

BAB I PENDAHULUAN

A. Latar Belakang

Pendidikan merupakan unsur penting dalam kehidupan manusia yang tidak dapat dipisahkan dari keberlangsungan hidup dan pembangunan suatu negara. Pendidikan adalah sebuah cahaya penerang yang berusaha menuntun manusia dalam menentukan arah, makna, dan tujuan kehidupan ini. Melalui pendidikan, manusia dapat memperoleh pengetahuan dan keterampilan yang diperlukan untuk meningkatkan kualitas hidupnya, mengembangkan diri, dan mencapai tujuan hidupnya. Pendidikan juga merupakan sebuah kebutuhan mutlak yang harus dipenuhi oleh umat manusia sepanjang hayat. Tanpa adanya pendidikan, mustahil bagi suatu kelompok manusia untuk hidup berkembang sejalan dengan aspirasi (cita-cita) untuk maju. Oleh karena itu, penting bagi setiap individu untuk memiliki kesempatan yang sama dalam mendapatkan akses dan pendidikan yang berkualitas.

Pendidikan tidak hanya berhenti pada pengajaran di dalam kelas, namun juga terkait dengan proses penilaian dalam pembelajaran. Enny, dkk (2019) menyampaikan bahwa penilaian adalah salah satu aspek penting dalam pembelajaran. Informasi yang diperoleh dari hasil penilaian merupakan acuan untuk mengevaluasi keberhasilan proses pembelajaran dalam era 4.0, peserta didik dihadapkan dengan tantangan persaingan yang semakin ketat dan kompleks. Oleh karena itu, tugas guru dalam membantu peserta didik bersiap menghadapi persaingan di era 4.0 adalah dengan membantu mengajarkan berpikir kritis, berpikir kreatif, analitis dan kemampuan pemecahan masalah pada peserta didik

Namun, kemampuan berpikir tingkat tinggi pada siswa sekolah di Indonesia masih belum memadai. Berdasarkan hasil penilaian Program PISA (2023) Indonesia berada di peringkat 68. Dalam kategori Sains, Indonesia mendapatkan skor 383, turun dari nilai 396 pada penilaian tahun

2018. Selain itu, dalam kategori Matematika, Indonesia berada di peringkat ketujuh dari bawah dengan skor 366, turun dari nilai 379 pada tahun 2018. Skor terendah tercatat dalam kategori Membaca, yaitu 359, turun dari nilai 371 pada tahun 2018. Berdasarkan hasil survei ini, skor PISA Indonesia harus ditingkatkan melalui kegiatan pembelajaran dan penilaian yang berorientasi pada keterampilan berpikir tingkat tinggi.

Untuk mengembangkan kemampuan berfikir *high order thinking skill* (HOTS) pada siswa, guru harus menggunakan soal-soal sekolah yang dapat membantu mengembangkan pola *high order thinking skill* (HOTS) pada siswa. HOTS akan meningkatkan kemampuan siswa dalam mengkonstruksi atau menyusun argumen yang tepat dan efektif untuk membuat keputusan atau solusi yang rasional (Herawati, 2021). Fatimah & Pahlevi (2020) menjelaskan bahwa kemampuan berpikir kritis peserta didik dapat diasah dengan menggunakan logika dan kemampuan bernalar untuk menyelesaikan masalah dan membuat keputusan. Sedangkan, kemampuan berpikir kreatif peserta didik dapat dilatih dengan menciptakan ide, gagasan, dan konsep yang baru.

Penyusunan soal *high order thinking skill* (HOTS) menjadi suatu hal yang cukup sulit bagi para guru dalam proses pembelajaran siswa. Seiring dengan adanya kurikulum 2013 yang mengedepankan pengembangan kemampuan berpikir tingkat tinggi, guru dituntut untuk mampu menyusun soal HOTS dengan baik dan benar. Namun, hal ini tidak mudah dilakukan karena banyak faktor yang menjadi kendala bagi guru dalam menyusun soal HOTS. Kemampuan guru dalam membuat soal HOTS masih rendah, baik soal bentuk esai ataupun pilihan ganda (Chandra & Heryadi, 2020).

