

# NPFL123 Dialogue Systems

## 8. Dialog Authoring Tools

<https://ufal.cz/npfl123>

Ondřej Dušek, Mateusz Lango, Simone Balloccu, Kristýna Klesnilová & **Jan Cuřín**

10. 4. 2024



Charles University  
Faculty of Mathematics and Physics  
Institute of Formal and Applied Linguistics



unless otherwise stated

# Short Intro of Jan Cuřín

## Education

- MFF UK – PhD in 2006 at IFAL (Statistical Machine Translation)
- Johns Hopkins University, Baltimore, MD - 1999 NLP workshop at CLSP (MT: EGYPT -> GIZA)

## Work Experience

- MAMA AI, Telma AI – 2021- Co-Founder, CTO
  - Natural Language Processing, Conversational Agents, AI
- IBM – 2004-2021 – Research Scientist, Manager at IBM Watson R&D Lab
  - Conversational Systems, NLP Technologies
- IFAL, MFF UK – 2002-2004 – Researcher, PhD Student
  - Machine Translation, NLP Technologies
- Schemantix – 2000-2001 – Software Engineer
  - Machine Translation, XML Technologies
- IFAL, MFF UK – 1998-2000 – Researcher, MSc. Student
  - Machine Translation, Parallel corpora for Prague Dependency Treebank (PCEDT)
- Intax – 1996-1998 – Member of translation/localization team
  - Localization of apps and operating systems for IBM Czech Republic



### IBM Watson Services

- Watson Assistant
- Watson Speech (STT/TTS)
- Watson Language Translator

# Business Use Of Dialogue Systems

- Chat bots on web pages
  - Navigation through the content of the web pages (smart search)
  - Frequently asked question (FAQ), Product search
  - Customer support
- Mobile applications with open text input
  - Domain specific apps with chatting functionality, banking apps
  - ChatGPT, ChatGPT+, BING, Copilot, ...
- Assisting systems
  - Intra company “expert” system
  - Data analytics via chat - chat with your data
  - Support for human operators
- Speech based systems
  - Call centers automation – handling top x% of traffic
  - Outbound calls (to inform or collect feedback)
- Automotive applications
  - Search, calls, navigation, infotainment/entertainment (music, POIs)
- Infotainment systems
  - Infotainment systems for hotels, banks’ lobbies, home, etc ...
  - Marketing campaigns



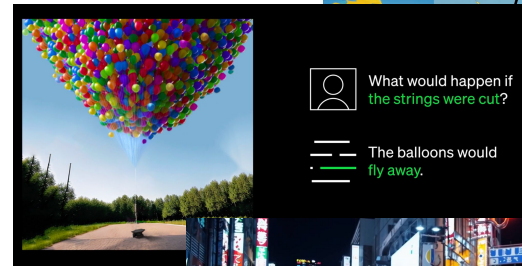
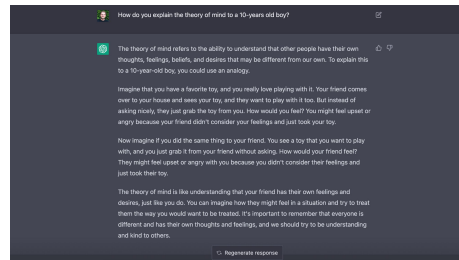
# Challenges of Creating Dialog Systems

- Data

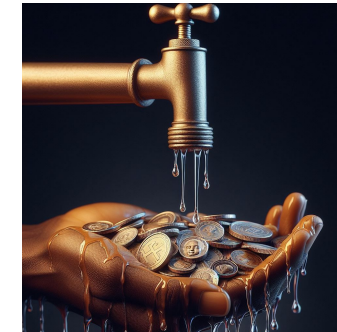
- Use of existing of human-to-human logs
- Use of intracompany structured and unstructured data
- No data, just ideas
- No idea at all

