

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

Zespół projektowy:

Maciej Dziewił 185755 - Kierownik

Paweł Sienkiewicz 185698

Opiekun Projektu: Agnieszka Czapiewska

Katedra Systemów i Sieci

Radiokomunikacyjnych

Zgłaszający Projekt: Paweł Czapiewski

(Solution sp. z o.o.)

ID Projektu - 196

Firma Solution

- Doradztwo personalne i szkolenia menedżerskie
- Ocenianie kompetencji miękkich kandydatów
- Prowadzenie szkoleń w tym zakresie
- Sesje AC/DC



Sesja AC/DC

- **ASSESSMENT CENTRE / DEVELOPMENT CENTRE**
- **Testowanie kompetencji miękkich kandydatów**
 - Ocena behawioralna
 - Symulacje
 - Ćwiczenia
 - Testy
- **Uczestnicy są obserwowani i oceniani przez zespół wyszkolonych asesorów**

Cel Projektu

Zaprojektowanie systemu, wykorzystującego duże modele językowe do oceny kompetencji miękkich kandydatów.

Osiągnięcia

- Zebranie danych testowych w postaci plików video
- Program dokonujący transkrypcji tekstu z plików video
- Program dokonujący oceny transkrypcji przy pomocy LLM
- Porównanie odpowiedzi modeli językowych z odpowiedziami profesjonalnego asesora
- Ocena jakości odpowiedzi modeli

Sposób oceny wyników modeli

		Klasa rzeczywista	
		pozytywna	negatywna
Klasa predykowana	pozytywna	prawdziwie pozytywna (TP)	fałszywie pozytywna (FP)
	negatywna	fałszywie negatywna (FN)	prawdziwie negatywna (TN)

$$\text{Dokładność} = \frac{TP + TN}{TP + TN + FP + FN}$$

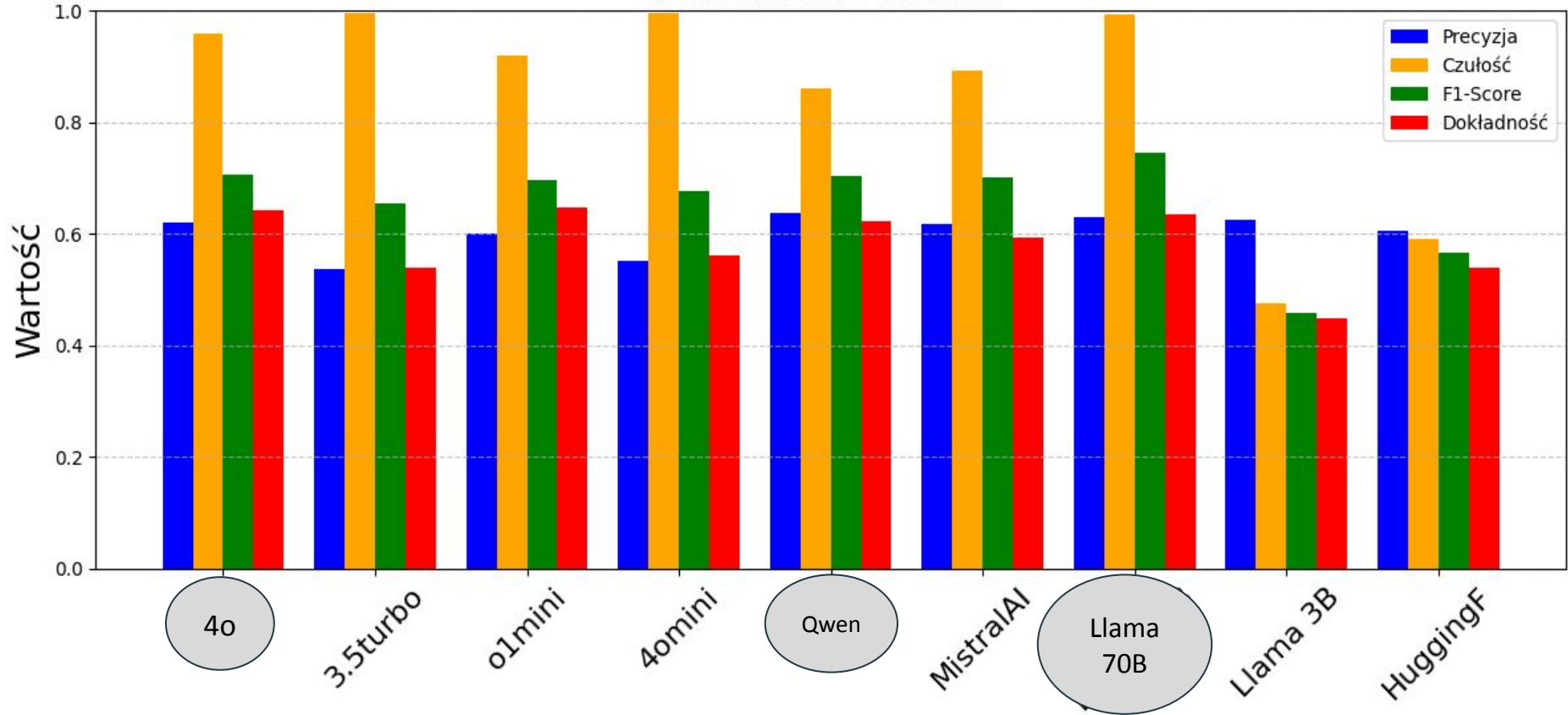
$$\text{Precyzja} = \frac{TP}{TP + FP}$$

$$\text{Czułość} = \frac{TP}{TP + FN}$$

$$F1 = 2 \times \frac{\text{Precyzja} \times \text{Czułość}}{\text{Precyzja} + \text{Czułość}}$$

