

## BAB II

### TINJAUAN PUSTAKA

#### 2.1 Penelitian Terdahulu

Bagian ini menguraikan kajian-kajian sebelumnya yang berkaitan dengan topik pengembangan generator soal pilihan ganda berbasis SLM dalam konteks pendidikan berbasis kearifan lokal. Sejumlah riset telah menyelidiki penggunaan model *GPT-2*, *ELECTRA-Small*, serta *ALBERT-Base-v2*. Hasil dari studi-studi tersebut mengindikasikan bahwa SLM yang telah melalui proses *fine-tuning* pada dataset khusus domain dapat menghasilkan performa optimal meskipun dengan keterbatasan sumber daya komputasi, sekaligus membuka peluang untuk interpretabilitas mekanistik serta tingkat efisiensi yang sangat tinggi. Akan tetapi, hampir semua penelitian tersebut masih terpusat pada bidang medis, berita, atau tugas-tugas umum, dan belum ada yang secara spesifik mengarahkannya untuk penyusunan soal pendidikan yang berakar pada nilai-nilai lokal desa dengan dataset berukuran sangat kecil yang bersumber dari transkrip video. Tabel 2.1 di bawah ini menyajikan ringkasan dari sepuluh penelitian terdahulu yang menjadi dasar teori dan metodologi bagi penelitian ini, termasuk temuan utama serta perbedaan dengan pendekatan yang diusulkan dalam studi ini.

Tabel 2. 1 Penelitian Terdahulu

No	Tahun	Judul	Peneliti	Model SLM	Hasil
1.	2025	<i>GPT-D Inducing Dementia-related Linguistic Anomalies by Deliberate Degradation of Artificial Neural</i>	Changye Li et al.	GPT-2	GPT-D dapat mengidentifikasi gangguan linguistik terkait demensia menggunakan pendekatan model degradasi, memberikan hasil yang efektif dalam mengidentifikasi perubahan bahasa pada penderita demensia.

No	Tahun	Judul	Peneliti	Model SLM	Hasil
		<i>Language Models</i> [11]			
2.	2025	<i>Distilling Reasoning Capabilities into Smaller Language Models</i> [12]	Bastian Bunzeck & Sina Zarrieß	GPT-Wee	Meningkatkan kemampuan reasoning multi-langkah pada model bahasa kecil melalui penyempurnaan spesialisasi model.
3.	2021	<i>It's Not Just Size That Matters: Small Language Models Are Also Few-Shot Learners</i> [13]	Timo Schick, Hinrich Schütze	ALBERT with PET/iPET	Mengungguli GPT-3 pada SuperGLUE dengan 32 contoh hanya menggunakan 0,1% parameternya; mencapai kinerja yang mirip dengan GPT-3 dengan model yang jauh lebih kecil melalui pertanyaan isian singkat dan optimasi berbasis gradien.
4.	2024	<i>Weak-to-Strong Search Align Large Language Models via Searching over Small Language Models</i> [14]	Zhanhui Zhou et al.	GPT-2, GPT-3.5	Meningkatkan kemampuan model besar dengan pencarian berbasis model kecil yang tidak diubah, meningkatkan kinerja tanpa pelatihan ulang.
5.	2022	<i>Prompting ELECTRA: Few-Shot Learning with Discriminative Pre-Trained Models</i> [15]	Mengzhou Xia, Mikel Artetxe, Jingfei Du, Danqi Chen, Ves Stoyanov	ELECTRA	Mengungguli BERT dan RoBERTa masing-masing sebesar 10,2 dan 3,1 poin secara rata-rata di 9 tugas dengan opsi token tunggal untuk model berukuran dasar; juga mengungguli RoBERTa pada 4