- Scenarios, use-cases

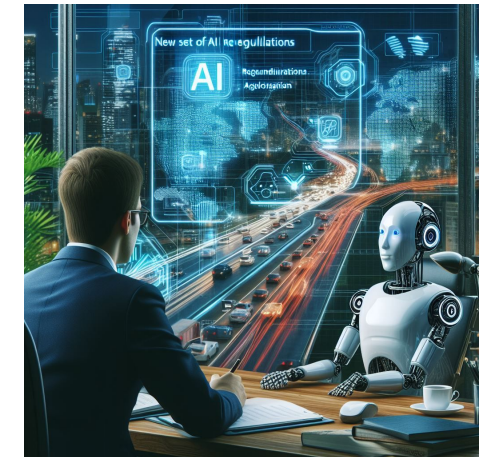
- Single domain
- Single domain with chit-chat capability
- Multi-domain



Generative AI



business



AI Regulation

# Authoring a Dialogue System “old style”

## Restaurant booking scenario

- System: Hello, this is Chez P  p   restaurant reservation system. How may I help you?
- User: I would like to book a table for tomorrow for 5 people
- System: From what time?
- User: From 7pm
- System: OK, I will make a reservation for tomorrow at 7pm, table for 5.
- User: Great!
- System: Looking forward to see you soon.

# Authoring a Dialogue System “old style”

## Restaurant booking scenario

### Intents

#### #reserve\_table

I would like to book a table for 5  
Can I make a reservation?  
I need a reservation for tomorrow

#### #opening\_hours

Until when are you open?  
What are the opening hours?

#### #cancel\_reservation

I made my reservation yesterday,  
I want to cancel it.  
We could not make it today,  
may I cancel the reservation?

### Entities

I need reservation for 5  
people for tomorrow at 7pm

Next Friday from 6pm

For two.

@date

@time

@number

@restaurant\_location

### Dialogue

Welcome  
Book a table

Entity	Context variable	Req.
@date	\$res_date	Y
@time	\$res_time	Y
@number	\$guests	Y

Opening hours  
Cancel reservation  
Yes  
No  
<default answer>

# Intents

- Collection of example how users will trigger the intent
- Usually corresponds to the actions supported by the dialog
- Intent model can be trained even on a small set of examples
- Word and sentence embeddings, stemmer, lemmatizer
- Bigger data collection needed for production system
- Data augmentation by large-language models (LLMs)
- Ordered n-best lists with confidences
- Use of intent n-bests in the dialog – disambiguation

## Intents

### #reserve\_table

I would like to reserve a table for 5  
Can I make a reservation?  
I need a reservation for tomorrow

### #opening\_hours

Until when are you open?  
What are the opening hours?

### #cancel\_reservation

I made my reservation yesterday,  
I want to cancel it.  
We could not make it today,  
may I cancel the reservation?

## ~ Named entities recognition (NER)

- Different type of entities

- Prebuilt (system) entities

- Numbers, dates, time, GEO location, person names, units, currency

- Domain catalogues

- User defined entities

- Gazetteers – fixed list of entities/synonyms
    - Regular expression based
    - Sequence labelling model based on sample annotations (contextual entities)

### Entities

I need reservation for 5  
people for tomorrow at 7pm

Next Friday from 6pm

For two.

