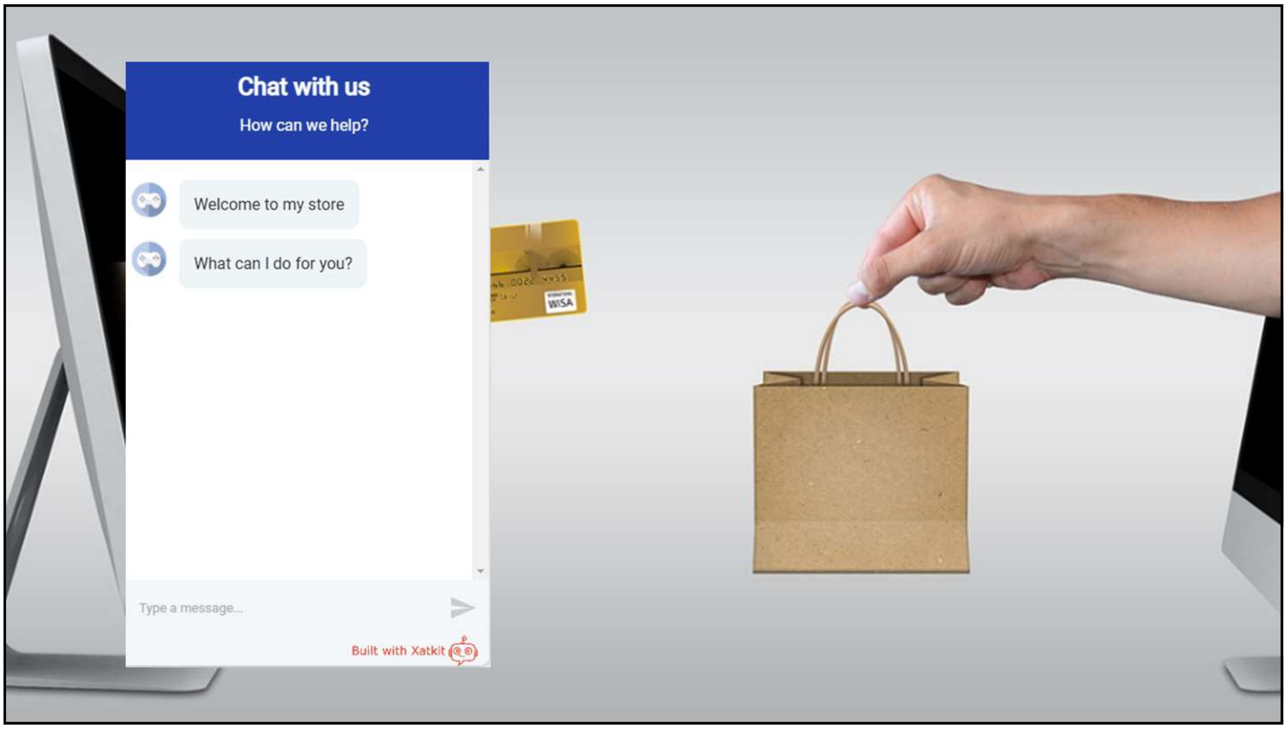
GitHub
Jawn78 - Overview
I am a Recruiter who believes in people loving their work. Technology is the future of freeing people to do meaningful engaging work! - Jawn78

Stargazer APP 8:29 AM
🎉 New star on xatkit by boticello 🎊
Current stargazer count: 130 👏 🍾 (edited)

GitHub
GitHub - xatkit-bot-platform/xatkit: The simplest way to build all types of smart chatbots and digital assistants
The simplest way to build all types of smart chatbots and digital assistants - GitHub - xatkit-bot-platform/xatkit: The simplest way to build all types of smart chatbots and digital assistants (60 kB) ▾

GitHub

22



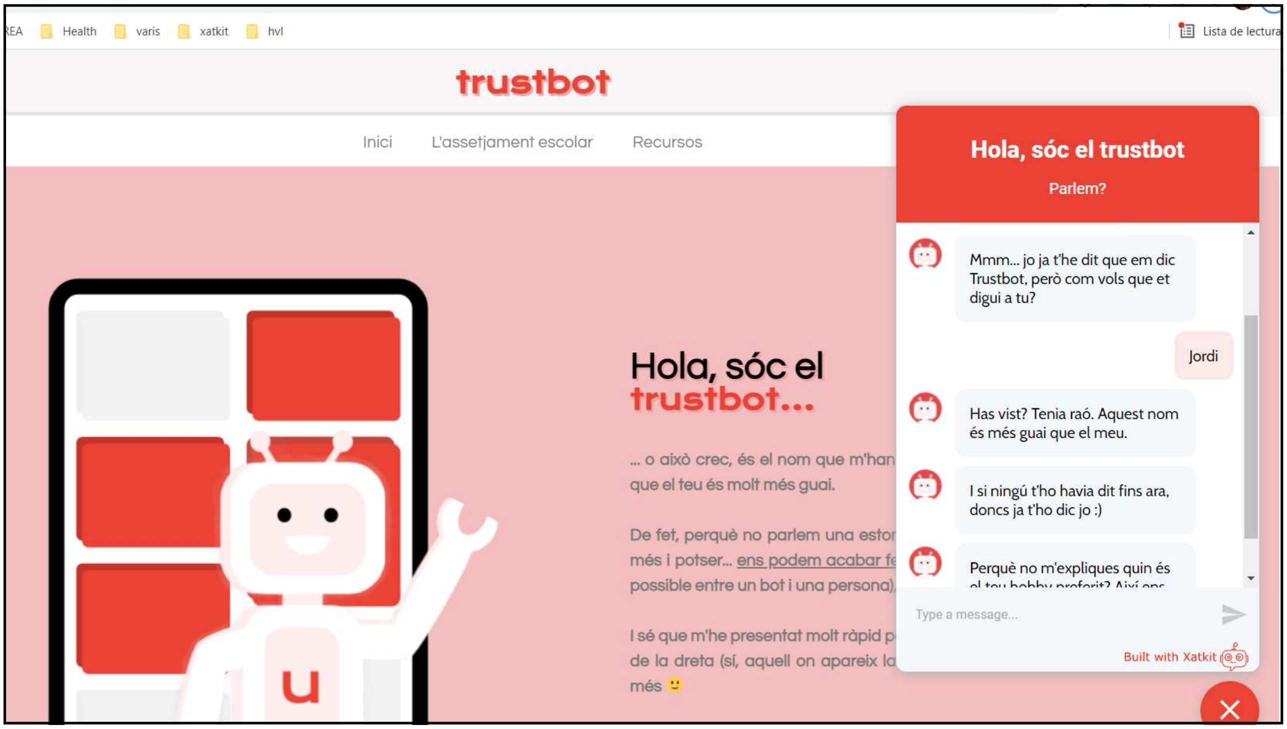
Chat with us

How can we help?

Welcome to my store

What can I do for you?

23



trustbot

Inici L'assetjament escolar Recursos

Hola, sóc el trustbot...

... o això crec, és el nom que m'han que el teu és molt més guai.

De fet, perquè no parlem una estor més i potser... ens podem acabar fe possible entre un bot i una persona),

I sé que m'he presentat molt ràpid p de la dreta (sí, aquell on apareix la més 🙂

Hola, sóc el trustbot

Parlem?

Mmm... jo ja t'he dit que em dic Trustbot, però com vols que et digui a tu?

Jordi

Has vist? Tenia raó. Aquest nom és més guai que el meu.

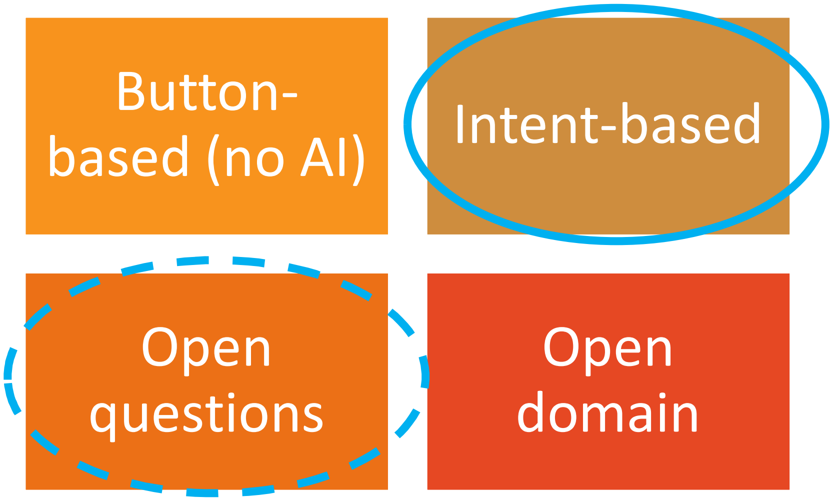
I si ningú t'ho havia dit fins ara, doncs ja t'ho dic jo :)

Perquè no m'expliques quin és el teu hobby preferit? Així ens

24

Types of chatbots

Types of chatbots



Button-based (no AI)

Intent-based

Open questions

Open domain

We ignore open-domain bots (e.g. LaMDA)

- They can chat about *everything* but
  - A complete black-box
  - ``Useless’’ for a specific business

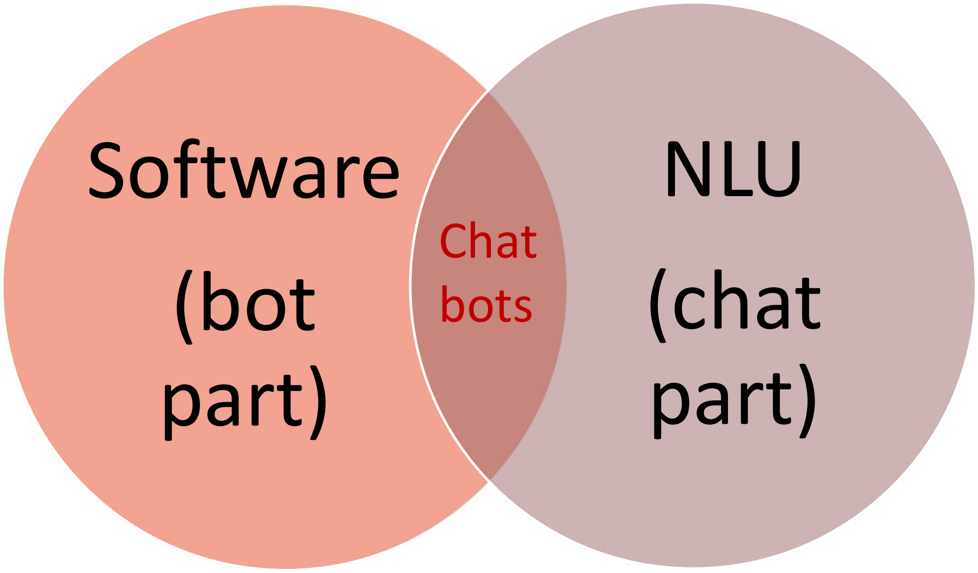
https://mindmatters.ai/2022/03/the-ai-illusion-state-of-the-art-chatbots-arent-what-they-seem/

27

Creating a chatbot 1 – Why is it so challeging?

