

# Integrasi Gemini AI Berbasis *Retrieval-Augmented Generation (RAG)* pada Moodle untuk Penilaian Esai Otomatis dengan Pendekatan *Human-in-the-Loop* di Pendidikan Tinggi

Rizki Adha<sup>1</sup>, Lusianto<sup>2</sup>, Dodi Syaripudin<sup>3</sup>, Bobi Kurniawan S<sup>4</sup>, Adam Mukharil Bachtiar<sup>5</sup>, Hanhan Maulana<sup>6</sup>, Ednawati Rainarli<sup>7</sup>

<sup>1-7</sup>Doktor Sistem Informasi, Universitas Komputer Indonesia, Indonesia

Email: <sup>1</sup>rizkiadha67@gmail.com, <sup>2</sup>maslusianto@gmail.com, <sup>3</sup>dodisyaripudin@gmail.com,

<sup>4</sup>bobi@email.unikom.ac.id, <sup>5</sup>adam@email.unikom.ac.id, <sup>6</sup>hanhan@email.unikom.ac.id,

<sup>7</sup>ednawati.rainarli@email.unikom.ac.id

**Abstrak** - Penilaian esai merupakan komponen penting dalam evaluasi pembelajaran di pendidikan tinggi, namun proses penilaian manual oleh dosen membutuhkan waktu yang besar dan berpotensi menimbulkan inkonsistensi, terutama pada kelas dengan jumlah mahasiswa yang besar. Penelitian ini bertujuan untuk mengembangkan dan mengevaluasi sistem penilaian esai otomatis berbasis Gemini AI yang terintegrasi dengan *Learning Management System (LMS)* Moodle menggunakan pendekatan *Retrieval-Augmented Generation (RAG)* dan mekanisme *Human-in-the-Loop (HIL)*. Penelitian menggunakan metode *Research and Development (R&D)* dengan model ADDEI, yang meliputi tahap analisis, perancangan, pengembangan, evaluasi, dan implementasi sistem. Evaluasi sistem dilakukan melalui pengujian fungsional (*black-box testing*), serta kuesioner *Human-in-the-Loop* yang melibatkan 24 dosen dari 14 perguruan tinggi swasta di wilayah Banten, DKI Jakarta, dan Jawa Barat. Hasil pengujian menunjukkan bahwa seluruh fungsi sistem berjalan sesuai dengan spesifikasi. Evaluasi HIL menunjukkan tingkat penerimaan yang tinggi hingga sangat tinggi, terutama pada indikator peran AI sebagai *decision-support system* dan tanggung jawab akademik dosen. Selain itu, hasil validasi dosen menunjukkan bahwa sebagian besar rekomendasi skor dan umpan balik yang dihasilkan oleh sistem dapat diterima, dengan dosen tetap memiliki kendali penuh dalam menentukan nilai akhir. Temuan ini menunjukkan bahwa integrasi Gemini AI berbasis RAG dengan mekanisme *Human-in-the-Loop* efektif sebagai sistem pendukung penilaian esai yang efisien, akuntabel, dan sesuai dengan kebutuhan penilaian akademik di pendidikan tinggi.

**Kata Kunci** - Automated Essay Scoring, Gemini AI, Moodle, Retrieval-Augmented Generation, Human-in-the-Loop.

**Abstract** - Essay assessment is a critical component of evaluation in higher education; however, manual grading by lecturers is time-consuming and prone to inconsistency, particularly in large-enrollment courses. This study aims to develop and evaluate an automated essay assessment system integrated into Moodle Learning Management System using Gemini AI with a Retrieval-Augmented Generation (RAG) approach and a Human-in-the-Loop (HIL) mechanism. The study employed a Research and Development (R&D) method following the ADDEI model, encompassing analysis, design, development, evaluation, and implementation phases. System evaluation was conducted through functional testing (*black-box testing*), User Acceptance Testing (UAT), and a Human-in-the-Loop questionnaire involving 24 lecturers from 14 private higher education institutions in Banten, Jakarta, and West Java,

