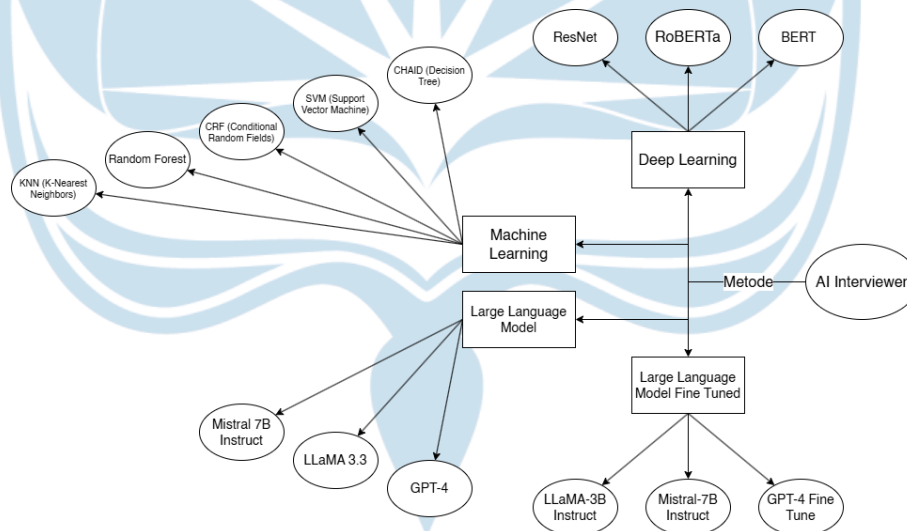


## BAB II

### TINJAUAN PUSTAKA

Pengembangan sistem rekrutmen magang berbasis web dengan integrasi *AI Interviewer* memerlukan pemahaman mendalam tentang berbagai metode dan teknologi yang relevan. Mulai dari konsep dasar dalam *Machine Learning* dan *Deep Learning*, hingga *Large Language Model* seperti GPT-4 dan LLaMA. Berbagai algoritma klasifikasi, seperti *Support Vector Machine* (SVM) dan *Random Forest*, *Bidirectional Encoder Representations from Transformers* (BERT), menjadi fondasi dalam pemrosesan data dan pengambilan keputusan. Untuk memberikan gambaran yang lebih jelas mengenai keterkaitan antara berbagai metode dan teknologi yang akan dieksplorasi dalam pengembangan sistem ini, pada Gambar 2.1.



**Gambar 2.1 Berbagai Teknologi dan Metode Pengembangan *AI Interviewer***

Gambar 2.1 di atas mengilustrasikan secara komprehensif berbagai metode dan model yang relevan dalam pengembangan sistem *AI Interviewer*. Secara garis besar, pendekatan yang dapat diterapkan terbagi menjadi tiga kategori utama: *Machine Learning*, *Deep Learning*, dan *Large Language Model*. Ketiga kategori ini memiliki sub-metode dan model yang spesifik, seperti yang digambarkan, yang

dapat dimanfaatkan untuk dalam tahapan pembuatan wawancara berbasis AI. Pembahasan selanjutnya akan menguraikan lebih lanjut setiap metode dan model yang disebutkan, serta relevansinya dalam konteks *AI Interviewer*.

BERT sebagai pendekatan baru dalam representasi bahasa alami dengan melakukan *pre-training* secara *bidirectional* dari teks tanpa label. Teknologi yang digunakan mencakup dua metode utama yaitu *Masked Language Modeling* (MLM) dan *Next Sentence Prediction* (NSP), yang dilatih pada korpus Wikipedia dan BookCorpus. Dalam pengujian pada *benchmark Natural Language Processing* (NLP), BERTLARGE mencetak skor F1 sebesar 93.2 pada SQuAD v1.1, mengungguli model terbaik sebelumnya dengan peningkatan +1.5 poin dalam *mode ensemble* dan +1.3 poin sebagai model tunggal. Pada SQuAD v2.0, BERTLARGE mencapai skor F1 83.1, meningkat +5.1 poin dibandingkan sistem sebelumnya. Sementara itu, pada *benchmark General Language Understanding Evaluation* (NSP), BERTLARGE meraih skor 80.5, meningkat +7.7 poin dari model sebelumnya, dan pada task *Multi-Genre Natural Language Inference* (MNLI), akurasi naik 4.6% menjadi 86.7% [4]. Pencapaian ini menunjukkan kemampuan BERT dalam memahami konteks dua arah secara mendalam dan menerapkannya dalam berbagai tugas NLP. Dalam konteks *AI interviewer*, kemampuan BERT dalam menangkap hubungan antar kalimat dan memahami makna dalam percakapan sangat bermanfaat untuk menginterpretasi respons kandidat secara akurat, merancang pertanyaan lanjutan yang relevan, serta mengevaluasi jawaban secara kontekstual dan objektif.

