



Development of a Formative Assessment Chatbot for a 100-Level Course in Cybersecurity

¹Keyamo, C. A., ²Edun, O. P., & ³Baale, A. A

¹Department of Computer Science, Dennis Osadebay University, Asaba, Nigeria

²Department of Cybersecurity, Dennis Osadebay University, Asaba, Nigeria. Edun.peace@dou.edu.ng

³Department of Information systems, Ladoke Akintola University of Technology, Ogbomosho, Nigeria.

*Corresponding author email: keyamo.clement@dou.edu.ng

Abstract

This study presents Cypanion: a formative assessment chatbot developed to support first-year undergraduate students in learning foundational cybersecurity concepts through interactive dialogue and real-time feedback. Classroom observations revealed limited prior exposure to cybersecurity among students, often leading to disengagement and poor performance. To address this, Cypanion was built using the DIET-enabled RASA framework for natural language understanding, supported by Python, SQLite3, and the PyCharm API. The system was trained on domain-specific datasets comprising instructional content, question-answer pairs, and conversational tasks including registration, assessment, and formative feedback. Evaluation focused on two natural language understanding subtasks: intent classification and entity extraction. The chatbot achieved 95% precision, recall, and F1-score for intent classification, and 85% precision, 75% recall, and 80% F1-score for entity extraction. These results demonstrate Cypanion's effectiveness in accurately interpreting student inputs and generating relevant educational responses. The findings suggest that formative assessment chatbots can enhance engagement, support personalized learning, and scale assessment delivery in technical domains like cybersecurity. Future improvements will include speech-based interfaces and exam security modules for deployment in remote learning environments.

Keywords: Natural Language Processing (NLP), Formative Assessment, Educational Chatbots, Conversational Agents, Dialogue-Based Assessment,

Introduction

Chatbots have recently emerged as a vital technology for teaching, assessment, and administration in higher education (Mageira et al., 2020; Okonkwo & Ade-Ibijola, 2020; Iniesto et al., 2021). They serve as round-the-clock backup tutors capable of presenting instructional materials, administering assessments, and offering feedback prepared by human teachers. Also they can serve as companions to isolated students in remote learning environments, providing 24/7 support services, including answering frequently asked questions (FAQs) and providing general guidance services.

Chatbots have been deployed as mobile applications, providing students with customized online services such as course content information, and replacing static form registration with an interactive dialogue-based registration (Iniesto et al., 2021). The applications of chatbots in higher learning not only enhances student learning and engagement, but can greatly reduce staff administrative workload through the convenience of dialogue-based automation (Okonkwo & Ade-Ibijola, 2020). One very important application area of chatbots in higher learning is in electronic assessment. This application, often called conversational-based assessment (CBA), has been suggested by researchers (Lopez et al., 2021; Lehman & Jackson, 2018) as useful in combining assessment with feedback, thereby improving student motivation to learn and overall learning, while evaluating the student in the process. This application area is the basis of this work.

In the Nigerian context, cybersecurity is a relatively new field at the undergraduate level, and many 100-level students have limited prior exposure to its foundational concepts. Based on direct classroom observations by the authors, this unfamiliarity often results in poor academic performance and disengagement among students. While global interest in educational chatbots is growing, localized challenges (such as inadequate student background knowledge) underscore the need for context-aware solutions. To address this, we propose Cypanion: a formative assessment chatbot designed to reinforce basic cybersecurity concepts through continuous, interactive dialogue.

Consequently, the aim of this work was to develop a formative assessment chatbot that can assist fresh undergraduate students learn cybersecurity concepts, taking advantage of the power of tireless automation and human-like dialogue that chatbots can present in the tutoring of students. This aim was achieved with the following objectives:

- 1 To collect and pre-process domain-specific dataset from relevant multiple sources;
- 2 To train a Natural Language Understanding (NLU) model of the proposed chatbot using the collected datasets and the RASA framework.
- 3 To implement an experimental formative assessment chatbot using RASA framework, python and associated libraries.
- 4 To evaluate the chatbot performance using F1 score.

The next section briefly outlines some chatbot schemes related to our proposed solution.

Chatbot application in learning and assessment

Researchers have taken advantage of the potentials in chatbots to build educational tools that facilitate learning and assessment in higher educational systems. Graesser et al. (2004), built AutoTutor as a dialogue-based system for teaching university students physics and computers. The experimental work increased the performance of students. Ruan et al. (2019) developed Quizbot, a chatbot designed to assist students learn factual knowledge in Science, English as well as Safety discipline. The system showed that students that used it increased their recollection of correct answers by as much as 20% higher than students that used the alternative learning aid.

