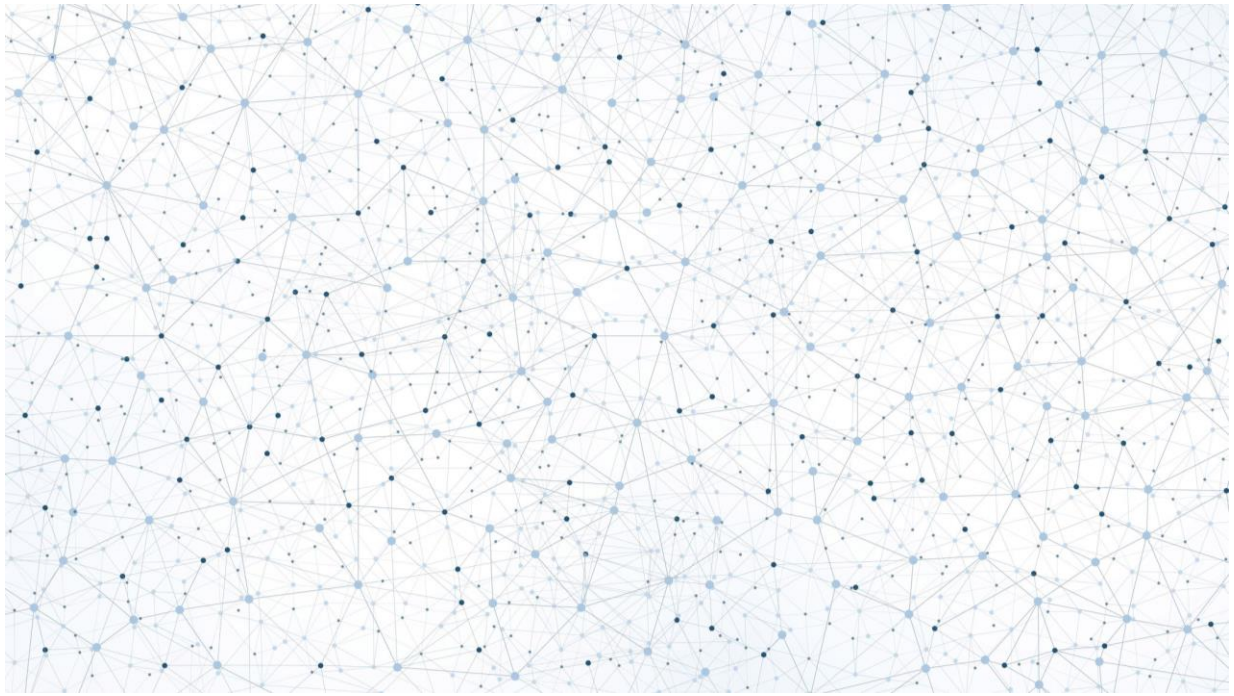
Sarunboti un RAG Tehnoloģijas Biznesā: Iespējas un Stratēģijas

