ÇANKAYA UNIVERSITY
FACULTY OF ENGINEERING
COMPUTER ENGINEERING DEPARTMENT

Test Plan, Test Design Specifications and Test Cases

Version 1

CENG 408

Innovative System Design and Development II

BelsisMIS-Intelligent-Customer-Support-Assistant-BICSA

Ali Emrecan Selvili 202111031
Ahmet Selçuk Özdil 202011040
Ege Beçin 202011067
Yusuf Tuna Üner 202011077
Mehmet Efe Kaya 202111024

Advisor: **Prof. Dr. Hayri SEVER**

BelsisMIS-Intelligent-Customer-Support-Assistant-BICSA

Table of Contents

İçindekiler Tablosu

BelsisMIS-Intelligent-Customer-Support-Assistant-BICSA

1. INTRODUCTION

1.1 Version Control

Version No	Description of Changes	Date
1.0	Initial version of the test plan.	March 24, 2025

1.2 Overview

This document outlines the testing strategy for the AI-powered chatbot developed using the OpenAI Assistant model. The chatbot is designed to respond to user prompts by referencing a set of HTML-based end-user documents, which serve as its primary knowledge base. The test plan aims to ensure that the chatbot accurately retrieves and presents information from this knowledge base in a coherent and contextually appropriate manner.

1.3 Scope

This test plan applies to all components of the chatbot that interact with the knowledge base [1], including:

- Prompt understanding
- Knowledge base document parsing
- Response generation
- Context retention within a session
- Response traceability to HTML content

It includes both manual and automated test cases, test design specifications, and validation strategies to ensure the chatbot performs correctly under various scenarios.

1.4 Terminology

Acronym	Definition
KB	Knowledge Base
HTML	HyperText Markup Language
TC	Test Case
AI	Artificial Intelligence
NL	Natural Language

2. FEATURES TO BE TESTED

2.1 Prompt Understanding (PU)

The chatbot should be able to parse and understand the intent of user input using natural language processing (NLP) techniques.

2.2 Knowledge Retrieval (KR)

The chatbot should extract relevant information from the uploaded HTML documents (knowledge base) based on the user's query.

2.3 Response Generation (RG)

The chatbot should provide human-like, coherent, and accurate responses grounded in the knowledge base content.

2.4 Source Reference Inclusion (SR)

The chatbot should reference specific sections of the knowledge base (e.g., using anchor tags or headings) to justify its response.

2.5 Session Context Management (SCM)

The chatbot should maintain context within a session and provide follow-up answers based on previous interactions.

3. FEATURES NOT TO BE TESTED

- External API integrations (not part of current implementation)
- User Login (not part of current implementation)
- Multilingual support (limited to Turkish for this phase)

4. ITEM PASS/FAIL CRITERIA

A test passes if the chatbot provides an accurate, relevant, and traceable response based on the knowledge base content.

A test fails if:

- The chatbot gives irrelevant or hallucinated content
- No citation/reference to HTML document is included
- The prompt is misunderstood

4.1 Exit Criteria

- 100% of the test cases executed
- At least 95% test case pass rate
- All high and medium priority test cases must pass

5. REFERENCES

[1] Ali Emrecan Selvili, Ahmet Selçuk Özdil, Ege Beçin, Yusuf Tuna Üner, Mehmet Efe Kaya, *Software Requirements Specification (SRS) for Level-0 Support Chatbot for BelsisMIS*, Çankaya University, March 2025.

6. TEST DESIGN SPECIFICATIONS

6.1 Prompt Understanding (PU)

6.1.1 Subfeatures to be tested

- **PU.INT** – Intent Detection: The chatbot correctly identifies the user's goal or information need.
- **PU.ENT** – Entity Recognition: The chatbot detects and isolates important entities within the input prompt.

6.1.2 Test Cases

Here list all the related test cases for this feature:

TC ID	Requirements	Priority	Scenario Description
PU.INT.01	3.1.4.1 – User Inquiry Resolution [1]	H	User asks a question directly based on a heading in the KB
PU.INT.02	3.1.4.1 – User Inquiry Resolution [1]	H	User asks an abstract or vague question to see if chatbot guesses intent
PU.ENT.01	3.1.4.3 – AI-Based Text Response [1]	M	User includes multiple entities; chatbot should resolve them correctly

6.2 Knowledge Retrieval (KR)

6.2.1 Subfeatures to be tested

KR.HDR – Heading-Based Retrieval: Retrieve based on HTML headings
KR.SEQ – Sequential Scanning: Scan full document for matching content beyond headers.