28

Complexity of chatbots



Software

(bot part)

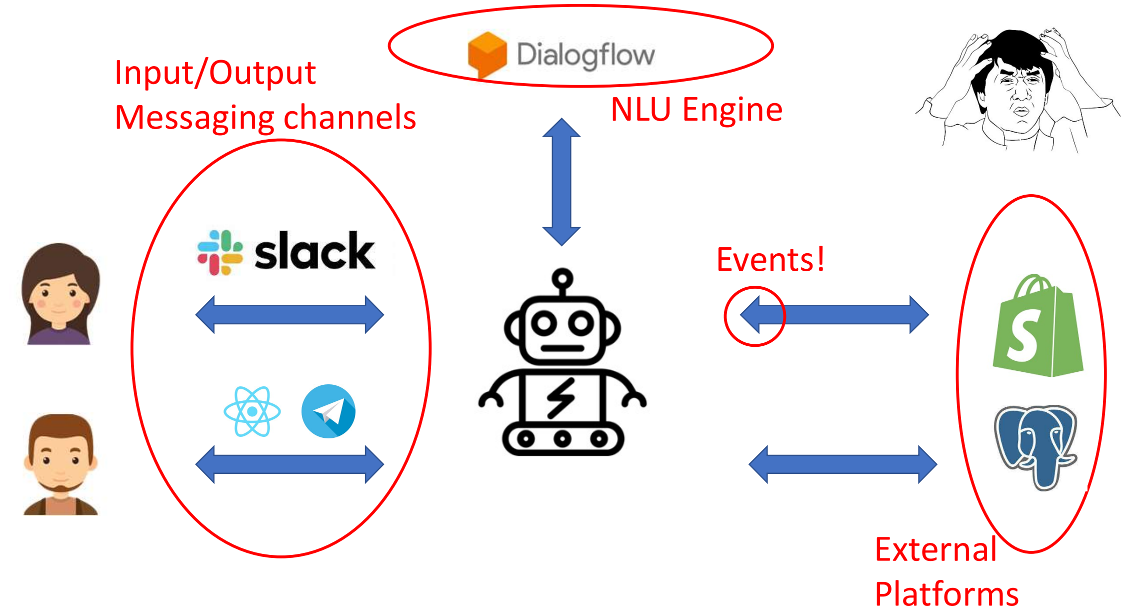
Chat bots

NLU

(chat part)

29

Many pieces to assemble



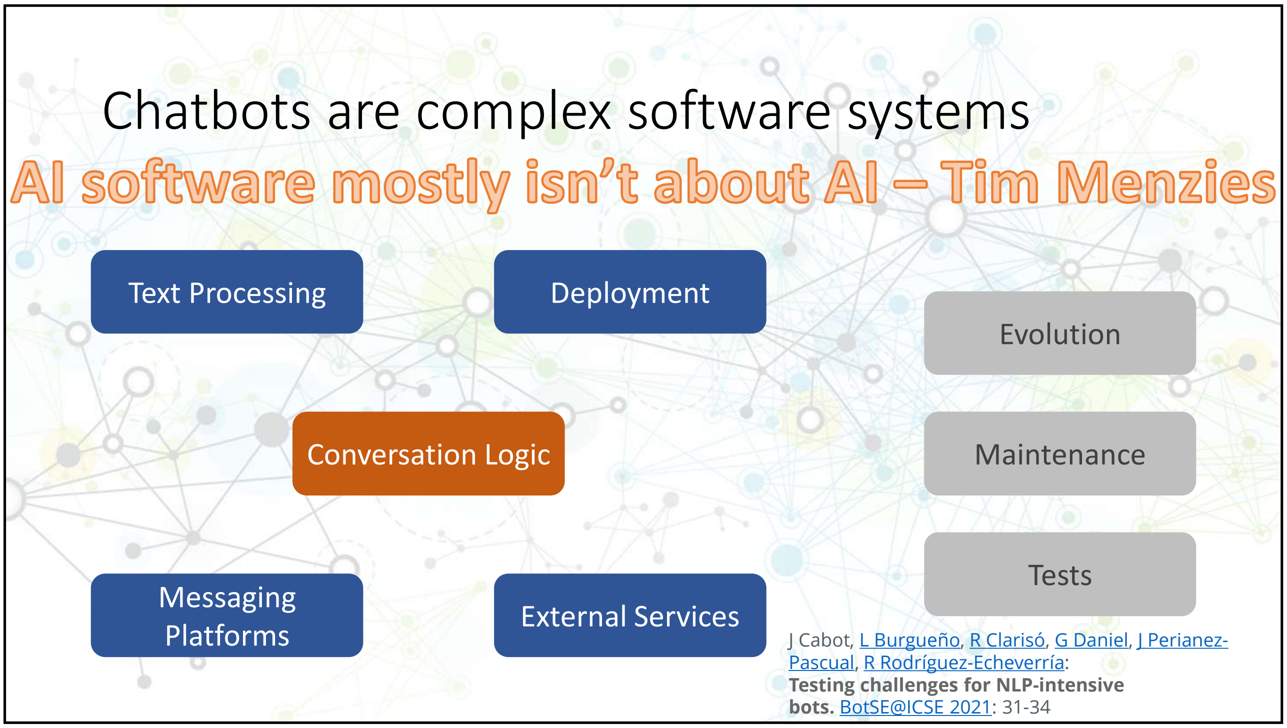
Input/Output Messaging channels

NLU Engine

Events!

External Platforms

30



Chatbots are complex software systems

AI software mostly isn't about AI – Tim Menzies

Text Processing

Deployment

Evolution

Conversation Logic

Maintenance

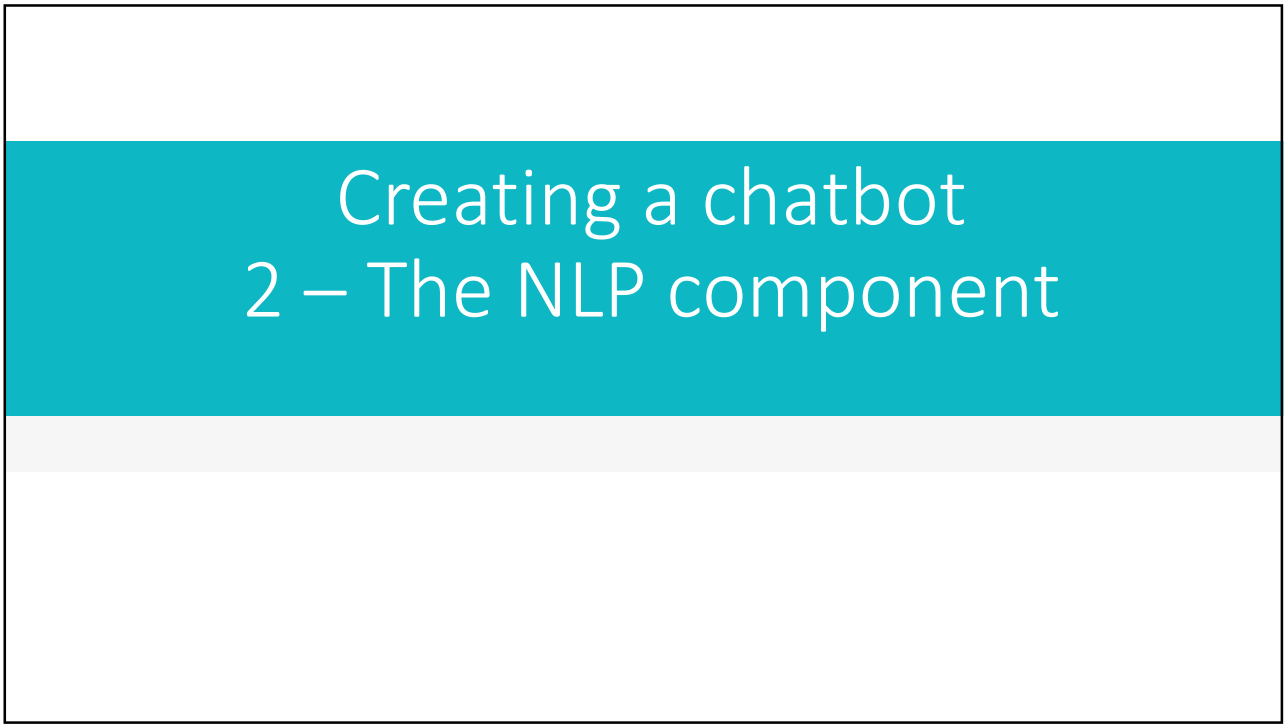
Tests

Messaging Platforms

External Services

J Cabot, L Burgueño, R Clarisó, G Daniel, J Perianez-Pascual, R Rodríguez-Echeverría: **Testing challenges for NLP-intensive bots.** BotSE@ICSE 2021: 31-34

