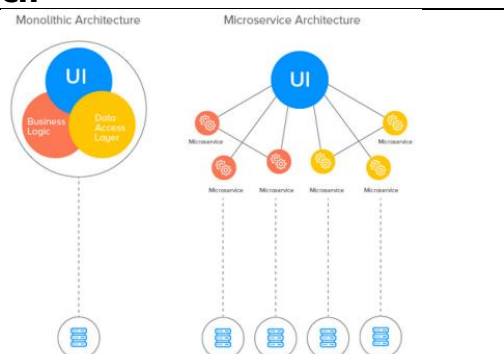




Katedra Systemów i Sieci Radiokomunikacyjnych

Zespół projektowy: ID-196	1. Maciej Dziewit - kierownik 2. Paweł Sienkiewicz
Opiekun:	dr inż. Agnieszka Czapiewska
Klient:	Paweł Czapiewski (Solution sp. z o.o.)
Data zakończenia:	24.01.2025
Słowa kluczowe:	LLM, ocena kompetencji, zadania, prompt engineering, miary, metoda S.T.A.R., wywiady behawioralne



TEMAT PROJEKTU:

System AI do weryfikacji umiejętności miękkich oparty na generacyjnej sztucznej inteligencji do tworzenia zadań w sesjach AC/DC

TEZA BADAWCZA, CELE I ZAKRES PROJEKTU:

Projekt ma na celu stworzenie, wdrożenie i ocenę systemu do weryfikacji umiejętności miękkich kandydatów. Ten system ma wykorzystywać generacyjną sztuczną inteligencję do automatycznego tworzenia zadań podczas sesji ACDC (ang. Assessment Center/Development Center).

Projekt jest realizowany dla firmy Solution.zoo firma zajmuje się w głównej mierze doradztwem personalnym i szkoleniami menedżerskimi. Jednym z ich produktów jest sesja AC/DC. Sesje ACDC to złożony proces oceny umiejętności, w którym uczestnicy są obserwowani i oceniani przez zespół wyszkolonych asesorów podczas wykonywania wielu zadań.

Celem projektu jest utworzenie systemu, który ma wykorzystywać generacyjną sztuczną inteligencję do automatycznego tworzenia zadań podczas sesji AC/DC. System ma się opierać na dużych modelach językowych

Docelowo system ma zastąpić asesorów obiektywną sztuczną inteligencją.

Proces oceny przez taki system wygląda w sposób następujący:

- Przesyłanie video przez kandydata**
Kandydat nagrywa trzy krótkie filmy (maksymalnie 3 minuty każdy), odpowiadając na następujące pytania:
 - Opisz trudny problem, który udało Ci się rozwiązać.
 - Opisz sytuację, w której Twoje sugestie usprawniły proces lub stworzyły optymalizację.
 - Opisz sytuację, w której musiałeś wykonać zadanie we współpracy z innymi osobami.
- Przesyłanie video**
Kandydat przesyła pliki video na Google Drive.
- Pobieranie plików**
Program uzyskuje dostęp do plików video na Google Drive za pomocą API i pobiera je.
- Transkrypcja**
Program transkrybuje pliki video (w formacie MP4) do plików tekstowych i zapisuje je w wyznaczonym folderze.

TEAM PROJECT INFORMATION FOLDER – JUNE 2006

5. Ocena kompetencji

Model językowy (LLM) ocenia odpowiednie kompetencje, analizując transkrypcje za pomocą starannie opracowanych promptów.

6. Zapisanie wyników

Program zapisuje wyniki oceny w pliku Excel (XLSX). Wyniki zawierają oceny LLM oraz dowody w postaci fragmentów odpowiedzi kandydata.

7. Porównanie z ocenami ludzkimi

Wyniki generowane przez różne modele LLM są porównywane z ocenami wykonanymi przez ludzkich oceniających. Porównanie to pomaga ocenić wydajność modeli w różnych metrykach.

OSIĄGNIĘTE REZULTATY:

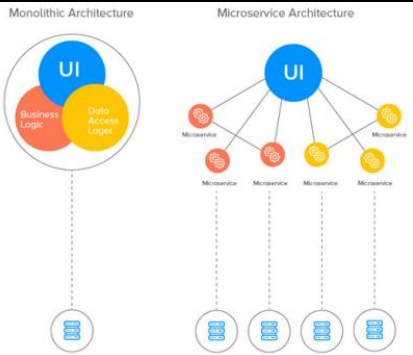
1. Analiza dostępnych modeli językowych
2. Wykonanie systematycznego przeglądu literatury, oraz sporządzenie tabeli z przejrzanymi pracami i ich krótkim opisem
3. Zapoznanie się oraz wzięcie udziału w sesji AC/DC
4. W ramach ćwiczenia wykonano fine-tuning modelu językowego
5. Przygotowanie środowiska i narzędzi programistycznych
6. Napisanie programu do pobierania i transkrypcji plików wideo na pliki tekstowe z dysku Google
7. Napisanie programu wyszukującego i oceniającego kompetencje miękkie kandydatów w ich wypowiedziach z wykorzystaniem LLM.
 - Przy użyciu platformy Open AI
 - Przy użyciu platformy HuggingFace
8. Porównanie oceny LLM z oceną asesora
9. Porównanie różnych LLM w miarach.
10. Wyznaczenie najlepszych LLM do tego zadania