No	Tahun	Judul	Peneliti	Model SLM	Hasil
					tugas dengan opsi multi-token; selaras lebih baik dengan tugas hilir melalui pra-pelatihan diskriminatif.
6.	2023	<i>Unleash GPT-2 Power for Event Detection</i> [16]	Zhang et al.	GPT-2 teacher-student (GPTEDOT )	Arsitektur teacher-student GPTEDOT melampaui baseline pada beberapa dataset: pada ACE 2005, precision 82,3, recall 76,3 dan F1 79,2; pada CySecED, precision 65,9, recall 64,1, F1 65,0; pada RAMS, precision 55,5, recall 78,6, F1 65,1.
7.	2023	<i>Generative approach to Aspect Based Sentiment Analysis with GPT Language Models</i> [17]	Stanislav Chumakov et al.	GPT-3, GPT-2	Mengembangkan pendekatan generatif untuk analisis sentimen berbasis aspek menggunakan GPT, meningkatkan kinerja analisis sentimen meskipun dengan dataset yang lebih kecil.
8.	2022	<i>Depression Classification From Tweets Using Small Deep Transfer Learning Language Models</i> [18]	Muhammad Rizwan, Muhammad Faheem Mushtaq, Urooj Akram, Arif Mehmood, Imran Ashraf,	Electra Small Generator (ESG), Electra Small Discriminator (ESD), XtremeDistil-L6 (XDL), Albert Base V2 (ABV)	ESG mencapai skor F1 terbaik sebesar 89% untuk klasifikasi intensitas depresi multikelas; mengungguli DistilBert dalam hal kinerja dan waktu pelatihan; cocok untuk perangkat berdaya rendah dengan <15 juta parameter.

No	Tahun	Judul	Peneliti	Model SLM	Hasil
			Benjamín Sahelices		
9.	2023	<i>GPT-Wee How Small Can a Small Language Model Really Get</i> [19]	Bastian Bunzeck & Sina Zarrieß	GPT-Wee	Menunjukkan bahwa model GPT kecil dapat melakukan tugas pemrosesan bahasa yang kompleks meskipun ukurannya jauh lebih kecil dibandingkan model GPT besar.
10.	2023	<i>Assessing the Efficacy of the ELECTRA Pre-Trained Language Model for Multi-Class Sarcasm Subcategory Classification</i> [20]	Imogen Jones	ELECTRA	Menunjukkan kemanjuran dalam klasifikasi subkategori sarkasme multi-kelas; ELECTRA yang disempurnakan mengungguli model dasar pada tugas deteksi sarkasme, mencapai akurasi tinggi dalam membedakan jenis sarkasme dengan kebutuhan sumber daya yang rendah.

C. Li et al., “*GPT-D: Inducing Dementia-related Linguistic Anomalies by Deliberate Degradation of Artificial Neural Language Models*,” in Proc. Conf. Assoc. Comput. Linguist. Meet., vol. 2022, pp. 1866–1877, May 2022, mengupas strategi degradasi model *GPT-2* untuk mendekripsi anomali linguistik terkait demensia. Dengan baseline *perplexity* pada teks sehat vs. demensia, model *GPT-D* yang didegradasi berhasil mencapai *state-of-the-art* pada tugas “*Cookie Theft*” dengan generalisasi baik ke percakapan spontan. Melalui pendekatan *perplexity* ratio antara model asli dan terdegradasi, penelitian ini berhasil menginduksi anomali bahasa demensia, dengan peningkatan performa hingga mencapai hasil kompetitif tanpa memerlukan

dataset besar. Hasil ini menguatkan ide bahwa degradasi model dapat mensimulasikan gangguan kognitif untuk deteksi dini, memberikan potensi besar dalam penelitian medis dan aplikasi klinis. Selain itu, kemampuan *GPT-D* dalam mendeteksi pola bahasa yang halus memungkinkan penerapan lebih lanjut dalam deteksi dini gangguan bahasa lainnya.

Keberhasilan pendekatan ini juga terlihat dalam beberapa studi lainnya, terutama yang memanfaatkan *GPT-2*. Penelitian oleh A. Pouran Ben Veyseh et al., “*Unleash GPT-2 Power for Event Detection*,” menunjukkan bagaimana arsitektur *teacher-student* berbasis *GPT-2* berhasil melampaui model *BERT-ED* dengan F1 79,2 pada dataset ACE 2005 melalui distilasi data sintetis. Bastian Bunzeck and Sina Zarrieß, “*GPT-wee: How Small Can a Small Language Model Really Get?*,” menunjukkan bahwa meskipun dengan ukuran yang lebih kecil, model *GPT* dapat melakukan *reasoning* multi-langkah dengan hasil yang kompetitif dibandingkan model yang lebih besar. Penggunaan teknik distilasi untuk menurunkan kemampuan *reasoning* dari model besar ke model kecil memperlihatkan bahwa model yang lebih kecil masih dapat menyelesaikan tugas yang kompleks, seperti *reasoning* matematis atau penalaran berbasis langkah. K. Shridhar et al., “*Distilling Reasoning Capabilities into Smaller Language Models*,” memperlihatkan bahwa distilasi *Socratic CoT* dari *GPT-2* ke model kecil meningkatkan performa *reasoning* lebih dari 70% pada dataset GSM8K. Z. Zhou et al., “*Weak-to-Strong Search: Align Large Language Models via Searching over Small Language Models*,” menunjukkan bahwa *weak-to-strong* search menggunakan *GPT-2* sebagai guide meningkatkan *win rates* hingga 37,9% pada *Llama-3-70B-Instruct*. Ketiga penelitian ini menunjukkan bahwa *GPT-2*, melalui *fine-tuning*, distilasi, dan interpretabilitas, mampu memberikan performa tinggi pada tugas yang sangat spesifik seperti deteksi peristiwa, prediksi akronim, dan rekomendasi, namun tetap efisien dalam penggunaan sumber daya.

