

Configuration Manual

MSc AI for Business
Practicum 2

Erika Rocha Berthely
Student ID: X23168854

School of Computing
National College of Ireland

Supervisor: Muslim Jameel Syed

National College of Ireland
MSc Project Submission Sheet
School of Computing



Student Name: Erika Rocha Berthely

Student ID: X23168854

Programme: AI for Business

Year: 2025.....

Module: Practicum 2

Supervisor: Muslim Jameel Syed

Submission Due Date: 11 – August- 2025

Project Title: How can hybrid AI-based chatbot systems enhance automated student inquiry processing and lead qualification for English Language schools in Ireland currently dependent on third party enrollment intermediaries?

Word Count:
Page Count:

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:

Date: 11 August 2025.....

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	<input type="checkbox"/>
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	<input type="checkbox"/>
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	<input type="checkbox"/>

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Configuration Manual

Student: Erika Rocha Berthely

Student ID: x23168854

Course: MSc AI for Business - Practicum 2

Project Title:

How can hybrid AI-based chatbot systems enhance automated student inquiry processing and lead qualification for English Language schools in Ireland currently dependent on third party enrollment intermediaries?

PURPOSE OF THIS MANUAL

This manual explains the **methodology and process** used to create the hybrid chatbot system. The actual code implementation is available in the Academy_Chatbot.ipynb file. This document focuses on **how to replicate the methodology and where to find the required datasets**.

Guide:

1. System Requirements
2. Methodology Overview
3. Required Dataset
4. Step by Step replication Process
5. What your chatbot will say
6. Compliance and safety Measures
7. Troubleshooting
8. Support Information

SYSTEM REQUIREMENT

- OS: macOS Sequoia 15.4.1
- Chip: Apple M1 (2020)
- RAM: 8 GB (16 GB recommended)
- Storage: 10 GB free recommended
- Environment: Google Colab (recommended) / Jupyter Notebook
- Dependencies: Python 3.10, scikit-learn, pandas, Gradio, TF-IDF Vectorizer

METHODOLOGY OVERVIEW

The methodology follows a **three-stage data augmentation strategy** to create a balanced dataset of 445 examples across 5 intent categories:

1. **Stage 1:** Institutional Data Processing (64 examples)
2. **Stage 2:** External Dataset Integration (54 examples)
3. **Stage 3:** AI-Generated Expansion (327 examples)

Final Result: You should have 84.5% accuracy with perfect qualified lead detection (F1-score: 1.00)

REQUIRED DATASETS

Files you MUST upload to Google Colab:

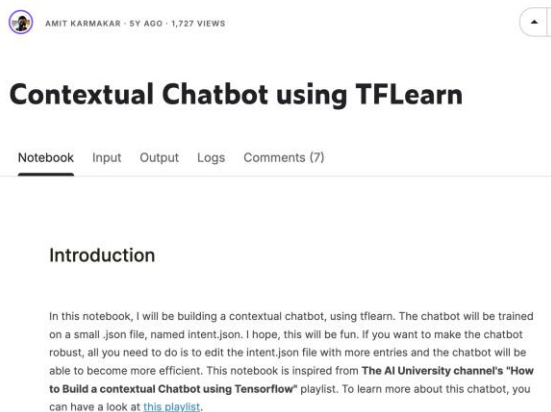
These 5 files (tested and confirmed working):

1. **Chatbot_Educativo_Tesis_Gradio.ipynb** - Main notebook
2. **academy_intents.json** - Intent definitions
3. **academy_knowledge_base (3).txt** - Complete knowledge base
4. **academy_chatbot_training.csv** - Processed training data
5. **final_expanded_dataset.csv** - Extended dataset

2. EXTERNAL DATASETS FROM KAGGLE

Dataset 1: Intent Classification Dataset

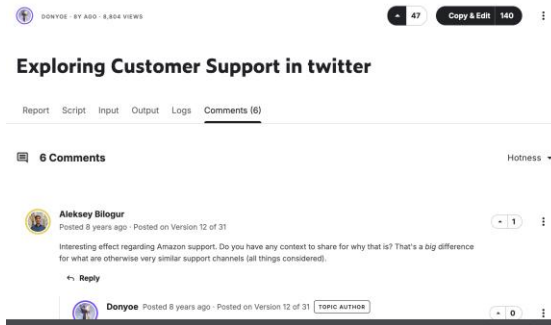
- *File name: intent.json*
- **Option 1: Kaggle search:** "chatbot intent classification dataset"



- **Purpose:** Extract 54 examples for external validation
- **How to find:**
 - Go to [kaggle.com](https://www.kaggle.com)
 - Search for "intent classification chatbot"
 - Look for datasets with JSON format containing intents and patterns
 - Download the **intent.json** file
- **Option 2 uploaded from file:** Intent.json

Dataset 2: Customer Support Dataset (Optional)