31



Creating a chatbot
2 – The NLP component

32

How NLP works in chatbots

33

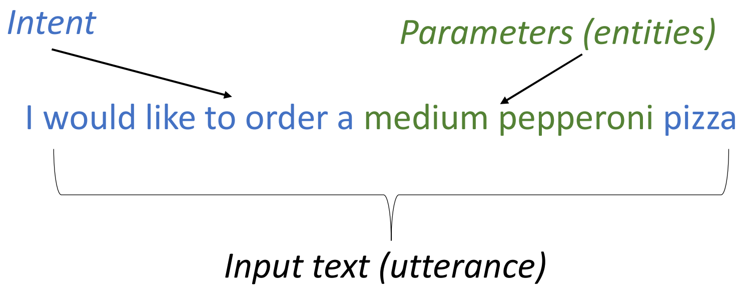
Bots and NLP/NLU

- At the core of all AI bots there is a NLU/NLP engine with two key missions
  - *Classify the user utterance* (i.e. input text) in one of the intents (set of possible questions/requests) the bot understands
  - *Identify the parameters* (called « entities ») from the utterance

34

NLP in a chatbot implies

- Understand what the user is trying to ask the bot
- Identify and gather the data in the request that the bot needs to construct the response

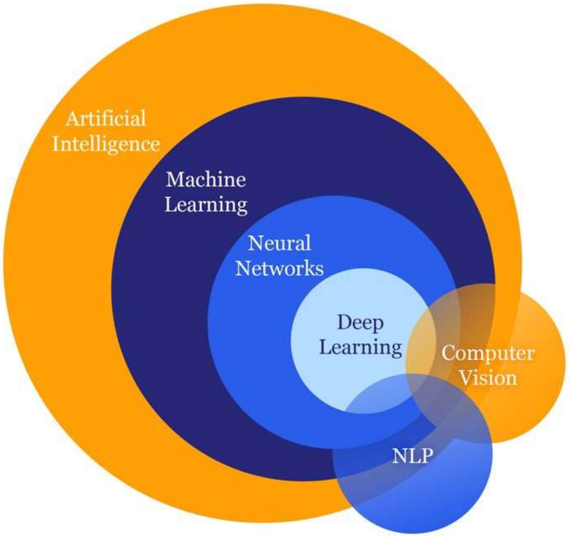


Intent

Parameters (entities)

I would like to order a medium pepperoni pizza

Input text (utterance)



slimmer AI

NLU approaches

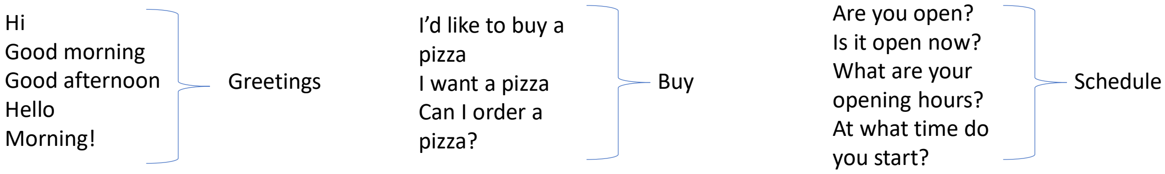
- Regular expressions
  - Require an almost exact match. Only useful for very concrete cases

- Grammar-based approaches
  - Useful for small domains with a very specific vocabulary and constructs
  - E.g. Stanford Parser

★

- Neural Networks - Multiclass classifiers
  1. You provide some *training sentences* for each intent (the "classes").
  2. The classifier learns from these sentences
  3. It assigns any new input text to one of the classes (with %confidence)

Example: bot that greets you, informs about the schedule and lets you buy a pizza



Hi
Good morning
Good afternoon
Hello
Morning!

Greetings

I'd like to buy a pizza
I want a pizza
Can I order a pizza?

Buy

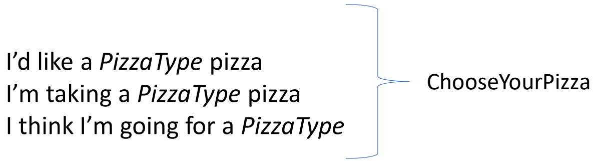
Are you open?
Is it open now?
What are your opening hours?
At what time do you start?

Schedule

If a user says, "I would like to order a pizza now" what is the closest intent?

Not that easy when intents start to overlap (e.g. offering a job vs looking for a job: "can I offer you a job?" vs "do you have any job offer")

Entity recognition

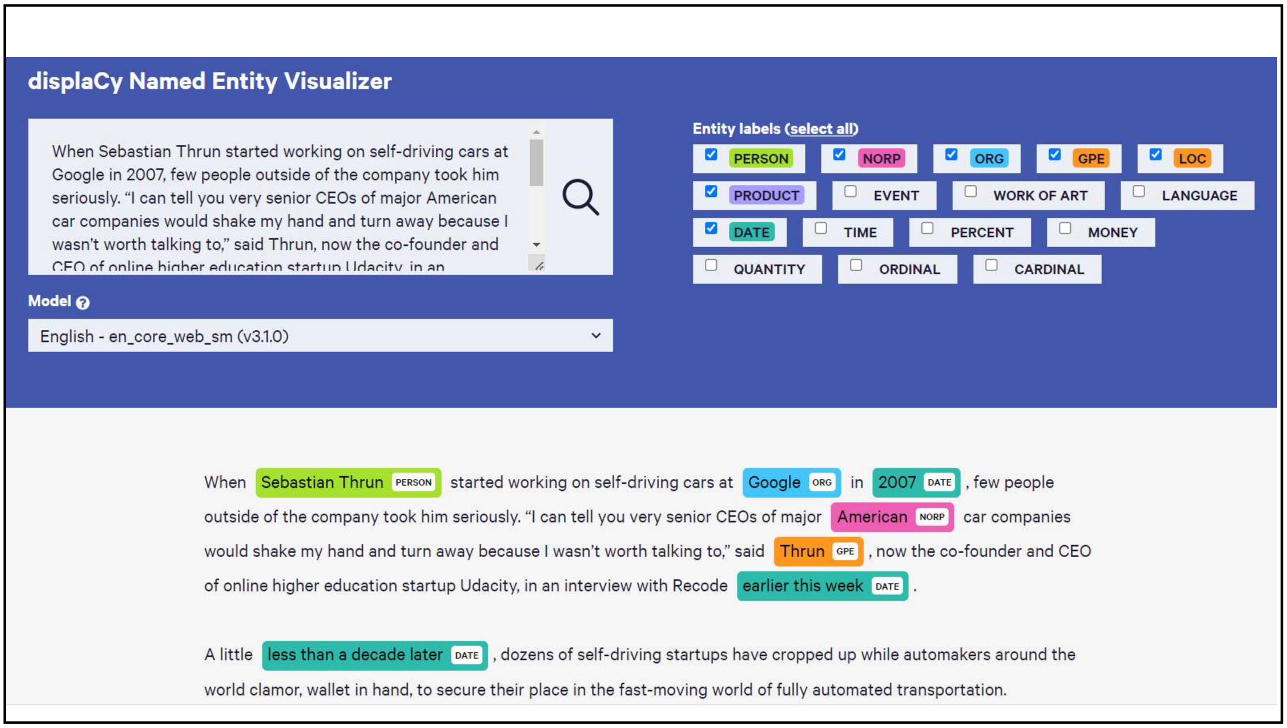


I'd like a *PizzaType* pizza
I'm taking a *PizzaType* pizza
I think I'm going for a *PizzaType*

ChooseYourPizza

- We need to understand the *intent* but also identify the *entities*
- Much easier if PizzaType is an enumeration ({ Pepperoni, Veggie, Hawaian ,…}) instead of the *any* string

39



displaCy Named Entity Visualizer

When Sebastian Thrun started working on self-driving cars at Google in 2007, few people outside of the company took him seriously. "I can tell you very senior CEOs of major American car companies would shake my hand and turn away because I wasn't worth talking to," said Thrun, now the co-founder and CEO of online higher education startup Udacity, in an

Entity labels (select all)

☑ PERSON ☑ NORP ☑ ORG ☑ GPE ☑ LOC
☑ PRODUCT □ EVENT □ WORK OF ART □ LANGUAGE
☑ DATE □ TIME □ PERCENT □ MONEY
□ QUANTITY □ ORDINAL □ CARDINAL

Model
English - en_core_web_sm (v3.1.0)

When Sebastian Thrun PERSON started working on self-driving cars at Google ORG in 2007 DATE , few people outside of the company took him seriously. "I can tell you very senior CEOs of major American NORP car companies would shake my hand and turn away because I wasn't worth talking to," said Thrun GPE , now the co-founder and CEO of online higher education startup Udacity, in an interview with Recode earlier this week DATE .

A little less than a decade later DATE , dozens of self-driving startups have cropped up while automakers around the world clamor, wallet in hand, to secure their place in the fast-moving world of fully automated transportation.

