

Conversational Systems Development: An Engineering Perspective

Zoraida Callejas, Universidad de Granada



RTTH Fall School, Nov 23

Speech technologies for daily
life: voice assistants, chatbots
and spoken dialogue systems

Universidad de Granada

Depto. Lenguajes y Sistemas Informáticos

Grupo de investigación en Sistemas de Diálogo
Hablado y Multimodal (SISDIAL)



The engineering lifecycle of conversational systems

- Requirement gathering and analysis
- System design and prototyping
- Iterative development
- Deployment
- Test and quality assurance

1. REQUIREMENTS

Gathering requirements for conversational systems

1. Speaking the same language

What is a dialogue system / chatbot / virtual assistant / conversational robot?

What can it do and cannot do?

It is important to manage the client's expectations, setting realistic boundaries for the system's capabilities.

Explain the current capabilities and limitations of the technology.

Gathering requirements for conversational systems

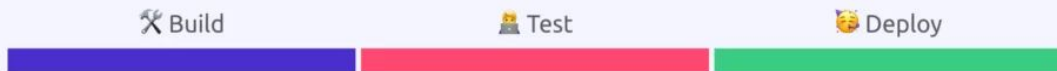
1. Speaking the same language

Familiarize clients with the development process

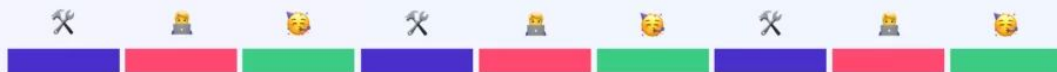
Conversational systems development is an iterative process.

It is more effective to start with a basic version and gradually enhance it based on user feedback.

How people think launching a bot works



How it actually works



Gathering requirements for conversational systems

2.1. Identifying customer needs: why?

Why do they want/need a chatbot?

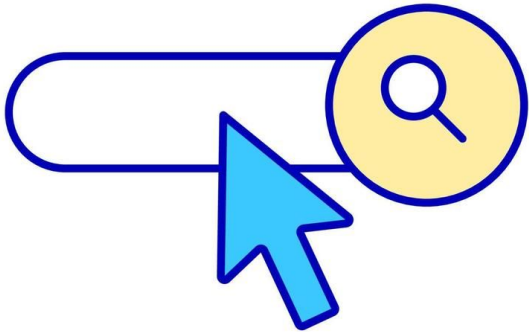
What will the chatbot do? which processes and activities must be adjusted to the chatbot?

Who are the users? Are they familiarized with this type of interface?

Gathering requirements for conversational systems

2.1. Identifying customer needs:

Is a chatbot even the best choice??



The case of the chatbot for employees in administration:

A chatbot to ease looking up for information in massive procedure manuals and written instructions.



Gathering requirements for conversational systems

2.2. Identifying customer needs: **scope**

Get the scope of the chatbot:

How do end-users communicate with the client now?

What do they ask for? What do they get?

| | | |
|-----------------------|---|---|
| |  |  |
| FAQ and documentation | Structured knowledge with precise questions and answers. | Sometimes they use the wording of the company and not of the users. |

Gathering requirements for conversational systems

2.2. Ideation

Get

FAQ and
documente

| UrbanYou FAQ | |
|---|---|
| How does UrbanYou Work? | + |
| Who are your Service Pros? | + |
| What is the UrbanYou Satisfaction Guarantee? | + |
| What does your Bond Back Guarantee cover? | + |
| Do Service Pros bring their own supplies? | + |
| Does UrbanYou offer office cleaning services? | + |
| I love UrbanYou! How can I tell my friends? | + |





Gathering requirements for conversational systems

2.2. Identifying customer needs: **scope**

Get the scope of the chatbot:

How do end-users communicate with the client now?

What do they ask for? What do they get?

| |  |  |
|------------------------|---|---|
| FAQ and documentation | Structured knowledge with precise questions and answers. | Sometimes they use the wording of the company and not of the users. |
| CAU / customer support | Human team that actually answers the users' questions. | Semi-structured knowledge / need to adequately explain chatbot capabilities. |



Real use case:

Example: FAQ chatbot
Chatbot for the Spanish Health Ministry - FAQ for the process of assignment of positions for medical internships

Chatbot MIR (April 2023)

The screenshot shows the official website of the Spanish Health Ministry (Ministerio de Sanidad). The header is yellow and contains the ministry's logo, the text 'MINISTERIO DE SANIDAD', and the 'U23' logo. Navigation links include 'Ministerio', 'Áreas', 'Prensa y comunicación', 'Sanidad en datos', 'Servicios a la Ciudadanía', and 'Participación Pública'. A search bar is located in the top right corner. Below the header, there is a navigation menu with 'Webs temáticas' and 'Sede Electrónica'. The main content area shows a breadcrumb trail: 'Traducir' > 'Está usted en: Inicio' > 'Prensa y comunicación' > 'Notas de Prensa'. A sidebar on the left lists 'Noticias', 'Ruedas de Prensa', 'Campañas informativas', and 'Agenda'. The main article is titled 'Notas de Prensa' and 'Un total de 21.682 aspirantes opta desde el lunes a la Formación Sanitaria Especializada 2022/2023'. The article text includes:

- El Ministerio de Sanidad ha adelantado una semana la elección de plazas y su adjudicación, volviendo así al calendario habitual de incorporación previo a la pandemia provocada por la COVID-19.
- La aplicación SIMULE es una herramienta informática que permite a los aspirantes disponer de información de las probabilidades de obtener una determinada plaza de acuerdo con las prioridades que han formulado los aspirantes que le preceden en número de orden.
- La convocatoria 2022/2023, fijada por la Comisión de Recursos Humanos del SNS, está conformada por un total de 11.171 plazas, en Medicina, Farmacia, Enfermería y los ámbitos de la Psicología, la Química, la Biología y la Física.
- Los actos de adjudicación de plaza se realizarán una vez finalizado el plazo de elección de plaza a la que los aspirantes han sido convocados y las personas a las que se adjudique plaza tomarán posesión de la misma en el centro o unidad docente correspondiente entre los días 22 y 23 de mayo.

A partir de la fecha de elección de plaza, los aspirantes contarán con dos vídeos explicativos, alojados en el [Canal de Youtube del Ministerio de Sanidad](#), desde el próximo lunes 10 de abril a las 8:00 horas; uno sobre la plataforma y otro sobre SIMULE. Asimismo, contarán con preguntas frecuentes y un chatbot.

El servicio de atención al aspirante estará disponible de lunes a domingo de 8:00 a 22:00 horas, salvo el día 1 de mayo que no hay elección de plaza.

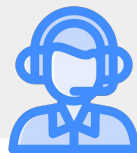


Real use case:

Chatbot MIR (April 2023)

Eliciting scope requirements

| Category | Intent | Examples of activation phrases | Comments |
|----------|--------|--------------------------------|----------|
| | | | |
| | | | |



*Iterative in collaboration
with CAU*



Real use case:

Chatbot MIR (April 2023)

72 intents

Eliciting scope requirements

¿Como puedo confirmar que mi solicitud esta tramitada?
Como compruebo que mi solicitud se ha tramitado correctamente?
como se que hice bien la solicitud de plaza? → 60 training examples



spelling, typos

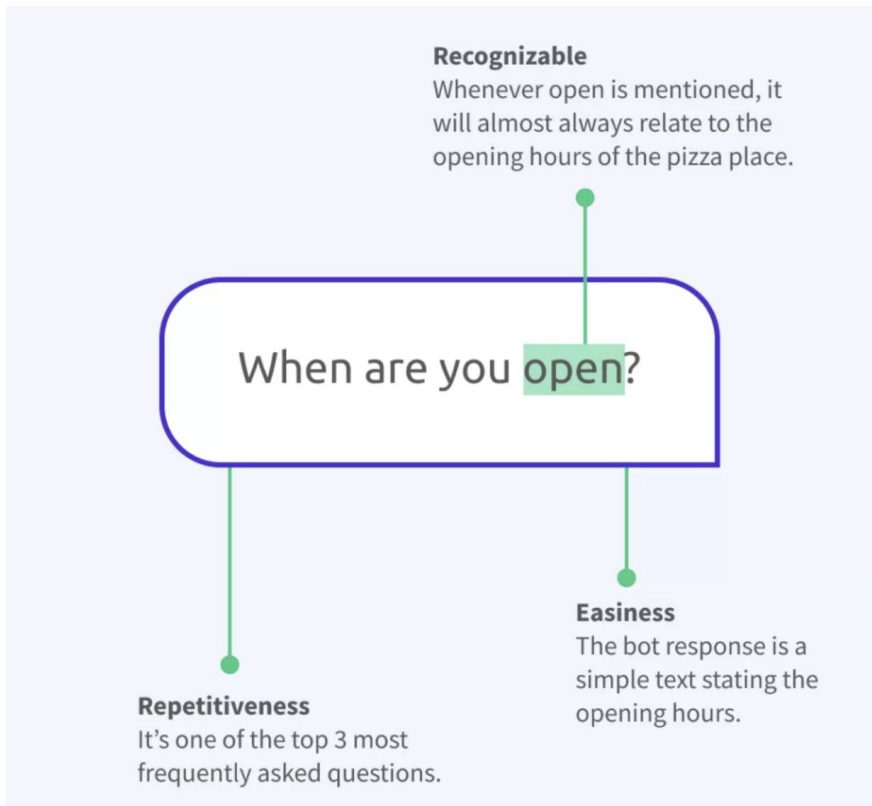
Para verificar que tu solicitud está correctamente tramitada, acude al menú "Solicitud de Plaza". Si figura como TRAMITADA significa que se ha gestionado correctamente. Además, en el momento en el que se ha realizado tu solicitud, se envía un correo con un PDF adjunto con la firma de Ministerio como justificante de tu tramitación, también puedes descargar este justificante desde la opción de Solicitud de plaza, en la parte inferior te aparecerá un botón que permite "Descargar el justificante".



legal-legal-legal

Gathering requirements for conversational systems

scope



👉 First pilot should start with highly recognizable, repetitive and easy intents

Image from:
<https://www.thecampfire.ai/post/how-to-scope-your-next-chatbot-project>

Gathering requirements for conversational systems

MIR chatbot for the Spanish Health Ministry

scope

Draft

Intent catalog with basic training phrases provided by the client, guided by the development team

Prototype

Intent catalog revised by development team: modification of intents, inclusion of training phrases.

OK by client.

Development.

