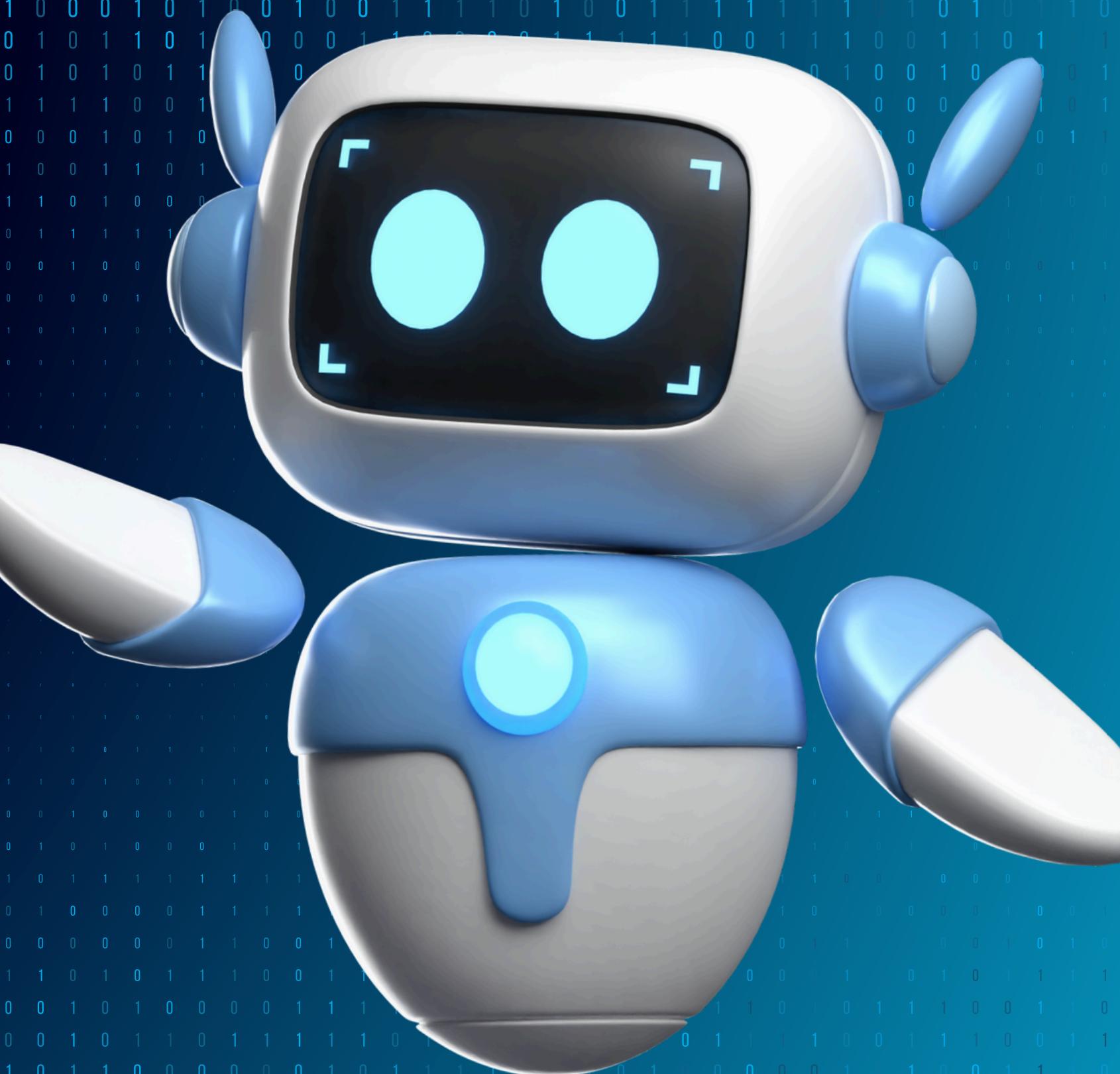