40

Common tricks in NLU Engines

- Stemmers identify the lemma of the word (e.g. to focus on the verb, not on the tense of the verb for matching purposes)
- Synonyms and paraphrasing are an automatic way to expand the training sentences
- Entity augmentation also generalize training sentences
    - E.g. What’s the weather in Barcelona -> What’s the weather in @city
    - Augmentation is possible for a large number of data types and common concepts (measures, countries, dates,…)

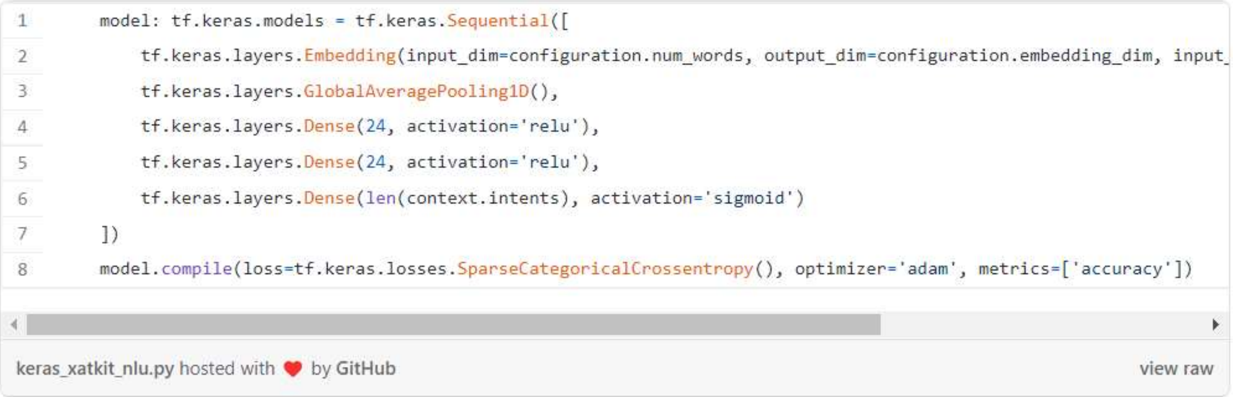
41

Taking a closer look at the NN

https://github.com/xatkit-bot-platform/xatkit-nlu-server

42

The ML model

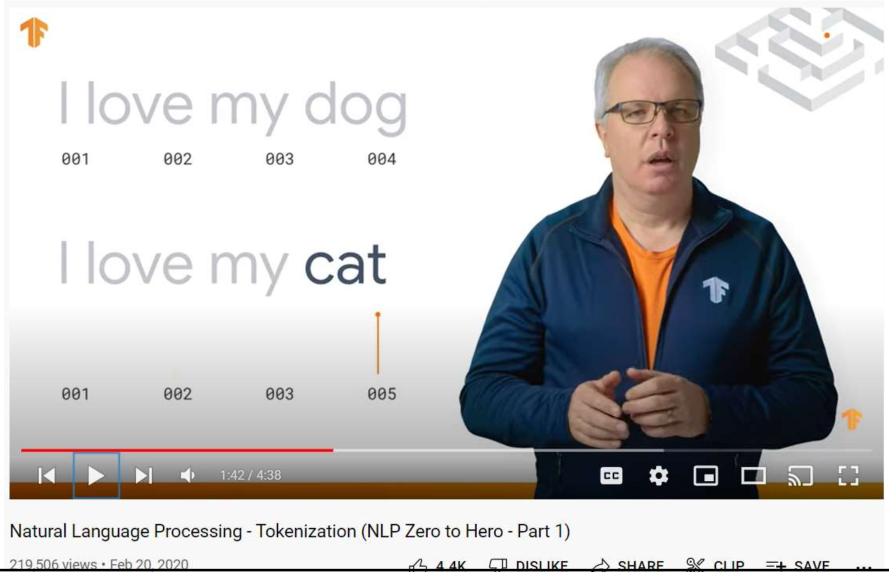


```
model: tf.keras.models = tf.keras.Sequential([
    tf.keras.layers.Embedding(input_dim=configuration.num_words, output_dim=configuration.embedding_dim, input_
    tf.keras.layers.GlobalAveragePooling1D(),
    tf.keras.layers.Dense(24, activation='relu'),
    tf.keras.layers.Dense(24, activation='relu'),
    tf.keras.layers.Dense(len(context.intents), activation='sigmoid')
])
model.compile(loss=tf.keras.losses.SparseCategoricalCrossentropy(), optimizer='adam', metrics=['accuracy'])
```

Even this "simple" arch easily gets over > 100.000 parameters!

43

Step 1 – From words to vectors of numbers

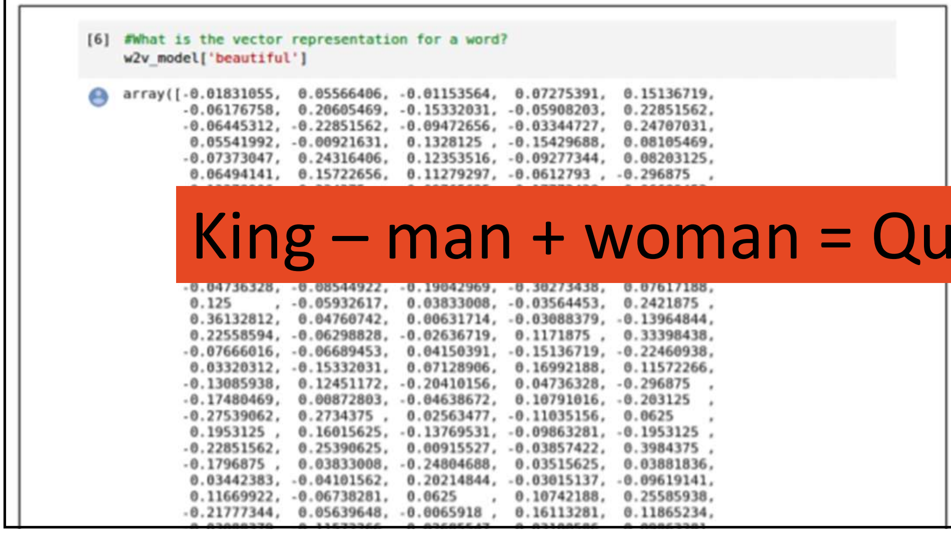


Natural Language Processing - Tokenization (NLP Zero to Hero - Part 1)

219,506 views • Feb 20, 2020

44

Step 2 Embedding -> similar words have similar vectors



```
[6] #What is the vector representation for a word?
    w2v_model['beautiful']
```

```
array([-0.01831055,  0.05566406, -0.01153564,  0.07275391,  0.15136719,
       -0.06176758,  0.20605469, -0.15332031, -0.05908203,  0.22851562,
       -0.06445312, -0.22851562, -0.09472656, -0.03344727,  0.24707031,
        0.05541992, -0.00921631,  0.1328125 , -0.15429688,  0.08105469,
       -0.07373047,  0.24316406,  0.12353516, -0.09277344,  0.08203125,
        0.06494141,  0.15722656,  0.11279297, -0.0612793 , -0.296875  ,
       -0.04736328, -0.08544922, -0.19042969, -0.30273438,  0.07617188,
        0.125     , -0.05932617,  0.03833008, -0.03564453,  0.2421875 ,
        0.36132812,  0.04760742,  0.00631714, -0.03088379, -0.13964844,
        0.22558594, -0.06298828, -0.02636719,  0.1171875 ,  0.33398438,
       -0.07666016, -0.06689453,  0.04150391, -0.15136719, -0.22460938,
        0.03320312, -0.15332031,  0.07128906,  0.16992188,  0.11572266,
       -0.13085938,  0.12451172, -0.20410156,  0.04736328, -0.296875  ,
       -0.17480469,  0.00872803, -0.04638672,  0.10791016, -0.203125  ,
       -0.27539062,  0.2734375 ,  0.02563477, -0.11035156,  0.0625    ,
        0.1953125 ,  0.16015625, -0.13769531, -0.09863281, -0.1953125 ,
       -0.22851562,  0.25390625,  0.00915527, -0.03857422,  0.3984375 ,
       -0.1796875 ,  0.03833008, -0.24804688,  0.03515625,  0.03881836,
        0.03442383, -0.04101562,  0.20214844, -0.03015137, -0.09619141,
        0.11669922, -0.06738281,  0.0625    ,  0.10742188,  0.25585938,
       -0.21777344,  0.05639648, -0.0065918 ,  0.16113281,  0.11865234,
```

King – man + woman = Queen

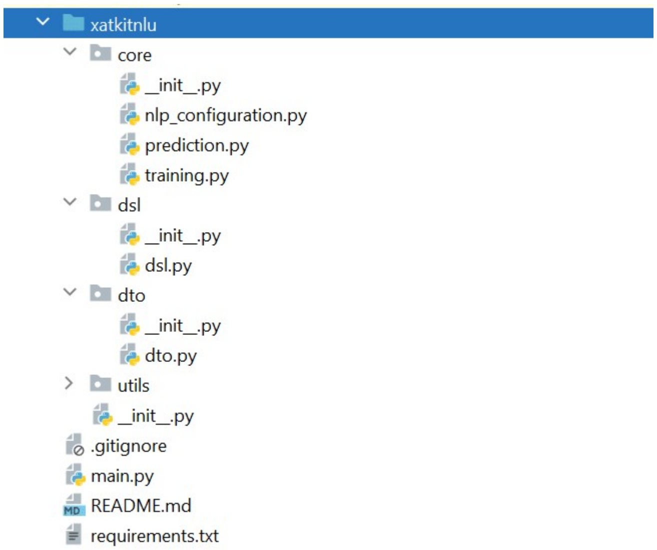
Training and prediction

```
history = model.fit(np.array(context.training_sequences), np.array(context.training_labels),
                                                            epochs=configuration.num_epochs, verbose=2)
```

```
full_prediction = context.nlp_model.predict(sequence_from_sentence_to_predict)
```

Tip: Predicting always via the NN is overengineering sometimes (e.g. Yes/No answers)

Remember: the SW part of any ML project



```
xatkitnlu
    core
        __init__.py
        nlp_configuration.py
        prediction.py
        training.py
    dsl
        __init__.py
        dsl.py
    dto
        __init__.py
        dto.py
    utils
    __init__.py
.gitignore
main.py
README.md
requirements.txt
```

47

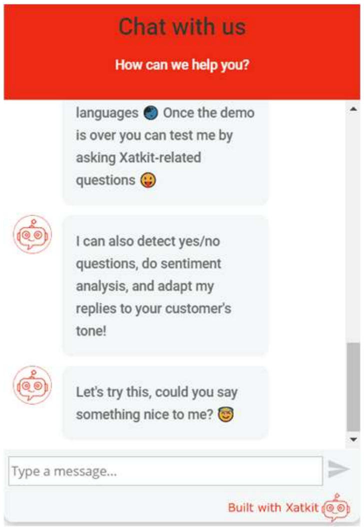
And we could go on and on ...