Indonesia. The results indicate that all core system functionalities operated as expected. The HIL evaluation demonstrated high to very high levels of lecturer acceptance, particularly regarding the role of AI as a decision-support system and the preservation of academic accountability. Furthermore, lecturer validation results show that most AI-generated scores and feedback were approved, while final grading authority remained fully under lecturer control. These findings suggest that integrating RAG-based Gemini AI with a Human-in-the-Loop mechanism provides an effective, reliable, and accountable solution for supporting essay assessment in higher education.

**Keywords** - Automated Essay Scoring, Gemini AI, Moodle, Retrieval-Augmented Generation, Human-in-the-Loop..

## I. PENDAHULUAN

Kemampuan berpikir kritis dan menulis analitis merupakan kompetensi esensial dalam pendidikan tinggi, sejalan dengan tuntutan Kurikulum Merdeka Belajar–Kampus Merdeka serta standar pembelajaran abad ke-21 [1]. Kompetensi tersebut menuntut mahasiswa tidak hanya menguasai konsep, tetapi juga mampu menyusun argumen, melakukan analisis, dan merefleksikan pemahaman secara logis dan terstruktur. Dalam konteks ini, penilaian berbasis esai dipandang sebagai instrumen evaluasi yang efektif karena mampu mengukur pemahaman konseptual dan kemampuan berpikir tingkat tinggi secara lebih komprehensif dibandingkan penilaian objektif [2].

Dalam praktiknya, penilaian esai pada sebagian besar perguruan tinggi di Indonesia masih dilakukan secara manual oleh dosen, dengan dosen bertindak sebagai penilai utama sekaligus pengambil keputusan akademik. Meskipun pendekatan ini menjaga kendali penuh dosen terhadap proses evaluasi, peningkatan jumlah mahasiswa dan kompleksitas tugas akademik seringkali berdampak pada tingginya beban kerja serta potensi inkonsistensi penilaian [3]. Beberapa studi juga menunjukkan bahwa penilaian esai manual bersifat subjektif dan dapat dipengaruhi oleh faktor non-akademik, sehingga berpotensi memengaruhi reliabilitas dan keadilan penilaian [4].

Seiring dengan percepatan digitalisasi pembelajaran, pemanfaatan Learning Management System (LMS) telah menjadi praktik umum di perguruan tinggi Indonesia. Moodle merupakan salah satu platform LMS yang paling banyak digunakan karena sifatnya yang fleksibel dan open-

source. Namun demikian, LMS umumnya masih berfungsi sebagai sarana pengelolaan pembelajaran dan belum menyediakan mekanisme penilaian esai otomatis yang mampu mengevaluasi jawaban mahasiswa secara kontekstual dan mendalam. Akibatnya, meskipun proses pembelajaran telah terdigitalisasi, penilaian esai tetap sepenuhnya bergantung pada penilaian manual oleh dosen.

Penelitian mengenai Automated Essay Scoring (AES) telah berkembang dari pendekatan Natural Language Processing (NLP) klasik hingga deep learning berbasis model pralatih seperti BERT dan RoBERTa [5]. Walaupun menunjukkan peningkatan performa, sistem AES konvensional masih memiliki keterbatasan dalam memahami kualitas argumentasi, koherensi naratif, dan konteks pembelajaran, khususnya untuk bahasa dengan sumber daya terbatas seperti Bahasa Indonesia [6]. Selain itu, sebagian besar penelitian AES lebih menekankan evaluasi performa algoritmik dan belum banyak mengkaji integrasi sistem secara langsung ke dalam LMS yang digunakan dalam praktik pendidikan tinggi.