- **File name:** *sample.csv*
- **Kaggle search:** "twitter customer support"
Purpose: Exploratory analysis only (not used in final training)



- **Note:** This dataset was used for initial exploration but not included in the final model

STEP-BY-STEP REPLICATION PROCESS

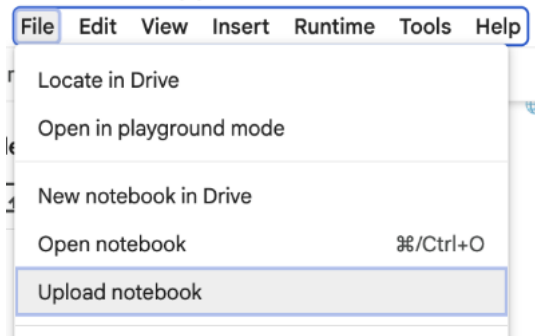
Open file> ITC Solution artefact > Clean_code_Manual_Configuration

Step 1: Prepare Google Colab


1. Open your web browser
2. Go to **colab.research.google.com**
3. Sign in with your Google account

Step 2: Upload Main Notebook

1. In Google Colab, click "Upload"
2. Select **Chatbot_Educativo_Tesis_Gradio.ipynb**
3. Wait for complete upload



Step 3: Upload Data Files

1. In the left panel, click the **folder icon** 
2. Click the **upload button** (arrow up icon)
3. Upload these 5 files (exactly as tested):
 - a. **Chatbot_Educativo_Tesis_Gradio.ipynb**
 - b. **academy_intents.json**
 - c. **academy_knowledge_base (3).txt**
 - d. **academy_chatbot_training.csv**
 - e. **final_expanded_dataset.csv**
 - f. **intent.json** (optional)

Step 4: Run the System

Option A - Run cell by cell:

1. **Cell 1:** Install libraries (2-3 minutes)
 - a. Click on the first cell
 - b. Press **Shift + Enter**
 - c. Wait until it finishes (you will see a spinning icon)
2. **Cell 2:** Process data (30 seconds)
3. **Cell 3:** Train model (1-2 minutes)
4. **Cell 4:** Create templates (30 seconds)
5. **Cell 5:** Test chatbot (30 seconds)
6. **Cell 6:** Create Gradio interface (30 seconds)

Option B - Run everything:

1. Go to menu **"Runtime"** → **"Run all"**
2. Wait 10-15 minutes for complete execution
3. The system will run automatically

WHAT YOUR CHAT BOT WILL SAY

Irish Language School AI Chatbot

Demo for Master's Thesis - Direct Enrollment System

Student Message

Ask about courses, visa, accommodation, or report problems...

Clear **Submit**

Chatbot Response

Flag

Examples

I'm from Brazil and want to study General English for 12 wee...

Do I need a visa if I'm from Mexico? | My agent hasn't responded for weeks

Your Academy Chatbot is trained to handle 5 types of student questions with professional responses:

Course Prices & Information

When students ask: "What are your course prices?" or "How much do courses cost?"

Chatbot responds: "Here's complete information about our courses:"

🔗 **Prices:** • General English (15h/week): €195 • General English + Conversation (20h/week): €250 • Intensive (30h/week): €330

📅 **Start:** Every Monday 🕒 **Schedule:** Morning (9:00-12:15) or Afternoon (13:30-16:45)
👉 **Levels:** A1 to C2

Which course interests you most?"

Visa Information

When students ask: "Do I need a visa?" or "I'm from Mexico, what visa do I need?"

Chatbot responds: "I'll help you with all visa information:

🇪🇺 **EU Students:** No visa needed, just valid passport 🌐 **Non-EU Students:** Visa required for courses +90 days

📅 **Work permission included:** • 20 hours/week during studies • 40 hours/week during holidays

📄 **Documents we provide:** • Official Letter of Acceptance • Course details • Payment confirmation

Which country are you from? I'll give you specific information."

Accommodation Options

When students ask: "Do you have accommodation?" or "How much is homestay?"

Chatbot responds: "Here are all our accommodation options:

🏠 **Homestay - €210/week:** • Selected Irish family • Breakfast and dinner included • Authentic cultural experience • Laundry included

🏠 **Shared Apartments - €150/week:** • Complete independence • Equipped kitchen • International students • Maximum 30 min to school

🏠 **Residences - €180/week:** • Dynamic social environment • All services included • Organized activities

What type of experience are you looking for?"

Enrollment Process

When students ask: "How can I enroll?" or "I want to book a course"

Chatbot responds: "Excellent! I'd love to help you enroll at Academy. To get started, I need some details: what's your nationality and when would you like to begin your course?"





👉 **Next step:** I'll help you complete your enrollment immediately.