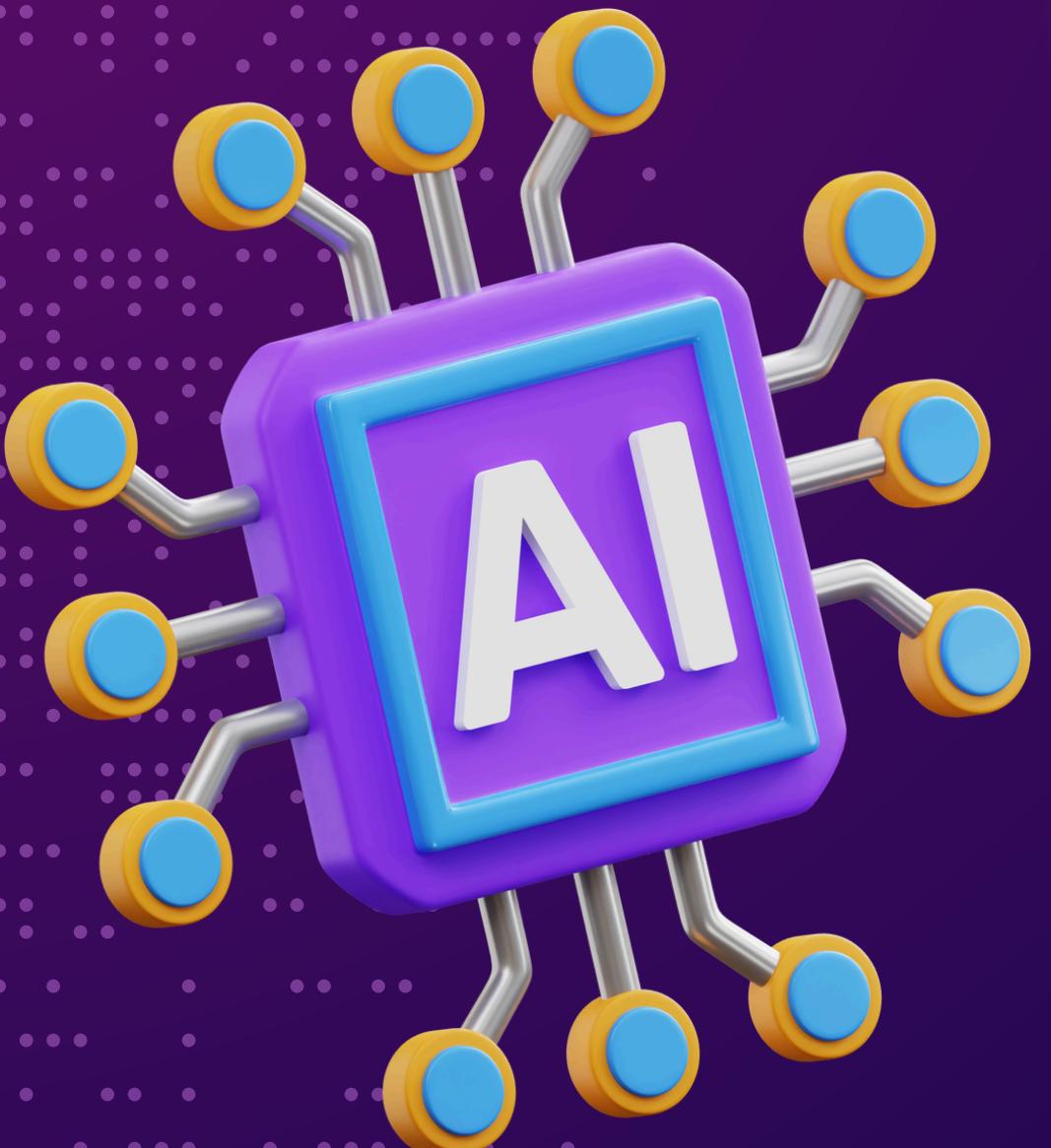