Perkembangan Generative Artificial Intelligence melalui Large Language Models (LLM) seperti Gemini AI membuka peluang baru dalam penilaian esai karena kemampuannya dalam memahami konteks dan menghasilkan analisis teks tingkat lanjut [7]. Namun, penggunaan LLM secara langsung untuk penilaian akademik masih menghadapi tantangan, seperti potensi hallucination, kurangnya transparansi dasar penilaian, dan risiko ketidaksesuaian dengan materi serta rubrik yang ditetapkan dosen.

Untuk mengatasi tantangan tersebut, pendekatan Retrieval-Augmented Generation (RAG) dikembangkan dengan mengombinasikan kemampuan generatif LLM dan pengambilan informasi dari sumber pengetahuan yang terkurasi [8]. Dalam konteks penilaian esai, RAG memungkinkan sistem memanfaatkan materi pembelajaran dan rubrik penilaian yang disediakan dosen sebagai konteks sebelum proses inferensi, sehingga hasil penilaian menjadi lebih selaras dengan standar akademik.

Selain itu, penerapan AI dalam penilaian akademik perlu memastikan bahwa peran dosen sebagai otoritas akademik tetap terjaga. Oleh karena itu, pendekatan Human-in-the-Loop (HIL) digunakan untuk menempatkan AI sebagai sistem pendukung keputusan, di mana Gemini AI memberikan rekomendasi skor dan umpan balik awal, sementara dosen memiliki kewenangan penuh untuk menyetujui atau merevisi hasil penilaian sebelum nilai akhir diberikan kepada mahasiswa.

Hingga saat ini, penelitian yang secara sistematis mengembangkan dan mengevaluasi sistem penilaian esai otomatis berbasis Gemini AI dengan pendekatan RAG dan Human-in-the-Loop yang terintegrasi langsung ke dalam LMS Moodle, khususnya untuk esai berbahasa Indonesia, masih terbatas. Berdasarkan kesenjangan tersebut, penelitian ini bertujuan untuk mengembangkan, mengimplementasikan, dan mengevaluasi sistem penilaian esai otomatis berbasis Gemini AI yang terintegrasi dengan Moodle dengan pendekatan RAG dan Human-in-the-Loop.

Pertanyaan penelitian yang diajukan meliputi:

(RQ1) Bagaimana merancang arsitektur dan mekanisme integrasi Gemini AI berbasis RAG ke dalam LMS Moodle untuk mendukung penilaian esai otomatis?

(RQ2) Sejauh mana rekomendasi skor dan umpan balik yang dihasilkan sistem dapat diterima oleh dosen dalam kerangka Human-in-the-Loop?

(RQ3) Bagaimana tingkat penerimaan dosen terhadap pendekatan Human-in-the-Loop serta kinerja fungsional sistem berdasarkan pengujian black-box?

Secara akademis, penelitian ini memberikan kontribusi awal dalam integrasi Gemini AI dan pendekatan RAG pada sistem penilaian esai berbahasa Indonesia dengan perspektif Human-in-the-Loop. Secara praktis, penelitian ini diharapkan dapat meningkatkan efisiensi dan konsistensi penilaian esai, tanpa mengurangi peran dan tanggung jawab akademik dosen.

## II. METODE PENELITIAN

Penelitian ini menggunakan pendekatan Research and Development (R&D), yaitu metode penelitian yang bertujuan untuk mengembangkan, mengimplementasikan, dan mengevaluasi suatu produk yang dapat digunakan secara praktis dalam konteks pendidikan. Pendekatan R&D banyak digunakan dalam penelitian teknologi pendidikan karena memungkinkan peneliti tidak hanya mengkaji fenomena pembelajaran, tetapi juga menghasilkan solusi konkret yang dapat diadopsi dalam praktik pembelajaran nyata [9], [10].

Pendekatan ini dipilih karena penelitian tidak berfokus pada pengujian hipotesis semata, melainkan menghasilkan sebuah produk berupa sistem penilaian esai otomatis yang terintegrasi dengan Learning Management System (LMS) Moodle, sehingga relevan dengan kebutuhan pembelajaran digital di perguruan tinggi [11]. Dengan pendekatan R&D, sistem yang dikembangkan dapat diuji secara fungsional serta dievaluasi dari sisi penerimaan pengguna.