Test

Lab test.
Test with client.

Deployment

Deployment of the 1st version with low threshold to derive to CAU.

Recording of user-system interactions.

Receiving feedback from CAU.



Real use case:

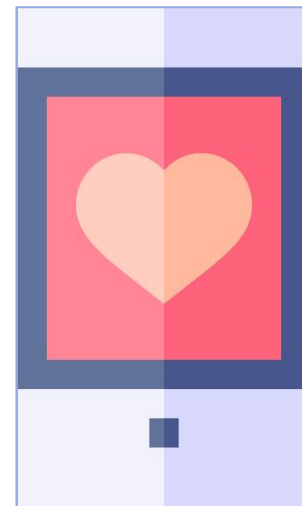
MENHIR project (2019-2024)

Conversational systems to promote mental health

- Improving access to mental healthcare
- Engaging and empowering users
- Increasing mental health literacy
- Providing additional sources of information to detect symptoms and relapse

Conversational systems are based on natural spoken communication:

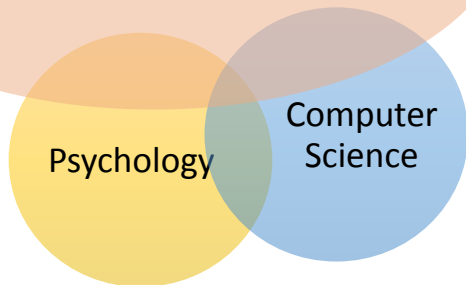
- breaking barriers for technology usage
- providing acoustic and linguistic cues for user state detection




<https://menhir-project.eu/>

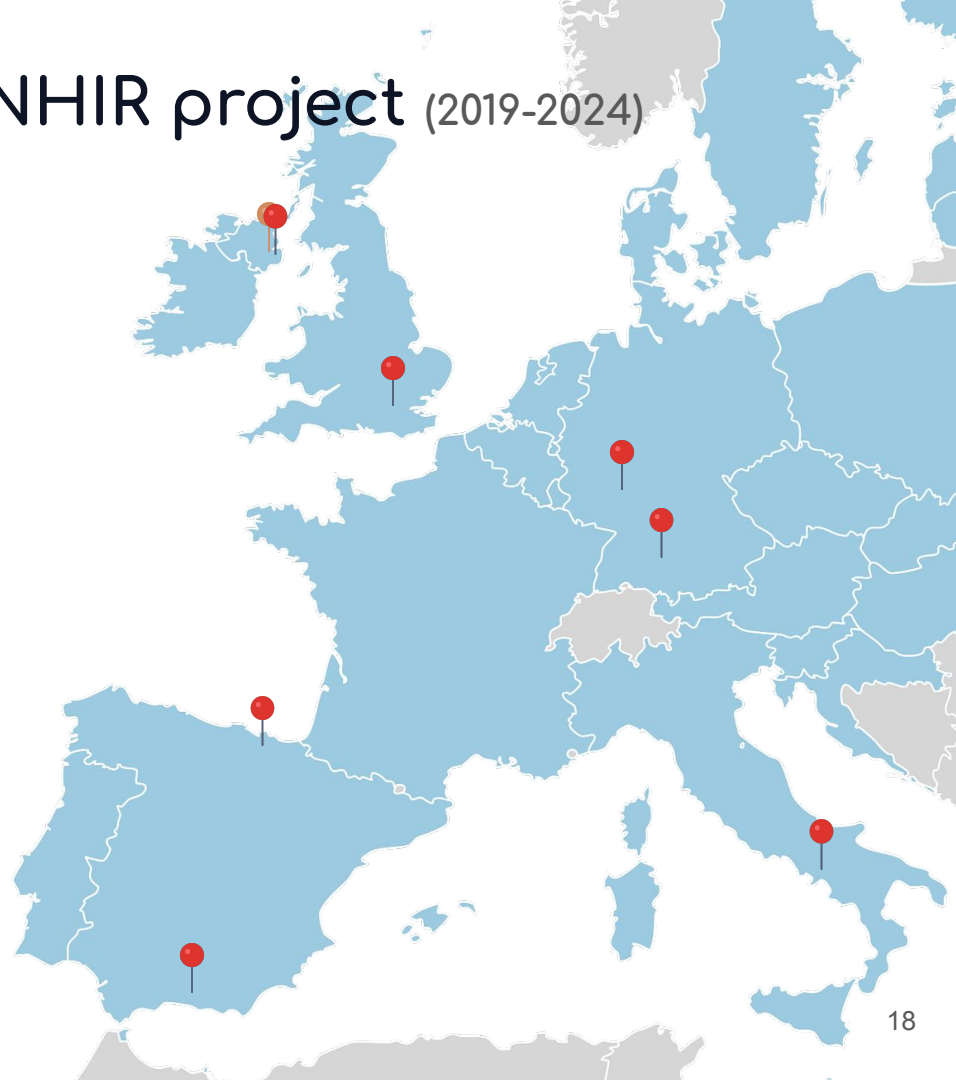


Real use case:



-  Universidad de Granada (coordinator)
-  Action Mental Health
-  Ulster University
-  Universidad del País Vasco
-  Intelligent Voice Ltd.
-  Ulm University
-  Univ. della Campania Luigi Vanvitelli
-  GLOBIT

MENHIR project (2019-2024)





Real use case:

MENHIR project (2019-2024)

Co-creating requirements and assessing end-user acceptability



Experts by experience: nobody knows better about their needs and preferences than users.



Co-creation: The technology studied and developed in MENHIR is designed together with persons with mental ill health, experts and caregivers.



Real use case:

MENHIR project (2019-2024)

To **assess** the **acceptability** and **co-create requirements** for a speech-based mental health chatbot for people with anxiety and/or mild depression.

We performed a **thematic analysis** of data collected at a **user-centred workshop** involving clients of Action Mental Health.



Real use case:

MENHIR project (2019-2024)



9 Action Mental Health (AMH) clients, 1 key worker from AMH,
3 MENHIR researchers

1. Welcome and brief introduction to the project.
2. Demonstration of a mental health chatbot and smart speaker by workshop facilitators.
3. Completion of consent form and study proforma.
4. Group discussion.

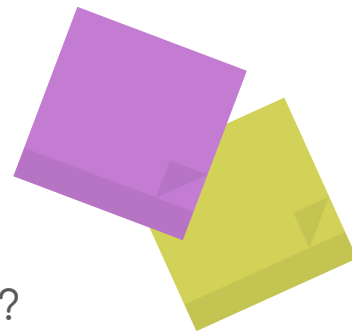


Real use case:

MENHIR project (2019-2024)

Group discussion:

- Strengths and limitations of this type of chatbot.
- Who would use this chatbot?
- Which type of people would the chatbot suit?
- Could it be used in conjunction with AMH mental health support services? (how?)
- What are the pros and cons of using chatbot in the context of mental ill health?
- What features should a mental health chatbot have?
- What kind of conversations would you want to have with a chatbot?
- What kind of persona/s should the chatbot have?





Real use case:

MENHIR project (2019-2024)

Themes identified

Challenges faced by people living with mental ill health:

1. Isolation
2. Difficulty for honest disclosure

When there is no-one to talk to or no-one they feel they can trust

They may have a key worker or somebody that they would talk to, but these people are not available after a certain time a day, so therefore they do not have anybody



I love the idea of the chatbot because a lot of the time with mental health you can feel very isolated, very much on your own, cut off from the world...

There are people that I would tell the whole story and there are people I would say yeah I am fine

I do not want to burden you, I do not want to put my problems on you

A lot of the time (...) if anybody even family asks how are you? I say ohh I am fine, I'd rather say I'm fine rather than go in detail about how you really feel



Real use case:

MENHIR project (2019-2024)

Themes identified

Chatbot functionalities:

3. Symptom recognition
4. Continuous monitoring
5. Disclosure facilitation
6. Companionship/active listening
7. Risk detection



“The chatbot would pick up an inflection in your voice to know that things aren’t right”

“I didn’t recognize (...) being so mentally unwell. In reflection we can see that (...) things would have been very different”

“You get a level of distress and you may not be able to recognize it”

“Your therapist gets to see you one day a week and you might be doing great (...) whereas in your worst moments he does not get to see those symptoms”

“Companion (...) somebody to confide in and again you are not worried about their opinion”

“I would activate it because (...) I need to get this out”

“Just generic, just prompting you, open discussion” - even if it does not make sense



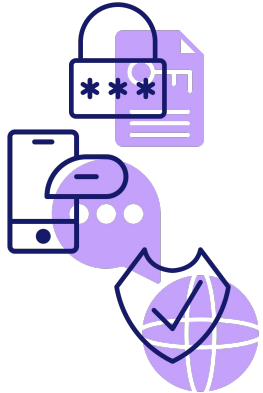
Real use case:

MENHIR project (2019-2024)

Themes identified

Chatbot characteristics:

- | | | | |
|-----|----------------------------------|-----|----------------|
| 8. | Personalization | 11. | Explainability |
| 9. | Configurable proactiveness | 12. | Privacy |
| 10. | User access to their information | 13. | Vulnerability |



“Selective menus, because I would find it very helpful a reminder to take my medications, for my appointments...”

“I would feel differently depending on the day (...) some days I would like it doing checking on me, some days that would really annoy me”

“Where does it get that information from? If I type that into Google I get a bunch of options and I can see this one is from a credible source, this one isn't from a credible source (...) and I can then use my judgement to choose the correct one, whereas if I use a chatbot (...) it gives me an answer”



Real use case:

MENHIR project (2019-2024)

Themes identified

Use of the chatbot:

14. Cost of the chatbot
15. Access to the chatbot



“My issue would be like how much is it gonna cost? (...) is it a subscription thing? Can I try it for a couple of months and then see if it works for me? This is the type of thing I can try”
“Widely free and easily available”

16. Intention to use the chatbot
17. Use of the chatbot as a complement to therapy



“I would definitely use it”
“I would certainly use the chatbot”
“I would certainly make use of a reminder”
“In the past two weeks how have you felt about this? (...) It is very hard to fill that in because you don't tend to spend a lot of time thinking ohh in the last two weeks I felt like this, you tend to answer how you felt that day (...) If there is a way that you could build that in“



Real use case:

MENHIR project (2019-2024)

The system should provide mechanisms to let the users decide whether they allow the system to **collect different types of data** (voice, demographic, hobbies/interests, daily habits, emotional state, emergency contacts)

The system should provide mechanisms to let the users decide which **pieces of information (if any) can be shared with their key workers** (counsellors, psychiatrists, etc) and under which conditions.