48

Beyond intent classification

- Great bots go beyond providing the right answer, e.g. they could show some empathy

- Several language models (for toxicity detection, for sentiment analysis, language identification, emoji support…) can be easily integrated in bots



Chat with us

How can we help you?

languages 🌏 Once the demo is over you can test me by asking Xatkit-related questions 😛

I can also detect yes/no questions, do sentiment analysis, and adapt my replies to your customer's tone!

Let's try this, could you say something nice to me? 😇

PRO tip: Monitor your bots

- You will fail to predict:
  - What users want to know
  - How they express their intentions

- Careful with self-learning
  - Filter out what the bot needs to add as training sentences to avoid strange derivations

Creating a chatbot
3 – The Software component

51

What bot building tool should I use?

52

>100 bot platforms

- Rasa, BotPress, Chatfuel, Inbenta, Botsify, Flow XO, Xatkit…
- And plenty NLP engines: DialogFlow, Lex, LUIS, Watson, NLP.js

53

It all depends

- What are your priorities
  - Open source?
  - Free?
  - Programming language of the bot?
  - Natural languages (e.g. Catalan)
  - Availability of connectors
  - On-premises option?
  - Plain textual bots or fully functional ones?
  - Low-code/ no-code or tech skills?
  - Possibility to migrate from one NLP engine to another?
  - Monitoring dashboard / reports ?
  - …

54

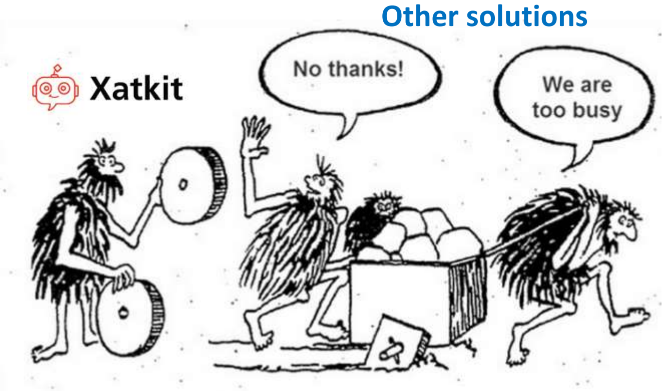
Building bots in Xatkit

https://github.com/xatkit-bot-platform

55

Xatkit's philosophy: Do not reinvent the wheel

Other solutions

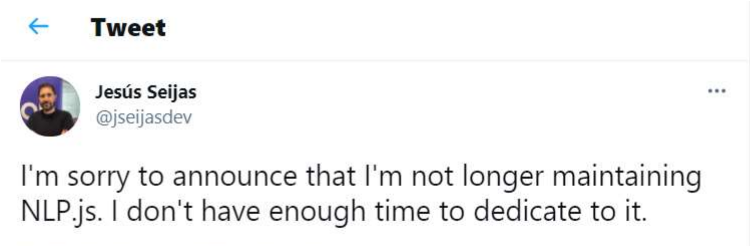


Plenty of great NLU engines and libraries, let's not waste our time building a new one but take the best of breed and focus on features to help you create better bots

Xatkit is a chatbot orchestration platform

56

Another reason to avoid strong coupling to specific NLU/NLP libraries

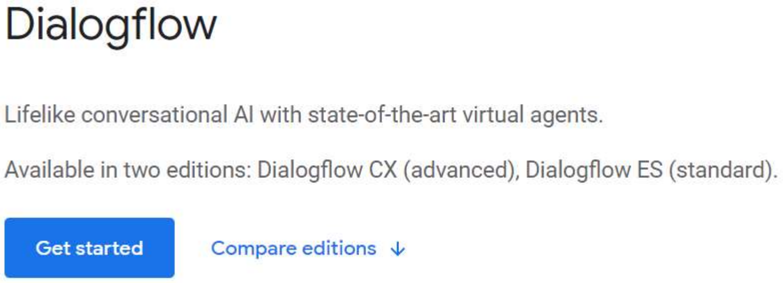


← **Tweet**

Jesús Seijas
@jseijasdev

I'm sorry to announce that I'm not longer maintaining NLP.js. I don't have enough time to dedicate to it.

10:36 AM · May 18, 2021 · Twitter Web Ap



Dialogflow

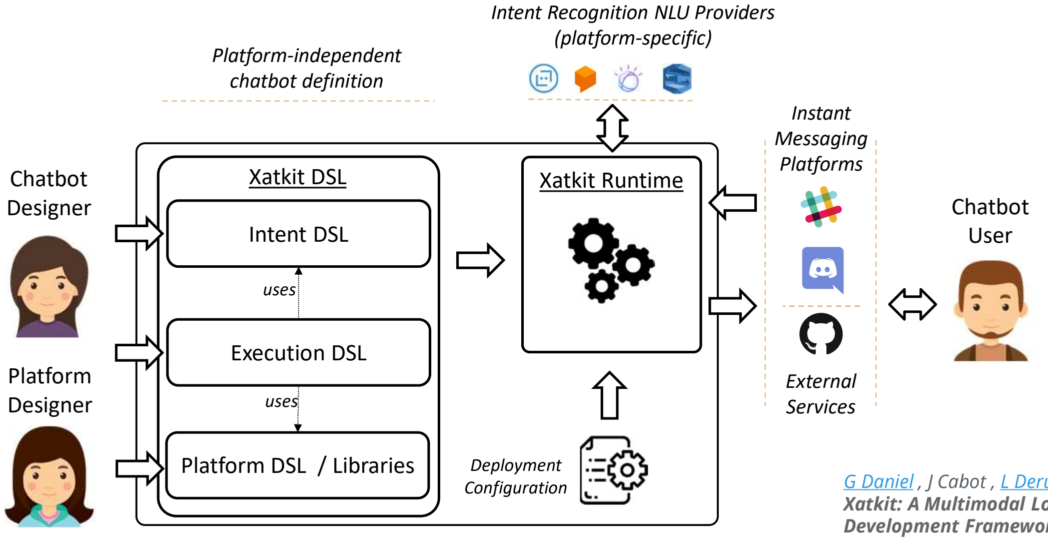
Lifelike conversational AI with state-of-the-art virtual agents.

Available in two editions: Dialogflow CX (advanced), Dialogflow ES (standard).

57

Xatkit architecture

(write your bot once, run it anywhere)



G Daniel , J Cabot , L Deruelle, M Derras : ***Xatkit: A Multimodal Low-Code Chatbot Development Framework.*** *IEEE Access 8: 15332-15346 (2020)*

58

Xatkit DSL

- Bots are created with the Xatkit DSL offering a chatbot-specific syntax for creating:
  - Intents the bot needs to match
  - The behavior to execute in response to the matched intents
- Our DSL aims to combine the benefits of:
  - Having a chatbot-specific DSL that facilitates the creation of chatbot applications by providing higher-level chatbot modeling primitives
  - The full power of a general programming language to create any type of bot, not just simple textual ones

59

Xatkit DSL

- Should we go for an external DSL or an internal one?
  - External one: independent language with its own compiler, editor,… You control everything
  - Internal ones: lives inside another language. You can reuse the host language but are also constrained by it
- We started with an external one and had to reimplement everything

In our scenario, Internal DSLs are the way to go (Martin Fowler)

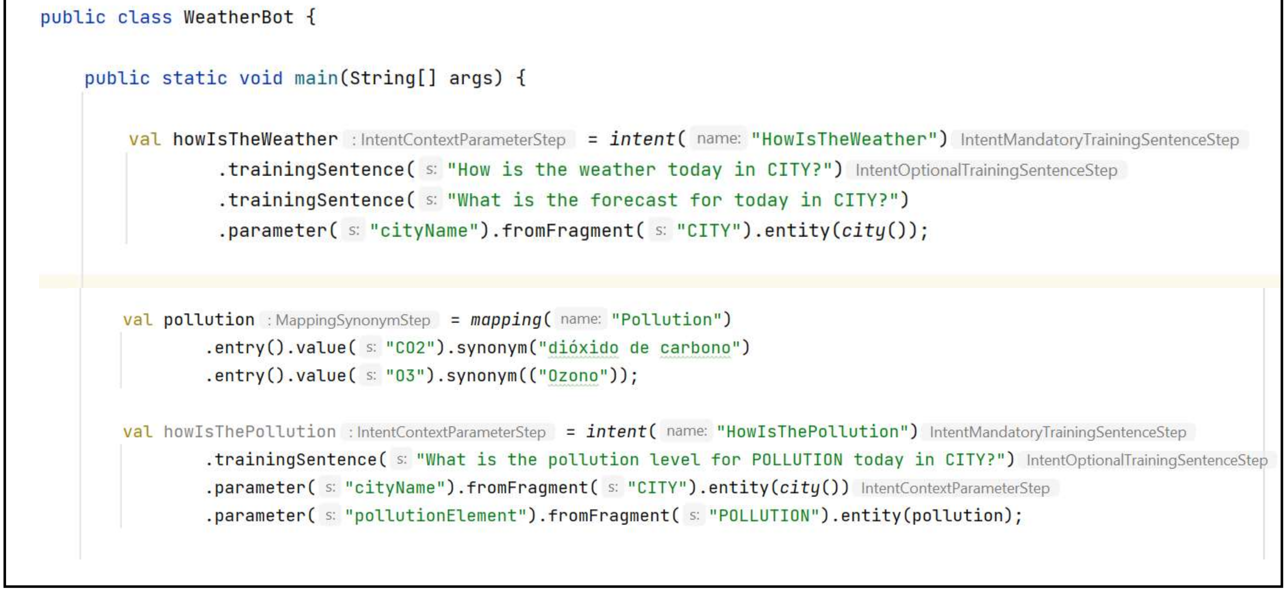
60

Our DSL is implemented as a Java Fluent API

- Create bots using your preferred Java editor
  - Benefit from all existing Java tooling when developing bots (e.g. debuggers)
  - Reuse any Java library for complex bot behaviors
  - Intuitive Fluent Interfaces to help you create advanced conversations
  - Based on state machine semantics to build any type of bot