Di sisi lain, penggunaan *GPT/ChatGPT* dalam generasi teks dan analisis juga telah terbukti efektif, seperti yang terlihat pada penelitian oleh Dian

Nurdiana et al., “*Generative Artificial Intelligence as an Evaluator and Feedback Tool in Distance Learning: a Case Study on Law Implementation*,”. Dalam penelitian ini, pendekatan *few-shot* dan *fine-tuning* berbasis *GPT* berhasil melakukan evaluasi tugas hukum dengan akurasi *binary* 80% dan F1-score tinggi. Sementara itu, Muhammad Rizwan et al., “*Depression Classification From Tweets Using Small Deep Transfer Learning Language Models*,” in Proc. IEEE Access, vol. 2022, pp. 1–14, 2022, membandingkan *GPT* dengan model lain untuk klasifikasi depresi, dengan *GPT* menunjukkan *BERTScore* tinggi dan menghindari bias. Kedua penelitian ini menegaskan bahwa *GPT/ChatGPT* memiliki fleksibilitas yang luar biasa dalam berbagai aplikasi, baik untuk analisis sentimen maupun ringkasan teks, dan pentingnya memahami mekanisme internal model ini untuk meningkatkan hasil yang diinginkan.

Selanjutnya, penelitian oleh Timo Schick and Hinrich Schütze, “*It’s Not Just Size That Matters: Small Language Models Are Also Few-Shot Learners*,” in Proc. Conf. Assoc. Comput. Linguist., 2021, pp. 1–14, tentang model kecil yang juga bisa *few-shot learner* menunjukkan bahwa meskipun dengan ukuran yang lebih kecil, model kecil dapat melakukan penalaran multi-langkah dengan hasil yang kompetitif dibandingkan model lebih besar. Penggunaan teknik distilasi untuk menurunkan kemampuan reasoning dari model besar ke model kecil memperlihatkan bahwa model yang lebih kecil masih dapat menyelesaikan tugas yang kompleks, seperti *reasoning* matematis atau penalaran berbasis langkah. Hal ini menjadi sangat relevan dalam konteks pengembangan model-model dengan efisiensi komputasi yang lebih baik. Keberhasilan ini menunjukkan potensi besar model kecil untuk digunakan dalam aplikasi dengan keterbatasan komputasi, seperti perangkat dengan kapasitas rendah atau aplikasi berbasis perangkat mobile, yang memungkinkan mereka untuk menyelesaikan tugas-tugas berat dengan sumber daya yang terbatas.

Penelitian oleh Mengzhou Xia et al., “*Prompting ELECTRA: Few-Shot Learning with Discriminative Pre-Trained Models*,” in Proc. Conf. Assoc. Comput. Linguist., 2022, pp. 1–11, dalam *Application of Transformer-Based Deep Learning Models for Predicting the Suitability of Water for Agricultural Purposes* mengadaptasi *ALBERT Base v2* untuk klasifikasi air layak minum, kemudian mengusulkan varian *ALBERT-WPD* yang dioptimalkan. Model *baseline* menghasilkan akurasi 91% dan F1-score 0,91. Setelah pelatihan lanjutan dan optimasi parameter, *ALBERT-WPD* mencapai akurasi 95,88% ( $\approx 96\%$ ) serta meningkatkan presisi dan *recall*. Ini menunjukkan bahwa modifikasi sederhana pada arsitektur SLM dapat meningkatkan akurasi prediksi potabilitas air.

Imogen Jones, “*Assessing the Efficacy of the ELECTRA Pre-Trained Language Model for Multi-Class Sarcasm Subcategory Classification*,” in Proc. Int. Conf. Comput. Linguist., 2023, pp. 1–62, menunjukkan bahwa *ELECTRA* efektif untuk klasifikasi sarkasme multi-kelas.