The system should be able to **start a conversation with the user proactively** according to an event, a programmed frequency or a learned routine.

The users should be able to **personalize the frequency** with which they engage in conversations with the system and whether they are initiated by them or by the system.

The system must be able to provide the users a space where they can express their thoughts and feelings acting as a companion in the form of an **interactive diary**.

The system should be able to **assess the emotional state** of the user (from voice input, interaction behaviour, standardized scales).

The system should be able to **track the changes** in the users' emotional state over time.



Real use case:

MENHIR project (2019-2024)

The system should be able to **detect and notify risks** in the user's interactions and during the mood tracking.

The system should be aware of the **users action plans and preferences**. This information can be provided by the users and/or their key workers or counsellors.

The system should be able to periodically **remind and notify** the user to engage in activities related to their action plan. These activities can be specific for each user or based on a generic model.

The system should be able to **motivate the user** explaining them reasons why they should attend or complete their activities, which can be pragmatic and/or emotional.

The system must be able to **explain the benefits** derived from attending the activities and be engaged on the action plan.

The system should be able to **suggest tasks, activities or advice** the user in certain kinds of topics related to mental health and general wellbeing.

The reminding and motivation functionalities can be **personalized for each user and optimized as the system gains a better understanding** of each particular user through sustained interactions over time.



Real use case:

MENHIR project (2019-2024)

Co-creation workshop with scientists and people living with mild anxiety and depression to understand the acceptability and requirements for the development of mental health chatbots.

Results of **thematic analysis** highlight 17 key themes related to challenges that could be addressed (isolation and honest disclosure), functionalities (symptom recognition and monitoring, companionship, risk detection), characteristics (including personalization, proactiveness, accessibility, and privacy) and usage conditions (cost, access, intention to use).



Real use case:

MENHIR project (2019-2024)

Second International Digital Mental Health and Wellbeing Conference

<https://www.ulster.ac.uk/conference/digital-mental-health-and-wellbeing>

19-21 June 2024

Derry-Londonderry



slido



Can you share a challenge you've faced to gather requirement from a client/user?
how did you address it?

① Click **Present with Slido** or install our [Chrome extension](#) to activate this poll while presenting.

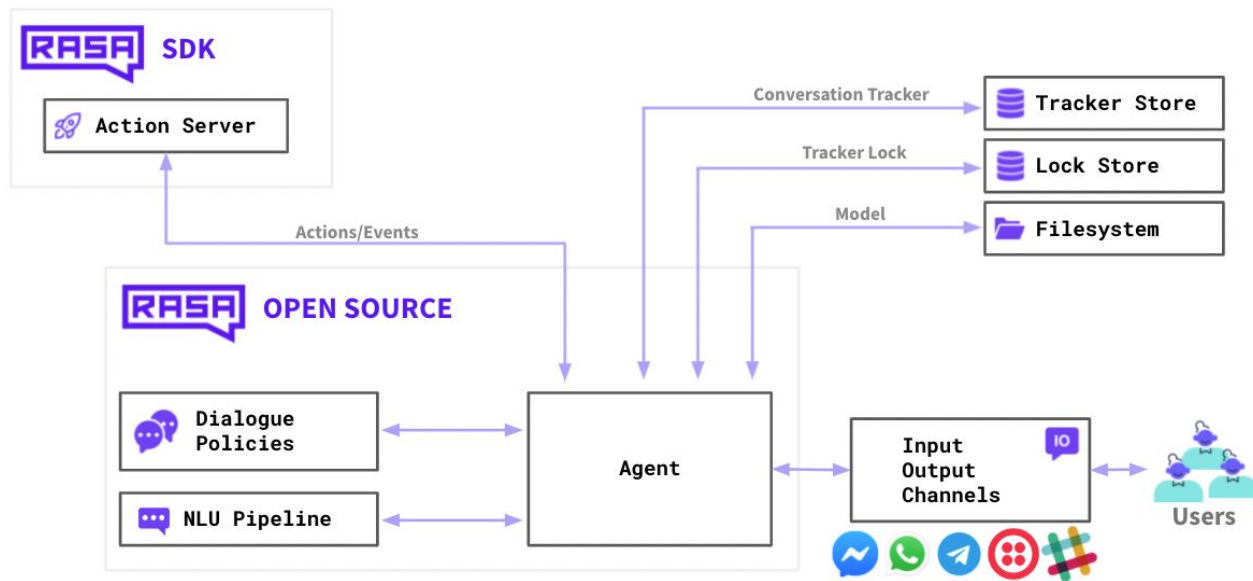
Gathering requirements for conversational systems

3. Identifying communication channels and connections

- Connection to business processes in the back-end
- Triggers
- Channels

Gathering requirements for conversational systems

3. Identifying communication channels and connections



Example of channel-specific responses in RASA:

responses:

utter ask game:

- text: "Which game would you like to play on Slack?"
channel: "slack"
- text: "Which game would you like to play?"



Real use case:

Cátedra RTVE-UGR (Started 2022)

RTVE-UGR Chair

in deep speech
synthesis and
conversational AI
and their
applications in
news verification

RESEARCH...

Deep speech
synthesis

Conversational
AI

Training

Applied to audio
deep fake
detection

Applied to news
verification

Dissemination
and
communication

... APPLIED

<https://catedrartve.ugr.es/>



Real use case:

Cátedra RTVE-UGR (Started 2022)

Applications of conversational systems:

- Assistants for journalists
- Debunk authoring assistants
- Fact check conversations
- Opposing news content
- Generative AI: tone, accessibility...
- Participative journalism, data collection and multimedia
- Information discovery



Real use case: 2023)

Example: Elections Skill

Pilot to cover electoral information in small towns (<https://www.rtveia.es>)

Alexa Skill for Elections (July

5.000 towns under 1k inhabitants

Received the IBC 2023 Award

Participants: Narrativa, Monoceros Labs, Universidad de Castilla-La Mancha, Universidad de Granada, ONCE, AWS y el Centro Territorial de Castilla-La Mancha de RTVE



rtve Narrativa monoceros LABS UCLM UNIVERSIDAD DE GRANADA cátedra rtve-ugr ONCE "Alexa, abre elecciones inteligencia artificial."

Elecciones Generales 23J Elecciones Municipales 28M

El contenido que se ofrece es el resultado de una prueba tecnológica y de concepto. Más información

Inicio

Resultados Elecciones Generales 23J

Selección comunidad: [dropdown] Selección provincia: [dropdown] Selección municipio: [dropdown] Busca un municipio [input]

Listado de municipios

| A | | Luciana | Lucillo |
|--------------|------------------|-----------|-----------|
| A Teixeira | A Veiga | Lucillos | Ludiente |
| Ababuj | Abades | Luelmo | Luenta |
| Abadía | Abáigar | Luesia | Luesma |
| Abajas | Ábalos | Lugros | Lójar |
| Abaltzisketa | Abánades | Lumbreras | Lumpiaque |
| Ahanto | Abarca de Campos | Luna | Lupiana |

slido



**What type of devices /
channels have you used for
your systems?**

① Click **Present with Slido** or install our [Chrome extension](#) to activate this poll while presenting.

A solid blue vertical bar is positioned on the left side of the slide, extending from the top to the bottom.

2. DESIGN

Practical dialogue management design considerations

Designing conversations: start with the **happy path**:

Ideal user-system interaction flow: the user interacts exactly as expected by the system, aligned with the system capabilities, thus obtaining the best possible outcome.



Automated testing using predefined paths, for example in RASA with test stories (stay tuned for the lab).

Baseline for successful interactions and serves as a reference point for testing and refining the chatbot's performance

Practical dialogue management design considerations

Designing conversations: start with the happy path:

Intent alignment:

- The user's input is clear and aligns with the chatbot's understanding of user intents.
- The chatbot accurately interprets the user's query.

Accurate responses:

- Information provided by the chatbot is relevant and accurate.

Smooth flow:

- The conversation progresses without interruptions or errors.
- Follow-up questions or prompts from the chatbot are contextually appropriate and easy for the user to understand.

Task completion:

- The user successfully completes their intended task or achieves their goal through the chatbot interaction.
- Any actions or transactions are executed successfully.

Practical dialogue management design considerations

Designing conversation flow

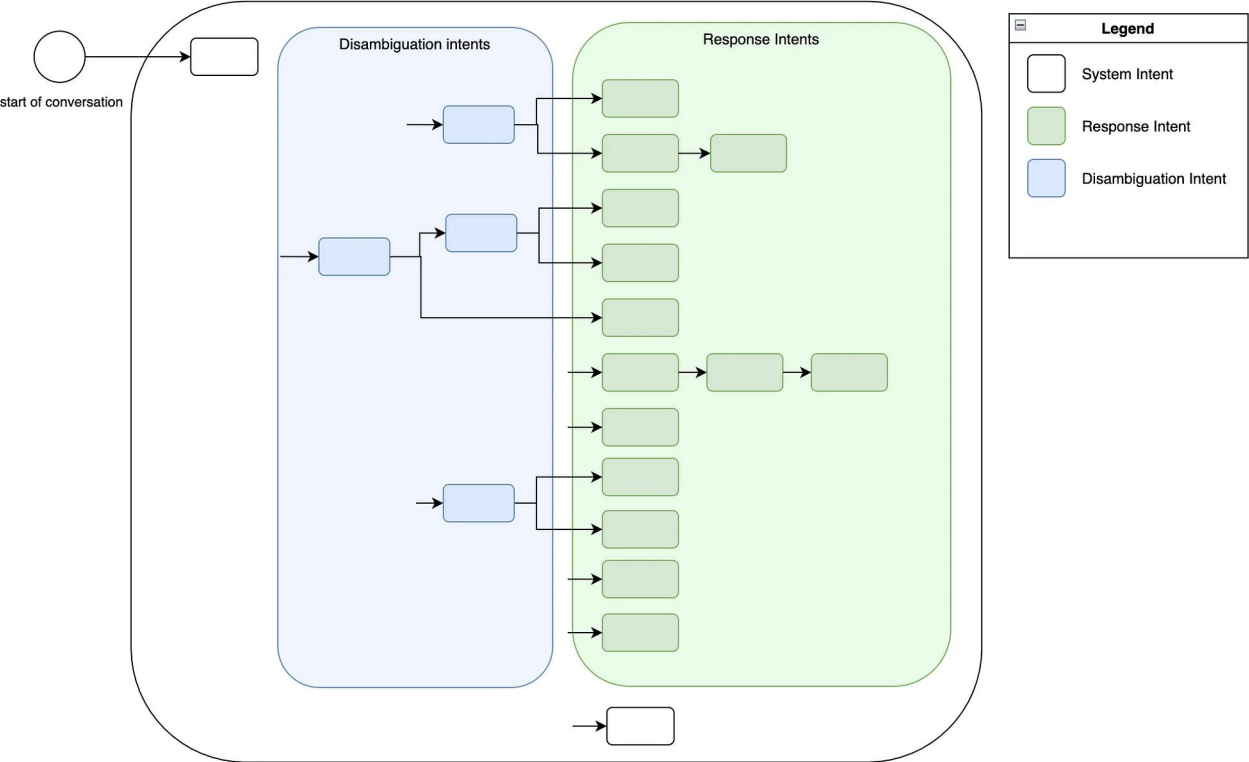


Image from:
<https://towardsdatascience.com/suggestions-on-how-to-structure-intents-in-chat-bots-and-gather-useful-feedbacks-f72f7e552090>