Penelitian oleh Winarti, dkk (2021) menunjukkan meskipun guru-guru sudah diberikan pelatihan pembuatan soal evaluasi berbasis HOTS, namun guru-guru masih terkendala dalam pembuatan soal HOTS. Guru juga kesulitan dalam membedakan tingkat berpikir aspek kognitif, misalnya membedakan apakah soal tersebut termasuk soal pemahaman (C2) atau analisis (C4). Penelitian lain oleh Rahmatul & Ernawati (2020) mengenai

analisis level kognitif soal UAS geografi semester gasal kelas XI IPS SMA Negeri tahun pelajaran 2019/2020 juga menunjukkan bahwa persentase soal HOTS masih sangat rendah, hanya sebanyak 11% yang berada pada kategori C4 (menganalisis), C5 (mengevaluasi), dan C6 (mencipta), sementara persentase soal LOTS (C1) sebanyak 27% dan persentase soal MOTS (C2 dan C3) sebanyak 62%. Hal ini menunjukkan bahwa guru masih kesulitan dalam membuat soal HOTS. Hal ini menunjukkan bahwa penyusunan soal HOTS tetap menjadi tantangan bagi guru dan berpengaruh terhadap hasil soal yang dibuat oleh guru.

Salah satu pendekatan untuk memecahkan masalah dalam pembuatan soal adalah dengan mengembangkan pembuat soal otomatis oleh komputer. Berbagai pendekatan digunakan untuk mengembangkan pembuat soal otomatis. Penelitian yang dilakukan oleh Wyse & Piwek (2000) dan Heilman & Smith (2010) menggunakan pendekatan *rule-based* untuk membuat pertanyaan. Pendekatan *rule-based* mengandalkan ekstraksi sintaks dan semantik dari suatu teks dan *template* untuk membuat pertanyaan. Penelitian oleh Mulla & Gharpure (2023) menunjukkan bahwa pembuatan soal dengan pendekatan ini hanya dapat menargetkan topik yang spesifik apabila topik yang digunakan cenderung luas maka teks tidak dapat dikonversi secara efektif menjadi pertanyaan.

Penelitian oleh Flor & Riordan (2018) menggunakan pendekatan lain yaitu teknologi pemrosesan bahasa alami (*natural language processing*) untuk pengembangan pembuat pertanyaan otomatis. Peneliti mengembangkan pembuat pertanyaan otomatis berdasarkan pelabelan peran semantik (*semantic role labeling*) dan jaringan saraf tiruan (*artificial neural network*). Pendekatan ini menghasilkan pertanyaan langsung dari analisis semantis, tanpa menggunakan *template*. Peneliti menggunakan jaringan saraf buatan berbasis algoritma *long short-term memory* (LSTM) yang dilatih menggunakan korpus pasangan pertanyaan dan jawaban dari Wikipedia dan dataset SQuAD. Namun pendekatan ini memiliki

keterbatasan dimana konteks dari pertanyaan yang dibuat hanya terkait dengan dataset yang diberikan.

Perkembangan teknologi yang pesat melahirkan sebuah teknologi pemrosesan bahasa alami (*natural language processing*) baru yaitu ChatGPT, sebuah perangkat lunak besutan OpenAI yang merupakan model *large language model* (LLM) berbasis chatbot yang mampu menghasilkan respon teks yang sangat mirip dengan respon alami manusia (Dale, 2021). ChatGPT adalah perangkat lunak canggih yang mampu memenuhi berbagai permintaan berbasis teks. Mulai dari menjawab pertanyaan sederhana hingga tugas yang lebih sulit seperti membuat surat terima kasih atau membimbing individu dalam diskusi yang sulit tentang suatu masalah spesifik (X. Liu, dkk 2021) Kemampuan ChatGPT untuk menghasilkan bahasa yang mirip dengan manusia dan menyelesaikan tugas kompleks membuatnya menjadi inovasi yang signifikan dalam bidang pemrosesan bahasa alami dan kecerdasan buatan. ChatGPT memanfaatkan basis data yang luas dan desain model yang efisien untuk memahami dan menafsirkan permintaan dari pengguna dan kemudian menghasilkan tanggapan yang sesuai dalam bahasa yang sangat mirip dengan tanggapan manusia (B. D. Lund & Wang, 2023).