@date

@time

@number

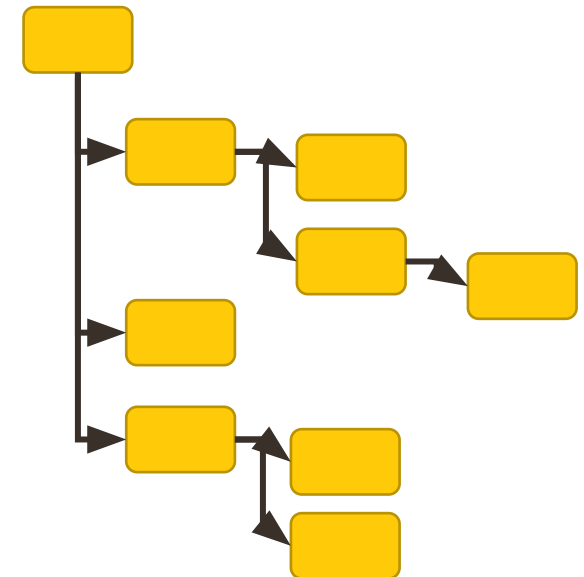
@restaurant\_location



# Dialogue Flow/Tree

- Slot filling style (linear dialog)
  - Set of slots to fill is (required/optional)
  - Able to fill all slots partially or at once
  - Asking just for missing information
  - Ability to customize questions and answer for a particular slot
  - Ability to correct already filled information
  - Tight to user variables
- Dialogues tree (non-linear dialog)
  - Dialogue flow driven by a tree or graph structure
  - Conditions to get to the individual nodes of the tree/graph
  - Fallback strategies (none of the conditions is specified)

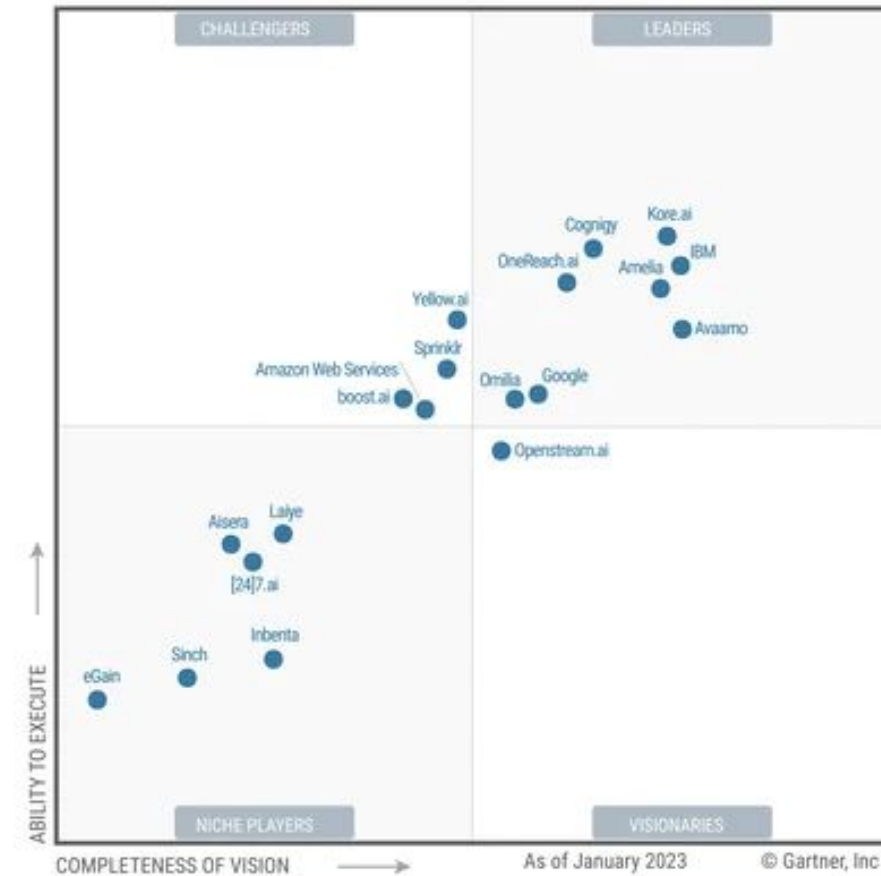
Entity	Context variable	Req.
@date	\$res_date	Y
@time	\$res_time	Y
@number	\$guests	Y



# Gartner Magic Quadrant for Enterprise Conversational AI Platforms 2023

- Evaluation of conversational AI platforms in 2023
- IBM is historically performing well, still among the best in the Completeness of vision

Figure 1. Magic Quadrant for Enterprise Conversational AI Platforms



Source: article in [CXToday](#)