Mageira et al. (2022) developed Asasarabot: an experimental chatbot tool for teaching high school students foreign language cultural contents. The results of the work show that chatbots are effective for learning both foreign languages and their cultural contents. Furthermore, Lenk and Schimdt (2022) developed an intelligent chatbot as a virtual tutor for online collaborative learning and assessment. The aim of their research was to support formative assessment in online learning (OCL) environments by combining learning analytic data with the communication interface of a chatbot. The work showed that chatbots could reduce work load of teachers and administrators and facilitates formative assessments of learners. Also, Yildirim-Erbasli et al. (2023) developed a chatbot with dual role of interface and e-assessment system. The aim of the work was to design and implement a new conversational-based assessment(CBA) system and investigate the functionality in answering students responses accurately and to investigate its usability. The results demonstrated the positive potential of CBA as a technology-enhanced assessment tool

Nemieboka et al. (2024) developed a dialogue agent for online electronic testing of visually impaired students. Their results show that chatbots integrated with speech-based interfaces can effectively grant the visually impaired access to online educational resources and serve as practical tools for assessment. Building on this approach, Cypanion is proposed in this study as a text-based formative assessment chatbot for teaching entry-level undergraduate cybersecurity courses. To lay the foundation for its development, the next section presents the general architecture and key components of chatbot systems.

Chatbot development

According to Adamopoulou and Moussiades (2020), the general chatbot architecture (Figure 1) consists of the following key components:

User Interface Component: This may include simple text-based interfaces or advanced options such as speech recognition systems, point-and-click interactions, images, videos, text boxes, and more.

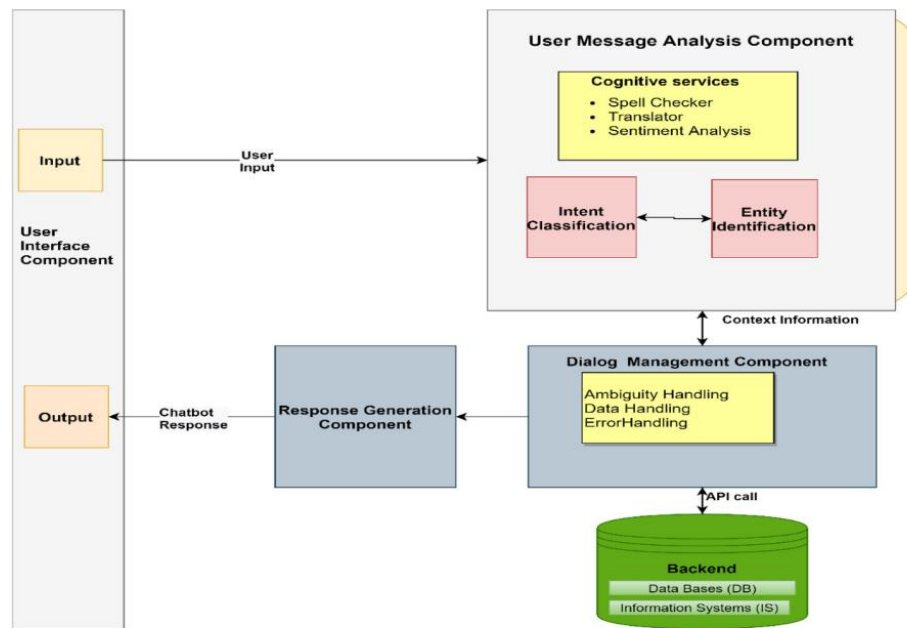


Figure 1: General architecture of a chatbot (source: Adamopoulou & Moussiades (2020))

Message analysis component: This component is responsible for understanding human communication, which constitutes the Natural Language Understanding (NLU) aspect of a chatbot. It consists of two parts: intent recognition and entity extraction. An intent represents the user's purpose for interacting with the system, while an entity is a key piece of information the system needs to fulfill the user's intent. Intents and entities are extracted from user inputs using machine learning and NLU techniques, enabling classification into predefined categories for processing.

Dialogue management component: This manages the conversational flow and context. It resolves ambiguity by using default responses when intents cannot be classified, maintains context by storing user inputs, and handles errors to ensure continuity.

Response generation component: This generates responses based on user intent using one of three models: rule-based (predefined rules), retrieval-based (external API queries), and generative (new output based on dialogue history). A hybrid approach may combine these models for enhanced performance.