Practical dialogue management design

Prototyping tools:

Voiceflow

<https://youtu.be/pb6kADbEFUQ>



<https://creator.voiceflow.com/project/65513a79650cc0e692e36dd3/domain/cl8um46q3058r07o11fb98t7o/canvas/63d6e471f4d193000763b6ff>

Practical dialogue management design

Not always more sophisticated is better

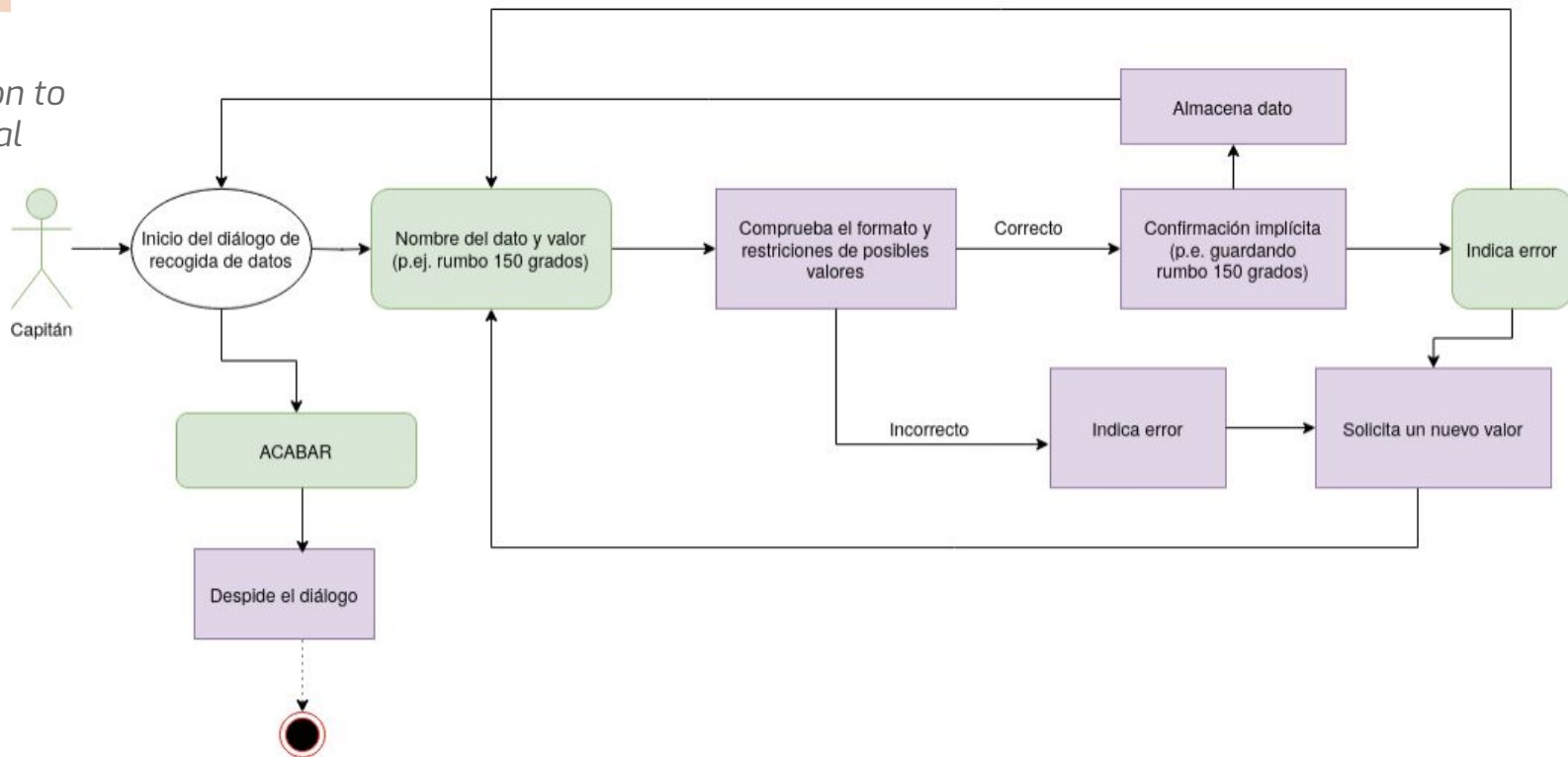


Real use case:

Bitacora project (2023)

Example:

Data collection to create a digital navigation logbook





Real use case:

MENHIR project (2019-2024)

B. Always remind all activities

| Pros | Cons |
|--|--|
| Totally automated | Less client autonomy |
| Decreased probability that the client forgets the activity | Client IT Skill level/ Use of Smartphone/Tablet/Computer |
| Simple design | Reliance on notifications |
| increased communication with client | notifications seen as intrusions |
| Increased programme attendance levels | |

C. Start reminding all the activities and progressively reduce the reminders.

C.1. Reduce the frequency of reminders after a certain number of days

| Pros | Cons |
|--------------------------|---|
| Supports client autonomy | It is not sensible to particular cases |
| Totally automated | Possible effect of the lack of reminder in class attendance |
| | The frequency of reminders only decreases |
| | Client becomes reliant on reminder |

Smart reminder app

C.3. A mix of C.1 and C.2: frequency of reminders is set according to time and attendance

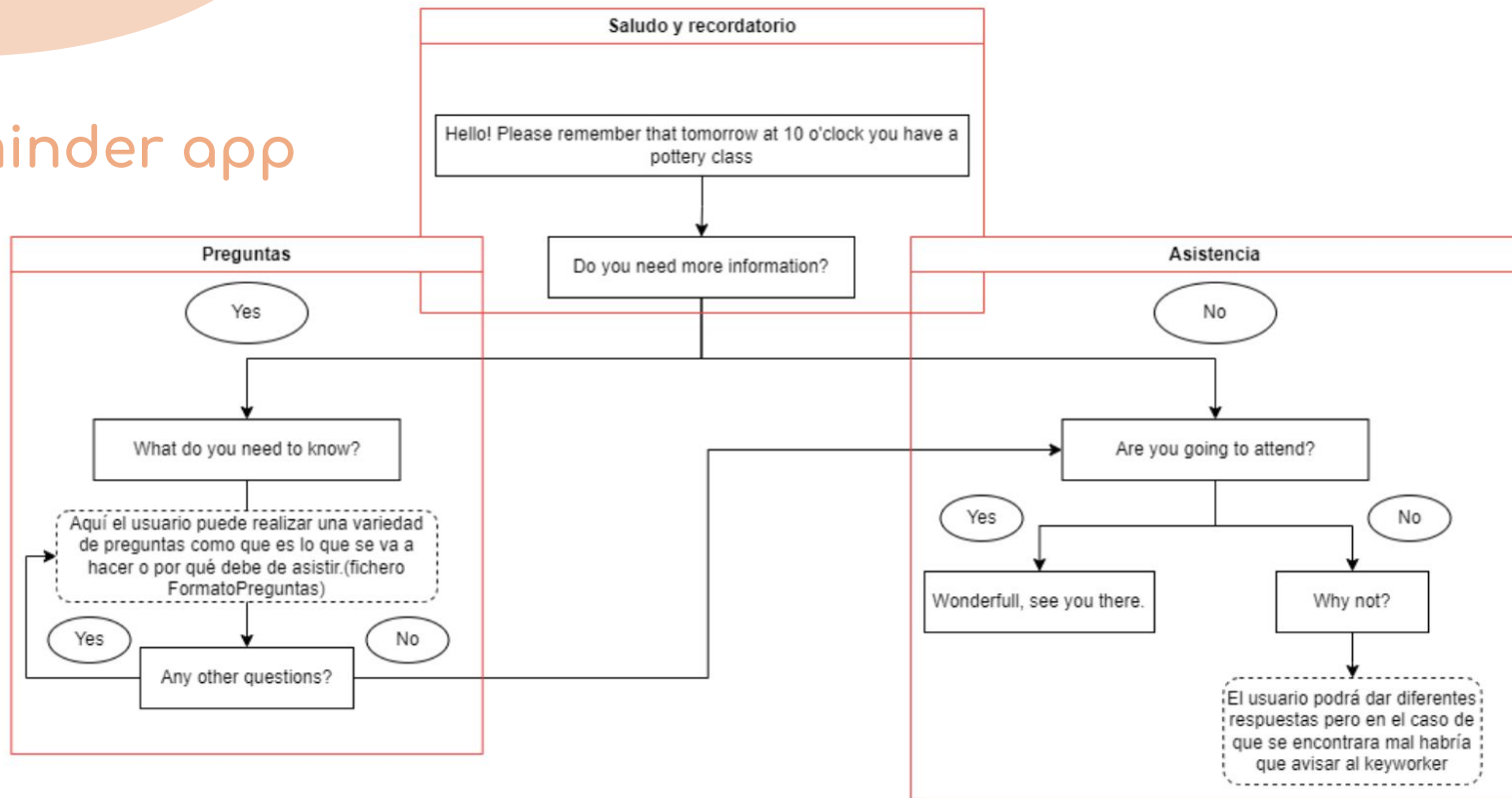
| Pros | Cons |
|---|---|
| Supports client autonomy | Possible effect of the lack of reminder in class attendance |
| Totally automated | Complex design: we could pay attention to the number of classes missed or to the evolution of class attendance over time (whether there is a sustained, increasing or decreasing attendance). |
| Sensible to client engagement in the activities | |
| The frequency of reminders can increase or decrease | |



Real use case:

MENHIR project (2019-2024)

Smart reminder app



Practical system prompt design considerations

Maxim of quantity: the speaker should provide the listener as much information as necessary to fulfill the purpose of interaction

| | | |
|--|--|--|
| <p>Clear and concise</p> | <p>User: "What's the weather like in Jaca?"</p> <p>Chatbot: "The current temperature in Jaca is 19°C with sunny sky."</p> | <p>User Prompt: "Can you recommend a good restaurant nearby?"</p> <p>Chatbot Response: "Sure! There's a Pizza restaurant called 'Little Italy' nearby. It has been in business for 10 years, has a diverse menu including pizza, pasta and antipasti. The ambiance is cozy, and they have a happy hour from 4 PM to 7 PM."</p> |
| <p>Good practice: progressive disclosure</p> | <p>User Prompt: "Tell me about your pricing plans."</p> <p>Chatbot Response (Initial): "We offer three pricing plans: Basic, Standard, and Premium."</p> <p>Chatbot Response (Upon Follow-up Inquiry): "The Basic plan starts at \$9.99 per month and includes..."</p> | |

Practical system prompt design considerations

More conversational first-level no-match:

System: And how many people are going?