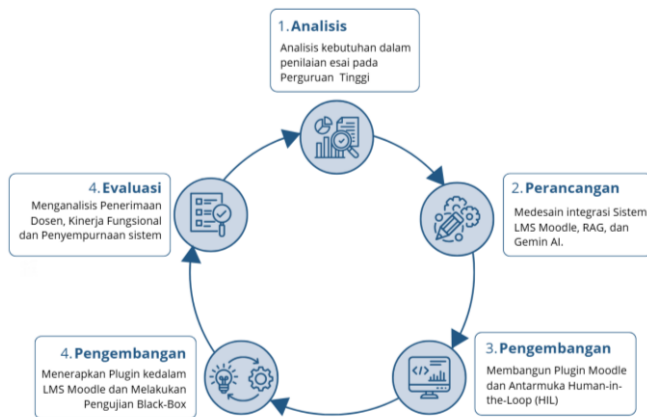
Dalam pengembangannya, penelitian ini mengadopsi model ADDIE (Analyze–Design–Develop–Implement–Evaluate) yang bersifat sistematis dan iteratif. Model ADDIE banyak digunakan dalam penelitian pengembangan sistem dan media pembelajaran karena mampu memastikan bahwa produk yang dikembangkan sesuai dengan kebutuhan pengguna dan dapat disempurnakan secara berkelanjutan berdasarkan hasil evaluasi [9], [12].

### A. Model Pengembangan ADDIE

Model ADDIE terdiri atas lima tahapan utama, yaitu Analyze, Design, Develop, Implement, dan Evaluate. Setiap tahapan dilakukan secara berurutan namun tetap memungkinkan adanya umpan balik untuk perbaikan pada tahap sebelumnya. Karakteristik iteratif ini menjadikan ADDIE relevan untuk pengembangan sistem berbasis teknologi yang memerlukan proses evaluasi dan penyempurnaan berulang [9], [12].

Penerapan model ADDIE dalam penelitian ini bertujuan untuk menghasilkan sistem penilaian esai otomatis yang

fungsional, reliabel, dan sesuai dengan konteks pembelajaran di perguruan tinggi, khususnya dalam mendukung peran dosen sebagai penilai utama. Model ADDID pada penelitian ini dapat digambarkan sebagai berikut :



Gambar 1. Pendekatan Research and Development (R&D) dengan model AADIE

#### a. Tahap Analyze

Tahap Analyze bertujuan untuk mengidentifikasi kebutuhan dan permasalahan yang menjadi dasar pengembangan sistem. Analisis dilakukan melalui:

1. Analisis proses penilaian esai yang berjalan di LMS Moodle.
2. Analisis kebutuhan dosen terhadap sistem penilaian esai otomatis.
3. Analisis karakteristik tugas esai dan rubrik penilaian yang digunakan.

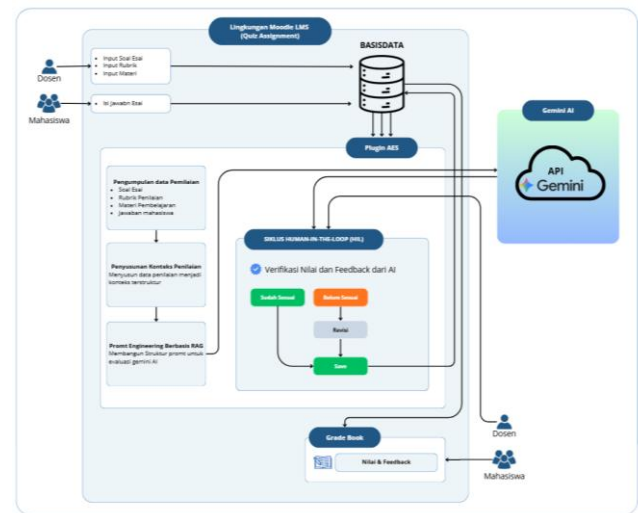
Hasil analisis menunjukkan bahwa penilaian esai masih dilakukan secara manual oleh dosen, sehingga berpotensi menimbulkan beban kerja yang tinggi, keterbatasan waktu, serta inkonsistensi penilaian [13]. Studi-studi sebelumnya juga menunjukkan bahwa penilaian esai manual memiliki tingkat subjektivitas yang relatif tinggi, terutama ketika jumlah mahasiswa meningkat [14]. Oleh karena itu, diperlukan sistem pendukung yang mampu memberikan rekomendasi penilaian tanpa menghilangkan peran dosen sebagai pengambil keputusan akademik [15].