Backend component: This component accesses stored responses or external APIs to fulfill user intents. Retrieved data is passed through the dialogue manager to the response generation module for user output. These components represent the foundational architecture of effective chatbot systems. The message analysis and dialogue management modules highlight the need for robust Natural Language Processing (NLP) capabilities in chatbot systems. The next section discusses this briefly.

Machine learning techniques for chatbot systems

To interpret user input accurately and manage conversations effectively, modern chatbots depend on advanced techniques from the fields of Natural Language Processing (NLP) and Machine Learning (ML) (Rasool & Shahri (2021). NLP is a collection of computational methods used for representing and analysing human language in order to transfer the capacity to communicate in human fashion, to machines. NLP consists of two divisions: Natural Language Understanding (NLU), that is concerned with the capacity of machines to understand human conversations; and the Natural Language Generation (NLG), that is concerned with the response of the machines in a human conversational manner.

Two subtask of NLU are Named Entity Recognition (NER) and Intent Classification or extraction tasks (Al-Tuama & Nasrawi, 2022). In the NER sub task, the entities are extracted from the conversation of the user in order to determine their intentions or needs for using the system. In Intent classification, the actual reason for using the system are classified by the system and used to generate an appropriate response. Intent classification is very important in dialogue systems like chatbots as it allows the chatbot understand what the user wants. It involves classifying the user utterances into predefined categories.

There are two data-driven ML approaches to building dialogue models that run chatbots: the end-to-end approach and the modular approach (Bunk et al., 2020). In the end-to-end approach, the user input can be directly connected to the dialogue engine for selection of the machine response, while in the modular approach, the NLU and NLG are separated. The advantage of the end-to-end approach over the modular is that the modular separates the classification and recognition tasks, which often results in error propagation over the modules. To ameliorate this, the Dual Intent and Entity Transformer (DIET) architecture (figure 2) was constructed by Bunk et al. (2020). DIET has capacity for both multi-task and single application in dialogue models' training. Its advantage is that it not only improves on the state of the art and speeds up the process, but it also can incorporate, seamlessly, some pre-trained models.

According to Bunk et al. (2020), DIET, allows for the training of intent and entity recognition models as a single process. The schematic diagram for the configuration of DIET for joint entity recognition and intent classification model training is presented with figure 2.

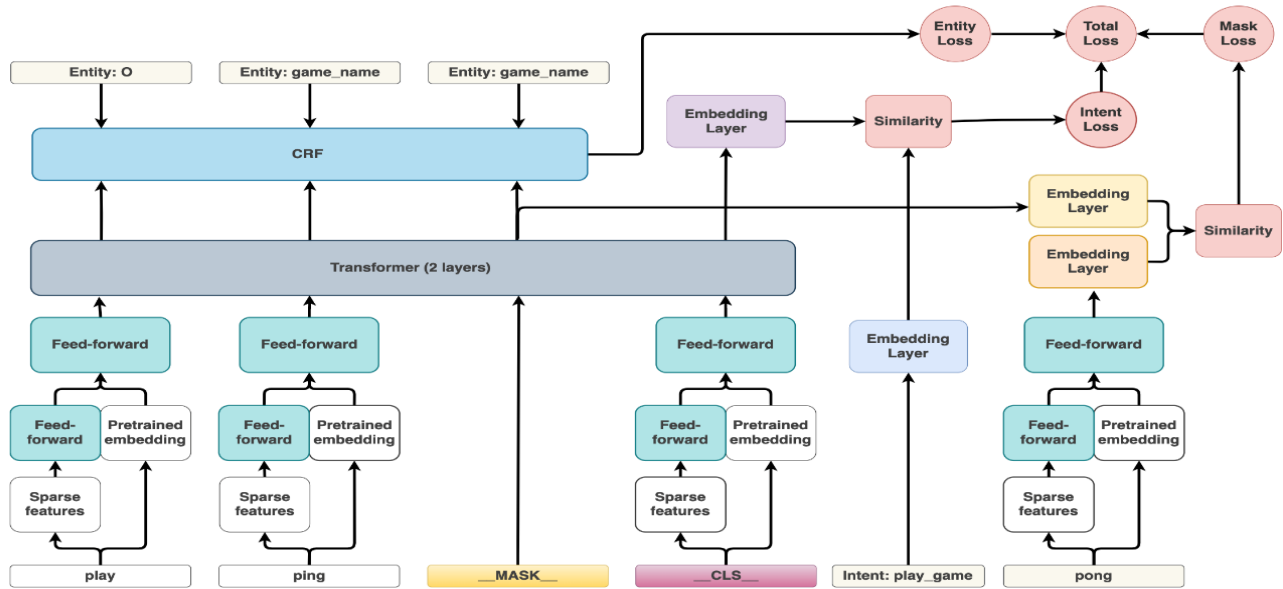


Figure 2 : Configuration of DIET with training data for the NER and Intent classification (source: Bunk et al., 2020)