CECHY CHARAKTERYSTYCZNE ROZWIĄZANIA, KIERUNKI DALSZYCH PRAC:

1. Wybór konkretnego modelu językowego
2. Fine-tuning modelu językowego
3. Analiza odpowiedzi modelu
4. Wywiady behawioralne, konstruowanie zadań do sesji AC/DC
5. Konstruowanie i ocenianie zadań przy użyciu metody S.T.A.R.
6. Transkrypcja plików wideo na pliki tekstowe przy użyciu modeli językowych
7. Prompt engineering – konstruowanie odpowiednich zapytań
8. Skorzystanie z różnych platform i bibliotek w celu przetestowania różnych LLM
9. Porównanie i wyznaczenie najlepszych modeli do takiego typu zadania



Department of Software Engineering

Project team: ID-196	1. Maciej Dziewit - leader 2. Paweł Sienkiewicz 3. 4.	
Supervisor:	dr inż. Agnieszka Czapiewska	
Client:	Paweł Czapiewski (Solution sp. z o.o.)	
Date:	24.01.2025	
Key words:	LLM, competency assessment, tasks, prompt engineering, measures, S.T.A.R. method , behavioral interviews	

PROJECT TITLE:

AI system for verifying soft skills based on generative artificial intelligence to create tasks in AC/DC sessions

RESEARCH THESIS, OBJECTIVES AND SCOPE:

The project aim is to create, implement and evaluate a system for verifying candidates soft skills. This system is intended to use generative artificial intelligence to automatically create tasks during AC/DC (Assessment Center/Development Center) sessions.

The project is carried out for Solution.zoo, a company that deals mainly with personnel consulting and managerial training. One of their products is the AC/DC session. These sessions are a complex skills assessment process in which participants are observed and assessed by a team of trained assessors as they perform multiple tasks.

The aim of the project is to create a system that will use generative artificial intelligence to automatically create tasks during AC/DC sessions. The system is to be based on large language models

As a result the system is intended to replace assessors with objective artificial intelligenceThe evaluation process is outlined below:

1. Candidate Video Submission

The candidate is tasked with recording three short videos (maximum 3 minutes each) answering the following questions:

- Describe a difficult problem that you managed to solve.
- Describe a situation in which your suggestions improved a process or created an optimization.
- Describe a situation where you had to complete a task in collaboration with other people.

2. Video Upload

The candidate uploads the video files to Google Drive.

3. File Download

The program accesses the video files on Google Drive using the API and downloads them.



4. Transcription

The program transcribes the video files (in MP4 format) into text files and saves them in a designated folder.

5. Competency Assessment

The LLM evaluates the relevant competencies by analyzing the transcriptions using carefully crafted prompts.

6. Results Compilation

The program saves the evaluation results in an Excel (XLSX) file. The results include the LLM's assessments and supporting evidence in the form of excerpts from the candidate's responses.

7. Comparison with Human Assessments

The results generated by different LLMs are compared with assessments made by human assessors. This comparison helps evaluate the performance of the models across various metrics.

RESULTS:

1. Analysis of available language models
2. Conducting a systematic review of the literature and preparing a table with the reviewed works and their short description
3. Getting to know and taking part in the AC/DC session
4. As part of the exercise, fine-tuning of the language model was performed
5. Preparation of the environment and programming tools
6. Writing a program to download and transcribe video files into text files from Google Drive
7. Writing a program that searches for and assesses candidates' soft skills in their statements using LLM.
 - Using the Open AI platform
 - Using the HuggingFace platform
8. Comparison of the LLM assessment with the assessor's assessment
9. Comparison of different LLMs in measures.
10. Appointing the best LLMs for the job

MAIN FEATURES, FUTURE WORKS:

1. Selection of a specific language model
2. Fine-tuning of the language model
3. Analysis of the model's responses
4. Behavioral interviews, designing tasks for AC/DC sessions
5. Designing and evaluating tasks using the S.T.A.R. method
6. Transcription of video files into text files using language models
7. Prompt engineering – constructing appropriate queries
8. Utilizing various platforms and libraries to test different LLMs



9. Comparison and identification of the best models for this type of task