User: It's my wife and myself.

System A: Sorry, how many was that?

System B: I am sorry, I didn't catch that, please say the number of people going to the party, for example "two"

Practical system prompt design considerations

More conversational first-level no input:

System: to get started, tell me your account number

<Noinput>

System A: I'm sorry, I didn't hear anything. What is your account number?

User: Well, that's because I didn't say anything, but here it is...

System B: Other system prompt: You account number is 10 digits long and you can find it in your bill...

Practical system prompt design considerations

Informative disconfirmations:

System: That's from Boston to New York on July 15th at 6PM. Did I get that right?

User: No, the 25th / Not Boston, Preston ...

System A: Sorry, did I get it right?

System B: If it is correct, say "yes", otherwise say "no"

All these examples are from James Giangola:

https://www.youtube.com/watch?v=wuDP_eygsvs

Other design considerations: bot persona

Elements: Name, voice, communication style, story and purpose.

Objectives: Brand representation, user engagement, memorability, cultural sensitivity.

Caution: can be misleading about system capabilities,

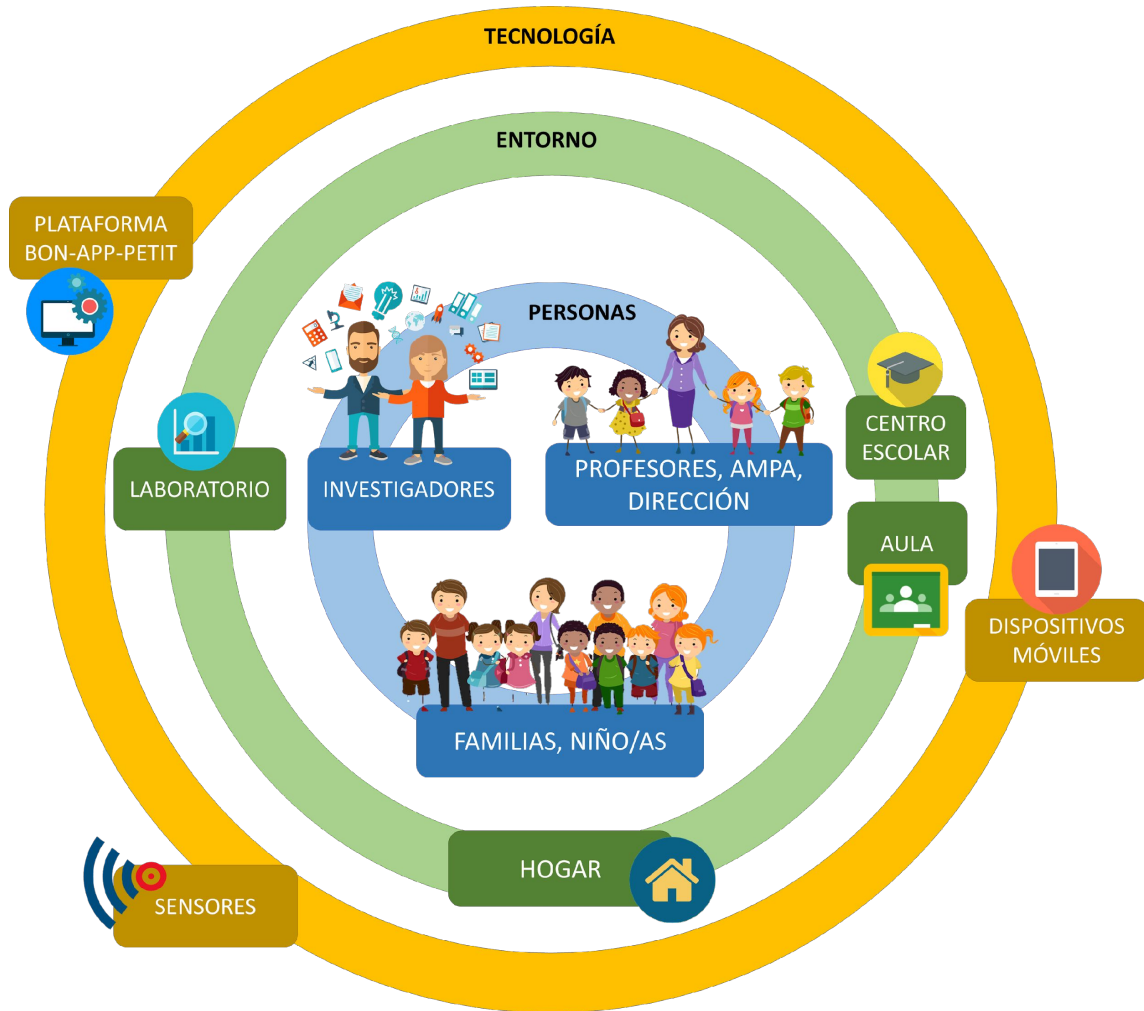


 Real use case:
Bon-App-Petit project
(2020-2023)

Example: Virtual assistant

Virtual assistant for the promotion of healthy habits among school children

<https://bon-app-petit.eu>





Real use case: Bon-App-Petit project (2020-2023)

Example: Virtual assistant

Virtual assistant for the promotion of healthy habits among school children



Orange-Tan = orange + orangutan

Orange-Tan is a defender of forests and mountains. What he enjoys the most is experiencing nature in its purest form and always staying active.

Don't be deceived by his appearance; besides being strong, he is very agile and can climb any tree to get his favorite food, fruit.

He knows perfectly well that fruit is a very healthy food that provides him with all the water, vitamins, minerals, fiber, and various beneficial compounds for the body. He eats oranges two at a time!

His favorite sport is athletics; you should see him in the obstacle race – he has no rival!



Real use case: Bon-App-Petit project (2020-2023)

Example: Virtual assistant

Virtual assistant for the promotion of healthy habits among school children



Atoona = toon + tuna

You'll find her **training** **on the coast** or reading a good book on the sand after her **workout**.

Thanks to her diet rich in **fish**, she is strong and intelligent, able to remember any date or event effortlessly.

Since **fish** provides her with proteins, vitamins, minerals, and fatty acids like omega-3, she has an incredible memory.

She loves **swimming or diving** in the **Mediterranean Sea** and water sports. In swimming, no one can go as far or as fast as Atoona!

3. DEVELOPMENT

The iterative development process

Highlight incremental improvements:

Emphasize that chatbot development is an **iterative process**.

Start with a basic version of the chatbot and gradually enhance its capabilities based on **user feedback**.

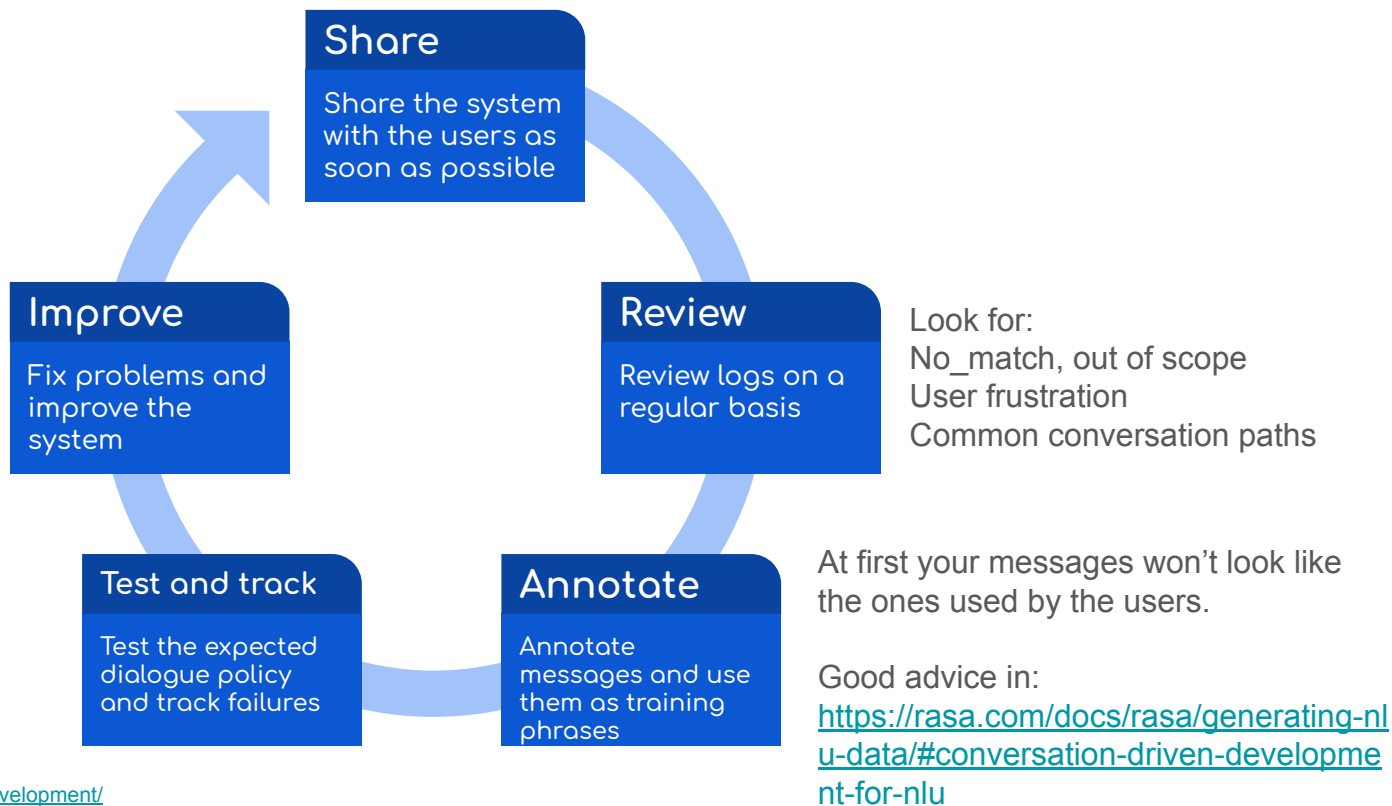
Provide demos and examples:

Offer live demonstrations or prototypes to **showcase** the system's capabilities.