*Answer Sentence Selection* (AS2) memanfaatkan informasi kontekstual lokal dan global dalam arsitektur *Transformer*, khususnya menggunakan model *Robustly Optimized BERT Pretraining Approach* (RoBERTa). Teknologi yang dikembangkan dalam penelitian ini mencakup penyisipan konteks lokal berupa kalimat sebelum dan sesudah kalimat kandidat jawaban. Model mereka diuji pada tiga dataset benchmark: ASNQ, WikiQA, dan SQuAD. Hasil eksperimen menunjukkan bahwa integrasi konteks lokal (LOC T) memberikan peningkatan akurasi yang signifikan dibanding model *baseline*, dengan peningkatan *Mean*

*Average Precision* sebesar +6% pada WikiQA dan +5% pada SQuAD dibandingkan dengan Transformer biasa. Selain itu, model kombinasi lokal dan global (DUAL-CTX) mencatat *Precision 1* sebesar 0.952 pada SquAD [5]. Dalam konteks *AI interviewer*, pendekatan ini sangat relevan karena memungkinkan sistem untuk tidak hanya memahami pertanyaan dan jawaban secara eksplisit, tetapi juga menangkap makna tersembunyi dari konteks sekitarnya.

*Multimodal feature extraction* mencakup ekspresi wajah, intonasi, serta fitur leksikal (jumlah kata, topik, *filler words*) yang diperoleh dari 138 sesi wawancara mahasiswa *Massachusetts Institute of Technology* (MIT). Teknologi yang digunakan mencakup kombinasi *Support Vector Regression* dan *LASSO regression*, serta analisis statistik korelasi dan AUC. Hasil eksperimen menunjukkan bahwa model mampu memprediksi performa wawancara secara keseluruhan dengan korelasi  $r > 0.65$  dan  $AUC = 0.81$ , serta prediksi pada aspek sosial seperti engagement, excitement, dan friendliness dengan korelasi  $r \geq 0.75$  dan  $AUC > 0.85$  [6]. Dalam konteks *AI interviewer*, pendekatan ini memberikan fondasi kuat untuk merancang sistem wawancara otomatis yang tidak hanya memahami isi jawaban, tetapi juga dapat mengevaluasi kualitas komunikasi interpersonal secara objektif, termasuk aspek ekspresi emosi, keterlibatan, dan kefasihan verbal kandidat.

*Multimodal behavioral analytics* merupakan teknologi yang memanfaatkan tiga sumber data utama, yaitu fitur audio, video. Teknologi yang digunakan mencakup *Random Forest*, *Support Vector Classifier*, *Multitask Lasso*, dan *Multilayer Perceptron* untuk memprediksi skor pada 9 label performa wawancara, seperti *eye contact*, *engagement*, dan *speaking rate*. Hasil eksperimen menunjukkan bahwa model Random Forest menghasilkan akurasi tertinggi untuk 8 dari 9 label, dengan akurasi mencapai 96.43% pada *speaking rate*, 82.14% pada *not stressed* dan *pauses*, serta 75% pada *engagement*. Pendekatan ini relevan untuk sistem *AI Interviewer*, karena memungkinkan penilaian wawancara yang lebih objektif, tidak hanya dari isi jawaban, tetapi juga dari sinyal-sinyal non-verbal dan gaya komunikasi kandidat secara keseluruhan.