2024

Valdis Saulespurens



RBS
RIGA BUSINESS SCHOOL
Riga Technical University

About Me

Valdis Saulespurens

30+ Years programming

Run retail and e-commerce store in 90s California

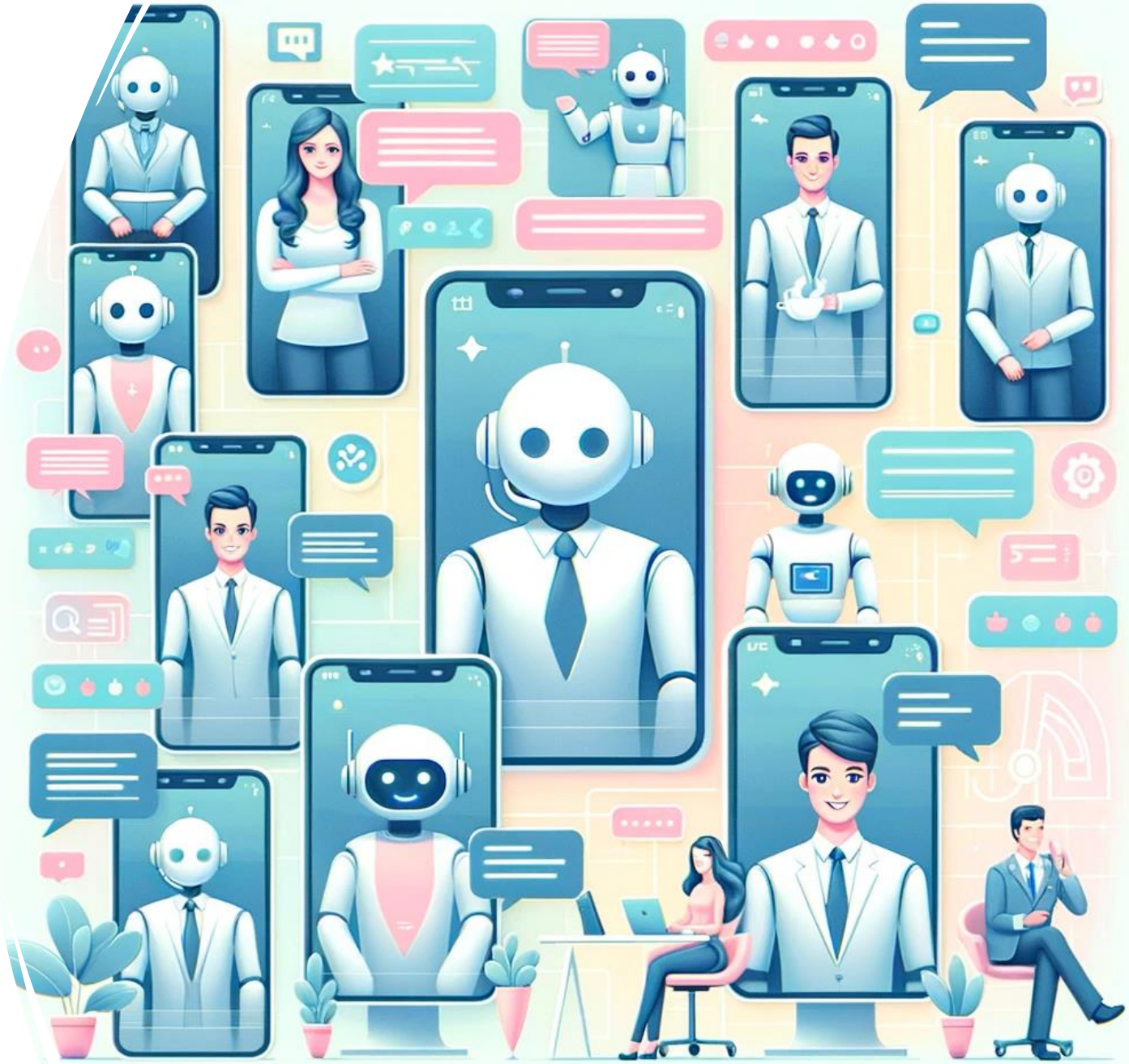
Lector RTU RBS

Reasearcher National Library of Latvia

valdis.saulespurens@rtu.lv

Introduction to Chatbots - Sarunboti

- Definition: Software applications designed to simulate human conversation.
- Interaction Modes: Text-based (e.g., messaging apps) and voice-based (e.g., virtual assistants).
- Core Technologies: Natural Language Processing (NLP), Machine Learning (ML), and Artificial Intelligence (AI) in general.