Integrate Human-in-the-Loop (HITL):

Human oversight can enhance accuracy and handle scenarios where automation might fall short (e.g. CAU).

Conversation-driven development (CDD)



More in:

<https://rasa.com/docs/rasa/conversation-driven-development/>

There exist tools,
e.g. RASA X or in
DialogFlow:

<https://cloud.google.com/dialogflow/cx/docs/concept/test-case>

Webhook: On

hi

Hello, this is a shirt ordering virtual agent. How can I help you?

buy shirt

Ok, let's start a new order. I'd like to collect a bit more information from you. What color would you like?

blue

I'd like to collect a bit more information from you. What size do you want?

1

large

You have selected a large, blue shirt. You can pick up your order for a large blue shirt in 7 to 10 business days. Goodbye.

2

Webhook: On

hi

Hello, this is a shirt ordering virtual agent. How can I help you?

buy shirt

Ok, let's start a new order. I'd like to collect a bit more information from you. What color would you like?

blue

I'd like to collect a bit more information from you. What size do you want?

1

large

You have selected a large, blue shirt. You can pick up your order for a large blue shirt in 7 to 10 business days. Goodbye.

2



Real use case:

Hispabot-COVID19 (2020)

Example:

Chatbot deployed on web and whatsapp to provide trustworthy information about COVID during lockdown

Initiative of the State Secretariat for Digitalization and Artificial Intelligence (SEDIA), through the Language Technologies Plan: coordinate the development of a conversational assistant to answer frequently asked questions about COVID-19 (March 27, 2020).

Multidisciplinary working group of public institutions, academic entities, and companies.

Access to information from official sources:

- Ministry of Health
- World Health Organization
- BOE

Objective: provide reliable data and information at any time to the public, preventing the overload of healthcare helpline numbers.



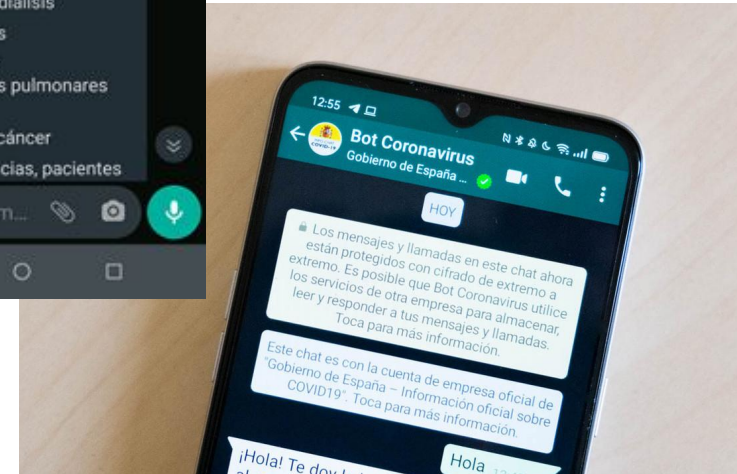
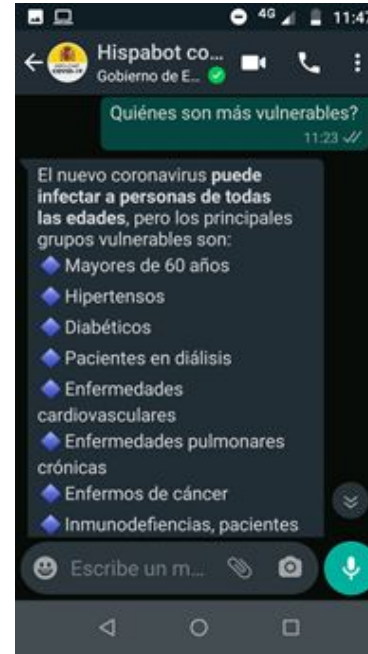
Real use case:

Hispabot-COVID19 (2020)

It responded to questions about:

- symptoms,
- vulnerable groups,
- relation to other diseases,
- how it spreads,
- how to prevent and protect oneself,
- debunking myths,
- living with infected individuals,
- conditions for quarantine and isolation,
- helpline numbers,
- application of the State of Alarm,
- Transition Plan towards the “New Normal”.

The assistant did not require or analyze personal data.





Real use case:

Hispabot-COVID19 (2020)

The iterative development process

Key issues in this use case:

Is a chatbot a valid resource?

Will it be used?

What will be the main questions? How will they be posed? Is it what we expected?



Real use case:

Hispabot-COVID19 (2020)

The iterative development process

Content creation and supervision:

- Medical specialists.
- Accessibility content specialists.

Development of the first version:

- Language understanding model: entities (e.g., autonomous communities, synonyms of coronavirus, names of related diseases, etc.), training phrases (>8000 phrases in the final model), question categories (200 categories in the final model).



Real use case:

Hispabot-COVID19 (2020)

The iterative development process

Deployment of the first pilot (3rd April):

- Integrated into the websites of the Government of La Rioja and Rioja Salud.
- From April 3rd to April 7th, the assistant received more than 5,000 queries on these portals.

The screenshot displays the website of the Government of La Rioja and Rioja Salud. The header includes the logo and navigation menu. The main content area features a 'Noticias' section with a large image of a street at night illuminated with green lights. A chatbot interface is overlaid on the right side, displaying a message: 'Este servicio almacena de forma anónima tus consultas para ir mejorando las respuestas y aprender a resolver nuevas preguntas. En ningún caso se solicitará ni debes facilitar ningún dato de carácter personal.' Below this, it offers updated information on COVID-19 and asks, '¡Hola! ¿Necesitas información sobre el coronavirus?'. The chatbot has a text input field labeled 'Escribe'.



Real use case:

Hispabot-COVID19 (2020)

The iterative development process

In parallel to the pilot in La Rioja:

- Study of logs, content improvement.

April 8th → Deployment on WhatsApp and Telegram.

- Continuous improvement of the system.
- Detection of understanding errors and restructuring of the knowledge base according to updates. Addition of entities, categories, training phrases.
- Information about the last update of the chatbot available to users.

The system was operative until the end of June. In the period from April to June it helped solve more than 350.000 citizen queries.





Real use case:

Hispabot-COVID19 (2020)

Daily reviews included hot issues or the most prominent news in the media:

- Symptoms of the disease and how to act if they occur, helpline numbers, and official figures.
- Conditions for going out and shopping.
- Use of masks.
- Time slots and mobility.

The daily analysis of queries provides valuable real-time information to understand the most common concerns of citizens, which varied greatly from March to June.

| Tipo de pregunta | Número de consultas | Tipo de pregunta | Número de consultas |
|---|---------------------|---|---------------------|
| Desplazamientos permitidos | 11.660 | Mascarillas | 2.250 |
| Plan de Transición | 11.190 | Fases y provincias | 1.920 |
| Franjas horarias | 8.330 | Síntomas de la enfermedad | 1.800 |
| Circulación en coche y transporte público | 5.540 | Bares y servicios de restauración | 1.720 |
| Cifras de contagiados | 4.790 | Uso de parques y playas | 1.070 |
| Condiciones para salir de casa | 3.780 | Qué es la enfermedad | 1.060 |
| Salidas con menores de 14 años | 3.360 | Permisos para trabajar | 995 |
| Tiendas abiertas | 3.340 | Realización de trámites administrativos | 883 |
| Salir y recibir visitas | 3.260 | Agricultura y huertos propios | 880 |
| Viajar fuera de España | 2.400 | Reuniones | 874 |



Real use case:

Bon-App-Petit project (2020-2023)

The iterative development process

Key issues in this use case:

A lot of uncertainty about how children would communicate with the system, wordings, names of meals, etc.

Need to collect information about the initial habits through validated questionnaires.

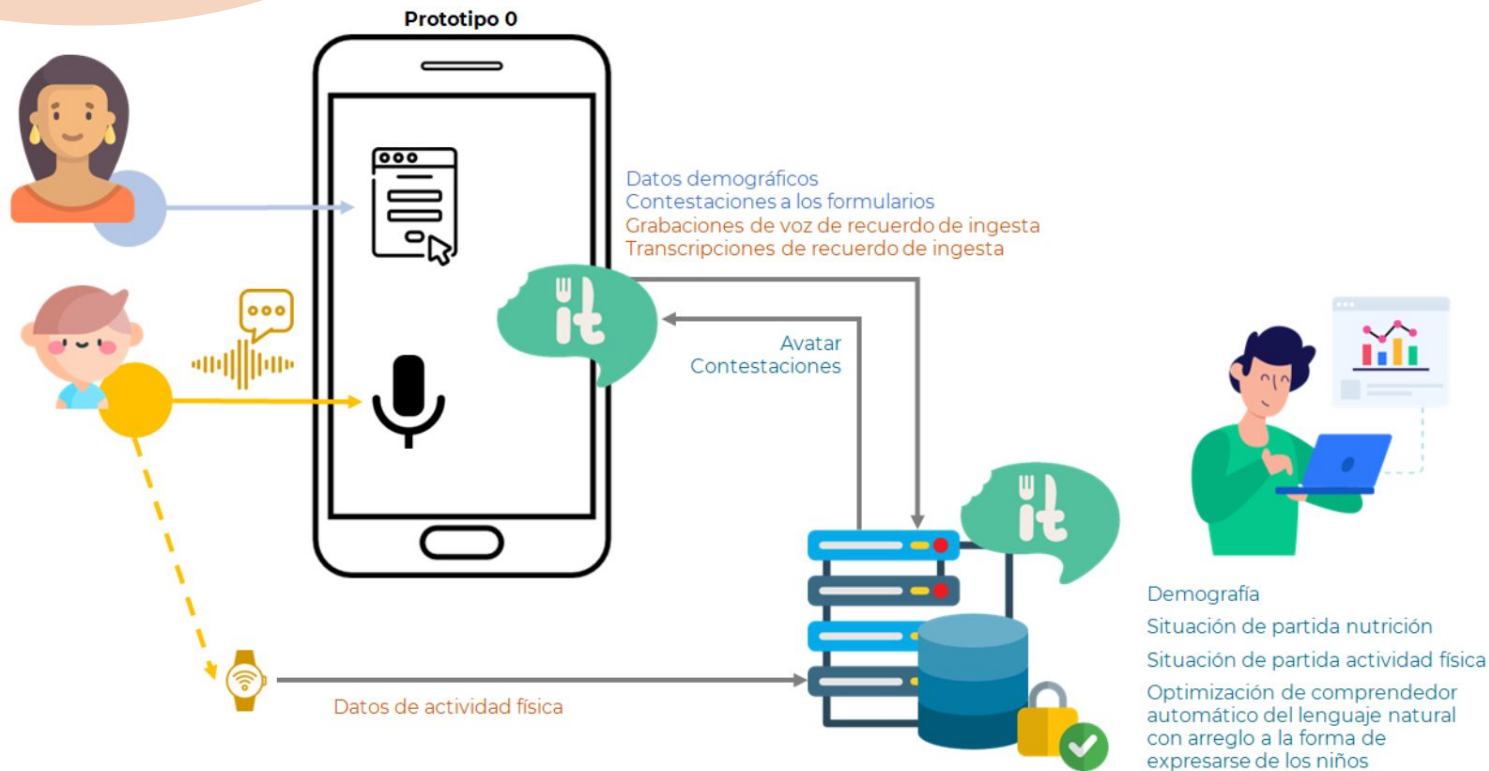
Potential problems with the first pilot (e.g. NLU errors) and “boring” questionnaires could undermine the positive effect of the avatars.