For the Named Entity Recognition (NER) model training, the input utterance, x_{entity} , and the associated entity labels, y_{entity} , and the special mask token, $_Mask_,$ generated by DIET from the training data randomly, are used to configure the DIET. The CRF layer of the DIET is used to predict the output entity labels, y , based on the transformer output sequence, α . this is given as:

$$L_E = L_{\text{CRF}}(\alpha, y_{\text{entity}}) \quad (1)$$

Where :

L_E is the loss function, L_{CRF} is the negative log likelihood of the CRF and α is the output sequence of the 2 layer DIET transformer and y is the training target or entity label.

For the intent classification model training, the training data intent labels, y_{intent} , $_masks_,$ training data utterance tokens, x_{intent} , and the special CLS tokens, $_CLS_,$ used to mark the end of each utterance sentence, are used to configure the DIET for the intent classification model.

The model training is carried out with the objective of minimizing the total loss, which is sum of the mask loss, intent loss and entity loss, given as:

$$L_{total} = L_{intent} + L_{entity} + L_{mask} \quad (2)$$

Where,

$$L_{intent} = - \langle S_i^+ - \text{Log} (e^{s^+} + \sum \Omega^{-1} e^{s^-}) \rangle \quad (3)$$

$$L_{mask} = - \langle S_i^+ - \text{Log} (e^{s^+} + \sum \Omega^{-1} e^{s^-}) \rangle \quad (4)$$

This work adapts the DIET architecture in building the dialogue model for the chatbot, Cypanion, in association with RASA framework and associated python libraries. This is covered in the next section: the methods and materials section.

Methods and materials

Figure 3 presents the architectural design of Cypanion: the proposed formative assessment chatbot for cybersecurity education. The system is designed as a text-based, retrieval-oriented chatbot that integrates formative testing with dynamic feedback delivery.

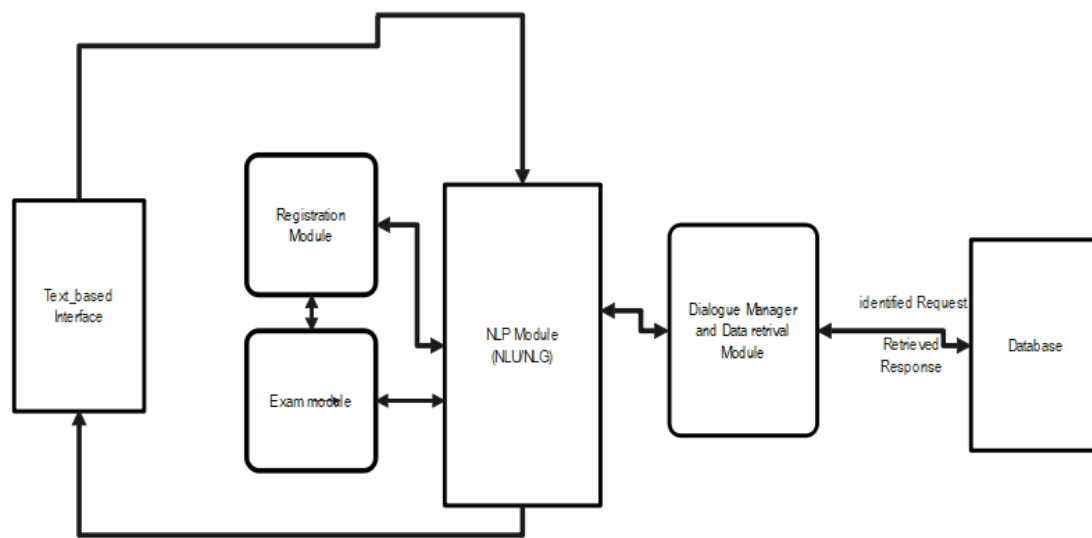


Figure 3: Architectural design of Cypanion

The chatbot consist of the following components:

Text-based interface: Enables conversational interaction between the user and the chatbot.