1960s-1970s: The Early Experiments

- **ELIZA (1966)** : Created by Joseph Weizenbaum at MIT, ELIZA was one of the first chatbots. It used pattern matching and substitution methodology to simulate conversation, most famously in the form of a Rogerian psychotherapist. Despite its simplicity, ELIZA demonstrated that machines could potentially interact with humans in meaningful ways.
- You: I like hockey.
- ELIZA: How do you feel about liking hockey?
- Try: https://web.njit.edu/~ronkowit/eliza.html
- Also Parry(1972) – ELIZA «with attitude» - paranoid

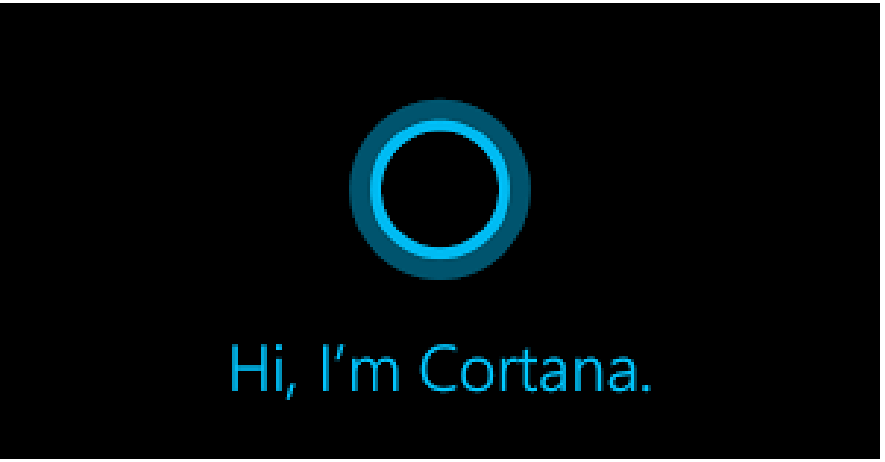
1980s: expert knowledge based systems

- **XCON** (also known as R1)(late 1970s-1980s) : A pioneering expert system developed by Digital Equipment Corporation in the 1980s to automate the configuration of VAX computer systems, significantly improving efficiency and accuracy in the assembly process. – not really a chatbot
- **Racter (1984)** : An artificial intelligence program capable of generating text, Racter was used in various writing projects. However, its actual intelligence and coherence were limited, often producing nonsensical or surreal output.

1990s-2000s: Commercial Use and Internet

- **Dr. Sbaitso (1992)** : Developed by Creative Labs, Dr. Sbaitso was a chatbot that ran on MS-DOS. It was designed as a digital psychologist, offering conversational interaction through text-to-speech technology.
- **ALICE (1995)** : Created by Richard Wallace, ALICE (Artificial Linguistic Internet Computer Entity) was based on AIML (Artificial Intelligence Markup Language). It became one of the most famous early chatbots, winning several Loebner Prize Turing Test contests.
- **SmarterChild (2001)** : Integrated into AOL Instant Messenger and MSN Messenger, SmarterChild provided users with information and entertainment, becoming an early example of a commercial chatbot with mass appeal.
- **Jabberwacky (1997-2003)** : Created by Rollo Carpenter, Jabberwacky aimed to simulate natural human chat in an entertaining manner. It learned from interactions with users, improving its responses over time.