#### b. Tahap Design

Tahap Design berfokus pada perancangan sistem penilaian esai otomatis yang akan dikembangkan. Pada tahap ini dilakukan:

1. Perancangan arsitektur sistem integrasi antara Moodle dan Gemini AI.
2. Perancangan alur kerja penilaian esai berbasis Retrieval-Augmented Generation (RAG).
3. Perancangan mekanisme Human-in-the-Loop (HIL) dalam proses penilaian.

Desain sistem dapat di gambarkan sebagai berikut :



Gambar 2. Arsitektur Pengembangan sistem AES

Sistem dirancang sebagai decision-support system, yaitu sistem yang memberikan rekomendasi berbasis analisis data untuk mendukung pengambilan keputusan manusia [15]. Dalam konteks penilaian esai, AI berfungsi menghasilkan rekomendasi skor dan umpan balik awal, sementara dosen tetap memiliki kewenangan penuh untuk menyetujui atau merevisi hasil penilaian sebelum nilai akhir ditetapkan. Pendekatan ini sejalan dengan praktik pemanfaatan AI yang menekankan kolaborasi manusia dan mesin dalam proses evaluasi akademik [16].

#### c. Tahap Develop

Tahap Develop merupakan tahap implementasi desain menjadi sistem yang dapat digunakan. Pada tahap ini dilakukan:

1. Pengembangan modul integrasi Moodle dengan layanan Gemini AI.
2. Implementasi mekanisme RAG dengan memanfaatkan materi pembelajaran dan rubrik penilaian sebagai sumber konteks.
3. Pengembangan antarmuka dosen untuk menampilkan rekomendasi skor dan umpan balik hasil penilaian AI.

Pendekatan Retrieval-Augmented Generation (RAG) digunakan untuk meningkatkan relevansi dan akurasi hasil penilaian dengan menyediakan konteks terkurasi sebelum proses inferensi model bahasa besar. Pendekatan ini terbukti mampu mengurangi risiko hallucination dan meningkatkan kesesuaian output AI dengan domain pembelajaran tertentu [16], [17].

#### d. Tahap Implement

Tahap Implement bertujuan untuk menerapkan sistem yang telah dikembangkan ke dalam lingkungan LMS Moodle. Implementasi dilakukan dalam skala terbatas dengan melibatkan dosen sebagai pengguna sistem. Pada tahap ini, dosen menggunakan sistem untuk menilai jawaban esai mahasiswa dengan bantuan rekomendasi yang dihasilkan oleh Gemini AI.

Pendekatan Human-in-the-Loop memastikan bahwa sistem berfungsi sebagai alat bantu keputusan, bukan sebagai pengganti dosen. Pendekatan ini penting untuk menjaga aspek pedagogis, etika, dan tanggung jawab akademik dalam proses penilaian berbasis AI [16], [18].