Penelitian ini berbeda dari penelitian-penelitian tersebut karena berfokus pada konteks lokal Desa Pete dan edukasi berbasis kearifan lokal. Dataset yang digunakan berasal dari tiga video tutorial budidaya cabai dengan total durasi sekitar 27 menit; transkripsinya dibersihkan dan dipilih sepuluh kalimat kunci untuk dijadikan pasangan konteks–pertanyaan. Model yang dilatih merupakan *Small Language Model GPT-2 Small*, *ELECTRA-Small*, dan *ALBERT-base-v2* yang ditune-ulang dengan parameter kecil *5 epoch*, *batch size 2*, *learning rate 5e-5* supaya cocok dijalankan pada perangkat dengan sumber daya terbatas. Selain itu, sistem ini diintegrasikan ke website resmi Desa Pete melalui framework *Laravel* dan *FastAPI* sehingga pengguna dapat mengakses generator soal MCQs langsung dari tombol di situs desa. Data lokal seperti kegiatan budidaya cabai dan penduduk dimasukkan ke dalam basis data kecamatan. Integrasi ini menjadikan laporan ini pionir dalam menerapkan pipeline *CRISP-DM* secara utuh untuk pembuatan soal MCQs berbasis kearifan lokal yang dapat berjalan di lingkungan low-resource, menawarkan kontribusi

praktis bagi Desa Pete serta bukti teoretis bahwa SLM kecil cukup andal untuk konteks tersebut.

## 2.2 Tinjauan Teori

### 2.2.1 Multiple Choice

Soal pilihan ganda merupakan salah satu bentuk penilaian yang paling sering dipakai dalam dunia pendidikan karena kemampuannya untuk mengukur beragam jenis pengetahuan serta tingkat kemampuan berpikir siswa secara efisien. Menurut Awan dkk. (2023), soal *MCQ* yang disusun dengan baik mampu mengevaluasi berbagai level keterampilan kognitif, mulai dari sekadar mengingat fakta hingga kemampuan analisis yang lebih rumit, yang selaras dengan kerangka *taksonomi Bloom* dalam perumusan tujuan pembelajaran [21]. Akan tetapi, mutu soal pilihan ganda menjadi faktor penentu keberhasilannya sebagai alat evaluasi. Kesalahan dalam penyusunan, seperti pilihan pengganggu yang tidak logis atau kalimat soal yang bersifat negatif, dapat mengurangi keabsahan hasil penilaian dan memengaruhi prestasi siswa secara tidak adil. Lebih lanjut, penelitian terkini oleh Oc dan Hassen (2025) menekankan perbedaan antara soal MCQ dengan satu jawaban benar dan soal dengan beberapa jawaban benar, di mana soal jenis MAQ terbukti lebih menantang serta mendorong pemahaman yang lebih dalam, meskipun sekaligus meningkatkan tingkat kecemasan dan kesulitan bagi peserta didik [22]. Walaupun terdapat kendala tersebut, soal MAQ mampu memberikan gambaran penilaian yang lebih menyeluruh karena merangsang kemampuan berpikir kritis dan penerapan ilmu pengetahuan, yang sangat dibutuhkan untuk mengembangkan keterampilan kognitif tingkat lanjut. Dengan demikian, penyusunan soal MCQ yang teliti menjadi hal yang sangat krusial agar soal tersebut tidak hanya bersifat adil, tetapi juga efektif secara pedagogis dalam mendukung pencapaian hasil belajar siswa.

### 2.2.2 Desa Wisata

Sejumlah desa kini aktif mengembangkan konsep desa wisata sebagai upaya untuk meningkatkan taraf ekonomi masyarakat sekaligus menjaga kelestarian budaya setempat. Di Desa Nyatnyono, pemanfaatan teknologi digital melalui pembuatan situs web dan akun media sosial telah berhasil digunakan untuk memasarkan wisata religi serta kuliner khas daerah tersebut [23]. Sementara itu, kajian di Desa Cupak menekankan betapa krusialnya keterlibatan aktif warga masyarakat, pemenuhan elemen-elemen utama pariwisata seperti atraksi, akses transportasi, fasilitas penginapan, pelayanan, serta infrastruktur pendukung, di samping pengelolaan yang baik terhadap tradisi lokal [24]. Pada Desa Giri di Gresik, hasil pemetaan sumber daya manusia mengungkapkan bahwa keberhasilan pengembangan desa wisata sangat ditentukan oleh kualitas sumber daya manusia, baik dari segi kecerdasan, nilai spiritual, maupun integritas yang mampu memberikan pelayanan ramah dan berbobot karakter [25]. Karenanya, pengembangan desa wisata perlu mengintegrasikan teknologi, memberdayakan masyarakat, serta terus meningkatkan kapasitas dan kualitas sumber daya manusia.