# EMPLOYEE ONBOARDING CHATBOT

Group 01





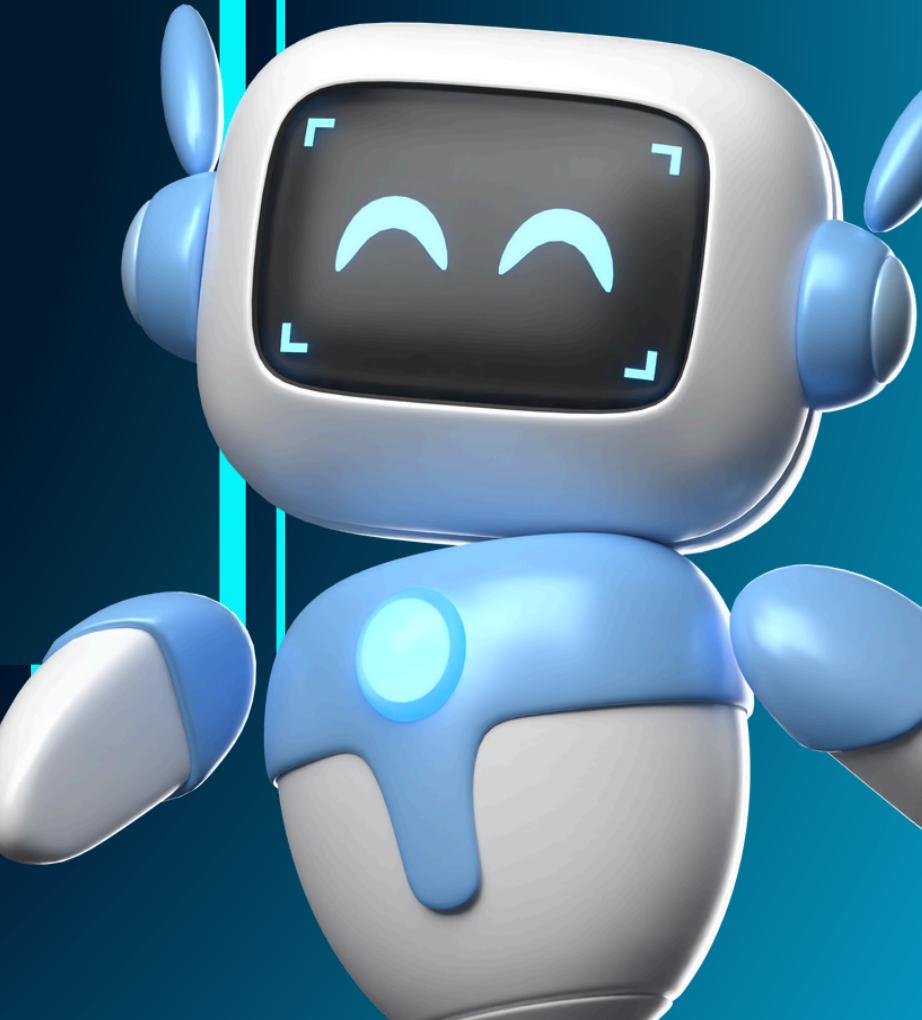
# CONTENTS

- ① Introduce
- ② Features
- ③ System Design
- ④ Workflow



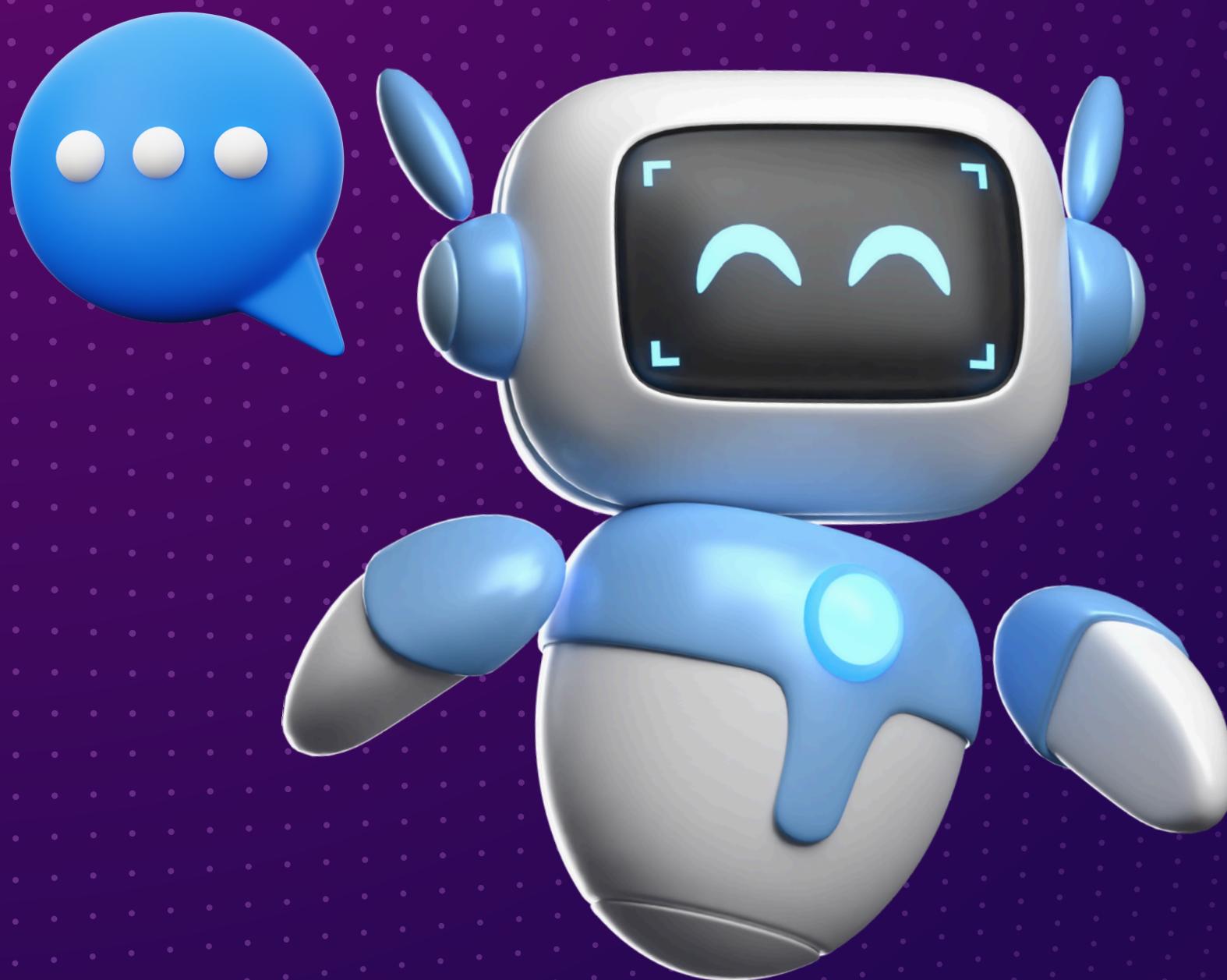
# WHAT IS THIS?