Src: https://botwiki.org/bot/smarterchild/

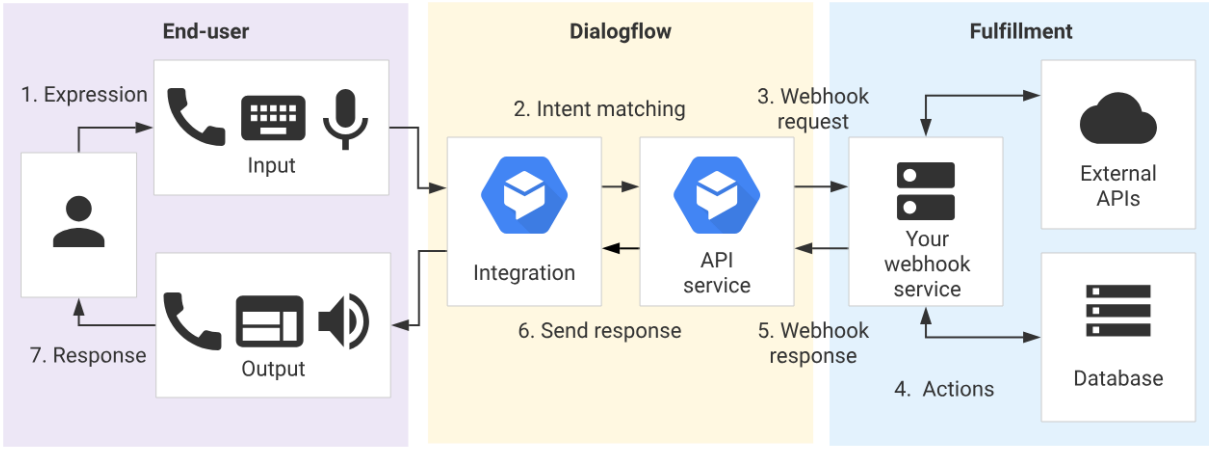


2010s: The AI Revolution

- **Siri (2011)** : Apple's Siri marked the beginning of virtual assistants integrated into smartphones. Using voice recognition and natural language processing, Siri could perform tasks, provide information, and interact with users conversationally.
- **Google Now (2012)** : Google's virtual assistant introduced predictive capabilities, providing users with information based on their habits and preferences.
- **Cortana (2014)** : Microsoft's entry into the virtual assistant space, Cortana, was designed to integrate with Windows and provide a wide range of services through voice and text interactions.
- **Alexa (2014)** : Amazon's Alexa, integrated into Echo devices, popularized the use of voice-activated assistants in the home, capable of controlling smart devices and providing a vast array of services.

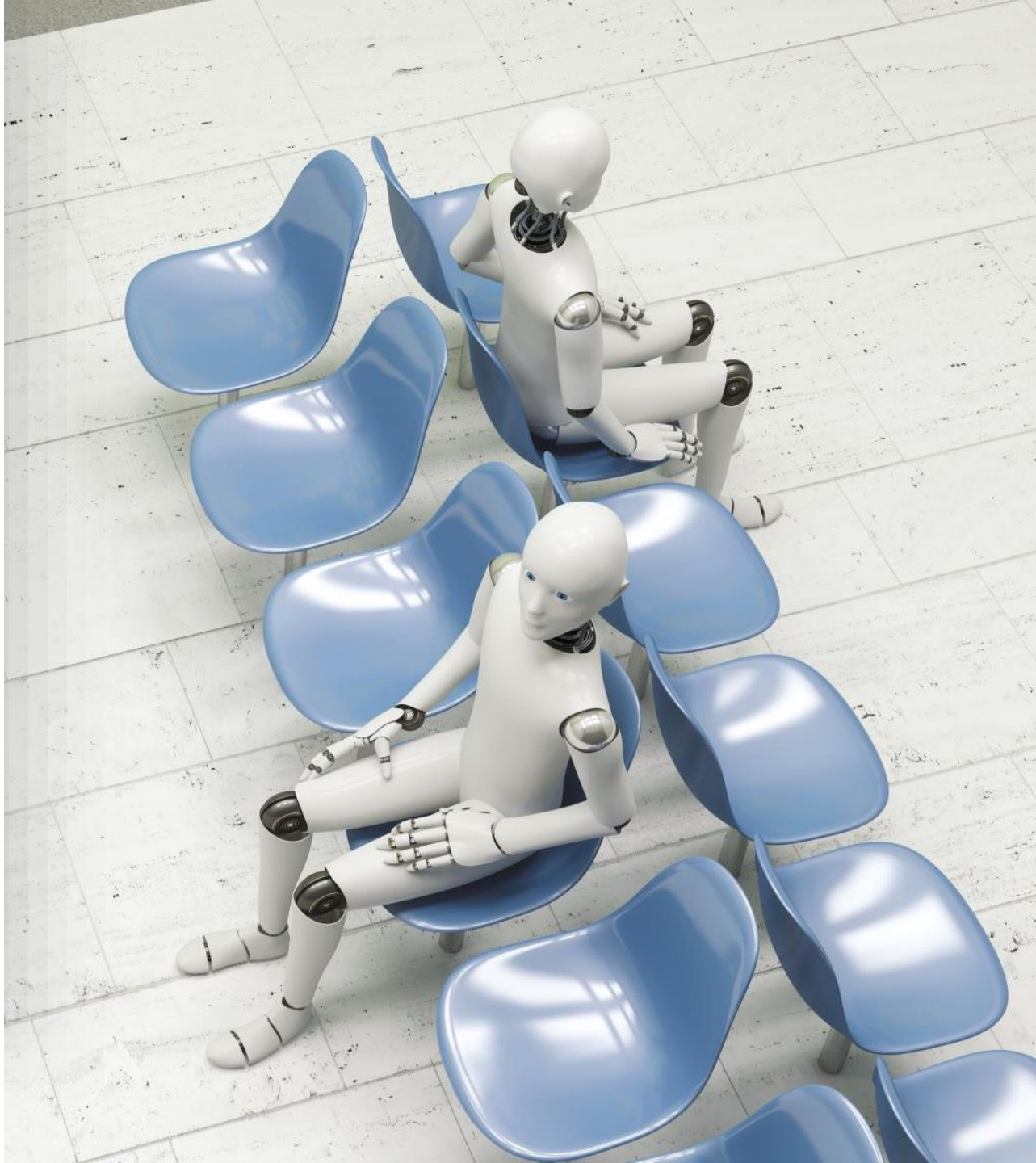
Late 2010s to 2020s: Advanced AI and Ubiquity