6.2.2 Test Cases

Here list all the related test cases for this feature:

TC ID	Requirements	Priority	Scenario Description
KR.HDR.01	3.1.4.5 – Step-by-Step Guidance [1]	H	Ask a question that maps directly to an HTML heading
KR.SEQ.01	3.1.4.5 – Step-by-Step Guidance [1]	M	Ask a question that requires scanning paragraph text

6.3 Response Generation (RG)

6.3.1 Subfeatures to be tested

- **RG.COH** – Coherent Response: Response must be grammatically correct and readable.
- **RG.COMP** – Complete Answer: Covers all parts of a multi-part question.

6.3.2 Test Cases

Here list all the related test cases for this feature:

TC ID	Requirements	Priority	Scenario Description
RG.COH.01	3.1.4.3 – AI-Based Text Response [1]	H	Provide a prompt and evaluate the linguistic quality of the output
RG.COMP.01	3.1.4.3 – AI-Based Text Response [1]	M	Ask a two-part question; chatbot should answer both parts clearly

6.4 Source Reference Inclusion (SR)

6.4.1 Subfeatures to be tested

- **SR.INL** – Inline References: Include direct reference (e.g., “As described under section 4.1”).

6.4.2 Test Cases

Here list all the related test cases for this feature:

TC ID	Requirements	Priority	Scenario Description
SR.INL.01	3.1.4.5 [1]	H	Answer includes inline mention of section or heading
SR.INL.02	3.1.4.5 [1]	M	Chatbot references two or more relevant HTML sections in a single response

6.5 Session Context Management (SCM)

6.5.1 Subfeatures to be tested

- **SCM.FUP** – Follow-Up Prompt Resolution: Chatbot retains context of previous message.
- **SCM.MUL** – Multi-turn Sessions: Handles 3+ interactions with consistent context.

6.5.2 Test Cases

Here list all the related test cases for this feature:

SCM.FUP.01	3.1.4.5 – Step-by-Step Guidance [1]	H	Ask follow-up question using pronouns or ellipsis
SCM.MUL.01	3.1.4.5 – Step-by-Step Guidance [1]	M	Conduct a 3-turn conversation and check for context retention

7. Detailed Test Cases

7.1 PU.INT.01 — Understand Direct Prompt with Clear Intent

TC_ID	PU.INT.01
Purpose	Test if the chatbot can understand a direct, clearly-intentioned question
Requirements	3.1.4.1
Priority	High.
Estimated Time Needed	3 Minutes
Dependency	Knowledge base must be loaded with HTML docs
Setup	Load an HTML file that contains a section titled "How to log in to the system"
Procedure	[A01] User enters the prompt: "How do I log in to the system?"
	[V01] Chatbot analyzes the input and matches it to the relevant HTML section
	[V02] Chatbot responds with accurate login steps
Cleanup	None

7.1 PU.INT.02 — Handle Vague or Abstract Prompt

TC_ID	PU.INT.02
Purpose	Test if chatbot can infer intent from vague input
Requirements	3.1.4.1
Priority	High.
Estimated Time Needed	4 Minutes
Dependency	None
Setup	HTML content loaded with topic coverage across modules
Procedure	[A01] User enters the prompt: "I'm having trouble with reports."
	[V01] Chatbot infers the user is referring to "report generation"
	[V02] Chatbot provides appropriate guidance from KB
Cleanup	None

7.2 PU.ENT.01 — Entity Recognition with Multiple Concepts

TC_ID	PU.ENT.01

Purpose	Test whether the chatbot can resolve multiple entities
Requirements	3.1.4.3
Priority	Medium.
Estimated Time Needed	5 Minutes
Dependency	None
Setup	Load HTML doc with sections on "Password Reset" and "Support"
Procedure	[A01] User enters: "Can I reset my password and contact support?"
	[V01] Chatbot extracts both entities: password reset and support
	[V02] Chatbot provides responses for both topics
Cleanup	None

7.3 KR.HDR.01 — Match Prompt to HTML Heading

TC_ID	KR.HDR.01
Purpose	Test heading-based retrieval
Requirements	3.1.4.5
Priority	High.
Estimated Time Needed	3 Minutes
Dependency	None
Setup	HTML contains How to upload a document
Procedure	[A01] User enters: "How can I upload documents?"
	[V01] Chatbot locates heading in KB
	[V02] Chatbot summarizes and responds based on that section
Cleanup	None

7.4 KR.SEQ.01 — Sequential Scanning Retrieval

TC_ID	KR.SEQ.02
Purpose	Test retrieval from paragraph-level, not heading-based
Requirements	3.1.4.5
Priority	Medium.
Estimated Time Needed	3 Minutes
Dependency	None
Setup	HTML document contains embedded paragraph on export limits
Procedure	[A01] User asks: "Is there a size limit when exporting files?"
	[V01] Chatbot searches non-header content
	[V02] Responds with info found in paragraph text
Cleanup	None