61

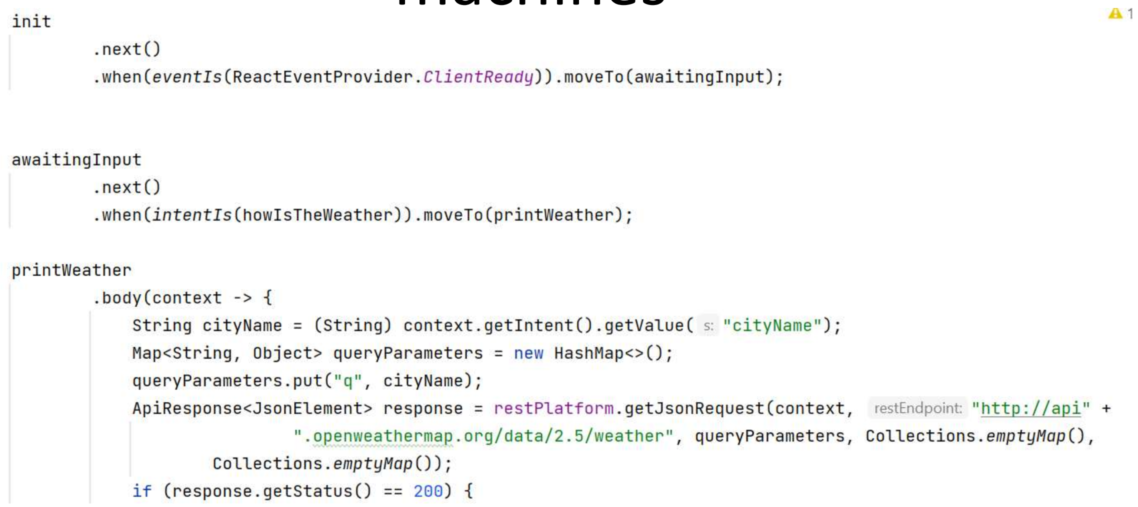
Intent DSL example



```
public class WeatherBot {

    public static void main(String[] args) {

        val howIsTheWeather : IntentContextParameterStep  = intent( name: "HowIsTheWeather") IntentMandatoryTrainingSentenceStep
                .trainingSentence( s: "How is the weather today in CITY?") IntentOptionalTrainingSentenceStep
                .trainingSentence( s: "What is the forecast for today in CITY?")
                .parameter( s: "cityName").fromFragment( s: "CITY").entity(city());


        val pollution : MappingSynonymStep  = mapping( name: "Pollution")
                .entry().value( s: "CO2").synonym("dióxido de carbono")
                .entry().value( s: "O3").synonym(("Ozono"));

        val howIsThePollution : IntentContextParameterStep  = intent( name: "HowIsThePollution") IntentMandatoryTrainingSentenceStep
                .trainingSentence( s: "What is the pollution level for POLLUTION today in CITY?") IntentOptionalTrainingSentenceStep
                .parameter( s: "cityName").fromFragment( s: "CITY").entity(city()) IntentContextParameterStep
                .parameter( s: "pollutionElement").fromFragment( s: "POLLUTION").entity(pollution);
```

62

Execution DSL example – State machines



```
init
        .next()
        .when(eventIs(ReactEventProvider.ClientReady)).moveTo(awaitingInput);


awaitingInput
        .next()
        .when(intentIs(howIsTheWeather)).moveTo(printWeather);

printWeather
        .body(context -> {
            String cityName = (String) context.getIntent().getValue( s: "cityName");
            Map<String, Object> queryParameters = new HashMap<>();
            queryParameters.put("q", cityName);
            ApiResponse<JsonElement> response = restPlatform.getJsonRequest(context,  restEndpoint: "http://api" +
                            ".openweathermap.org/data/2.5/weather", queryParameters, Collections.emptyMap(),
                    Collections.emptyMap());
            if (response.getStatus() == 200) {
```

63

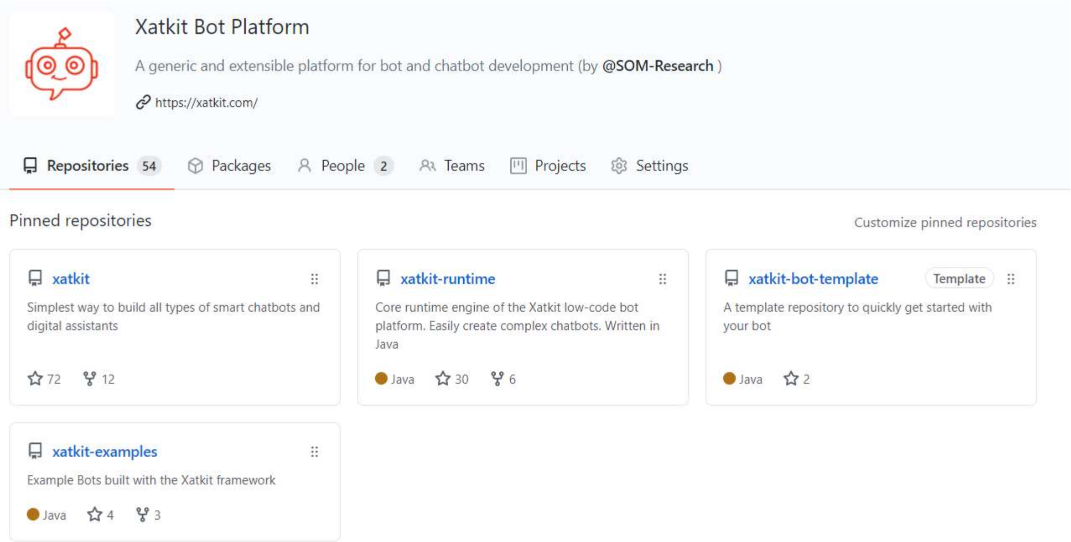
Easy deployments

A Xatkit bot can be packaged as a single jar with no external dependencies. Just execute the jar to get the bot up and running

Via a simple configuration file you can connect your bot with external services, e.g. NLP Engines and communication channels

64

Xatkit is available on GitHub



Xatkit Bot Platform

A generic and extensible platform for bot and chatbot development (by @SOM-Research)

https://xatkit.com/

Pinned repositories

xatkit
Simplest way to build all types of smart chatbots and digital assistants
☆ 72 · 12

xatkit-runtime
Core runtime engine of the Xatkit low-code bot platform. Easily create complex chatbots. Written in Java
● Java · ☆ 30 · 6

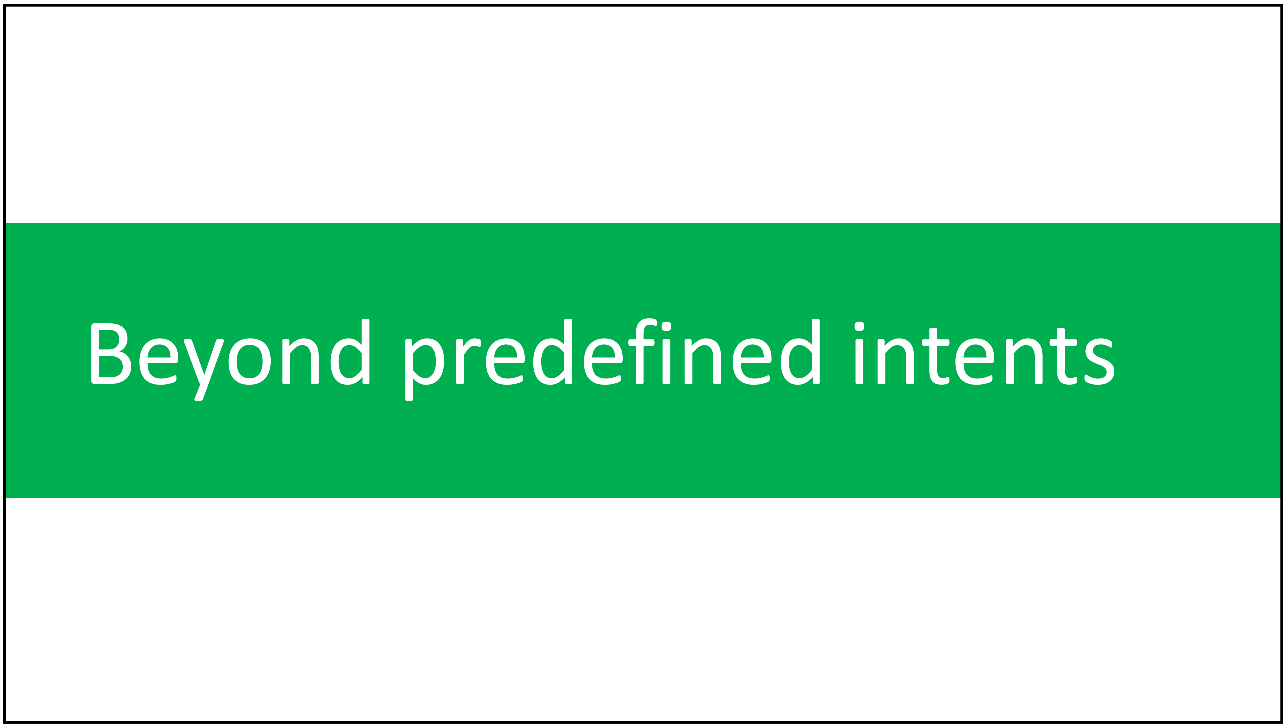
xatkit-bot-template Template
A template repository to quickly get started with your bot
● Java · ☆ 2

