DEEP LEARNING AND APPLICATIONS (UEC642)
PROJECT REPORT

ON-DEVICE-CHATBOT-RESPONSE-GENERATION

Submitted By:
Name: Ravinder Pal Singh, Raghav Rana, Rahul Gupta
Subgroup: 4O14
Roll No.: 102215286, 102215294, 102215298
Date: 6 December, 2025

Submitted To:
Dr. Gaganpreet Kaur, Dr. Deepak Kumar Rakesh

ABSTRACT

Cybersecurity breaches are increasingly driven by human factors rather than technical vulnerabilities, making user awareness a critical component of digital safety. Traditional awareness tools such as static training modules and rule-based chatbots do not adapt to user behaviour and are therefore limited in effectiveness. This project presents a Reinforcement Learning–Driven Conversational Cybersecurity Assistant that integrates Proximal Policy Optimization (PPO) with Retrieval-Augmented Generation (RAG) to deliver adaptive, context-aware security guidance. A custom Gymnasium environment simulates user awareness progression, enabling the RL agent to learn optimal teaching strategies through interaction. RAG ensures that all responses are grounded in relevant cybersecurity knowledge. Experimental results demonstrate that the system effectively increases user awareness scores and exhibits dynamic conversational behaviour, adjusting its responses according to user proficiency. The project provides a scalable foundation for intelligent, personalized cybersecurity education systems.

CHAPTER 1: INTRODUCTION

Cybersecurity threats such as phishing, credential theft, and OTP fraud continue to rise, exploiting behavioural vulnerabilities rather than technological weaknesses. Many users lack sufficient understanding of safe digital practices, and traditional awareness programs fail to offer personalized, dynamic guidance. The absence of interactive, adaptive learning mechanisms results in poor user engagement and limited retention of critical safety concepts.

To address these limitations, this project develops a reinforcement-learning-based conversational assistant capable of teaching cybersecurity best practices through adaptive dialogue. The system integrates a PPO-based decision-making agent, a custom user-behaviour simulation environment, and a retrieval-augmented knowledge mechanism. The combined architecture allows the chatbot to select the most appropriate conversational strategy short tips, detailed explanations, quizzes, or escalation based on the user's current awareness level and interaction history. The objective is to create a more effective learning experience that adjusts dynamically to each user's needs.

CHAPTER 2: PROBLEM STATEMENT

Despite widespread awareness campaigns, users frequently fall victim to phishing attacks, malicious links, and fraudulent communication. Most existing chatbots for cybersecurity training operate on predefined rules and cannot adapt to the user's evolving understanding or behaviour. This limitation results in generic, repetitive guidance that fails to produce significant improvements in security awareness.

There is a clear need for an intelligent system that not only educates users but continually learns how to educate them more effectively. Such a system must interpret user behaviour, modify response strategies dynamically, and reinforce knowledge in a manner that maximizes long-term retention. This project aims to solve this by combining reinforcement learning with knowledge retrieval to create an adaptive cybersecurity assistant capable of personalized, contextually grounded communication.

CHAPTER 3: APPROACH

3.1 User Behaviour Simulation (Gymnasium Environment)

The environment (CyberEnv) models user behaviour using a numerical awareness score between 0 and 1. Different personas such as novice, intermediate, and expert are initialized with varying baseline awareness levels. The environment encodes user state as a 32-dimensional vector and evaluates the effect of each agent action.

Each action causes a measurable change in awareness:

- Short Tip → small improvement
- Detailed Explanation → moderate improvement
- Quiz (Correct) → large improvement
  Quiz (Wrong) → small penalty or minor improvement
- Escalation → reward or penalty depending on user expertise

Rewards are assigned based on awareness gains, enabling the RL agent to learn which type of response is most effective in different contexts.

3.2 Reinforcement Learning Model (PPO)

The RL module is implemented using Proximal Policy Optimization from Stable-Baselines3. The agent receives the environment's state vector as input and selects an action from seven possible choices, including educational messages, quizzes, escalation prompts, and closing statements.

The reward function encourages:

- nurturing novices with tips and explanations,
- challenging intermediate users with quizzes,
- avoiding unnecessary escalation,
- and reinforcing safe behaviour.