Solution adopted = development of 2 prototypes: with and without avatar.



Real use case:

Bon-App-Petit project (2020-2023)





Real use case:

Bon-App-Petit project (2020-2023)

Phase 1

Duration 1 week

Description

The first use of the application is done by a parent. The application displays a form that can be completed in a few minutes.

When prompted by the application, the tablet is passed to the child, who first fills out a short form and can then interact with the character, which presents itself as a virtual egg. Every day, the child tells the egg what they ate for breakfast, lunch, and snack on that day, as well as dinner from the previous day.



Real use case:

Bon-App-Petit project (2020-2023)

Test IPA-C

Actividad física en tu tiempo libre: ¿Has hecho alguna de estas actividades en los últimos 7 días (última semana)? Si tu respuesta es sí, ¿cuántas veces lo has hecho?



| Actividad | No | 1-2 veces | 3-4 veces | 5-6 veces | 7 veces o más |
|---|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Saltar a la comba <small>*Obligatorio</small> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Patinar <small>*Obligatorio</small> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Jugar a juegos como el pilla-pilla <small>*Obligatorio</small> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Montar en bicicleta <small>*Obligatorio</small> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

⏪ Volver
Chat con Atoona ⓘ

¿Recuerdas qué cenaste anoche?



ESPERA PARA RESPONDER




Real use case:

Bon-App-Petit project (2020-2023)

Phase 2

Duration 2 weeks

Description

The virtual egg hatches and the character appears (Orange-Tan or Atoona according to each child's choice). This phase consists of continuing to talk about what they eat and also solve daily activities about what the character explains to them.

By solving the activities and eating healthy in phases 1 and 2, points will be obtained that can be exchanged for accessories for the characters (hats, T-shirts, etc.) in a virtual closet that will be activated in phase 2.

At the end of the second week, the application asks the parent to answer a questionnaire to evaluate the application.

Considerations for LLM-based development

- Obtaining data (crawling, specific data sources, language identification)
- Cleaning and deduplication
- Toxicity and bias removal
- Model training
- Model tuning

slido



Main topics you are working on related to LLMs

① Click **Present with Slido** or install our [Chrome extension](#) to activate this poll while presenting.



Research:

CONVERSA project (2022 - 2024)

TED2021-132470I00 - Effective and Efficient Resources and Models for Transformative Conversational AI in Spanish and co-official languages

CONVERSA aims to contribute to the democratic access to conversational AI in Spain through the provision of open data and open-source conversational technology in Spanish and co-official languages.

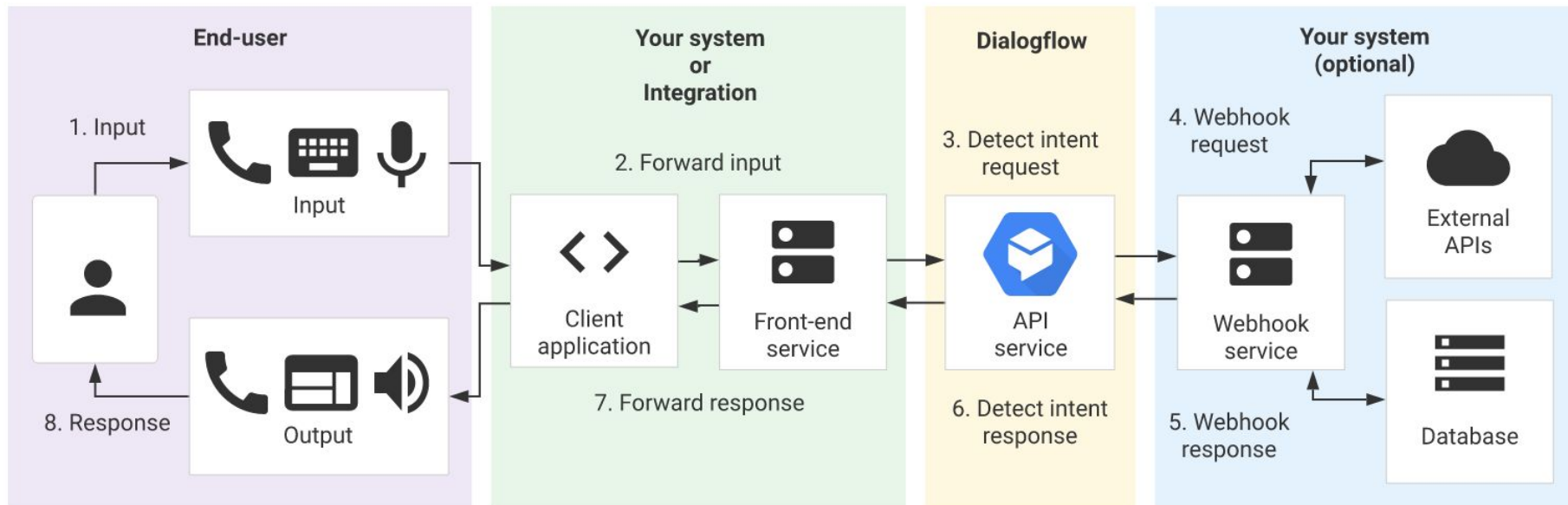
- Massive multilingual corpora
- Description of cleaning pipeline
- Open source tools
- Open models

<https://github.com/conversa-ai>

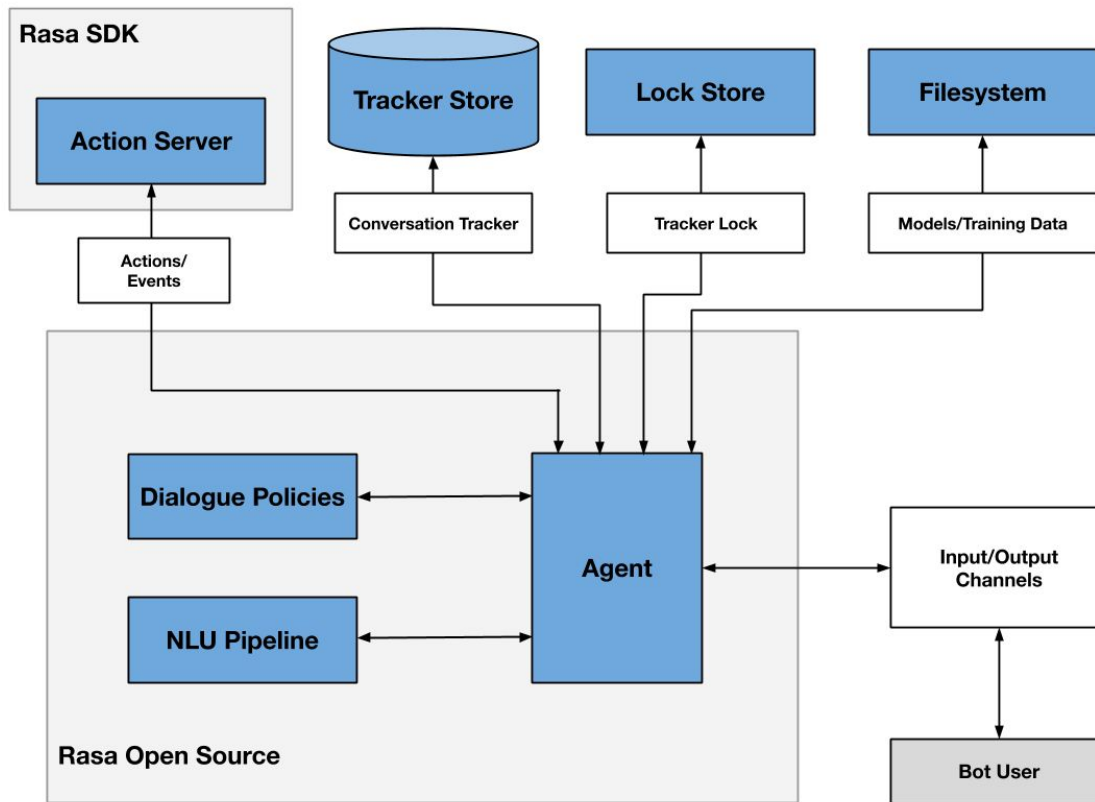
4. DEPLOYMENT

(and a bit of design again)

Example architecture in Dialogflow



Example architecture in RASA



<https://rasa.com/docs/rasa/arch-overview>

Deployment in RASA

Best practice: Helm Charts → Help you define, install, and upgrade Kubernetes applications.