- **Chatbots in Messaging Apps** : Platforms like Facebook Messenger, WhatsApp, and WeChat integrated chatbots for customer service, e-commerce, and entertainment, making them widely accessible.
- **Conversational AI Platforms** : Companies like IBM Watson, Google Dialogflow, and Microsoft Bot Framework offered robust tools for developing sophisticated chatbots capable of handling complex interactions and integrating with various services.
- **GPT-3 (2020)** : OpenAI's GPT-3, a state-of-the-art language model, brought significant advancements in natural language understanding and generation. It demonstrated the potential for creating highly coherent and contextually relevant conversational agents.



Present Day: Ubiquitous and Sophisticated Chatbots

- **Healthcare** : Chatbots are used for preliminary diagnosis, mental health support, and patient engagement.
- **Customer Service** : Businesses use chatbots to handle customer inquiries, streamline operations, and provide 24/7 support.
- **Education** : Chatbots assist in tutoring, language learning, and administrative tasks.
- **Entertainment** : Chatbots are part of gaming experiences, interactive storytelling, and social platforms.



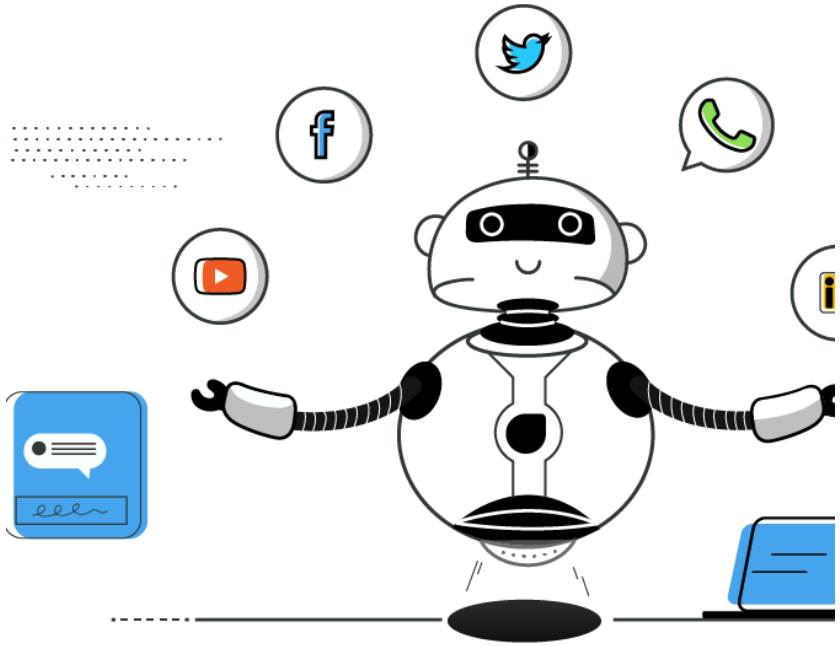
Chatbots in Customer Service

- **24/7 Availability** : Provides round-the-clock support for customer inquiries.
- Ensures customers can get assistance anytime, reducing wait times.
- Improves customer satisfaction and loyalty.
- **Instant Responses** : Delivers quick answers to frequently asked questions.
- Handles common issues and queries without human intervention.
- Frees up human agents to focus on more complex problems.
- **Reduces Operational Costs** : Minimizes the need for a large customer service team.
- **Efficient Query Management** : Manages large volumes of inquiries simultaneously.

