National College of Ireland

National College of Ireland

Project Submission Sheet

Student Name:	Haameem Shimar Yusuff Nazzer
Student ID:	x23330155
Programme:	MSc in AI **Year:** 2025
Module:	Intelligent Agents and Process Automation
Lecturer:	**Shreyas Setlur Arun**
Submission Due Date:	22 April 2025
Project Title:	Skill Gap Identifier & Training Recommender using Python Automation and AI
Word Count:	969

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 references section. Students are encouraged to use the Harvard Referencing Standard supplied by the Library. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action. Students may be required to undergo a viva (oral examination) if there is suspicion about the validity of their submitted work.

Signature:	Haameem Shimar Yusuff Nazzer
Date:	22 April 2025

PLEASE READ THE FOLLOWING INSTRUCTIONS:

1. Please attach a completed copy of this sheet to each project (including multiple copies).
2. Projects should be submitted to your Programme Coordinator.
3. **You must ensure that you retain a HARD COPY of ALL projects**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. Please do not bind projects or place in covers unless specifically requested.
4. You must ensure that all projects are submitted to your Programme Coordinator on or before the required submission date. **Late submissions will incur penalties.**
5. All projects must be submitted and passed in order to successfully complete the year. **Any project/assignment not submitted will be marked as a fail.**

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

AI Acknowledgement Supplement

Intelligent Agents and Process Automation

Skill Gap Identifier & Training Recommender using Python Automation and AI

Your Name/Student Number	Course	Date
Haameem Shimar Yusuff Nazzer/ x23330155	Msc in AI	22-04-2025

This section is a supplement to the main assignment, to be used if AI was used in any capacity in the creation of your assignment; if you have queries about how to do this, please contact your lecturer. For an example of how to fill these sections out, please click here.

AI Acknowledgment

This section acknowledges the AI tools that were utilized in the process of completing this assignment.

Tool Name	Brief Description	Link to tool

Description of AI Usage

This section provides a more detailed description of how the AI tools were used in the assignment. It includes information about the prompts given to the AI tool, the responses received, and how these responses were utilized or modified in the assignment. **One table should be used for each tool used**.

[Insert Tool Name]	
-	
-	-

Evidence of AI Usage

This section includes evidence of significant prompts and responses used or generated through the AI tool. It should provide a clear understanding of the extent to which the AI tool was used in the assignment. Evidence may be attached via screenshots or text.

Additional Evidence:

-

Additional Evidence:

-

Skill Gap Identifier & Training Recommender using Python Automation and AI

Haameem Shimar Yusuff Nazzer (x23330155)

Intelligent Agents and Process Automation– H9IAPI

MSCAI1

School of Computing

National College of Ireland

1. Introduction

The need to equip organizations with the skilled employees relevant in an ever-expanding corporate ecosystem is greater than ever. An organized sector has a major contributor, a well-skilled workforce which plays a vital role in the success of any organization. But identifying which skills are needed and pairing those with an organization's existing talent is a complex and lengthy process. Finding out the skill gaps in employees is generally a time-consuming activity done manually by Human Resources (HR) departments and thus prone to human error. This includes individual performance reviews for each employee that assess what skills they already have, as well as recommending the most appropriate training programs to bridge those gaps. In this project, we aim to mitigate these issues using an automated Skill Gap Identifier and Training Recommender system via state-of-art i.e, − Artificial Intelligence (AI), Natural Language Processing (NLP), and web automation. This technology works synergistically to make the whole process simpler, better, quicker and scalable. It mines employee information new skills, performance data and training suggestions according to the demands of jobs across the company. The system not only discovers gaps of skills but also creates personalized suggestions for training from Coursera - one of the most widely used platforms for online learning using smart algorithms. By incorporating the Coursera API, the system also remains up-to-date with the latest programs, ensuring that the suggested training is relevant and aligned with both the needs of each individual employee, as well as the overarching goals and plans of the organization with regards to employee development.

In the end, this returns HR professionals to a more value-added activity and lightens the administrative load. This also empowers organizations to upskill their workforce in a focused, efficient way that is critical to competitiveness and long-term sustainability in a rapidly evolving business landscape.

2. Project Objective

The objective of this project is:

- To automate the process of identifying employee skill gaps.
- To recommend personalized online training courses based on missing skills.
- To send automated personalized emails to employees with their skill gap reports.
- To model the entire business process using BPMN (Business Process Modelling Notation).
- To investigate automation opportunities and implement intelligent automation using Python and AI technologies.