Mistral-7B-*Instruct* yang dirancang untuk menjawab pertanyaan berbasis teks secara efisien menggunakan model LLM. Teknologi yang digunakan mencakup berbagai pendekatan seperti *supervised fine-tuning*, serta *Retrieval Augmented Generation*. Hasil eksperimen menunjukkan bahwa akurasi meningkat dari baseline 30,89% menjadi 49,13% setelah dilakukan kombinasi *fine-tuning* pada IT Datasets [7]. Dalam konteks *AI Interviewer*, pendekatan seperti ini sangat berguna untuk menghasilkan pertanyaan wawancara yang akurat, serta mengevaluasi jawaban kandidat secara kontekstual.

Sistem dengan GPT-4 *Fine-Tuned* bernama *Interview AI-assistant* yang dirancang untuk mendukung kolaborasi manusia-AI secara *real-time* selama proses wawancara, baik dalam tahap persiapan maupun pelaksanaan. Penelitian ini menggunakan pendekatan beberapa tahap, mencakup empat studi terintegrasi: eksplorasi kebutuhan pewawancara, pengembangan *prototype* sistem berbasis AI untuk latihan wawancara, eksperimen evaluasi efektivitas sistem AI selama wawancara nyata, dan studi lapangan untuk pengujian sistem dalam konteks dunia nyata. Teknologi yang digunakan memanfaatkan *Large Language Models* untuk mendukung tugas-tugas kognitif pewawancara seperti menyusun pertanyaan lanjutan, menjaga kelancaran dialog, dan memproses informasi. Hasil awal dari studi pertama mengungkap bahwa pewawancara mengharapkan untuk mendukung tujuan penelitian dan membantu dalam komunikasi interpersonal. Selain itu, hasil eksperimen menunjukkan bahwa *AI assistant* mampu meningkatkan kualitas pertanyaan, efektivitas *follow-up*, serta efisiensi wawancara, khususnya bagi pewawancara dalam mengelola dinamika percakapan secara *real-time* [8].

Pada Tabel 2.1 disajikan keseluruhan perbandingan pada penelitian sebelumnya yang telah dijelaskan.

**Tabel 2.1 Perbandingan Penelitian Sebelumnya**

| Sumber | Metode dan Model           | Kelebihan Utama                                     | Kekurangan                                                             | Keunggulan Penelitian                                             |
|--------|----------------------------|-----------------------------------------------------|------------------------------------------------------------------------|-------------------------------------------------------------------|
| [4]    | BERT                       | Pemahaman konteks dua arah ( <i>bidirectional</i> ) | Butuh banyak data label; tidak langsung untuk <i>generation</i> dialog | Fine-tuned pada dialog dan dapat melakukan wawancara              |
| [5]    | RoBERTa                    | Memanfaatkan konteks paragraf & dokumen             | Fokus pada seleksi jawaban (AS2), bukan generasi pertanyaan            | Menyediakan <i>generation &amp; evaluasi</i> jawaban terintegrasi |
| [6]    | SVM Multimodal             | Analisis perilaku verbal & non-verbal               | Hanya prediksi performa, tidak mensimulasikan dialog                   | Menghasilkan dialog interaktif sekaligus menilai performa         |
| [6]    | CRF + RF/SVM               | Ekstraksi entitas & kutipan kuat dalam transkrip    | Perlu <i>pipeline</i> multimodal kompleks;                             | Pipeline <i>one-stop: generation</i>                              |
| [7]    | Mistral-7B <i>Instruct</i> | Model ringan; fine-tuning domain-spesifik tanpa RAG | Skor QA masih di kisaran 30–40%; belum dirancang untuk wawancara       | Dialog natural lebih panjang                                      |
| [8]    | GPT-4 <i>Fine-Tuned</i>    | Kolaborasi <i>real-time</i> ;                       | Sumber daya mahal;                                                     | Dijalankan <i>on-premise</i>                                      |

| Sumber | Metode dan Model | Kelebihan Utama              | Kekurangan               | Keunggulan Penelitian                          |
|--------|------------------|------------------------------|--------------------------|------------------------------------------------|
|        |                  | <i>summarization support</i> | bergantung API eksternal | (LLaMA), hemat <i>resource</i> , kontrol penuh |