Img src: ScienceSoft

Chatbots in Marketing and Sales

- **Engagement with Website Visitors** : Interacts with visitors in real-time to capture their interest.
- **Automated Lead Qualification** : Uses predefined criteria to assess lead quality.
- **Tailored Recommendations** : Analyzes customer data to provide personalized product suggestions.
- **Dynamic Content Delivery** : Delivers personalized content and offers based on user preferences.
- **Guided Selling** : Assists customers through the buying process with interactive guidance.
- **Automated Follow-Ups** : Sends follow-up messages to nurture leads and encourage conversions.
- **Feedback Collection** : Gathers customer feedback post-purchase.



Src: freshworks.com

Chatbots in Healthcare

- **24/7 Availability** : Provides round-the-clock support for patients.
- **Symptom Checking and Advice** : Guides patients through symptom checkers to provide preliminary advice.
- **Medication Reminders** : Sends reminders to patients to take their medications.
- **Educational Resources** : Offers access to educational materials and resources.
- **Virtual Triage** : Directs patients to the appropriate level of care.
- **Appointment Scheduling** : Automates the process of booking appointments.
- **Follow-Up Care** : Coordinates follow-up appointments and care plans.

Src: Washington Post

Chatbots in Education

- **Personalized Tutoring** : Provides customized learning experiences for students.
- **Interactive Lessons** : Engages students with interactive and multimedia content.
- **Language Practice** : Facilitates language learning through conversational practice.
- **24/7 Availability** : Offers round-the-clock access to learning resources and support.
- **Information Retrieval** : Provides quick access to information about courses, schedules, and campus resources.
- **Feedback Collection** : Gathers student feedback on courses, instructors, and campus services.