### 2.2.3 Artificial Intelligence

Kecerdasan Buatan (*Artificial Intelligence*) merupakan cabang ilmu yang berupaya membangun sistem atau mesin yang mampu meniru kemampuan manusia dalam belajar, memecahkan masalah, serta mengenali pola-pola tertentu. Tinjauan literatur menunjukkan bahwa teknologi seperti jaringan saraf tiruan, logika *fuzzy*, sistem hibrida, serta berbagai bidang pendukung seperti sistem pakar, pemrosesan bahasa alami, dan robotika, menjadi dasar utama dalam mewujudkan kemampuan tersebut; bahkan AI sering dianggap lebih konsisten dan tahan lama dibandingkan kecerdasan manusia dalam menjalankan tugas tertentu [26]. Dalam ranah pemasaran digital, AI memanfaatkan teknologi-teknologi tersebut untuk mengolah data dalam jumlah besar, menyesuaikan konten secara personal, serta menargetkan iklan dengan

tepat guna meningkatkan interaksi dan keterlibatan pelanggan [27]. Institusi pendidikan dan organisasi lainnya dapat memanfaatkan AI untuk menciptakan proses pembelajaran yang lebih personal dan adaptif, sementara AI sendiri mendapatkan manfaat dari pengelolaan serta penyediaan data oleh institusi tersebut [28].

Tinjauan pustaka juga mengungkapkan bahwa AI telah mengubah lanskap pelayanan kesehatan melalui otomatisasi tugas-tugas administratif, percepatan proses diagnosis, serta peningkatan mutu perawatan secara keseluruhan [29]. Namun, penerapan teknologi ini juga membawa tantangan etika dan sosial, khususnya terkait dengan perlindungan privasi serta keamanan data pasien. Oleh sebab itu, Villegas-Ch beserta timnya mengusulkan sebuah kerangka kerja yang menekankan pentingnya perlindungan data, penguatan sistem keamanan, serta penyesuaian dengan peraturan yang berlaku guna mencegah manipulasi dan penyalahgunaan data [30]. Secara keseluruhan, berbagai literatur menggarisbawahi bahwa manfaat besar dari AI harus diimbangi dengan pendekatan yang kokoh dalam aspek keamanan, privasi, serta pertimbangan etika.

#### 2.2.4 Fine Tuning

Fine-tuning merupakan teknik penyesuaian model yang sudah *pre-trained model* menggunakan dataset besar, agar dapat disesuaikan dengan tugas yang lebih spesifik melalui dataset yang lebih kecil namun relevan. Pendekatan ini mengubah model yang telah menguasai pola-pola umum dari data dalam jumlah besar menjadi lebih sesuai dan optimal untuk kebutuhan tertentu, seperti dalam bidang pemrosesan bahasa alami maupun pengenalan gambar. Contohnya, model-model seperti *BERT* atau *GPT* yang telah dilatih dengan korpus data masif dapat diadaptasi lebih lanjut melalui *fine-tuning* guna memenuhi persyaratan khusus, sehingga mampu menghasilkan output yang lebih akurat dalam

ruang lingkup yang lebih terbatas, misalnya penyusunan soal pilihan ganda di dunia pendidikan atau aplikasi di sektor kesehatan [31].