Registration module: Handles new user enrolment and account identification.

Assessment module: Presents multi-choice formative questions, processes user responses, and delivers feedback for incorrect answers.

Natural Language Processing (NLP) module: Comprising the NLU and response generation layers; this component is trained to interpret cybersecurity-related conversational input and produce context-appropriate replies.

Dialogue manager: Manages conversation flow, tracks user input history, and handles ambiguity or classification errors.

SQLite database: Stores user profiles, question-answer pairs, feedback messages, and other domain-specific content.

Data collection and processing

To train the chatbot, a domain-specific dataset was compiled from multiple sources including:

Intent-annotated user utterances: Manually prepared conversational examples for typical interactions (e.g., "I want to register", "Start exam") were annotated with predefined intents and entities relevant to registration, assessment, and help tasks. Approximately 275 unique utterances were annotated and distributed across 14 intent classes.

Formative assessment dataset: A custom-built dataset was extracted from the course CYB111: Introduction to Cybersecurity, taught at Dennis Osadebay University (DOU), Nigeria. The dataset included over 120 multiple-choice questions with associated correct answers and explanatory feedback. These were incorporated into the chatbot's response bank to simulate real-world assessment behavior.

Entity patterns and sample variations: Entities such as `student_id`, `question_number`, and `matriculation_number` were defined with multiple representative examples to enhance generalization across different input forms. All data were preprocessed to normalize text, remove noise, and balance class representation. Dataset augmentation and initial annotation support were partially assisted by ChatGPT (OpenAI, 2024), particularly to generate paraphrased utterances for greater linguistic diversity. The complete dataset was split using an 80:20 training-validation ratio. This is consistent with common practice in supervised learning to ensure the model generalizes well without overfitting. Potential bias was mitigated by ensuring diverse phrasing and including variations in input structure for each intent.

Table 1 presents example utterances used to train the chatbot's intent recognition system. The table is divided into six main intent categories: `take_exam`, `provide_student_id`, `greet`, `provide_surname`, `register_exam`, and `provide_password`. Each column contains typical phrases or sentences that a student might use when interacting with the chatbot. For the `take_exam` intent, the training examples include natural expressions like "I want to take the exam" and "Begin my exam," helping the model recognize when a user wants to start an exam session. Similarly, for `provide_student_id`, the chatbot is trained to detect phrases like "Here is my ID: [67890]{student_id}," where the numeric portion is marked as an entity to extract the student ID.

The `greet` intent includes simple greetings such as "hello," "hi," and "good morning," while the `provide_surname` intent includes responses like "Surname is [Baale]{surname}" and "My surname is [Keyamo]{surname}," where the surname is tagged for extraction. Under `register_exam`, examples like "How do I register" and "Register me" guide the chatbot to understand requests related to exam registration. Lastly, the `provide_password` intent contains sample inputs like "Here is my password: [securepass]{password}," which show how users might submit login credentials. The annotated examples using curly braces indicate key entities such as `student_id`, `surname`, and `password`, which the chatbot extracts and uses during the conversation. These examples help train the model to identify both the user's intent and the relevant information needed to carry out that intent. Table 1 is significant because it provides a foundation for the chatbot's ability to interpret varied user input during formative assessments. It also illustrates the design of training data that balances linguistic diversity, clarity, and functionality, all critical for building an effective conversational system.

Table 1: Sample training data

Intent	take_exam	provide_student_id	greet	provide_surname	register_exam	provide_password
Training Examples	- I want to take the exam - Start my exam - Begin my exam - I am ready for the exam	My student ID is [12345](student_id) - Here is my ID: [67890](student_id)	hello - hi - hey - good morning - good evening	- Surname is [Onashoga](name) - Surname is [Baale](surname) My surname is [Keyamo](surname)	I want to register - How do I register - Register me	- Here is my password: [securepass](password) - It's [testpassword](password)

Experimental setup and development of the chatbot

To build the NLU core of the chatbot, the DIET-enabled RASA framework was configured in a dense+sparse pipeline with DistilBERT as the dense featurizer; this architecture is illustrated in Figure 4. It shows how user input is tokenized, processed through both count-based and transformer-based featurizers, and then passed through the DIET classifier and response selector to generate context-appropriate replies. The architecture begins with training examples, which are tokenized using a Whitespace Tokenizer. Features are then extracted through both a Count Vector Featurizer (for word frequency) and a Language Model Featurizer (DistilBERT, for contextual semantics). These are passed to the DIET Classifier, which handles both intent classification and entity recognition. Entities are normalized using the Entity Synonym Mapper to maintain consistency. The predicted intent then flows to the Response Selector, which chooses an appropriate reply. In cases of uncertainty, the FallBack Classifier generates a default or clarifying response. The output is a predicted intent and corresponding response, forming the chatbot's interaction logic.