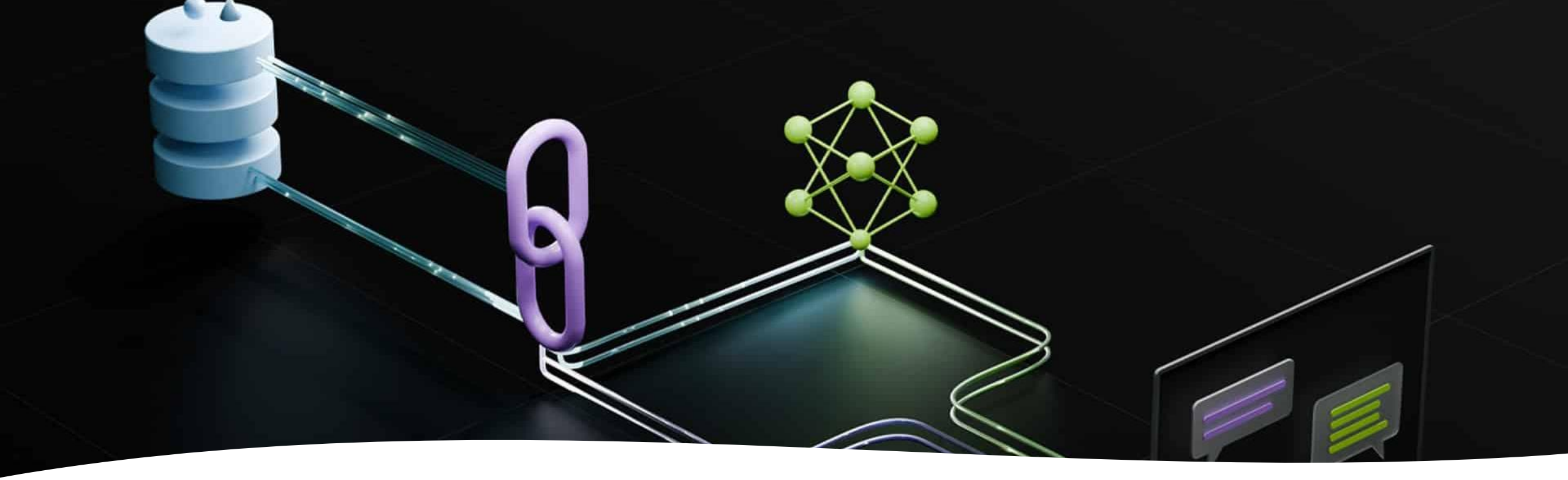
Implementing Chatbots in Your Business

- **Define Clear Objectives** : Identify specific goals for your chatbot.
- **Understand Your Audience** : Tailor the chatbot experience to your target users.
- **User-Friendly Design** : Ensure an intuitive and accessible chatbot interface.
- **Start Small and Scale** : Implement the chatbot in phases.
- **Integration with Existing Systems** : Connect the chatbot with your current CRM, ERP, and other systems.
- **Feature Set** : Evaluate the platform’s capabilities.
- **Customization and Flexibility** : Ensure the platform allows for customization.

What are the limitations of chat bots ?

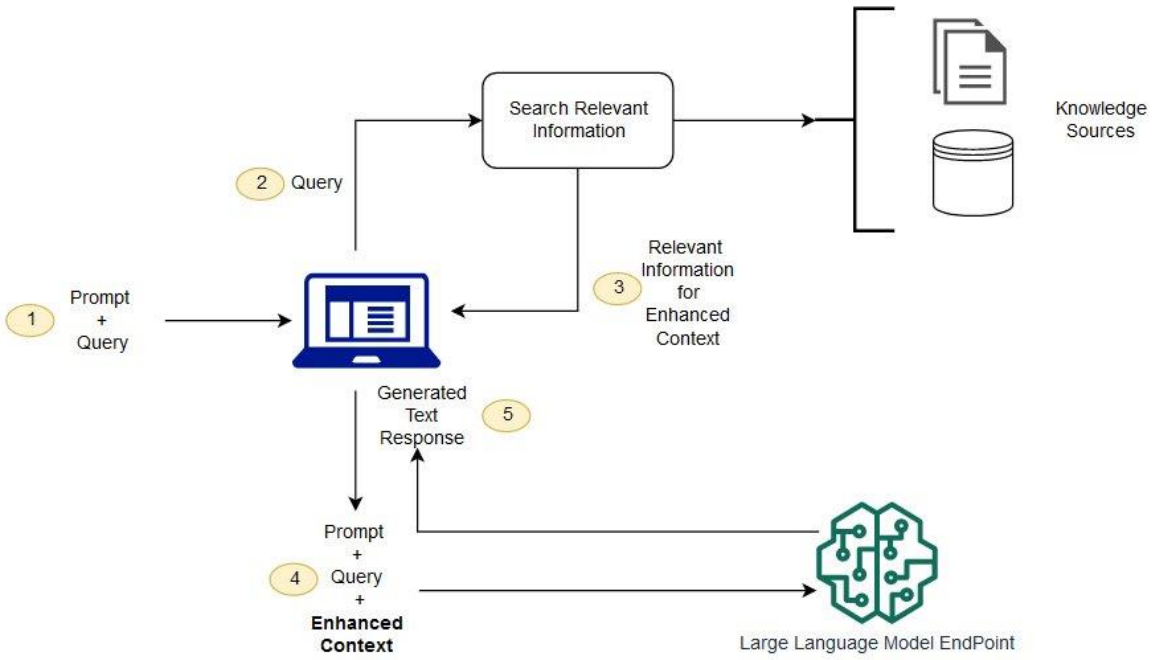
Challenges and Limitations of Chatbots

- **Limited Understanding and Context** : Chatbots may struggle with complex queries or contextual understanding.
- **Handling Ambiguity and Nuance** : Difficulty in interpreting ambiguous or nuanced language.
- **Integration with Legacy Systems** : Challenges in integrating chatbots with existing IT infrastructure.
- **User Frustration** : Ineffective chatbots can lead to user frustration and dissatisfaction.
- **Compliance with Regulations** : Adhere to data privacy regulations such as GDPR, HIPAA, etc.