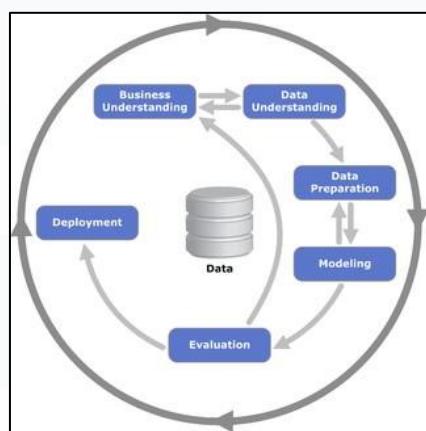
1. *Pre-trained Model* berfungsi sebagai fondasi utama dalam proses *fine-tuning* karena model tersebut sudah memiliki pemahaman dasar yang kokoh terhadap pola bahasa atau gambar secara umum [32]. Dengan memanfaatkan model yang sudah siap pakai ini, kita tidak perlu melakukan pelatihan dari nol, yang jelas membutuhkan waktu lebih lama serta sumber daya komputasi yang jauh lebih besar. *Fine-tuning* memungkinkan kita untuk menyesuaikan model tersebut hanya dengan data yang lebih sedikit namun lebih relevan dengan tugas yang diinginkan, sehingga menghasilkan model yang lebih efektif tanpa harus membangun semuanya dari awal.
2. *Data Efficient Fine-Tuning* menitikberatkan pada optimalisasi penggunaan data yang minim dengan memilih sampel-sampel yang paling representatif dan berpengaruh besar [31]. Teknik ini bertujuan untuk memangkas biaya pelatihan dengan hanya menggunakan data yang benar-benar memberikan dampak signifikan terhadap performa model, sehingga proses *fine-tuning* menjadi lebih cepat dan hemat sumber daya. Pendekatan ini memungkinkan pengembangan model berkualitas tinggi meskipun data yang tersedia terbatas, sekaligus memperpendek durasi pelatihan tanpa mengorbankan kualitas hasil akhir model.

## 2.3 Framework/Algoritma yang digunakan

### 2.3.1 Crisp DM

Metode *Cross-Industry Standard Process for Data Mining* (CRISP-DM) telah diterapkan secara meluas di berbagai sektor untuk mengatur alur kerja dalam kegiatan data mining. Sebagai contoh, Abdurazzaq dkk.

(2025) memanfaatkan CRISP-DM untuk menyempurnakan sistem pencarian informasi kesehatan, dengan mengintegrasikan kemiripan semantik dan *Large Language Models* guna mendeteksi penyakit [33]. Mereka mengikuti enam tahapan CRISP-DM, mulai dari pemahaman bisnis untuk menentukan konteks serta sasaran aplikasi, hingga tahap penerapan model dalam deployment.



Gambar 2. 1 Kerangka CRISP-DM [32]

Gambar 2.1 mengilustrasikan siklus CRISP-DM yang dimulai dari pemahaman bisnis hingga tahap penerapan, yang dapat diterapkan secara langsung dalam penelitian ini. Melalui pendekatan tersebut, transkrip dapat diolah untuk menghasilkan soal yang sesuai dengan kurikulum serta tujuan pembelajaran, tanpa bergantung pada sistem SLM eksternal. Di samping itu, penelitian Cevallos (2024) tentang pengklasifikasian berita digital dengan teknik machine learning juga menggunakan CRISP-DM sebagai kerangka metodologi untuk mengelola keseluruhan proses dari awal hingga selesai [34]. Tahapannya dimulai dengan pemahaman data melalui pengumpulan berita dari beragam sumber, penyimpanan dalam data lake, pra-pemrosesan, pemodelan, serta evaluasi. Pendekatan ini relevan dengan penelitian saya, di mana CRISP-DM diterapkan untuk mengekstraksi informasi dari transkrip budidaya cabai, yang kemudian diolah lebih lanjut guna menghasilkan soal MCQ. Tahapan persiapan data dan pemodelan dalam CRISP-DM akan menjadi landasan utama

dalam menyiapkan data transkrip serta menerapkan model untuk menciptakan soal yang sesuai dengan tingkat kesulitan dan sasaran pembelajaran siswa.

### 2.3.2 Small Language Model

*Small Language Models* hadir sebagai jawaban atas kendala komputasi serta isu privasi yang sering dihadapi oleh *Large Language Models*. SLM dibuat dengan ukuran yang jauh lebih ringkas, sehingga dapat dijalankan secara lebih hemat dan efektif pada perangkat yang memiliki kapasitas terbatas [35]. Berbagai teknik kompresi seperti *pruning*, *knowledge distillation*, serta kuantisasi diterapkan untuk mengurangi ukuran model tanpa mengorbankan performa secara berarti. Dalam penggunaan praktis, SLM menawarkan kelebihan berupa kecepatan inferensi yang lebih tinggi dan biaya operasional yang lebih murah, menjadikannya pilihan yang sangat cocok untuk tugas-tugas yang membutuhkan adaptasi cepat dan efisien, terutama pada perangkat seluler atau sistem berbasis *edge computing* [36]. Meskipun SLM tidak memiliki kemampuan emergent yang dimiliki oleh LLM berskala besar, mereka tetap memberikan nilai tambah signifikan dalam hal penghematan sumber daya dan perlindungan privasi data, yang semakin krusial dalam aplikasi khusus seperti sektor kesehatan maupun keuangan.