3. Why We Are Implementing

Manual skill gap analysis is:

- Slow and inconsistent across employees.
- Dependent on manual HR interventions and subjective judgments.
- Hard to scale across large organizations.

Implementing automation ensures:

- Speed (real-time processing of hundreds of employees).
- Accuracy (AI-driven skill matching).
- Personalization (each employee gets specific course recommendations).
- Scalability (works for any number of employees).

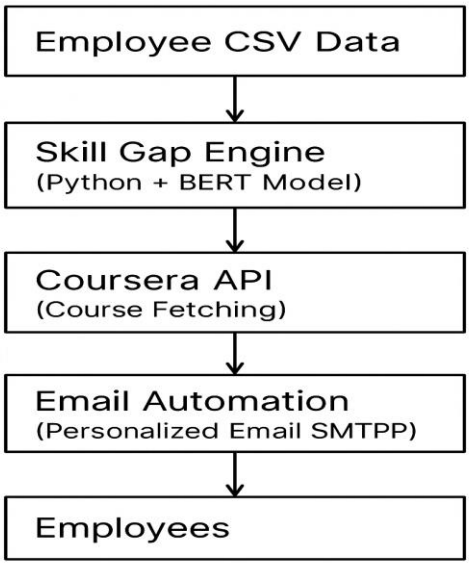
4. Technologies Used

Technology	Purpose
Python	Main programming language
Pandas	Data loading and processing (CSV handling)
Sentence-Transformers (BERT)	AI-based semantic matching of skills
Requests	Live course search via Coursera API
SMTP + EmailMessage	Sending automated emails to employees
BPMN Standard	Business process modeling (diagram provided)

5. Business Process Breakdown

Step Task

1. Load Employee and Role Requirement CSV files
2. Preprocess skill data (clean, tokenize skills)
3. Compare employee skills to role requirements using BERT model
4. Identify missing skills per employee
5. Search Coursera API for recommended courses for each missing skill
6. Save personalized training recommendations into a CSV report
7. Generate personalized emails for each employee
8. Send emails automatically through SMTP
9. Process complete



5.1. BPMN

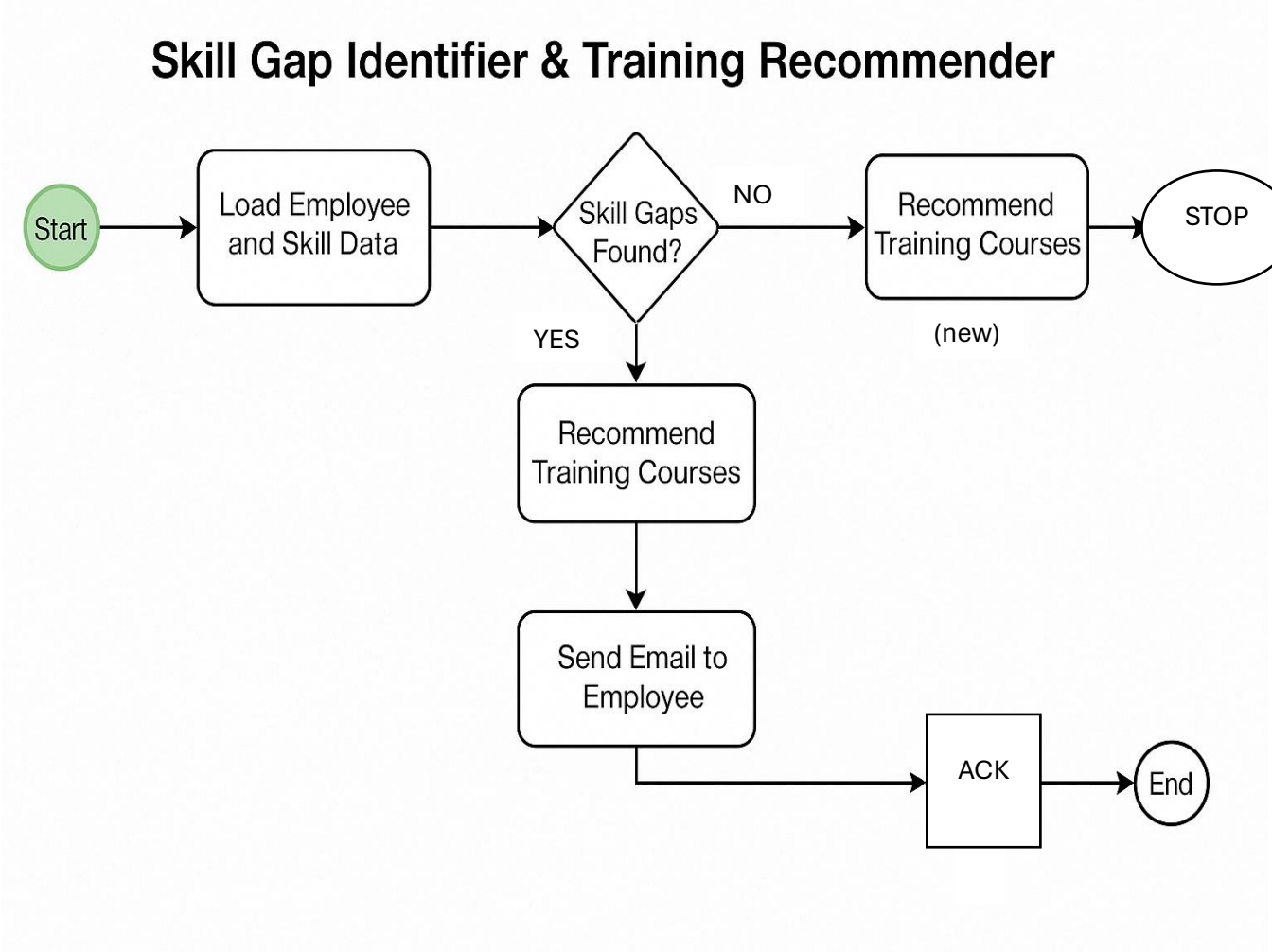
This process was visualized using the Business Process Model and Notation (BPMN) diagram, which outlines how the Skill Gap Identifier and Training Recommender system works as a whole. It is a widely accepted standard and it helps to formulate the business processes in a simplified and standard way to be understood by all stakeholders.