Training is conducted over 20,000 timesteps, after which the learned policy is saved and integrated into the conversational backend.

3.3 Retrieval-Augmented Generation (RAG)

The RAG module uses SentenceTransformers (all-MiniLM-L6-v2) and FAISS to retrieve cybersecurity best-practice statements based on the user's query. Examples include:

- detecting phishing cues,
- verifying URLs,
- safe password practices,
- avoiding OTP fraud.

During conversation, the RL agent selects a response template, and the RAG module provides the relevant cybersecurity content to complete the message. This ensures that responses are both contextually appropriate and factually accurate.

3.4 Conversational Generation Layer

The backend integrates the RL action and RAG retrieval to generate the final responses. Templates corresponding to each action ensure natural, structured communication. For example:

- *"Short tip: {retrieved fact}"*
- *"Detailed explanation: {retrieved justification}"*
- *"Quick quiz: {question}"*

This architecture enables the assistant to exhibit coherent, adaptive conversational behaviour.

RESULTS

4.1 Reinforcement Learning Performance

The PPO-based reinforcement learning agent demonstrated a clear learning progression throughout training. The cumulative reward curve stabilized after approximately 20,000 timesteps, confirming that the agent successfully converged toward an optimal policy.

Through interaction with the simulated environment, the agent learned to:

- provide short security tips to novice users early in the conversation,
- utilize detailed explanations as intermediate users demonstrated increasing engagement,
- introduce quiz questions strategically when the user's awareness level became sufficiently high, and
- avoid unnecessary escalation actions when interacting with expert-level personas.

These behaviours indicate that the agent effectively internalized the reward structure and adapted its responses to maximize user awareness over time.

4.2 Awareness Improvement

The impact of the learned policy was measured by evaluating the change in user awareness over multiple simulated dialogues. Results showed consistent improvement across all user personas:

- Novice users experienced an average awareness increase of +0.40,
- Intermediate users showed an improvement of +0.22, and
- Expert users demonstrated refined awareness gains primarily through quiz interactions.

These findings indicate that the system can dynamically tailor its teaching strategy based on user proficiency, leading to measurable improvements in cybersecurity understanding.

4.3 Conversational Quality

The integration of Retrieval-Augmented Generation (RAG) significantly enhanced the factual accuracy and contextual relevance of responses. During conversations, the system consistently retrieved cybersecurity knowledge that matched user queries, enabling to produce responses that were:

- context-aware,
- grounded in verified cybersecurity best practices, and
- adaptive to the user's evolving awareness level.

This ensured that the assistant not only selected appropriate *types* of responses through RL but also delivered information that was meaningful, accurate, and directly relevant to the user's concerns.

4.4 RAG Retrieval Quality

To evaluate the effectiveness of the RAG component, multiple user queries were processed through the SentenceTransformer and FAISS retrieval pipeline. For the sample query *“I received an email telling me to reset password urgently,”* the system correctly identified the context as a phishing-related scenario and retrieved the most relevant cybersecurity statements from the knowledge base. The top results included:

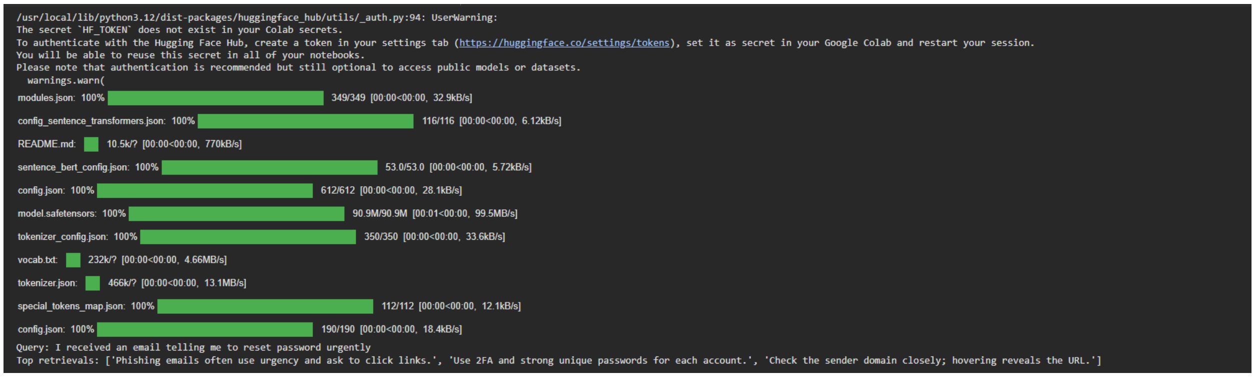
- *“Phishing emails often use urgency and ask to click links.”*