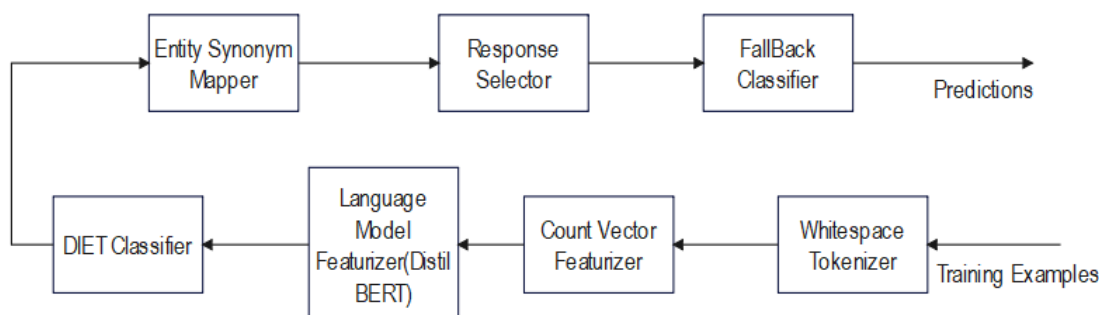


Figure 4: RASA pipeline with DistilBERT featurizer

The trained model was then deployed to implement an experimental formative assessment chatbot. Tools used for the implementation and evaluation tasks included the Python programming language, SQLite for database development, the RASA framework, and relevant supporting libraries. All software components were run on a Windows 10 Intel Pentium IV laptop PC with 4GB RAM. To assess the chatbot's performance, the test dataset was used to evaluate the dialogue model under controlled laboratory conditions. The results of this evaluation, along with a discussion of the findings, are presented in the next section.

Results

Figure 5 shows the Precision, Recall, and F1 average scores obtained for the overall setup across the entity extraction and intent classification sub-tasks using the evaluation datasets. The evaluation was conducted to assess the chatbot's capacity to correctly interpret user requests and respond appropriately. Three metrics were used: Precision, which measures the accuracy of positive predictions; Recall, which reflects the system's capacity to retrieve all relevant instances and ; F1 score, the harmonic mean of precision and recall, offering a balanced performance indicator.

In the intent classification task, the chatbot recorded an average of 95% across all three metrics (Precision, Recall, and F1). This indicates a strong ability to accurately interpret user intent and perform appropriate actions, such as handling registration, assessment, and feedback requests. For example, when a user initiates a registration or seeks their performance report, the chatbot reliably detects the correct intent and triggers the corresponding system function.

In the entity extraction sub-task, the chatbot achieved 85% Precision, 75% Recall, and an F1 score of 80%. This demonstrates a reasonably strong performance in identifying key variables like student names, test scores, and test types from user inputs. However, the slightly lower recall suggests that the system occasionally misses relevant entities, particularly when user input is ambiguous or loosely structured.

Figures 6 and 7 demonstrate the chatbot in action. Figure 6 presents a snapshot of the registration process, where the chatbot interacts with a user to collect account setup information and update the SQLite database. Figure 7 shows an assessment session, with the chatbot delivering cybersecurity-related questions, interpreting responses, and providing immediate feedback, especially for incorrect answers, which reinforces student learning.