Beberapa penelitian mencoba mengimplementasikan ChatGPT ke dalam berbagai bidang. Sebuah artikel yang diterbitkan di *Nature* pada Februari 2023 (van Dis, dkk 2023) menjelaskan penggunaan ChatGPT oleh para ahli biologi komputasi sebagai reviewer makalah penelitian yang telah selesai. Dalam waktu lima menit, para ahli biologi menerima tinjauan naskah mereka yang meningkatkan tingkat keterbacaan naskah dan saran mengenai kesalahan dalam penulisan. Saat menguji naskah peneliti, ChatGPT tidak selalu merespon dengan baik, namun naskah yang dihasilkan lebih baik dibandingkan dengan naskah acuan. Penelitian lain oleh Akhter & Cooper (2023) menunjukkan bahwa ChatGPT dapat digunakan membuat rangkuman umum mengenai pengobatan oksigen hiperbarik untuk penyakit edema paru akut. Meskipun demikian terdapat

beberapa kekurangan ChatGPT, seperti pembuatan sitasi yang tidak dapat diakses dan kekurangan untuk mengakses informasi terbaru.

Penerapan ChatGPT dalam dunia pendidikan, khususnya untuk pembuatan soal otomatis dengan tingkat kemampuan berpikir tinggi (HOTS) pada mata pelajaran Geografi masih jarang dikaji. Oleh karena itu, penelitian mengenai pengembangan pembuat soal otomatis untuk menyusun soal *higher order thinking skill* pada mata pelajaran geografi sangat penting untuk memahami dan mengidentifikasi cara-cara untuk mengatasi hambatan yang ada serta memaksimalkan potensi teknologi ini dalam meningkatkan pembelajaran di lingkungan pendidikan. Dari seluruh penjelasan yang telah dijabarkan penulis melakukan penelitian dengan judul "Pengembangan Aplikasi Web EdTek Berbasis ChatGPT Sebagai Pembuat Soal Otomatis Untuk Menyusun Soal *High Order Thinking Skill* Pada Mata Pelajaran Geografi."

B. Rumusan Masalah

Berdasarkan latar belakang diatas, maka rumusan masalah dalam penelitian ini adalah sebagai berikut

1. Bagaiman penerapan model pengembangan ADDIE untuk mengembangkan aplikasi EdTek berbasis ChatGPT sebagai pembuat soal otomatis untuk menyusun soal *high order thinking skill* pada mata pelajaran geografi?
2. Bagaimana gambaran aspek *high order thinking skill* (HOTS) dari soal yang dihasilkan oleh aplikasi EdTek?
3. Bagaimana kualitas dari soal *high order thinking skill* yang dihasilkan oleh aplikasi EdTek?
4. Bagaimana penilaian aspek *usability* aplikasi EdTek dalam proses pembuatan dan penyusunan soal *high order thinking skill*?

C. Tujuan Penelitian

Berdasarkan pada rumusan penelitian yang telah dijabarkan, maka tujuan pada penelitian ini adalah sebagai berikut:

1. Mengetahui penerapan model pengembangan ADDIE untuk mengembangkan aplikasi EdTek berbasis ChatGPT sebagai pembuat soal otomatis untuk menyusun soal *high order thinking skill* pada mata pelajaran geografi
2. Menganalisis gambaran aspek *high order thinking skill* (HOTS) dari soal yang dihasilkan oleh aplikasi EdTek.
3. Menganalisis kualitas dari soal *high order thinking skill* yang dihasilkan oleh aplikasi EdTek
4. Menganalisis penilaian aspek *usability* aplikasi EdTek dalam proses pembuatan dan penyusunan soal *high order thinking skill*

D. Manfaat Penelitian

Penelitian ini memiliki 2 manfaat yaitu manfaat teoritis dan manfaat praktis. Adapun manfaat penelitian ini adalah sebagai berikut:

1. Manfaat Teoritis

Penelitian ini dapat memberikan wawasan mendalam tentang kualitas soal-soal HOTS yang dihasilkan oleh aplikasi EdTek, sehingga pengembang aplikasi dapat melakukan perbaikan dan penyesuaian yang diperlukan. Selain itu juga penelitian ini dapat memperoleh data empiris tentang bagaimana soal-soal yang dihasilkan sesuai dengan standar pendidikan, membantu meningkatkan kualitas pengajaran dan evaluasi di sekolah

2. Manfaat Praktis

- a. Guru

penelitian ini akan memberikan pemahaman tentang penggunaan aplikasi EdTek dalam pendidikan, serta memudahkan guru dalam membuat soal HOTS.