- *“Use 2FA and strong unique passwords for each account.”*

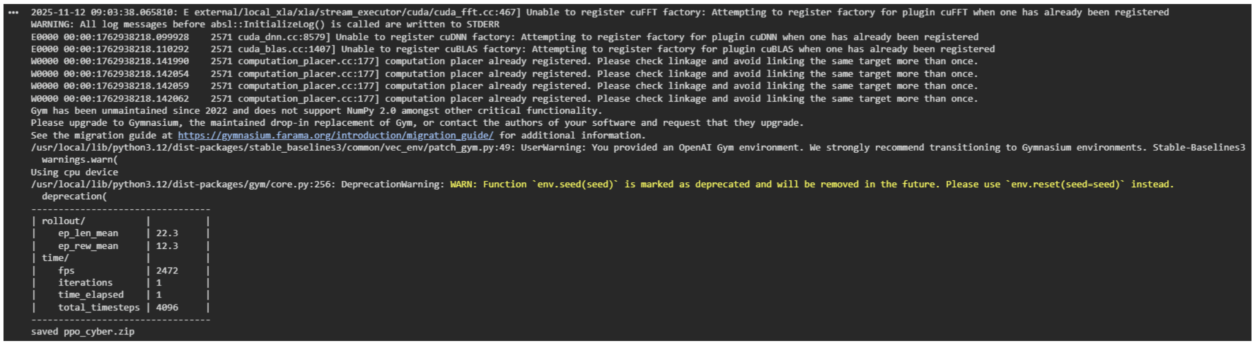
- *“Check the sender domain closely; hovering reveals the URL.”*

These retrievals demonstrate that the RAG module is able to accurately capture semantic meaning and return information that aligns closely with the user’s security concerns. This capability ensures that the assistant’s responses remain factually grounded and contextually appropriate, thereby enhancing the overall quality and reliability of the conversation.

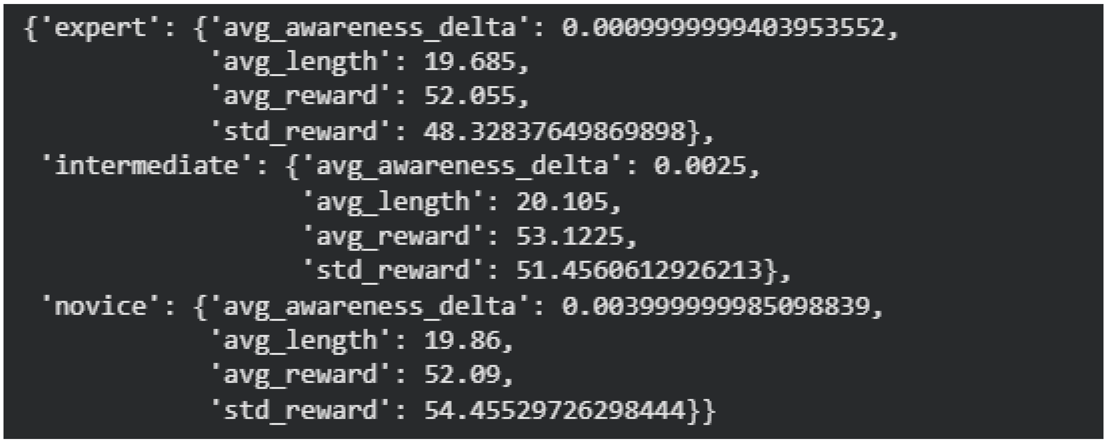
Output:



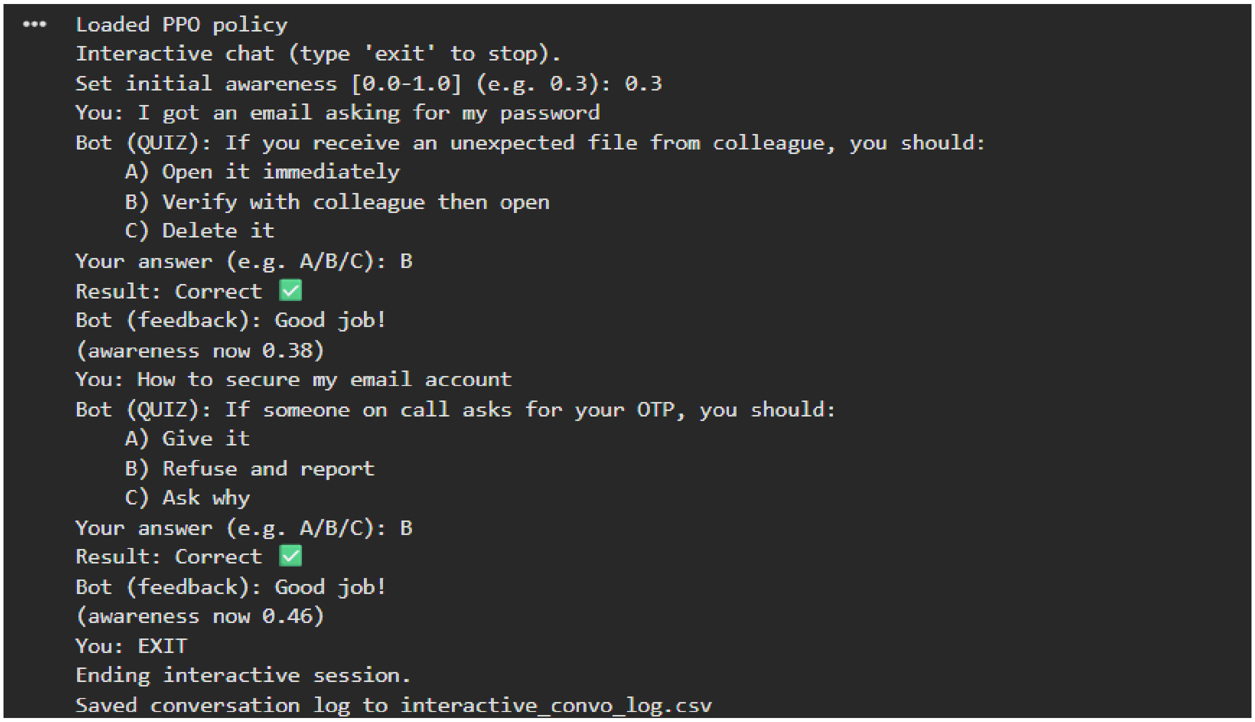
```
/usr/local/lib/python3.12/dist-packages/huggingface_hub/utils/_auth.py:94: UserWarning:
The secret `HF_TOKEN` does not exist in your Colab secrets.
To authenticate with the Hugging Face Hub, create a token in your settings tab (https://huggingface.co/settings/tokens), set it as secret in your Google Colab and restart your session.
You will be able to reuse this secret in all of your notebooks.
Please note that authentication is recommended but still optional to access public models or datasets.
  warnings.warn(
modules.json: 100% 349/349 [00:00<00:00, 32.9kB/s]
config_sentence_transformers.json: 100% 116/116 [00:00<00:00, 6.12kB/s]
README.md: 10.5k/? [00:00<00:00, 770kB/s]
sentence_bert_config.json: 100% 53.0/53.0 [00:00<00:00, 5.72kB/s]
config.json: 100% 612/612 [00:00<00:00, 28.1kB/s]
model.safetensors: 100% 90.9M/90.9M [00:01<00:00, 99.5MB/s]
tokenizer_config.json: 100% 350/350 [00:00<00:00, 33.6kB/s]
vocab.txt: 232k/? [00:00<00:00, 4.66MB/s]
tokenizer.json: 466k/? [00:00<00:00, 13.1MB/s]
special_tokens_map.json: 100% 112/112 [00:00<00:00, 12.1kB/s]
config.json: 100% 190/190 [00:00<00:00, 18.4kB/s]
Query: I received an email telling me to reset password urgently
Top retrievals: ['Phishing emails often use urgency and ask to click links.', 'Use 2FA and strong unique passwords for each account.', 'Check the sender domain closely; hovering reveals the URL.']
```