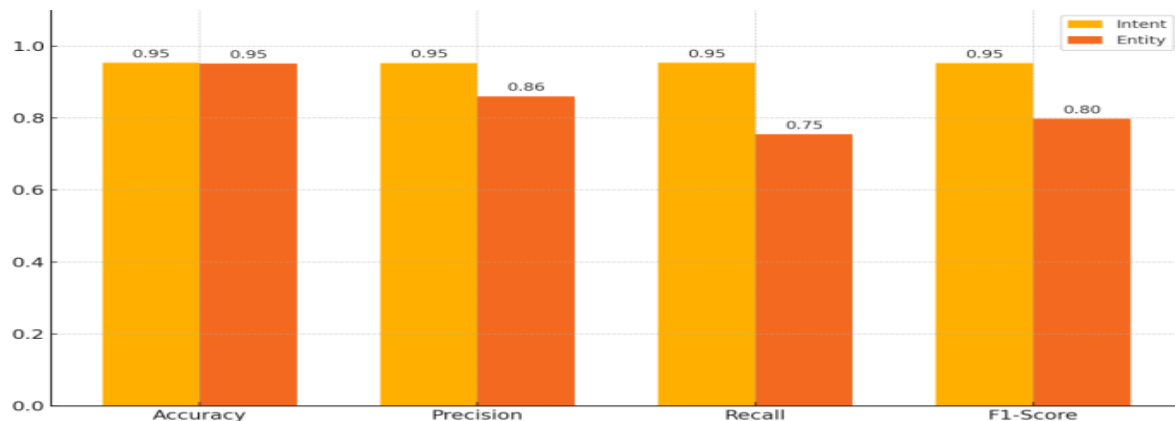


Figure 5: Performance report for the proposed chatbot

Discussion

The high performance in intent classification (95% across metrics) demonstrates the robustness of the chatbot's underlying DIET architecture. This aligns with findings from related works (Bunk et al., 2020; Arevalillo-Herráez et al., 2022), which highlight DIET models as efficient and capable of handling multi-class classification tasks even with lightweight configurations. Furthermore, the result supports the insights of Graesser et al. (2004), who emphasize the critical role of accurate intent recognition in sustaining productive educational dialogue, as seen in intelligent tutoring systems like AutoTutor.

While entity extraction results were also promising, with an F1 score of 80%, the lower recall value (75%) indicates occasional lapses in detecting entities, especially in cases where user input lacks clarity or deviates from expected formats. This challenge is consistent with other research findings (Iniesto et al., 2021; Nemieboka et al., 2024), which report that natural language variability among students can hinder entity recognition performance.

The interaction snapshots in Figures 6 and 7 validate the chatbot's practical usability. In the registration phase, the system effectively guides users through the process, while the assessment session showcases its instructional value, providing formative feedback that can enhance learning outcomes. These observations confirm that beyond performance metrics, the chatbot exhibits meaningful functionality in real-world educational settings.

```

Bot loaded. Type a message and press enter (use '/stop' to exit):
Your input -> Good morning
Hey! How are you? Welcome! I am Cypanion, your friendly formative assessment bot. I can register you or assist you learn CYB111 if you are already registered. Please submit the requested information.
What is your matriculation number, please?
Your input -> 0803745474
What is your surname, please?
Your input -> Keyamo
What is your first name, please?
Your input -> Clement
What is your middle name, please?
Your input -> Amone
What is your password, please?
Your input -> testpassword123
thanks! this is you saved biodata:Mat Number:0803745474,Surname:Keyamo,First name:Clement,Middle name:Amone, and your selected password:testpassword123
You can now learn cyb111 using the formative assessment questions?
Your input ->

```

Figure 6: Snapshot of registration process in the experimental assessment chatbot

```

Your input -> take exam
Here is your test. Please reply with the correct option letter (A, B, or C).
Q1: Which of these is a strong password?
A. 123456
B. qwerty
C. P@ssw0rd123!
Your input -> c
✅Correct!
Q2: What does phishing usually involve?
A. Stealing devices
B. Fraudulent emails
C. Logging in twice
Your input -> c
❌That's incorrect. The correct answer is B.
💡 Phishing tricks users through fake emails to reveal information.
Q3: What is malware?
A. A security update
B. Malicious software
C. Firewall settings

```

Figure 7: Snapshot of assessment process in experimental assessment chatbot

Conclusion

The formative assessment chatbot, Cypanion, was developed to support entry-level cybersecurity students in learning foundational concepts and principles through a more engaging and interactive experience. By integrating natural language processing and machine learning techniques, the system delivers formative assessments conversationally and provides immediate feedback to reinforce learning. The experimental prototype was implemented using the DIET-enabled RASA framework, PyCharm API, Python programming language, SQLite3, and relevant libraries. Evaluation using Precision, Recall, and F1-score metrics demonstrated strong performance in intent classification (95% across all metrics) and satisfactory results in entity extraction (Precision: 85%, Recall: 75%, F1: 80%). These results indicate the chatbot's potential effectiveness in accurately interpreting user input and providing relevant educational feedback. The findings support the feasibility of using conversational agents to facilitate formative assessment in cybersecurity education.