# Authoring tools

- IBM watsonx Assistant
  - <https://www.ibm.com/cloud/watson-assistant/>
  - Video tutorial: <https://console.bluemix.net/docs/services/assistant/tool-overview.html>
- Google Dialog Flow
  - <https://dialogflow.com/>
  - Video tutorials: <https://cloud.google.com/dialogflow/docs/video>
- Amazon Alexa Skills
  - <https://developer.amazon.com/alexa-skills-kit>
  - Video tutorial: <https://www.alphavoice.io/video/alexa-developers/alexa-skills-kit-developer-console-build>
- Microsoft Cortana Skills
  - <https://developer.microsoft.com/en-us/cortana>
- Apple SiriKit (Siri-enabled iOS apps)
  - <https://developer.apple.com/sirikit/>

- Sample chatbot in IBM watsonx Assistant

Restaurant booking scenario

<http://www.bienvenuechezpepe.com/>

# Features/Concepts

- Dialogue context / history
  - Condition on context variables collected in previous turns
  - Reference/anaphora resolution using collected variables
- Fallback strategies / Digression
  - Allow "jumping" to different topic for a while and then return back
- Disambiguation support
  - Similar confidence of multiple choices – ask user to select
- Calling external APIs
  - Webhooks/Cloud functions ...

~ Use of LLMs for “natural” response generation and chitchat

# Maintaining and improving chatbot in production

- Automatically
  - Learning from user selections
  - Statistics on user selections – automated "pre-selection" for next users
- Semi-automatically or manually
  - Chat log analysis → model update
  - Used Measures:
    - **Coverage** ... rate at which your chatbot is confident that it can address the user's request (per dialogue turn)
    - **Containment** ... rate at which your chatbot can satisfy a user's request without human intervention, i.e. connect to human agent not requested (per conversation)
- Metrics
  - Top 22 Metrics for Chatbot Analytics in 2024:  
<https://research.aimultiple.com/chatbot-analytics/>

# Authoring a Dialogue System “new style”

Welcome to MAMA Hotel, a fictional hotel of TELMA AI!

Because we build bespoke state-of-the-art AI solutions, we also demonstrate our FAQ Chatbot on a website of a fictional 5-star hotel. Where else would our client be taken care of by a top-notch AI than in a MAMA (AI) hotel?

On this page you can try out the functionality of our unique chatbot, which only needs the data presented below and it will be able to answer customers' questions in a naturally-looking language. You can, for example, ask what happens if you arrive at the hotel after midnight and will need to check-in.

You can find the list of the data the chatbot works with down below. You can ask our FAQ Chatbot questions in any other language. Try it for yourself.

Show the data

**Telma<sup>AI</sup>**

MAMA Hotel

Hello and welcome to the website of Hotel MAMA, I am virtual assistant Telma.

How may I help you?

I will arrive after midnight, is it still possible to check in?

Yes, it is possible to check in at Hotel MAMA after midnight as the reception services are available 24 hours a day.

Start typing...



# Authoring a Dialogue System “new style”

The screenshot shows a website for the Czech Army recruitment portal. The header includes the Czech coat of arms, the text 'Vaše cesta do armády', and navigation links: 'Koho hledáme', 'Hry', 'Příprava', 'O armádě', and 'Kontakty'. A search icon and a user profile icon are also present. The main content area features the hashtag '#ZaKariérou' and the large text 'DO ARMÁDY'. Below this, a paragraph encourages users to choose their path and explore offered positions. Another paragraph mentions the possibility of starting a career through a scholarship program. A 'Přidejte se' button is located at the bottom left. On the right side, a chatbot window for 'Adéla' is open, displaying a welcome message and a list of suggested topics for interaction.

**Vaše cesta do armády**

Koho hledáme Hry Příprava O armádě Kontakty

#ZaKariérou

# DO ARMÁDY

Staňte se vojákem! Vyberte si svou cestu a dokažte sobě i ostatním, že vy na to máte. Zde se nejen dozvíte vše potřebné, ale rovnou se můžete podívat na **příklady nabízených pozic**.