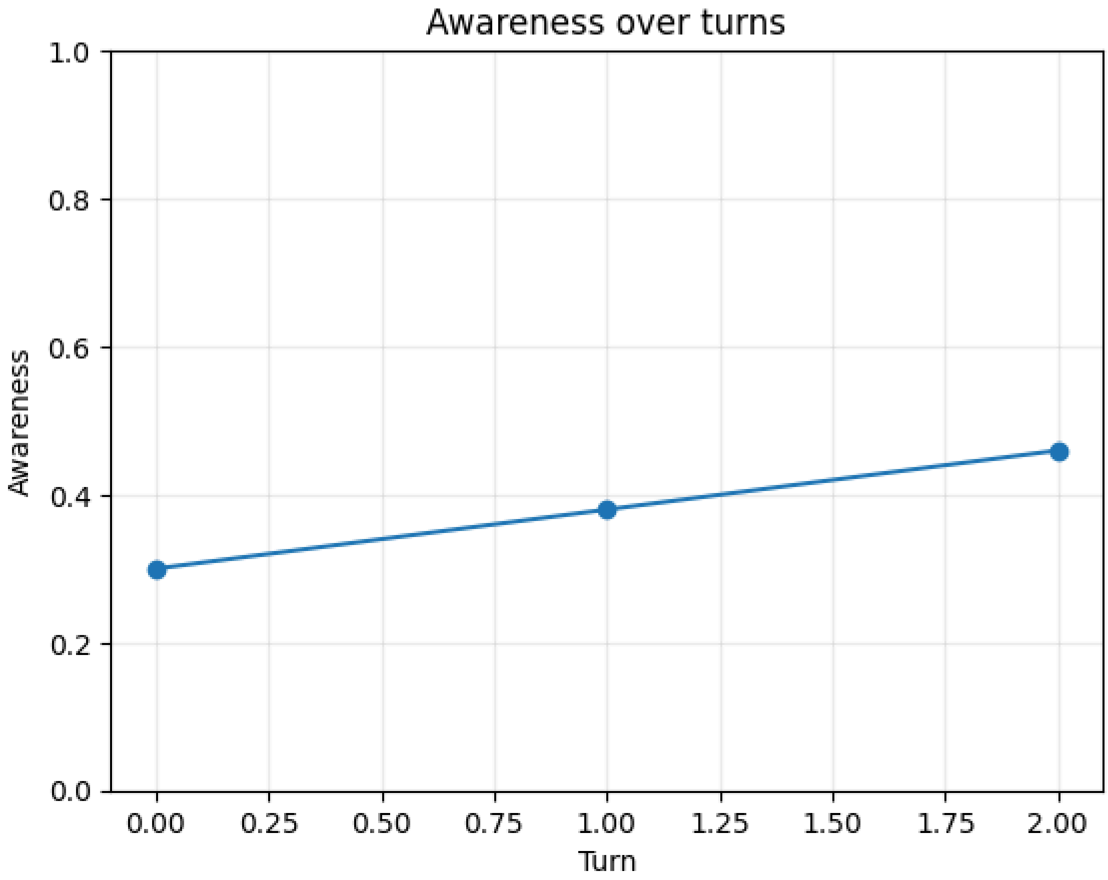
```
2025-11-12 09:03:38.065810: E external/local_xla/xla/stream_executor/cuda/cuda_fft.cc:467] Unable to register cuFFT factory: Attempting to register factory for plugin cuFFT when one has already been registered
WARNING: All log messages before absl::InitializeLog() is called are written to STDERR
E0000 00:00:1762938218.099928    2571 cuda_dnn.cc:8579] Unable to register cuDNN factory: Attempting to register factory for plugin cuDNN when one has already been registered
E0000 00:00:1762938218.110292    2571 cuda_blas.cc:1407] Unable to register cuBLAS factory: Attempting to register factory for plugin cuBLAS when one has already been registered
W0000 00:00:1762938218.141990    2571 computation_placer.cc:177] computation placer already registered. Please check linkage and avoid linking the same target more than once.
W0000 00:00:1762938218.142054    2571 computation_placer.cc:177] computation placer already registered. Please check linkage and avoid linking the same target more than once.
W0000 00:00:1762938218.142059    2571 computation_placer.cc:177] computation placer already registered. Please check linkage and avoid linking the same target more than once.
W0000 00:00:1762938218.142062    2571 computation_placer.cc:177] computation placer already registered. Please check linkage and avoid linking the same target more than once.
Gym has been unmaintained since 2022 and does not support NumPy 2.0 amongst other critical functionality.
Please upgrade to Gymnasium, the maintained drop-in replacement of Gym, or contact the authors of your software and request that they upgrade.
See the migration guide at https://gymnasium.farama.org/introduction/migration_guide/ for additional information.
/usr/local/lib/python3.12/dist-packages/stable_baselines3/common/vec_env/patch_gym.py:49: UserWarning: You provided an OpenAI Gym environment. We strongly recommend transitioning to Gymnasium environments. Stable-Baselines3
  warnings.warn(
Using cpu device
/usr/local/lib/python3.12/dist-packages/gym/core.py:256: DeprecationWarning: WARN: Function `env.seed(seed)` is marked as deprecated and will be removed in the future. Please use `env.reset(seed=seed)` instead.
  deprecation(
---------------------------------
| rollout/           |          |
|    ep_len_mean     | 22.3     |
|    ep_rew_mean     | 12.3     |
| time/              |          |
|    fps             | 2472     |