The BPMN diagram in this project presents a clear picture of how the system works end to end. The main components of BPMN are expressed in a BPMN diagram:

Start Event: The process starts with the activation of the system, which causes the workflow to execute. This is where the data (our employees, their current skill sets, and their roles) is fed to the system.

- **Service Tasks:** These verbs represent the true functionalities of the system:
- **Load Data:** In the first step we load employee and role data (The data is normally in CSV format which we would use in the upcoming steps).
- **Identify tasks :** The system identifies skills through the use of NLP (Workshop-like large language models, BERT models, etc…) comparing current vs required for a given position
- **Suggested Courses:** According to the identified skill gaps, the system calls the Coursera API to suggest relevant courses that can help employees fill these gaps. These courses are updated in real-time.
- **Email:** The system generates a personalized email for each employee showing its summary of the skills it is lacking along with the list of suggested courses to improve it.

End Event: The process ends when all steps have been performed successfully, and employees receive their personalized recommendations via email.To develop a clear outline of the automation, this diagram would help. Making the workflow in BPMN, allows to document the process in a structured way that makes sense to both technical and non-technical people, guaranteeing that all the stakeholders understand the way of working, creating alignment and everybody is one page.



(BPMN Diagram)

BPMN 2.0: BPMN (Business Process Model and Notation) is an established process modelling,standard for companies and models that are shared together or are solo and shared together in a collaborative way. This standard allows various stakeholders across departments to align and ensure understanding of how tasks, decision points, and subprocesses are sequenced, thereby providing clarity and consistency in process mapping in the design of the Skill Gap Identifier and Training Recommender system. It provides a more unified approach for the different users to have more visibility between teams and systems.

The design is REST API compliant to improve the system's performance and scalability. Hence, it also lets the system interact with third-party services, like coursera, in real-time, allowing fetching of live data to provide them with the latest relevant courses. RESTful APIs allow the system to easily communicate with additional services or data sources, should the need arise in the future, allowing for flexibility and expansion without interfering with existing workflows.

On the secure communication front, the system uses email via SMTP on SSL (Secure Sockets Layer). It encrypts email communications this way the sensitive data related to employee records and training recommendations on the network. Secure email handling helps

to class profiles; this minimizes the risk of spilling confidential information to other trusted users or external actors, preserving the system's integrity and users' privacy.

The system also has a salient feature being built with an AI-based mechanism to achieve semantic matching. Sentence-Transformers, built on top of pre-trained BERT (Bidirectional Encoder Representations from Transformers) models, are employed for model semantics that enable high accuracy in matching semantically. It's able to do this with such extreme precision that even when the terminology used to describe employees existing skills is slightly different from the terminology used in the role skills, it can identify the missing skills. The use of Artificial Intelligence is what makes training recommendations highly relevant and precise for every employee’s needs.