Vaše cesta do armády může začít už na střední škole využitím **stipendijního programu Ministerstva obrany**.  
Dejte své kariéře správný směr již dnes.

**Přidejte se**

**Adéla**

Dobrý den, jsem virtuální náborářka Adéla. V naší interakci po vás nebudu chtít žádné osobní údaje ani mi takové informace o sobě prosím nesdělujte. Na své nové pozici se stále učím, buďte ke mě prosím shovívaví.

Co by Vás zajímalo ohledně náboru do Armády České republiky?

**Na co se můžu zeptat**

**Kde začít**

**Volné pozice**

Můžete psát...



# Semantic Search, RAG

- Semantic Search (mSearch)

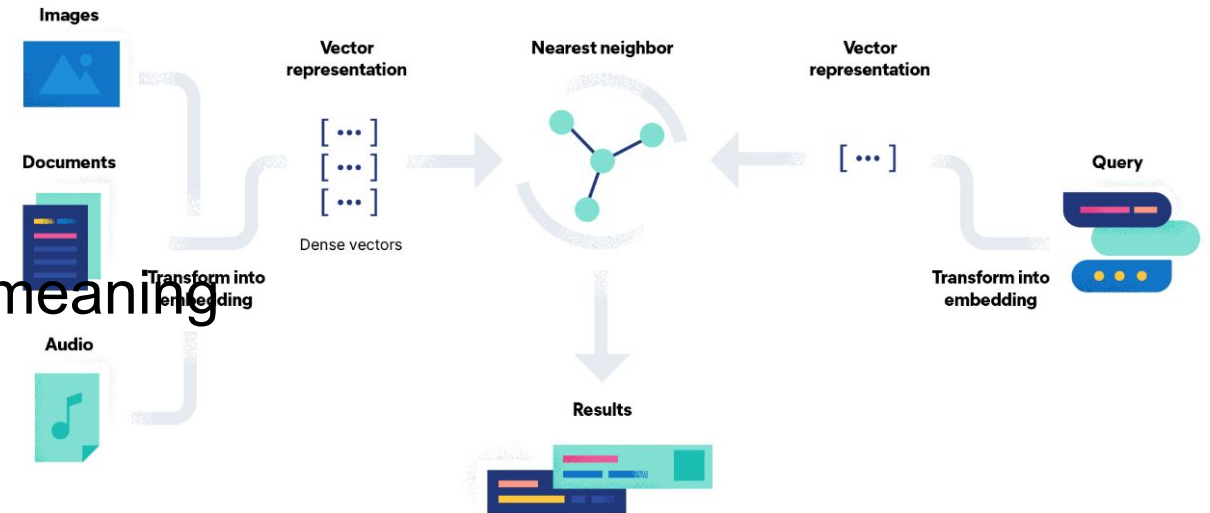
## Semantic search

~ search technology that interprets the meaning of words and phrases

- RAG (mDataChat)

## Retrieval-Augmented Generation

~ process of optimizing the output of a large language model, so it references an authoritative knowledge base outside of its training data sources before generating a response



Source: Elastic <https://www.elastic.co/what-is/semantic-search>

**AI WORKS FOR YOU:**  
**mDataChat**  
your unique AI buddy

- ✓ my data is safe in my data center
- ✓ I use natural language to search for information in documents and databases
- ✓ I get a response in the language of my preference, no matter what language my information is in

90% OF DATA IS UNSTRUCTURED  
ONLY 18% OF COMPANIES CAN USE IT  
JOIN THEM!