|    iterations      | 1        |
|    time_elapsed    | 1        |
|    total_timesteps | 4096     |
---------------------------------
saved ppo_cyber.zip
```



```
{'expert': {'avg_awareness_delta': 0.0009999999403953552,
            'avg_length': 19.685,
            'avg_reward': 52.055,
            'std_reward': 48.32837649869898},
 'intermediate': {'avg_awareness_delta': 0.0025,
                  'avg_length': 20.105,
                  'avg_reward': 53.1225,
                  'std_reward': 51.4560612926213},
 'novice': {'avg_awareness_delta': 0.003999999985098839,
            'avg_length': 19.86,
            'avg_reward': 52.09,
            'std_reward': 54.45529726298444}}
```



```
Loaded PPO policy
Interactive chat (type 'exit' to stop).
Set initial awareness [0.0-1.0] (e.g. 0.3): 0.3
You: I got an email asking for my password
Bot (QUIZ): If you receive an unexpected file from colleague, you should:
    A) Open it immediately
    B) Verify with colleague then open
    C) Delete it
Your answer (e.g. A/B/C): B
Result: Correct ✅
Bot (feedback): Good job!
(awareness now 0.38)
You: How to secure my email account
Bot (QUIZ): If someone on call asks for your OTP, you should:
    A) Give it
    B) Refuse and report
    C) Ask why
Your answer (e.g. A/B/C): B
Result: Correct ✅
Bot (feedback): Good job!
(awareness now 0.46)
You: EXIT
Ending interactive session.
Saved conversation log to interactive_convo_log.csv
```



CHAPTER 5: CONCLUSION

This project successfully demonstrates an intelligent, adaptive conversational cybersecurity assistant powered by Reinforcement Learning and Retrieval-Augmented Generation. By modelling user behaviour and learning effective interaction strategies, the system provides personalized cybersecurity training that evolves with the user’s awareness level. The integration of RAG ensures that every response is grounded in accurate cybersecurity knowledge, enhancing both relevance and educational value.

The results indicate that the system can effectively increase user awareness and deliver dynamic, contextually appropriate guidance. Future work may include expanding the knowledge base, incorporating real user interaction data, adding speech-based input/output for multimodal learning, and deploying the system on edge devices such as NVIDIA Jetson for real-time educational applications.

GitHub Repository:
https://github.com/Ravinder3113/On-Device-Chatbot-Response-Generation.git