Moreover, data privacy is built into the architecture of the system. No sensitive employee information is hardcoded into the system and data privacy best practices guidelines are followed. This method substantially minimizes the risk of revealing sensitive information, thus protecting employees' private data and adhering to industry regulations like the GDPR (General Data Protection Regulation). Outside of the ID-0ST FloorPlate that requires KYC verification at the point of access, this means no direct access to sensitive data; a guarantee of this framework builds trust and a safe environment for all users, offering peace of mind to employees and HR professionals alike.

6. Steps Involved in the Automation

Step	How It Was Implemented
1	Loading CSV using Pandas
2	Skill preprocessing (lowercasing, splitting)
3	Semantic comparison using BERT embedding and cosine similarity
4	Skill gap identification if similarity score < 0.7
5	Course searching using live Coursera API
6	Email generation with point-wise list of missing skills and courses
7	SMTP mailing using secure Gmail app passwords

7. Automation Potential Analysis

Task	Automation Potential	Commentary
Load employee and role data	Fully Automated	Loading CSV files is handled automatically with no manual intervention.
Skill comparison using AI	Fully Automated	BERT models automate semantic similarity checks between skills.
Skill gap identification	Fully Automated	Missing skills are detected automatically based on threshold values.
Course search using Coursera API	Fully Automated	Real-time API integration eliminates manual course searching.
Generating training recommendations	Fully Automated	Compiling recommendations is automatic based on API responses.
Sending personalized emails	Fully Automated	Each employee receives a customized email automatically.
Tracking course completion	Manual	Currently, course completion must be confirmed manually by employees (future automation possible).

Most tasks in the process have been successfully automated using intelligent technologies. Only post-training progress tracking remains a manual task, which can be automated in future work through Coursera's extended APIs or employee feedback loops.

8. Automation Proposal

This project utilized Python due to its flexibility, wide library support, and simplicity for rapid automation development. AI and NLP technologies like BERT (through Sentence-Transformers) were used for intelligent skill comparison, ensuring that minor differences in terminology (e.g., "OOP" vs "Object-Oriented Programming") were handled automatically.

Course recommendations were automated by integrating with the Coursera open API, ensuring real-time training suggestions based on up-to-date online content.

SMTP-based email automation enabled personalized communication to employees at scale, maintaining professionalism.

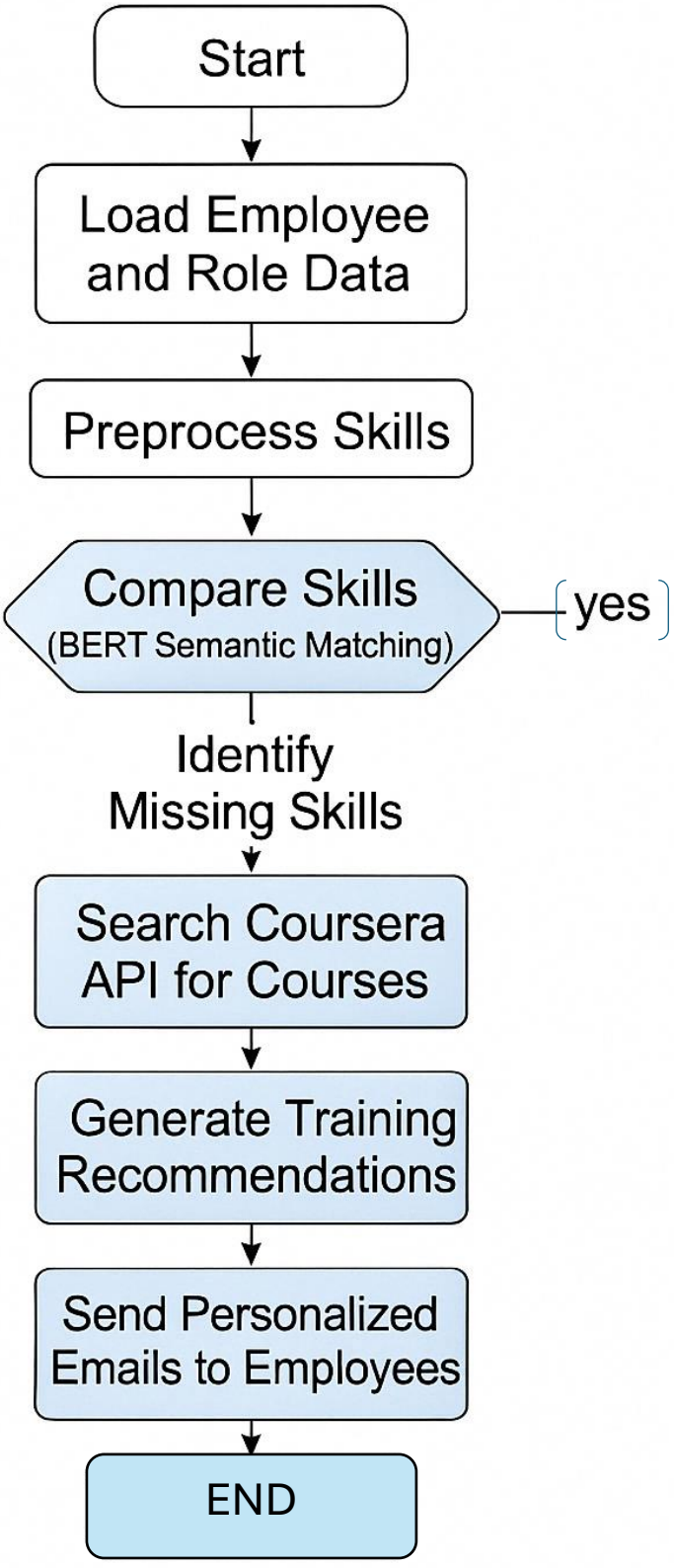
If scaled, the system could evolve into a full adaptive learning path generator using machine learning techniques, recommending future career moves based on completed courses and employee interests. In future iterations, progress tracking, dynamic reskilling suggestions, and integrated performance improvement dashboards could further enhance the automation impact.