# Authoring tools by Mama AI/Telma AI

- Mama Telma AI tooling for outbound calls

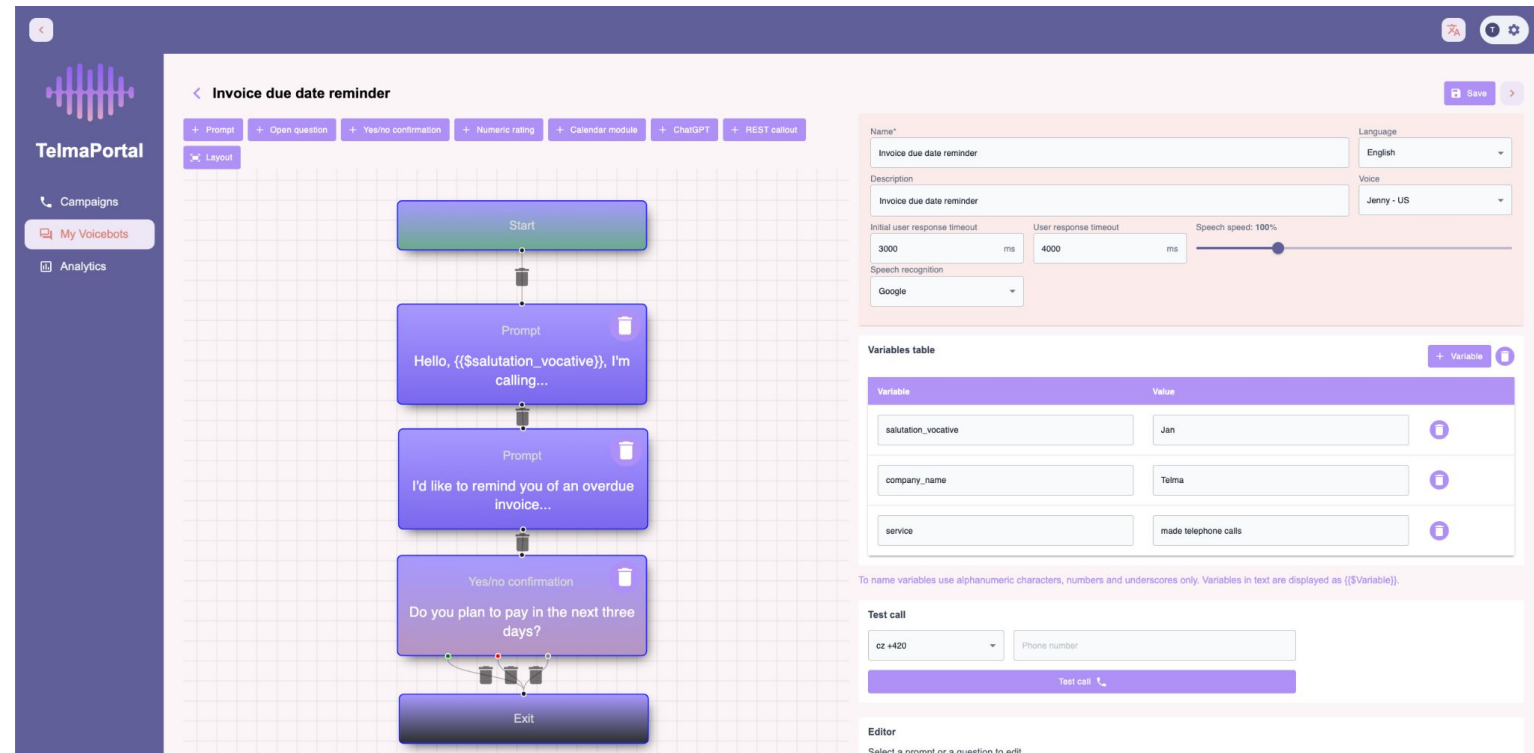
- No-code/low-code
- Modularization
  - Yes/No/Don't know
  - Open question
  - Calendar
- Webhook, SMS integration
- GPT module, semantic search

Examples: <https://telma.ai/products/outbound>

- Voice lounge concept

- Python implementation
- Modules
  - (longer) Number dictation
  - Address dictation (RÚIAN)
- Guess animal game on Alexa (see [youtube](https://www.youtube.com/watch?v=...))

Examples: <https://telma.ai/products/inbound>



# Thank you for you attention



Jan Cuřín <jan.curin@themama.ai>

<https://themama.ai>

<https://telma.ai>



 Jan Cuřín