xatkit-examples
Example Bots built with the Xatkit framework
● Java · ☆ 4 · 3

65

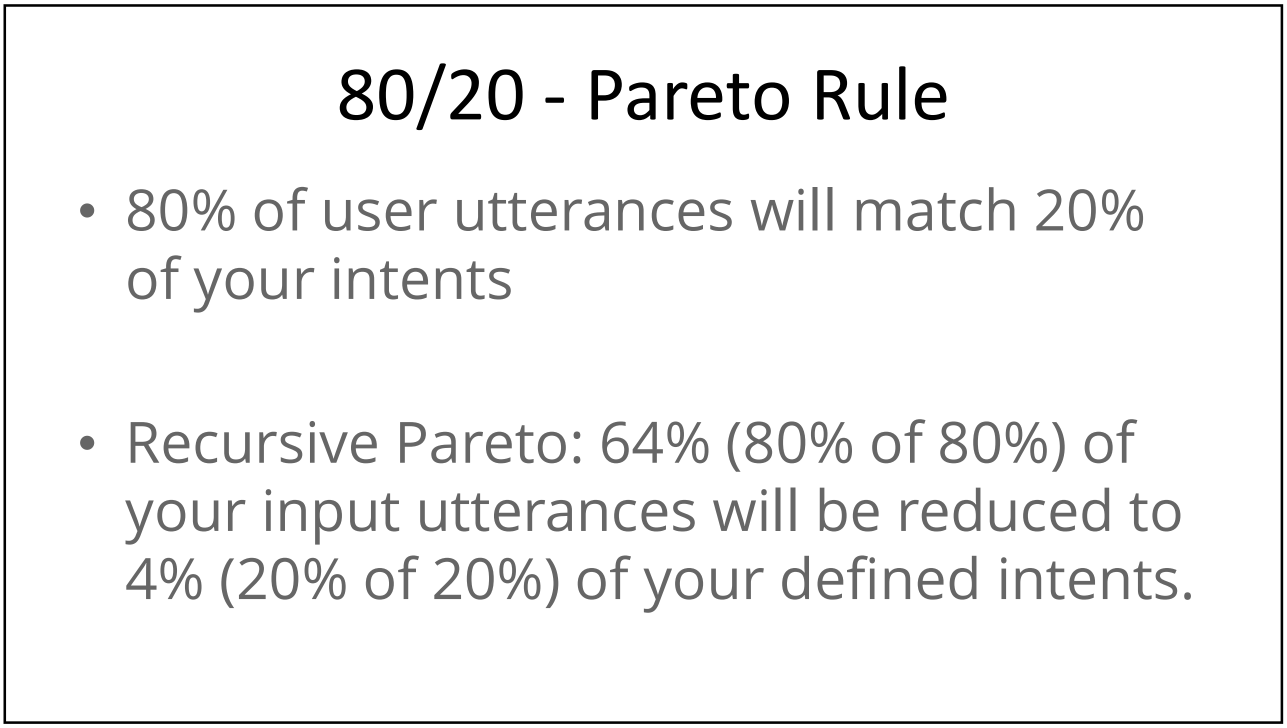
Advanced topics

66



Beyond predefined intents

67

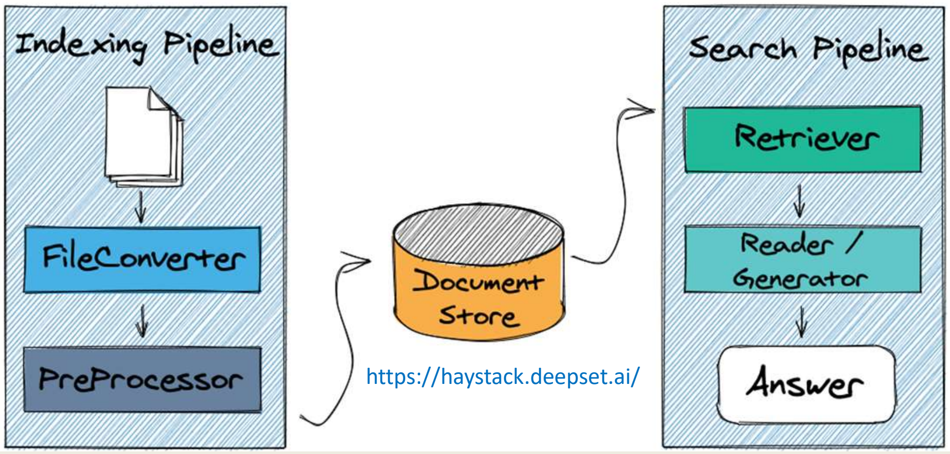


80/20 - Pareto Rule

- 80% of user utterances will match 20% of your intents
- Recursive Pareto: 64% (80% of 80%) of your input utterances will be reduced to 4% (20% of 20%) of your defined intents.

68

Open questions -> Plan B in your default fallback



https://haystack.deepset.ai/

BODI (Bots for Open Data) we use text to SQL models to try to answer a larger variety of questions

69

No-learn / No-work chatbot building

70

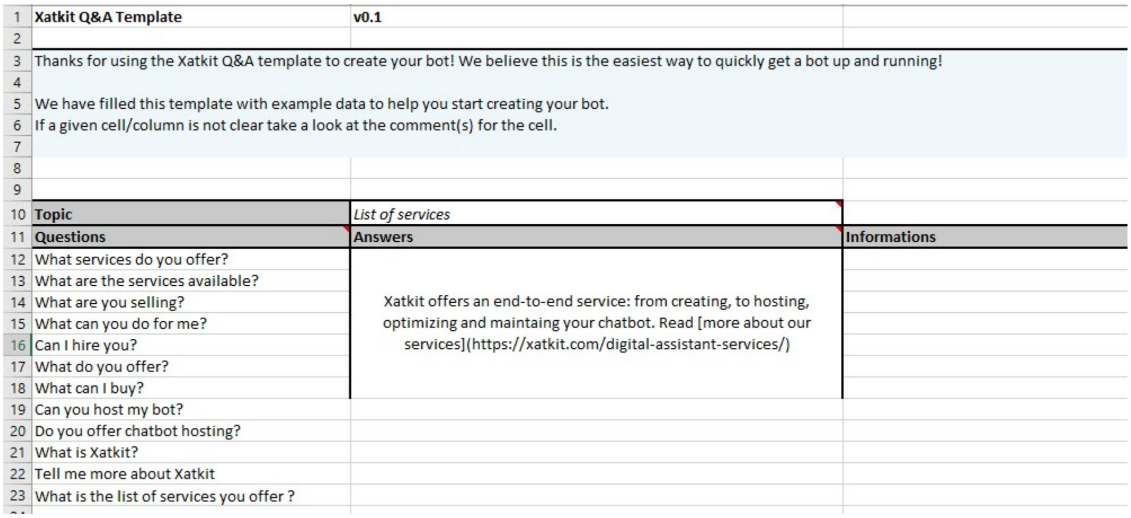
Low-code / no-code

- Many bot platforms try to offer low-code / no-code interfaces to define bots (e.g. GUIs) writing *less* code

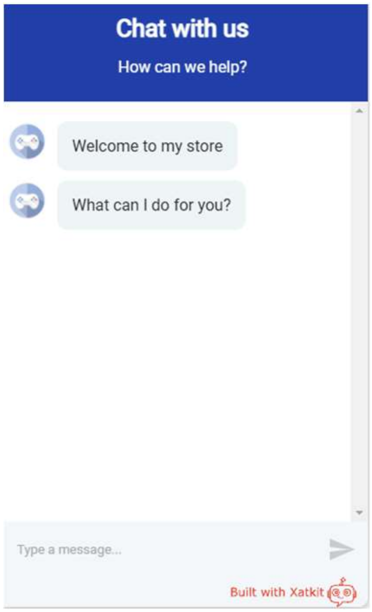
- But this is not enough for many non-tech users

Keep in mind: 1 language -> ++notations

No-Learn (e.g. bots from Excel)



1	**Xatkit Q&A Template**	**v0.1**	
2			
3	Thanks for using the Xatkit Q&A template to create your bot! We believe this is the easiest way to quickly get a bot up and running!		
4			
5	We have filled this template with example data to help you start creating your bot.		
6	If a given cell/column is not clear take a look at the comment(s) for the cell.		
7			
8			
9			
10	**Topic**	*List of services*	
11	**Questions**	**Answers**	**Informations**
12	What services do you offer?	Xatkit offers an end-to-end service: from creating, to hosting, optimizing and maintaining your chatbot. Read [more about our services](https://xatkit.com/digital-assistant-services/)	
13	What are the services available?		
14	What are you selling?		
15	What can you do for me?		
16	Can I hire you?		
17	What do you offer?		
18	What can I buy?		
19	Can you host my bot?		
20	Do you offer chatbot hosting?		
21	What is Xatkit?		
22	Tell me more about Xatkit		
23	What is the list of services you offer ?		



Chat with us

How can we help?

Welcome to my store

What can I do for you?

Type a message...