Wyniki

Porównanie modeli



Napotkane problemy

- **Używanie dużych modeli lokalnie = ogromne wymagania sprzętowe**
- **Próba samodzielnego Fine Tuningu modeli = ogromne wymagania sprzętowe**
- **Rozwiązaniem było skorzystanie z API i gotowych modeli**

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

Maciej Dziewit, Paweł Sienkiewicz, Agnieszka Czapiwska, Paweł Czapiwski

Abstract—To evaluate the soft skills of high-level job candidates, a system is needed that reliably assesses their wide range of competencies based on standardized tasks. Assessment Center/Development Center (AC/DC) sessions are precisely such a system. These sessions consist of tasks based on hypothetical situations, and assessors assess competencies based on the candidate's responses. This process is quite lengthy and costly because of the involvement of several qualified individuals who can assess the candidate's skills.

The development of artificial intelligence and various large language models allows for the assessment of people's competencies without human involvement. We utilized LLM's (such as ChatGPT) to evaluate candidates' responses and determine whether they possess a particular skill. The model was provided with knowledge necessary for assessing a person and checking if he has a given skill.

The LLMs were tested using a sample task that could be part of an AC/DC session. The candidate was required to record 3 videos, each answering one of three questions.

For each task, specific traits were assigned to determine whether the candidate demonstrated a given soft skill. Based on an assessment guide provided by the client and knowledge of behavioral interviews and the S.T.A.R. method, appropriate prompts were designed for the LLMs.

The tasks were also evaluated by an assessor, whose evaluations served as a reference standard against which all models were compared. Based on this comparison, the best performing model was identified by evaluating the LLMs using established metrics.

Index Terms—LLM, natural language processing, assessment, prompt engineering.

I. INTRODUCTION TO LARGE LANGUAGE MODELS

Since the invention of the Transformer architecture, the development of neural networks has accelerated significantly. The Transformer architecture, with its self-attention mechanism, enabled parallel data processing and efficient handling of long-term dependencies in text. This groundbreaking solution replaced the previously dominant Recurrent Neural Networks and Long Short-Term Memory, which were limited in terms of scalability and computational efficiency [1]. This enabled the development of Large Language Models (LLMs).

LLMs are distinguished by their ability to acquire so-called emergent capabilities, which were not present in smaller models. These include:

In-context Learning: The ability to learn new tasks from a small set of examples provided within the query. Instruction Following: After fine-tuning, LLMs can execute new types

of tasks without requiring additional training. Multi-step Reasoning: The capability to solve complex problems by breaking them down into logical steps [3].

Solutions such as BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer) gained popularity due to their outstanding performance in natural language processing (NLP) tasks. These models made it possible to perform tasks such as translation, text generation, sentiment analysis, and question answering with unprecedented effectiveness [2].

II. DIFFERENCE BETWEEN TRADITIONAL INTERVIEW AND BEHAVIORAL INTERVIEW

Traditional interviews and behavioral interviews are distinct methods of evaluating job candidates, each with its own focus and methodology. Traditional interviews focus on general job descriptions, resumes, technical credentials, and hypothetical questions. They often rely on unstructured formats, with questions varying between candidates. This approach can lead to subjective assessments and inconsistent comparisons, making it challenging to predict future performance accurately. Candidates may provide rehearsed answers, further complicating the evaluation process [4].

Behavioral interviews, on the other hand, focus on assessing the past behaviors of candidates to predict their future performance. This method uses structured questions directly related to the competencies required for the role, based on the notion that past behavior is the best predictor of future behavior. Instead of hypothetical scenarios, candidates are asked to recount specific experiences using the STAR (Situation, Task, Action, Result) method, which provides concrete insight into their skills, decision-making processes, and final results [4]. The time distribution between these approaches differs considerably. Traditional interviews devote substantial time to discussing credentials and hypothetical questions, while behavioral interviews prioritize probing past behaviors and relevant experiences. The structured nature of behavioral interviews reduces biases, ensures consistency, and helps uncover deeper insights into the motivations and work style of a candidate [4].

Compared to traditional interviews, behavioral interviews align employees with the company's tactical and strategic goals, effectively communicating the expectations for both individual and organizational success throughout the organization. Despite that behavioral interviews demand more time



Dziękujemy za uwagę

Maciej Dziewit 185755

Paweł Sienkiewicz 185698