Future development will focus on enhancing the system's capabilities by incorporating a speech-based interface and expanding its security modules. These modules will address both system security (including data privacy, secure communications, and adversarial defense techniques) and examination security (such as anti-cheating and speaker verification for impersonation prevention). This will enable the deployment of the chatbot in high-stakes assessment scenarios, including remote or online examinations.

Recommendations

1. It is recommended that future research explore the design of fully functional, domain-specific chatbots for formative assessment in cybersecurity and other disciplines.
2. Such systems could significantly improve student engagement, provide scalable feedback, and support continuous assessment in undergraduate education.

References

- Arevalillo-Herráez, M., Arnau-González, P., & Ramzan, N. (2022). On adapting the DIET architecture and the Rasa conversational toolkit for the sentiment analysis task. *IEEE Access*, 10, 107477-107487.
- Bunk, T., Varshneya, D., Vlasov, V., & Nichol, A. (2020). DIET: Lightweight language understanding for dialogue systems. *arXiv preprint arXiv:2004.09936*. <https://arxiv.org/abs/2004.09936>
- Candello, H., & Pinhanez, C. (2016). Designing conversational interfaces. *Simpósio Brasileiro sobre Fatores Humanos em Sistemas Computacionais-IHC*, 100. Graesser, A. C., Lu, S., Jackson, G. T., Mitchell, H. H., Ventura, M., Olney, A., & Louwerse, M. M. (2004). AutoTutor: A tutor with dialogue in natural language. *Behavior Research Methods, Instruments, & Computers*, 36(2), 180–192. <https://doi.org/10.3758/BF03195563>
- Hsu Myat Mo, S., Sann Su Su Yee, & Soe, K. M. (2023). On using DIET architecture for users' comments in Myanmar language [Paper presentation]. International Conference on Communication and Computer Research (ICCR 2023). Retrieved from https://www.researchgate.net/publication/379829156_DIET_Architecture_for_Users'_Comments_in_Myanmar_Language
- Iniesto, F., Coughlan, T., & Lister, K. (2021, April). Implementing an accessible conversational user interface: Applying feedback from university students and disability support advisors. In *Proceedings of the 18th International Web for All Conference* (pp. 1–5).
- Lehman, B., & Jackson, G. T. (2018). Eliciting Deeper Evidence through Conversation-Based Assessments. In *Deep comprehension* (pp. 209-223). Routledge.
- Lenk, F., & Schmidt, S. (2022, January). The virtual tutor: Combining conversational agents with learning analytics to support formative assessment in online collaborative learning. In *Proceedings of the 55th Hawaii International Conference on System Sciences* (pp. 1–10). University of Hawai'i at Mānoa. <https://doi.org/10.24251/HICSS.2022.009>
- Lopez, A. A., Guzman-Orth, D., Zapata-Rivera, D., Forsyth, C. M., & Luce, C. (2021). Examining the accuracy of a conversation-based assessment in interpreting English learners' written responses. *ETS Research Report Series*, 2021(1), 1–15. <https://doi.org/10.1002/ets2.12315>
- Magiera, K., Pittou, D., Papasalouros, A., Kotis, K., Zangogianni, P., & Daradoumis, A. (2022). Educational AI chatbots for content and language integrated learning. *Applied Sciences*, 12(7), 3239. <https://doi.org/10.3390/app12073239>
- Nemieboka, T. F., Onyenwe, I. E., & Asogwa, D. C. (2024). Development of an NLP-driven computer-based test guide for visually impaired students. *arXiv preprint arXiv:2401.12375*.
- Okonkwo, C. W., & Ade-Ibijola, A. (2020). Python-bot: A chatbot for teaching Python programming. *Engineering Letters*, 29(1), 346–355.
- Rosruen, N., & Samanchuen, T. (2018). Chatbot utilization for medical consultant system. In *Proceedings of the 3rd Technology Innovation Management and Engineering Science International Conference (TIMES-iCON)*, 1–5. IEEE.
- Ruan, S., Jiang, L., Xu, J., Tham, B. J., Qiu, Z., Zhu, Y., Murnane, E. L., Brunskill, E., & Landay, J. A. (2019). QuizBot: A dialogue-based adaptive learning system for factual knowledge. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–13). ACM. <https://doi.org/10.1145/3290605.3300587>

- Ruan, S., Willis, A., Xu, Q., Davis, G. M., Jiang, L., Brunskill, E., & Landay, J. A. (2019). BookBuddy: Turning digital materials into interactive foreign language lessons through a voice chatbot. *In Proceedings of the Sixth (2019) ACM Conference on Learning@ Scale* (pp. 1–4). ACM.