<https://github.com/RasaHQ/helm-charts/tree/main/charts/rasa>

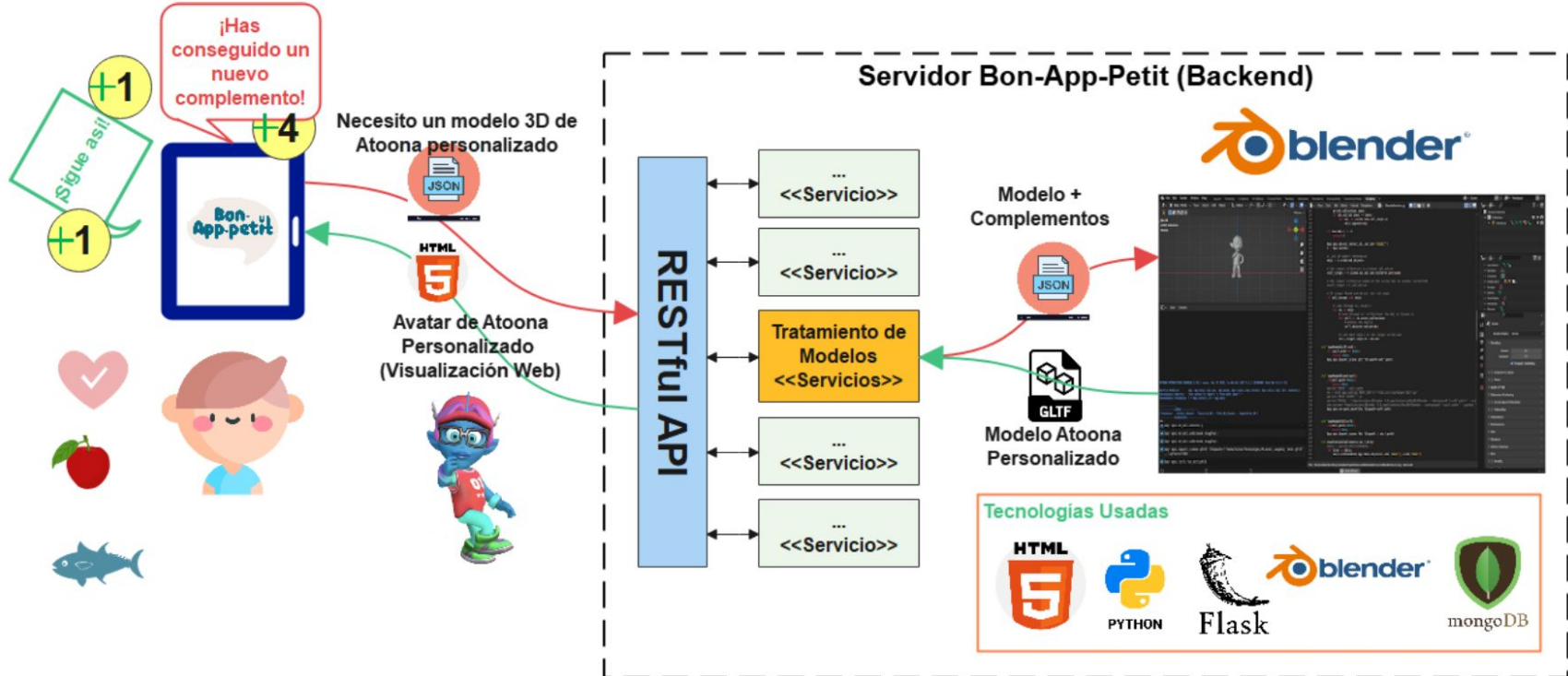
<https://rasa.com/docs/rasa/deploy/deploy-rasa>

Why? <https://www.youtube.com/watch?v=ISkfwJ4b2ss>



Real use case:

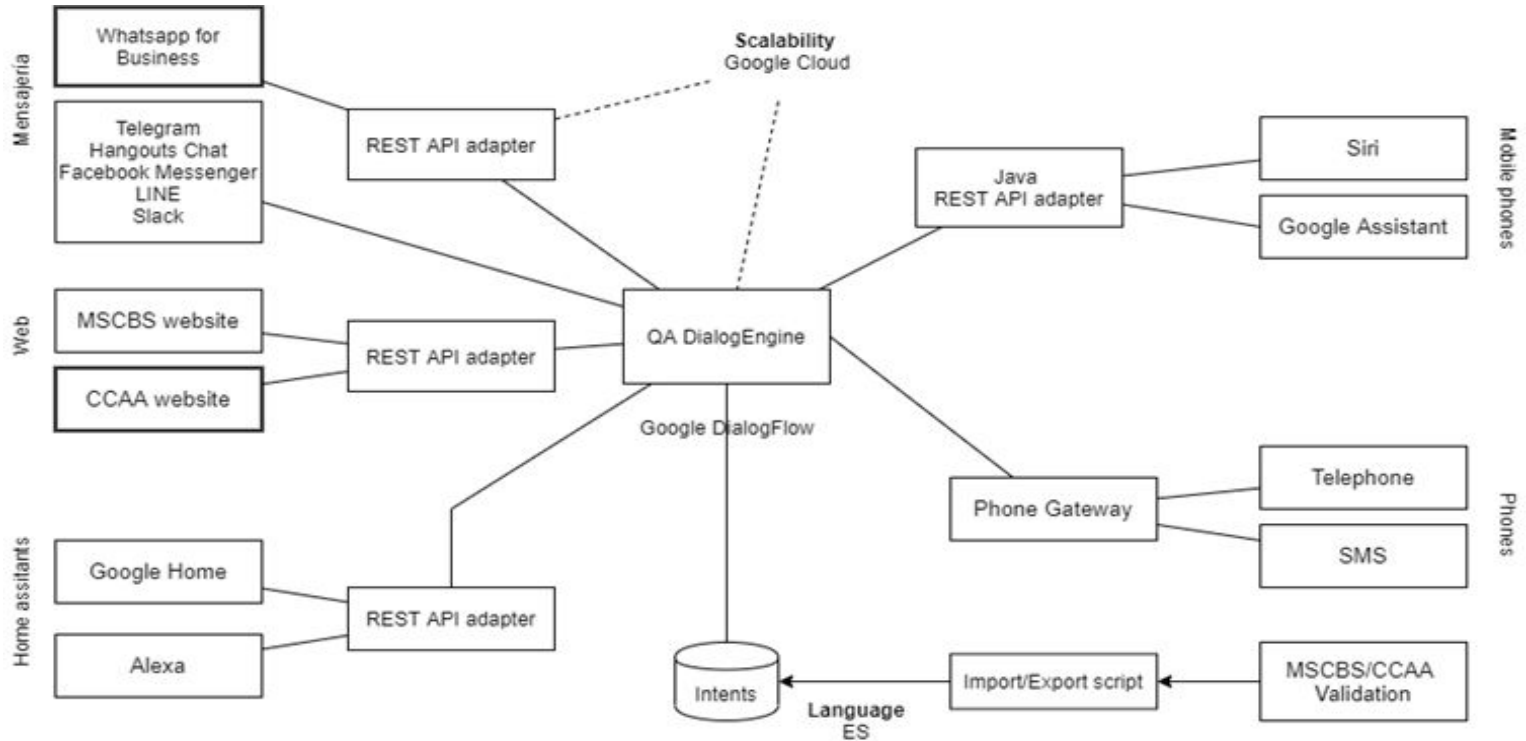
Bon-App-Petit project (2020-2023)





Real use case:

Hispabot-COVID19 (2020)





Research:

GOMINOLA project (2021 - 2024)

PID2020-118112RB-C21 and C22 - Goal-driven multimodal interaction based on microservice orchestration for socio-affective conversational AI

Hybrid approaches to multi-task dialog management

Microservices architecture - Personalized model servitization

5. TEST AND QA

Important KPIs

Volume:

- Interaction rate: % of users who visit website / see robot... and interact with the system.
- Bounce rate: % users who leave without interacting.
- User initiated actions.
- Avg. conversation time.
- User retention: repeated visitors.

Important KPIs

Engagement

Automatic log data:

- #logins, #daily activities, #activities completed, #multimedia elements played, duration of sessions, minutes of use, frequency of sessions,
- user traffic, user retention after a month.

Information analyzed over a period of time:

- Depth of use (number of activities completed)
- Breadth of use (variation in the use of different types of activities, e.g. those who require more involvement vs. passive activities)
- Comparison with intended use.

Effective engagement / efficiency: sufficient engagement to achieve intended incomes.

In some application domains, e.g. mental health, not always more use is better

Important KPIs

Quality:

- Task completion rate
- Improved user experience (qualitative feedback through surveys)

Company:

- Cost savings for customer support
- After-hours support
- Waiting time reduction



Real use case:

Bon-App-Petit project (2020-2023)

Example: Virtual assistant

Virtual assistant for the promotion of healthy habits among school children

It is challenging to define KPIs

E.g. Engagement vs. controlled use of tablets

Collaboration with experts is key for domain-specific KPIs (in our case a collaboration between: Nutrition and Bromatology, Software Engineering, Sport Sciences and a Public Primary School)

E.g. Healthy habits


[Volver](#) **Chat con Atoona** ⓘ



¿Qué tomaste anoche para cenar?

[Volver](#) **Ejercicio 3**

Calcio



El calcio lo puedes encontrar en los lácteos: leche, queso yogur, cuajada,... Es un mineral que ayuda a formar nuestros huesos y dientes.

[Continuar](#) ⓘ

Healthy nutrition habits: adoption vs. knowledge

[Volver](#) **Complementos del personaje**



Animar ⓘ



45

ⓘ No tienes suficientes manzanas para esto



Equipar



15

ⓘ No tienes suficientes manzanas para esto



25

ⓘ No tienes suficientes manzanas para esto

Ejercicio 1

Pan:

| | |
|------------------------------|----------------------|
| Alimento rico en... | Hidratos de carbono |
| que pertenece al grupo de... | Cereales y derivados |

Nueces:

| | |
|------------------------------|-------------------------|
| Alimento rico en... | Selecciona un nutriente |
| que pertenece al grupo de... | Selecciona un grupo |

Uvas:

| | |
|------------------------------|-------------------------|
| Alimento rico en... | Selecciona un nutriente |
| que pertenece al grupo de... | Selecciona un grupo |

Corregir ⓘ

The engineering lifecycle of conversational systems

- Requirement gathering and analysis
- System design and prototyping
- Iterative development
- Deployment
- Test and quality assurance

Coming soon...

“A cute chatbot learning to speak
in the Pyrenees”...

Could only be super-happy!

Continuation in the hands-on lab!





Thank you!