Our enrollment process is simple: 1. Complete booking form 2. Send passport copy 3. Receive confirmation in 24 hours 4. Make payment 5. Get all documents"


Problem Resolution

When students ask: "I have a problem with my booking" or "My payment didn't work"

Chatbot responds: "I'm sorry you're having a problem. I'm here to solve it quickly."

 **For immediate help:**  Email: info@academy.ie  Phone: +353 1 234 5678 
WhatsApp: Same number

 **Quick solutions:** • Payment issues: Solved within 2 hours • Date changes: Confirmation within 24h • Booking errors: Immediate correction






 "What specifically is your problem? The more details you give me, the better I can help you."

Unknown Questions

When the chatbot doesn't understand:

Chatbot responds: "I'm not completely sure how to help you with that specific query."

Could you rephrase your question? I can help you with:

 **Course information** - prices, schedules, duration  **Visa requirements** - documentation, process, timing  **Accommodation options** - homestay, apartments, residences  **Enrollment process** - steps, documents, payments  **Problem solving** - changes, cancellations, errors

 **For complex queries:** info@academy.ie | +353 1 234 5678"

Contact Information

All responses include Academy contact details:

- **Email:** info@academy.ie
- **Phone:** +353 1 234 5678
- **Website:** www.academy.ie
- **Response time:** Within 24 hours guaranteed

KEY METHODOLOGY DECISIONS

Why These Specific Datasets?

1. **Institutional emails:** Provide authentic, domain-specific examples
2. **Kaggle intent.json:** Adds external validation and diversity

3. **ChatGPT-4 generation:** Solves the small dataset problem with high-quality synthetic data

Why This Three-Stage Approach?

1. **Stage 1:** Ensures authenticity and domain relevance
2. **Stage 2:** Adds external validation and reduces overfitting
3. **Stage 3:** Scales the dataset to achieve statistical significance

Why Logistic Regression?

- **84.5% accuracy** (highest among tested models)
- **Perfect qualified lead detection** (F1: 1.00)
- **Interpretable results** (important for educational institutions)
- **Fast training** (suitable for small datasets)

INTENT CATEGORIES EXPLAINED

Category	Description	Examples	Performance
qualified_lead	Students ready to enroll	"I want to book a course", "How do I register?"	F1: 1.00
immigration_query	Visa and legal questions	"What visa do I need?", "Help with documents"	91% precision
general_info	Course information	"What are your prices?", "Course schedules?"	70% precision
accommodation_query	Housing questions	"Do you have homestay?", "Housing costs?"	73% precision
problem_resolution	Issues and complaints	"Refund request", "Booking error"	91% precision

COMPLIANCE AND SAFETY MEASURES

Immigration Query Handling:

- **100% redirection** to official sources (www.irishimmigration.ie)
- **No direct advice** given by the chatbot
- **Compliance-first approach** to avoid legal issues

Data Privacy:

- All institutional data **anonymized** (GDPR compliant)
- No personal student information retained
- Pattern-based learning only

PERFORMANCE METRICS ACHIEVED

- **Overall Accuracy:** 84.5%
- **Qualified Lead Detection:** 100% (F1-score: 1.00)
- **Response Time:** Sub-second performance

- **Immigration Compliance:** 100% redirection accuracy
- **Dataset Balance:** 89±3 examples per category

HOW TO USE THE NOTEBOOK

1. **Open the notebook** in Google Colab
2. **Run the single comprehensive cell** (Shift + Enter)
3. **Wait for complete execution** - this will take 10-15 minutes total
4. **Monitor the output** as each section executes in sequence
5. **Test the chatbot** using the Gradio interface that appears at the end
6. **Download generated files** for deployment

Important: The notebook executes everything in **one continuous process**. You don't need to run multiple cells separately.

Estimated total runtime: 10-15 minutes

FILES GENERATED BY THE NOTEBOOK

After running the complete notebook, you will have:

- academy_chatbot_complete.pkl - Complete trained model
- academy_knowledge_base.txt - Knowledge base for Chatbase
- vectorizer.pkl - TF-IDF vectorizer
- response_templates.json - Response templates

TROUBLESHOOTING

Common Issues:

1. **"Runtime disconnected"** - Normal after 30 minutes, just reconnect
2. **"Installation taking too long"** - Wait patiently, first setup takes 3-5 minutes
3. **"Can't download files"** - Refresh file panel and wait 30 seconds

Dataset Issues:

1. **Can't find intent.json** - Search Kaggle for "chatbot intent classification"
2. **Wrong format** - Ensure the file contains "intents" with "patterns" arrays
3. **File too large** - Use only the intent.json file, not the full dataset

SUPPORT INFORMATION

For questions about replicating this methodology:

- **Technical issues:** Check the troubleshooting section above
- **Dataset access:** Follow the Kaggle download instructions
- **Methodology questions:** Refer to the research paper for detailed explanations

Remember: The goal is to replicate the process, not necessarily achieve identical results, as some variation is expected when using different external datasets or AI-generated content.