What is RAG?

- **Definition** : A hybrid approach combining retrieval-based and generative models.
- **Core Components** : Utilizes a retriever model to fetch relevant documents and a generator model to produce coherent answers.
- **Improved Relevance** : Provides more accurate and contextually appropriate responses.
- **Handling Complex Queries** : Capable of addressing more intricate and detailed questions.
- **User Experience** : Enhances the overall user experience with more informative and engaging interactions.



How RAG works

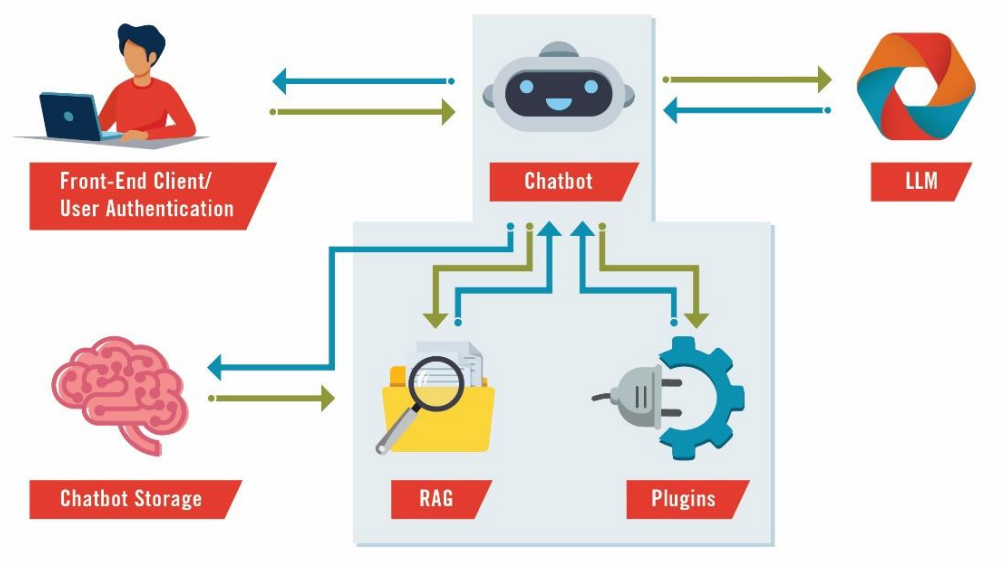
Retriever Model : Fetches relevant documents or information from a large dataset.

Generator Model : Produces human-like responses based on the retrieved documents.

Integration : Combines outputs from both models to deliver final responses.

Implementing RAG in Chatbot Systems

- **Retriever Model** : Dense Passage Retrieval (DPR), ElasticSearch, etc.
- **Generator Model** : GPT-3,3.5,4,, BERT, T5, Gemini, OpenLLAMA, Claude,etc.
- **Integration Frameworks** : APIs, Middleware, and SDKs.
- **Data Preparation** : Compile and preprocess the dataset for retrieval.
- **Model Training** : Train retriever and generator models.
- **Integration** : Implement models within the chatbot framework.
- **Testing and Optimization** : Continuously test and refine the system.



Resources to play around:

- https://chat.lmsys.org/ - explore various Large Language Models
- https://openrouter.ai/ - similar but offers control of which LLM model to test – costs about same as direct access
- https://chatgpt.com/ - The leading LLM, Latvian Language not so great, paid version is markedly better than free
- https://gemini.google.com/ - playing catchup
- https://www.heygen.com/ - VIDEO PRODUCER AI – expensive
- https://suno.com/ - Make a Song about Anything

Conclusion: humans first ☺

Questions

Thank You!

https://github.com/ValRCS/AI_Studio_24