No-Work

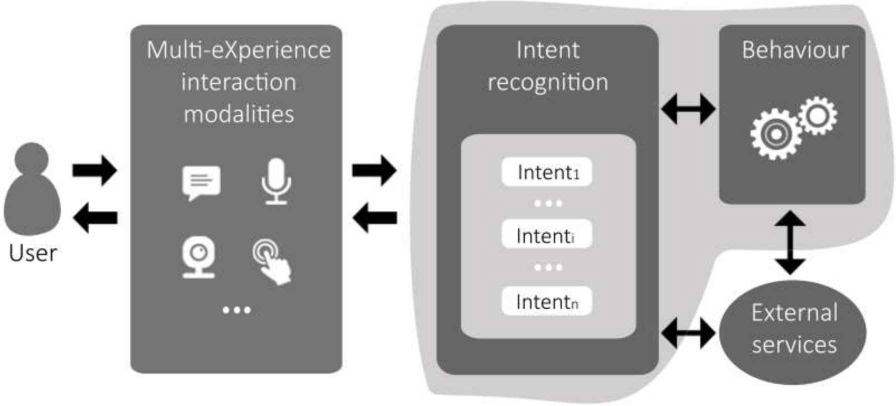
- Bots derived from your db data
  - E.g. an eCommerce bot that answers questions about your products, shipping options,…

73

Multiexperience Development Platforms (MXDP)

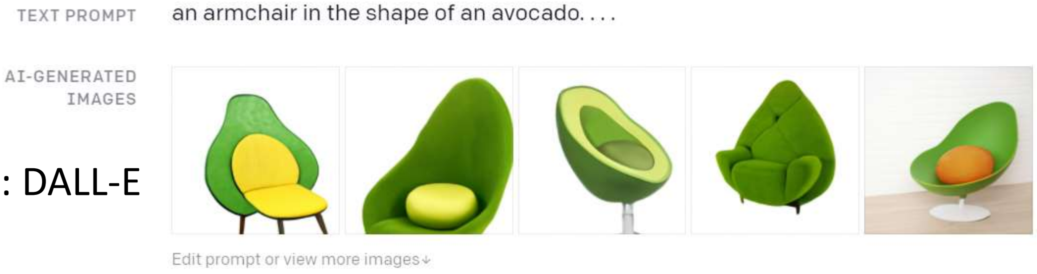
74

One app -> Multiple interfaces



75

Related also to Multimodal AI



TEXT PROMPT an armchair in the shape of an avocado. . . .

AI-GENERATED IMAGES

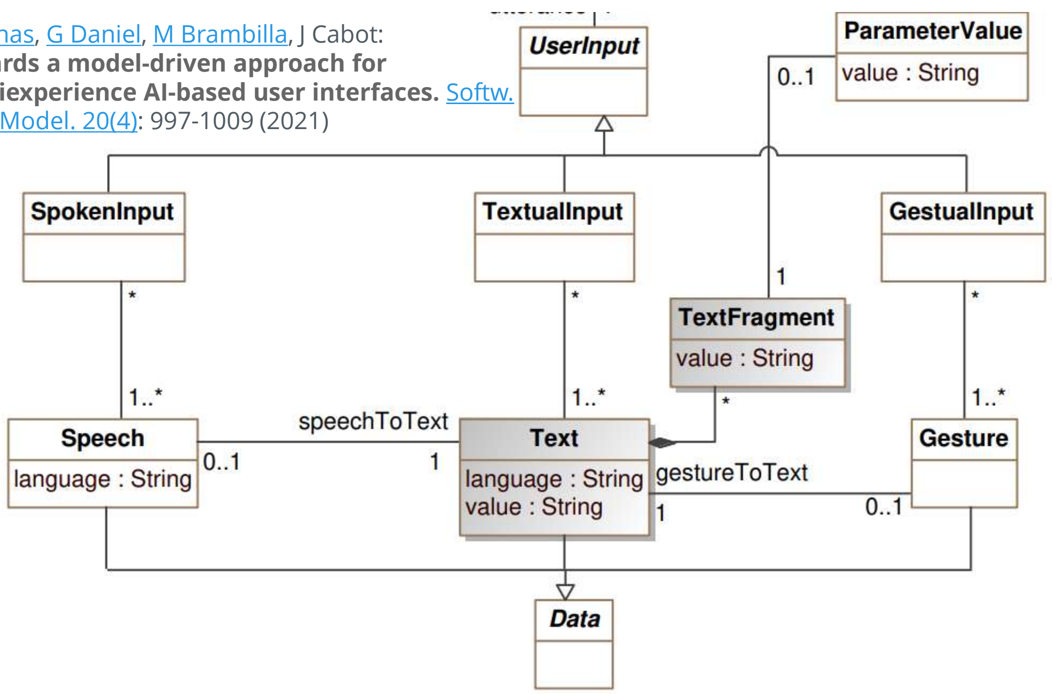
Edit prompt or view more images↓

E.g.: DALL-E

MMF

76

E Planas, G Daniel, M Brambilla, J Cabot: **Towards a model-driven approach for multiexperience AI-based user interfaces.** Softw. Syst. Model. 20(4): 997-1009 (2021)



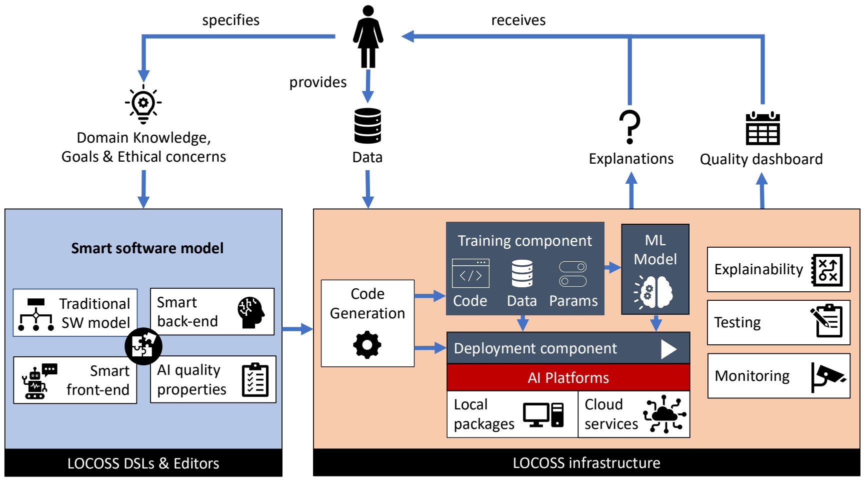
But this brings plenty of new challenges

- A chatbot that knows where you're looking at (e.g. the product in the eCommerce shop) and takes in that info
- How do you test that the different UIs behave in a consistent way?

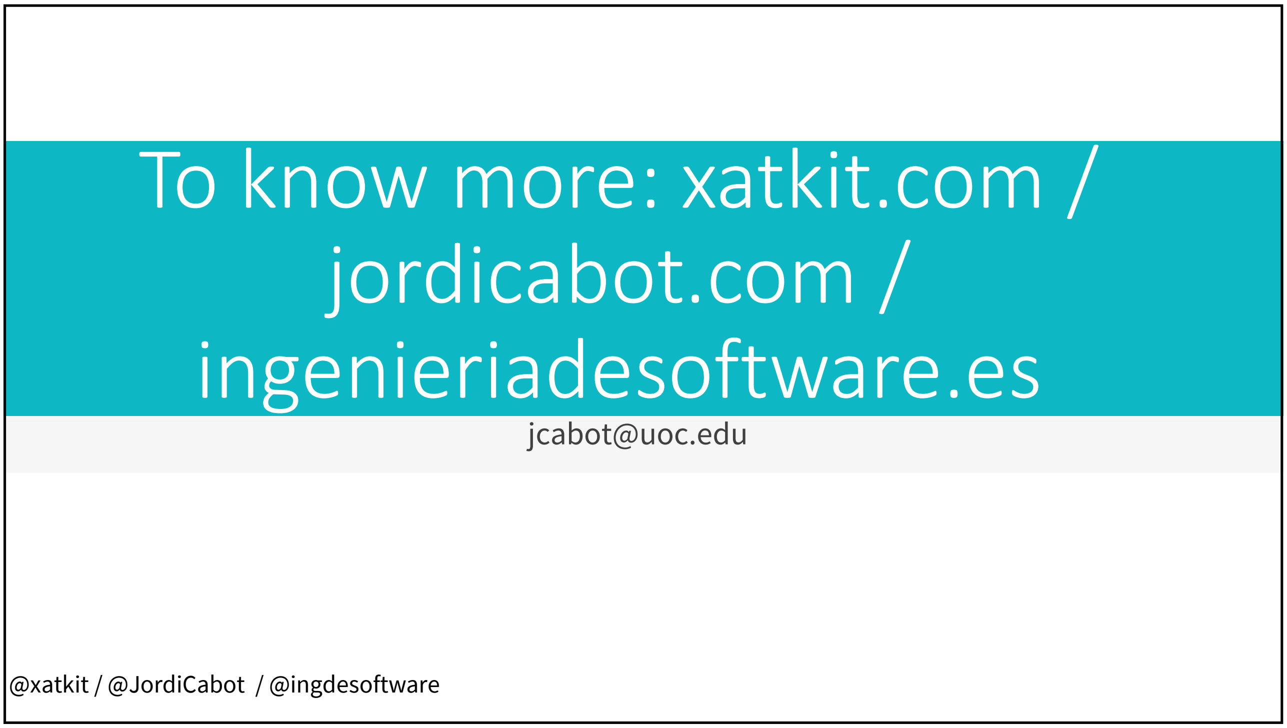
Exciting times ahead

79

The LOCOSS project (Low-code development of smart software)



80



To know more: xatkit.com / jordicabot.com / ingenieriadesoftware.es

jcabot@uoc.edu

@xatkit / @JordiCabot / @ingdesoftware

81