- b. Pemerintah

temuan dari penelitian ini dapat dijadikan dasar dalam pengambilan kebijakan yang mendukung penggunaan teknologi dalam pendidikan, serta membantu dalam pengembangan program

pelatihan bagi guru, yang pada akhirnya meningkatkan kualitas pendidikan secara keseluruhan

c. Masyarakat

Penelitian ini dapat mendorong adopsi teknologi pendidikan yang lebih luas, memberikan akses yang lebih merata kepada berbagai lapisan masyarakat, serta meningkatkan kualitas dan hasil belajar siswa.

E. Penelitian Terdahulu

1. Pertama, penelitian yang dilakukan oleh (Li & Xing, 2021) dengan judul "*Natural Language Generation Using Deep Learning to Support MOOC Learners*". Penelitian tersebut bertujuan untuk mengeksplorasi perkembangan model generasi bahasa alami (*natural language generation*) berbasis *deep learning*, khususnya *recurrent neural network* (RNN) dan *generative pretrained transformer 2* (GPT-2), dalam pemberian tanggapan yang mirip dengan tanggapan yang dibuat oleh manusia kepada siswa di forum diskusi MOOC. Penelitian tersebut dilakukan dalam kerangka teori dukungan sosial, dengan fokus pada penyediaan dukungan informasi, emosional, dan komunitas kepada siswa melalui model NLG di forum diskusi. Penelitian tersebut melibatkan pelatihan model RNN dan GPT-2 dengan 13.850 pasangan entri pesan-balasan dari forum diskusi MOOC. Evaluasi kuantitatif dilakukan terhadap kinerja model menggunakan metrik seperti kerancuan kata, keterbacaan, dan koherensi. Hasil pengukuran menunjukkan bahwa GPT-2 lebih unggul daripada RNN. Selain itu, penelitian ini juga melakukan survei terhadap partisipan untuk memvalidasi temuan-temuan yang ditemukan dalam penelitian ini. Hasil survei menunjukkan bahwa model GPT-2 dapat memberikan tanggapan yang mendukung dan kontekstual dengan tingkat yang sebanding dengan manusia.

2. Kedua, penelitian yang dilakukan oleh (Hu, dkk 2018). Penelitian yang berjudul “*Aspect-Based Question Generation*” bertujuan untuk membuat model jaringan saraf tiruan yang dapat menghasilkan pertanyaan sesuai dengan konteks yang ditanyakan sehingga dapat lebih relevan. Untuk menghasilkan pertanyaan peneliti menggunakan dataset dari korpus pertanyaan dan jawaban Amazon. Dataset ini berisi 1,4 juta pasangan pertanyaan dan jawaban tentang layanan dan produk Amazon. Peneliti kemudian melakukan pemilahan data dan menggunakan sekitar 900 ribu pasangan pertanyaan dan jawaban yang dibagi menjadi seribu pasangan pertanyaan dan jawaban untuk data tes, seribu pasangan pertanyaan dan jawaban untuk data pengembangan dan sisanya digunakan sebagai data latih. Peneliti mengevaluasi pertanyaan yang dihasilkan dengan metrik BLUE 1-4, METEOR, ROUGE_L serta evaluasi oleh penguji manusia. Peneliti mendapatkan skor 42,30 untuk BLUE 1, 26,23 untuk BLUE 2, 18,16 untuk BLUE 3, 13,09 untuk BLUE 4, 17,5 untuk METEOR, dan 42,48 Untuk ROUGE_L. Kemudian peneliti memilih 100 pasangan pertanyaan dan jawaban secara acak dan menyajikannya kepada 3 penguji manusia untuk menilai tata bahasa, kefasihan bahasa dan relevansi. dari pertanyaan yang dihasilkan. Nilai yang didapat adalah 68 dari 100. Hasil ini menunjukkan bahwa model yang digunakan memiliki kemampuan yang lebih baik dibandingkan dengan model sebelumnya.
3. Ketiga, penelitian yang dilakukan oleh (Basuki, S. & Kusuma, 2018) dengan judul “*Automatic Question Generation for 5W-1H Open Domain of Indonesian Questions By Using Syntactical Template-Based Features From Academic Textbooks*” Pada penelitian tersebut, peneliti mengembangkan Pembuat Soal Otomatis untuk soal bertipe W-H (*What, Why, When, Who, Where dan How*) dari materi berbahasa Indonesia. Peneliti menggunakan