### 2.3.3 GPT 2

*GPT-2* merupakan model bahasa berbasis arsitektur *Transformer* yang telah terbukti andal dalam menangani beragam tugas pemrosesan bahasa alami, termasuk prediksi teks dan bahkan kemampuan dasar dalam perhitungan matematika seperti membandingkan angka (“lebih besar dari”), walaupun model ini tidak dilatih secara khusus untuk tugas tersebut [37]. Kajian menunjukkan bahwa *GPT-2* mampu mengenali hubungan antar angka dengan memanfaatkan komponen internalnya, seperti *multi-layer perceptrons* dan mekanisme *attention* [38]. Kemampuan ini sangat bermanfaat untuk menyusun soal pilihan ganda

bagi siswa SD, khususnya pada soal-soal yang menguji logika serta konsep matematika sederhana seperti perbandingan bilangan. Dengan demikian, *GPT-2* dapat dimanfaatkan untuk menciptakan pertanyaan yang mampu mengukur kemampuan berpikir kritis dan analitis anak-anak, sekaligus memudahkan guru dalam menghasilkan soal yang sesuai dengan tingkat pemahaman mereka.

#### 2.3.4 Electra-small

*ELECTRA-small* merupakan model bahasa yang dirancang khusus untuk *pre-training* dengan tingkat efisiensi tinggi dalam memanfaatkan data pelatihan. Berbeda dengan model seperti *BERT* yang menerapkan *masked language modeling*, *ELECTRA-small* menggunakan teknik *replaced token detection*, di mana sejumlah token pada input sengaja diganti dengan token yang tidak tepat, dan model dilatih untuk mengenali mana token asli dan mana yang telah diganti. Strategi ini memungkinkan model belajar representasi kata dengan lebih cepat karena token yang diganti memberikan sinyal pelatihan yang lebih kaya dibandingkan hanya memprediksi token yang yang disembunyikan [39]. Dalam penerapannya untuk penyusunan soal pendidikan, *ELECTRA-small* sangat cocok untuk menghasilkan pertanyaan *Multiple Choice* yang menuntut analisis serta pemahaman mendalam. Model ini mampu menangkap konteks dengan baik dan menciptakan soal yang menguji kemampuan berpikir kritis siswa, termasuk jenis soal esai atau analisis yang melibatkan konsep-konsep abstrak [40]. Kemampuannya dalam menangani beragam tugas pemrosesan bahasa alami, seperti pengklasifikasian teks dan analisis konteks, menjadikannya instrumen yang sangat efektif untuk membangun sistem generator soal otomatis, sekaligus memberikan efisiensi baik dari segi durasi pelatihan maupun penggunaan sumber daya komputasi.

#### 2.3.5 Albert-Base-V2

*ALBERT-base-v2* adalah model *transformer* yang dibuat dengan fokus utama pada efisiensi komputasi melalui pengurangan jumlah parameter tanpa mengorbankan performa secara signifikan. Model ini telah terbukti sangat andal dalam berbagai tugas pemrosesan bahasa alami, termasuk deteksi kualitas air untuk keperluan pertanian dengan tingkat akurasi mencapai 96% [41]. Keunggulan efisiensinya dalam mengolah teks membuat *ALBERT-base-v2* sangat sesuai untuk menghasilkan soal pilihan ganda bagi siswa sekolah dasar. Model ini mampu menciptakan pertanyaan yang menguji pemahaman serta kemampuan analisis dalam bentuk pilihan ganda, dengan menggunakan bahasa yang sederhana dan sesuai dengan tingkat kognitif anak-anak, sehingga mendukung proses pembelajaran yang lebih bermakna. Selain itu, berkat kemampuannya memahami konteks secara mendalam, *ALBERT-base-v2* dapat menyusun soal yang tidak sekadar menguji hafalan, melainkan juga merangsang siswa untuk berpikir kritis dan kreatif, yang sangat esensial bagi pengembangan keterampilan berpikir tingkat tinggi sejak usia dini [42].