9. Industry Standards Followed

Standard	Compliance
BPMN 2.0	Used for business process modeling
REST API practices	Used for Coursera live data fetching
Secure Email Handling	SMTP over SSL used
AI Semantic Matching	Sentence-Transformers (BERT-based models)
Data Privacy	No sensitive employee information hardcoded

BPMN 2.0 seeing as Business Process Modeling providing a standardized and clear mechanism to map the mechanism leading to enhanced communication and understanding across the team. The integration with Coursera's live data fetching adheres to REST API practices, which allows the system to be both scalable and efficient, facilitating smooth communication across various components. SMTP over SSL is used for secure handling of email, while preventing sensitive communication from being compromised in transit and adding an extra layer of security. With the implementation of AI Semantic matching using Sentence-Transformers based on BERT models, Natural Language Processing capabilities further improve the semantic matches to be more accurate and contextually aligned. Additionally, by ensuring that no sensitive employee information is hardcoded into the system, the approach maintains compliance with data privacy best practices, minimizing the risk of exposing confidential data and fostering a secure environment for all users.

Skill Gap Identifier & Training Recommender



10. Future Improvements

Area	Improvement Idea
Course search	Further improve search term optimization dynamically
Progress tracking	Fetch course completion status automatically using APIs
Feedback loop	Employee ratings on recommended courses to refine AI models
Dashboarding	Develop real-time monitoring dashboards (Streamlit/PowerBI)
Adaptive Learning	Recommend future courses based on employee learning patterns

Subsequent developments in the system will enhance functionality and user experience in specific key areas. For course search, dynamic optimizing search terms based on available data and user behaviour and preferences enables improved accuracy in its search results. In progress tracking, utilizing APIs to fetch course completion status automatically will facilitate real-time updates, guaranteeing updated progress tracking of the learners. With employee ratings on prescribed courses, will further help to 'teach' AI models to (more accurately) personalising the output over time. Following the adopted systems mentioned above, we would see the rise of real time monitoring dashboards where the case will flow from a tool (Streamlit, PowerBI) .Finally, adaptive learning will be greatly improved by recommending employees what to learn next considering their past learning patterns, enabling employees to upskill their careers continuously. These advancements are geared toward developing a more tailored, reactive, and insightful educational experience.

11. Conclusion

Skill Gap Identifier & training recommender project also was a successful outcome to show how traditional HR processes can be transformed using intelligent automation. The system provided an end-to-end automated solution which was implemented through APIs, Python, AI and secured communication protocols to detect and fill the gaps of employee skills. We even accomplished real-time analysis, personalized course recommendations, and automated employee communication with the product delivery, which created a scalable and efficient HR support system. This work paves a way forward for future enhancements like real-time tracking, adaptive learning recommendations, and sophisticated employee development analytics.