978 kalimat yang terbagi ke dalam 8 jenis kalimat dan berasal dari 13 buku pembelajaran yaitu: Biologi, Fisika, Geografi, Sejarah, Bahasa Indonesia, Sosiologi, Ilmu Wayang, Teknik Sistem Antena, Produksi Pangan Buatan, Agribisnis, Pakan Ternak, Konservasi Tanah dan Air serta Teknik Penetasan Ikan. Sistem yang terdiri dari 2 tahap. Tahap pertama merupakan tahap klasifikasi menggunakan *machine learning* dan tahap kedua menggunakan pendekatan aturan tertulis. Pertanyaan yang dihasilkan kemudian dievaluasi berdasarkan format pertanyaan yang telah dibuat. Jika pertanyaan mengikuti format pertanyaan maka akan dinilai sebagai TRUE, jika tidak mengikuti maka akan dinilai sebagai FALSE dan jika pertanyaan tersebut tidak masuk dalam format pertanyaan namun dapat dimengerti oleh manusia maka dinilai sebagai *Understandable*. Peneliti menghasilkan 386 pertanyaan dengan tingkat akurasi sebesar 89,79%.

F. Definisi Operasional

Penelitian dengan judul “Pengembangan Aplikasi Web EdTek Berbasis ChatGPT Sebagai Pembuat Soal Otomatis Untuk Menyusun *Soal High Order Thinking Skill* Pada Mata Pelajaran Geografi” ini memiliki definisi pembahasan untuk menghindari kesalahpahaman dan perbedaan penafsiran. definisi tersebut dijabarkan antara lain sebagai berikut:

1. Kesesuaian Soal dengan Aspek *High Order Thinking Skill*.

Soal *high order thinking skill* merupakan instrumen pengukuran yang digunakan untuk mengukur kemampuan berpikir tingkat tinggi. Jawaban soal-soal HOTS tidak tersurat secara eksplisit dalam stimulus (Setiawati, dkk 2019). Aspek kemampuan berpikir tingkat tinggi termasuk kemampuan untuk memecahkan masalah, menganalisis, mengevaluasi, mencipta, penalaran dan logika, pengambilan keputusan, dan berpikir kreatif (Brookhart, 2010). Untuk mengukur kesesuaian digunakan instrumen penilaian

checklist terhadap kriteria dari aspek *high order thinking skill* menurut Brookhart.

2. Kualitas Soal dari Aplikasi EdTek

Pembuatan soal otomatis merupakan sebuah sistem dimana pertanyaan dihasilkan secara otomatis berdasarkan topik, ide, atau konteks yang terkandung dalam teks atau gambar. Tujuan utama sistem ini adalah mampu menghasilkan pertanyaan-pertanyaan yang bermakna dan benar baik secara sintaksis maupun semantik sesuai konteks penggunaannya (Mulla & Gharpure, 2023). Aplikasi EdTek dirancang sebagai aplikasi yang dapat menghasilkan soal secara otomatis berdasarkan panduan pengguna. Untuk mengukur kualitas pertanyaan yang dihasilkan pada tiga indikator yaitu tata bahasa, kefasihan bahasa dan relevansi. Pengguna diminta untuk memberikan berdasarkan skala 1 hingga 5 (nilai 5 menunjukkan nilai terbaik). Untuk penilaian keseluruhan menggunakan persentase antara 0-100

3. Aspek *Usability* dari Aplikasi EdTek

Pengujian aspek *usability* adalah pengujian terhadap aplikasi perangkat lunak yang berfokus pada kemudahan penggunaan sistem oleh pengguna (Wagner, 2013). Tujuan pengujian *usability* adalah memastikan bahwa setiap kategori mendukung antarmuka sistem dan memungkinkan pengguna untuk mempelajari serta menggunakan semua navigasi dan semantik. Untuk mengukur aspek *usability* menggunakan *USE Questionnaire*. *USE Questionnaire* mencakup empat aspek *usability* yaitu *usefulness*, *satisfaction*, *ease of learning*, dan *ease of use*. Kuesioner ini terdiri dari 30 pertanyaan yang terbagi ke dalam empat aspek tersebut. Pengguna diminta untuk memberikan berdasarkan skala 1 hingga 5 (nilai 5 menunjukkan nilai terbaik). Untuk penilaian keseluruhan menggunakan persentase antara 0-100.