A Chatbot system that supports the onboarding process for new employees, helping them quickly become familiar with the company and understand its policies and regulations.



2

## OBJECTIVE



AUTOMATE

STANDARDIZE

PERSONALIZE

②

## FEATURES

### AI-POWERED ONBOARDING CHATBOT

Answers questions about company policies, procedures, benefits, etc.

### ENHANCED SPEAKING SKILLS

Suggests relevant documents and courses based on the employee's job role.

②

## FEATURES

### AUTOMATED CONTENT GENERATION

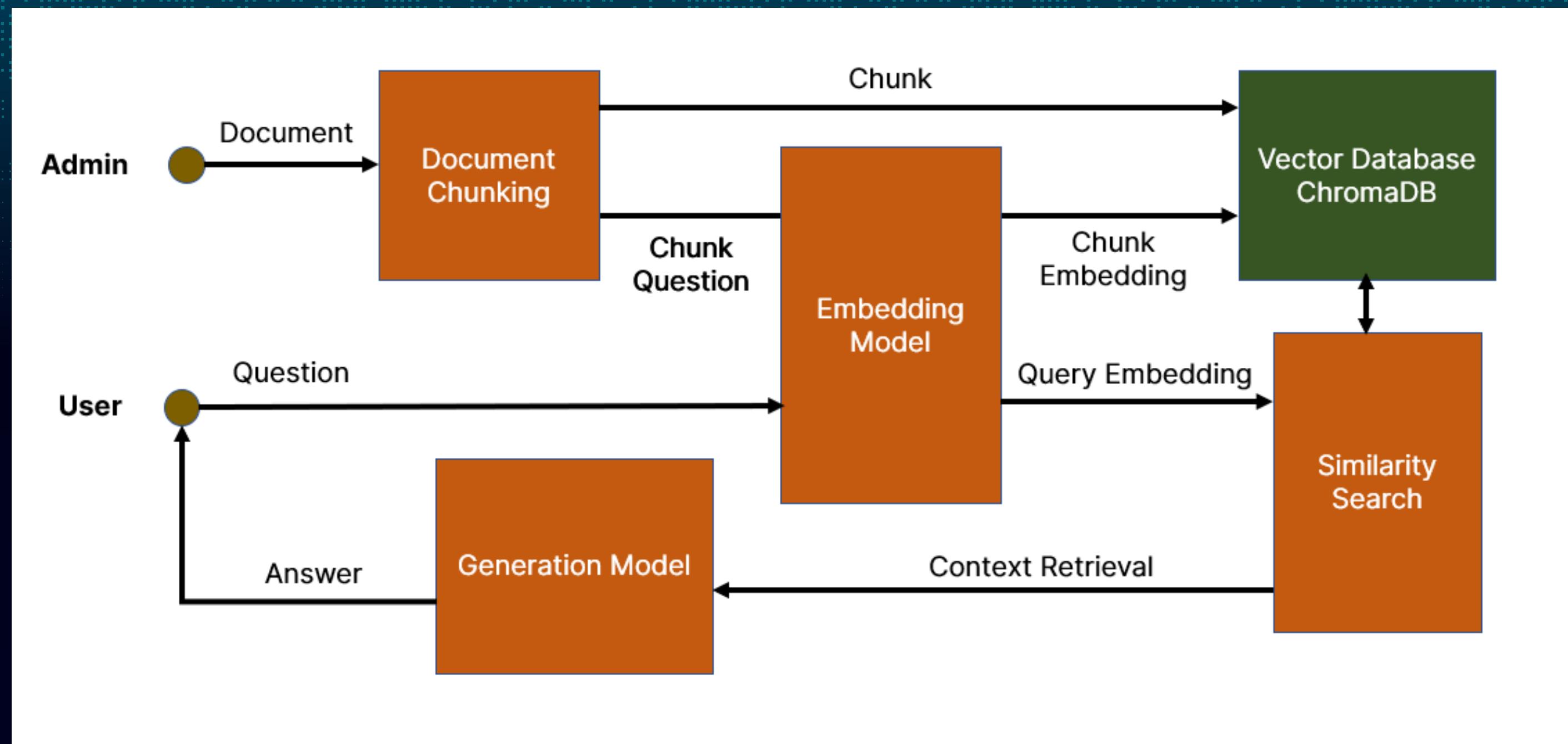
Drafts welcome emails,  
summarizes documents, and  
generates training questions.