## 2.4 Tools/software yang digunakan

### 2.4.1 Jupyter Notebook

*Jupyter Notebook* adalah lingkungan komputasi interaktif yang menyatukan kode pemrograman, visualisasi data, serta narasi teks dalam satu file terpadu, menjadikannya alat yang sangat praktis untuk keperluan pengajaran dan eksplorasi ilmiah. Bascuña dkk. (2023) menjelaskan bahwa *Jupyter Notebook* mampu memperkaya proses belajar di ranah teknik kimia melalui pengalaman langsung yang mencakup pemodelan, representasi visual, dan komputasi *real-time*, sehingga membantu mahasiswa menangkap ide-ide abstrak dengan lebih mudah [43]. Tak hanya itu, Owusu dkk. (2022) dalam pengembangan *PyGEE-SWToolbox* membuktikan bahwa *Jupyter Notebook* juga sangat efektif untuk pemetaan serta pengolahan sumber daya air dengan memanfaatkan

*Google Earth Engine*, memungkinkan pengguna tanpa pengalaman *coding* mendalam untuk mengakses dan memproses data satelit secara sederhana [44]. Berkat fleksibilitasnya yang tinggi, *Jupyter Notebook* menjadi pilihan tepat dalam penelitian ini sebagai instrumen utama untuk mengelola transkrip budidaya cabai melalui siklus CRISP-DM guna menciptakan soal pilihan ganda yang relevan dengan kebutuhan pembelajaran di Desa Pete.

#### 2.4.2 Python

*Python* merupakan bahasa pemrograman tingkat tinggi yang populer di kalangan analis data dan pembuat perangkat lunak berkat sintaksnya yang intuitif serta koleksi pustaka yang sangat komprehensif. Correa de Barros dan Martins (2024) menyebutkan bahwa *Python* menjadi prioritas utama dalam pembuatan aplikasi karena kemampuannya menyederhanakan pengembangan solusi AI, terutama melalui otomatisasi kode dengan *Small Language Models* untuk mengoptimalkan produktivitas pengkodean [45]. Selain itu, Öztürk (2024) menguraikan penggunaan *Python* dalam *tools* seperti *XCompress*, yang mengintegrasikan kompresi teks berbasis SLM untuk menentukan algoritma kompresi paling optimal sesuai karakteristik data, menunjukkan adaptabilitas *Python* dalam solusi kompresi data yang lebih cepat dan hemat [46]. Dengan beragam pustaka siap pakai, seperti *NumPy* untuk operasi numerik dan *SciPy* untuk komputasi saintifik, *Python* memfasilitasi penerapan teknik analisis serta pengolahan data yang disesuaikan dengan tuntutan khusus di berbagai bidang, termasuk riset dan pengajaran [47]. Kelebihan ini sangat selaras dengan pembuatan *tools* serta sistem dalam penelitian saya, seperti *generator* soal MCQ dari transkrip, yang mengandalkan kemudahan penyatuannya dengan pustaka dan algoritma Python untuk pengolahan data serta pembangunan model.

#### 2.4.3 Visual Studio Code

*Visual Studio Code* adalah editor kode *open-source* yang ringan dengan fitur *IntelliSense*, pemantauan kesalahan terpadu, serta ribuan *plugin* pendukung, menjadikannya favorit untuk menggarap aplikasi *Laravel* dengan pola *MVC* [48]. *VS Code* juga memfasilitasi arsitektur berbasis model melalui pengeditan metamodel *Ecore* dan aturan konversi *ATL* guna menghasilkan kode *Laravel* secara otomatis sambil tetap taat pada prinsip *MVC2* [49]. Pada penelitian ini, *VS Code* dipakai sebagai lingkungan inti untuk membangun backend *FastAPI*, mengatur model *GPT-2 Small*, serta menyambungkan *API* ke tampilan *Streamlit*, sekaligus menjadi basis transisi ke *Laravel* demi memudahkan kelanjutan pengembangan aplikasi “Belajar Seru Bersama AI” oleh mahasiswa Program Studi Sistem Informasi.

#### 2.4.4 Laragon

*Laragon* adalah *software* pengembangan web lokal untuk sistem *Windows* yang menyediakan *server all-in-one* untuk menjalankan *PHP*, *MySQL*, serta *Apache/Nginx* secara bersinergi, memungkinkan pengembang menyusun dan menguji aplikasi web seperti pengelolaan data anggota organisasi tanpa ketergantungan pada server luar pada fase pengembangan awal [50]. Lewat kemampuan seperti *auto-virtualhost*, *URL* ramah, dan *SSL* otomatis, *Laragon* mempercepat siklus pengembangan lengkap dari tahap perencanaan sampai pengujian, serta menyederhanakan perpindahan ke *hosting* produksi berkat pengaturan mudah dan portabilitas unggul, sehingga sangat sesuai untuk proyek web berbasis pengumpulan data yang menekankan keamanan serta kelincahan dalam pengelolaan database *MySQL/MariaDB*.