7.5 RG.COH.01 — Generate Coherent Response

TC_ID	RG.COH.01
Purpose	Ensure chatbot’s reply is grammatically correct and readable
Requirements	3.1.4.3
Priority	High.
Estimated Time Needed	2 Minutes
Dependency	None
Setup	Knowledge base loaded
Procedure	[A01] User enters: "What does the red button do?"
	[V01] Chatbot analyzes intent and locates related info
	[V02] Chatbot responds in full, coherent sentences
Cleanup	None

7.6 RG.COMP.01 — Multi-Part Answer Completion

TC_ID	RG.COH.01
Purpose	Ensure chatbot handles multi-part queries completely
Requirements	3.1.4.3
Priority	Medium.
Estimated Time Needed	3 Minutes
Dependency	None
Setup	HTML has info on password reset and email change
Procedure	[A01] User asks: "How do I reset my password and change my email?"
	[V01] Chatbot parses both tasks
	[V02] Provides responses to both in one message
Cleanup	None

7.7 SR.INL.01 — Inline Reference in Response

TC_ID	SR.INL.01
Purpose	Validate inline reference to a section
Requirements	3.1.4.5
Priority	High.
Estimated Time Needed	3 Minutes
Dependency	None
Setup	HTML contains section "Exporting Reports"
Procedure	[A01] User enters: "How can I export reports?"
	[V01] Chatbot finds relevant section
	[V02] Chatbot includes phrase like “As described in the 'Exporting Reports' section…”
Cleanup	None

7.7 SR.INL.02 — Inline Reference for Multiple Sections

TC_ID	SR.INL.02
Purpose	Reference two HTML sections in one response
Requirements	3.1.4.5
Priority	Medium
Estimated Time Needed	3 Minutes
Dependency	None
Setup	HTML contains "Data Export" and "Email Reports" sections
Procedure	[A01] User enters: "How can I export reports?"
	[V01] Chatbot finds relevant section
	[V02] Chatbot includes phrase like “As described in the 'Exporting Reports' section…”
Cleanup	None

7.8 SCM.FUP.01 — Handle Follow-Up Prompt

TC_ID	SCM.FUP.01
Purpose	Test context management with follow-up prompts
Requirements	3.1.4.5
Priority	High.
Estimated Time Needed	4 Minutes
Dependency	None
Setup	HTML includes user profile and password management
Procedure	[A01] User enters: "How do I change my email?"
	[V01] Chatbot responds with correct steps
	[A02] User follows with: "What if I forgot my password?"
	[V02] Chatbot understands context and provides password reset steps
Cleanup	None

7.9 SCM.MUL.01 — Multi-Turn Conversation with Context

TC_ID	SCM.FUP.02
Purpose	Verify chatbot retains context for multiple exchanges
Requirements	3.1.4.5
Priority	Medium
Estimated Time Needed	4 Minutes
Dependency	None
Setup	Simulate 3-question user session
Procedure	[A01] User: "How do I create a report?"
	[V01] Chatbot replies
	[A02] User: "How can I export it?"
	[V02] Chatbot continues from same context
	[A03] User: "Can I email it too?"
	[V03] Chatbot delivers cohesive answer integrating all 3 prompts
Cleanup	None

8. Test Results

8.1 Test Result Table

TC_ID	Priority	Result	Explanation
PU.INT.01	High	Pass	The chatbot correctly identifies the user's goal or information need.
PU.INT.02	High	Pass	The chatbot correctly identifies the user's goal or information need.
PU.ENT.01	Medium	Pass	The chatbot detects and isolates important entities within the input prompt.
KR.HDR.01	High	Pass	Retrieve based on HTML headings
KR.SEQ.01	Medium	Pass	Scan full document for matching content beyond headers.
RG.COH.01	High	Pass	Response must be grammatically correct and readable.
RG.COMP.01	Medium	Pass	Covers all parts of a multi-part question.
SR.INL.01	High	Pass	Include direct reference (e.g., "As described under section 4.1").
SR.INL.02	Medium	Pass	Include direct reference (e.g., "As described under section 4.1").
SCM.FUP.01	High	Fail	Chatbot retains context of previous message.
SCM.MUL.01	Medium	Pass	Handles 3+ interactions with consistent context.

8.2 Exit Criteria

- All test cases executed.
- All essential functionalities pass testing
- 90% or more test cases passed.
- All high priority test cases passed.
- Essential functionalities pass without critical issues.

8.3 Conclusion

After the comprehensive testing phase, our chatbot system has met the predefined success thresholds. All high priority test cases and the majority of medium priority cases have passed. Only one high-priority case (context memory) failed, which is already flagged for further development. Despite this, the chatbot fulfills core requirements and demonstrates acceptable performance in functional and response-related tasks. Therefore, the system can be considered ready for the next phase of development or deployment, with the exception of minor issues being tracked separately.