### AUTOMATIC INFORMATION EXTRACTION

Auto-fills forms using the new  
employee's CV and personal  
documents.

# 3

# SYSTEM DESIGN



CHATBOT MODULE

4

## WORKFLOW

**1. COLLECT DOCUMENT,  
CHUNKING**

Splitting the document into segments

**2. GENERATE QUESTION  
FROM CHUNK TEXT**

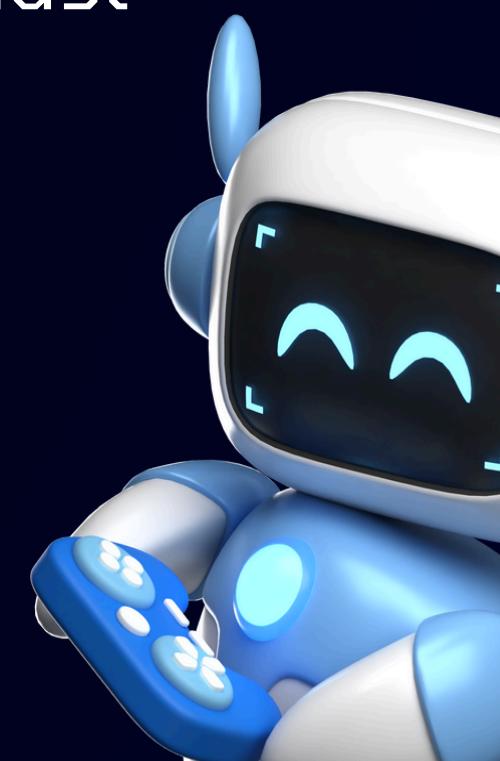
Generate FAQ with OpenAI.  
Enhancing retrieval-augmented generation (RAG)

**3. EMBEDDING & INDEXING  
(VECTORIZE)**

Convert questions and chunked texts into numeric vectors, supports fast similarity search

**4. PROMPT LLM WITH  
RELEVANT CONTEXT**

Use the most relevant document chunks to generate a precise answer



## PREPROCESSING

Admins manage and add documents to the chatbot's knowledge base. When a new document is added, it is reformatted, content is extracted, and then split into chunks.

Each chunk includes the document title and heading as context. Only plain text and tables from DOCX files are processed. Chunks are embedded and stored in a database (both raw text and vector form).

## RETRIEVAL

When the system receives a user question, it combines the question with chat history and uses an LLM to rewrite it for better context alignment and to remove inappropriate content. The rewritten question is then embedded and sent to the search module.

During search from Databasse, combining text matching and semantic search to find the most relevant text chunks.

**DEMO**