#### e. Tahap Evaluate

Tahap Evaluate dilakukan untuk menilai kinerja dan penerimaan sistem yang dikembangkan. Evaluasi dilakukan melalui:

1. Pengujian fungsional (black-box testing) untuk memastikan seluruh fitur sistem berjalan sesuai dengan kebutuhan pengguna.
2. Evaluasi penerimaan dosen terhadap rekomendasi skor dan umpan balik yang dihasilkan sistem.
3. Identifikasi kelebihan dan keterbatasan sistem sebagai dasar pengembangan lanjutan.

Hasil evaluasi digunakan untuk melakukan perbaikan sistem secara iteratif sesuai dengan prinsip model ADDIE [9], [12].

#### B. Peran Human-in-the-Loop dalam Sistem

Dalam penelitian ini, pendekatan Human-in-the-Loop (HIL) diterapkan untuk memastikan bahwa kecerdasan buatan tidak menggantikan peran dosen dalam proses penilaian. AI berfungsi sebagai sistem pendukung keputusan yang memberikan rekomendasi berbasis analisis teks, sementara keputusan akhir tetap berada pada dosen sebagai otoritas akademik.

Pendekatan ini sejalan dengan prinsip tata kelola penggunaan AI dalam pendidikan tinggi yang menekankan transparansi, akuntabilitas, dan pengawasan manusia, terutama dalam konteks evaluasi akademik yang berdampak langsung pada mahasiswa [16], [18].

#### C. Teknik Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan melalui:

1. Dokumentasi hasil penilaian esai sebelum dan sesudah penggunaan sistem, yang bertujuan untuk menggambarkan perubahan alur dan proses penilaian esai oleh dosen.
2. Kuesioner penerimaan dosen, yang digunakan untuk mengukur tingkat penerimaan dosen terhadap sistem penilaian esai otomatis berdasarkan:
3. Dimensi Human-in-the-Loop (HIL)
4. Dimensi pedagogis

Kuesioner diberikan kepada dosen setelah mereka menggunakan sistem penilaian esai otomatis dalam lingkungan LMS Moodle.

#### D. Instrumen Penelitian

Instrumen utama dalam penelitian ini adalah kuesioner penerimaan dosen yang disusun menggunakan skala Likert lima tingkat, mulai dari sangat tidak setuju hingga sangat setuju.

#### a. Dimensi Human-in-the-Loop (HIL)

Dimensi ini mengukur persepsi dosen terhadap:

1. *Human Control* (tingkat kendali dosen terhadap keputusan nilai akhir)
2. *Academic Accountability* (tanggung jawab akademik dosen atas hasil penilaian)
3. *Perceived Efficiency with Human Oversight* (efisiensi proses penilaian dengan tetap melibatkan dosen)
4. *Usability of Human Validation Mechanism* (kemudahan penggunaan mekanisme validasi oleh dosen)
5. *AI Role as Decision Support* (persepsi AI sebagai alat bantu, bukan pengambil keputusan)
6. *Ease of Human Intervention / Override* (kemudahan dosen mengoreksi dan mengubah hasil AI)

### III. HASIL DAN PEMBAHASAN

#### A. Hasil Tahap Analyze

Pada tahap analisis, diperoleh temuan bahwa proses penilaian esai di lingkungan penelitian masih sepenuhnya dilakukan secara manual oleh dosen. Dosen bertindak sebagai penilai utama sekaligus pengambil keputusan akademik, sementara LMS Moodle hanya berfungsi sebagai media pengumpulan tugas.

1. Hasil analisis kebutuhan menunjukkan bahwa dosen membutuhkan sistem yang:
2. Mampu membantu mempercepat proses penilaian esai,
3. Tetap mempertahankan kendali penuh dosen terhadap penentuan nilai akhir,
4. Memberikan umpan balik awal yang konsisten dan selaras dengan materi serta rubrik yang diajarkan.

Temuan ini menjadi dasar pemilihan pendekatan Retrieval-Augmented Generation (RAG) dan Human-in-the-Loop (HIL) dalam perancangan sistem.

#### B. Hasil Tahap Design dan Develop

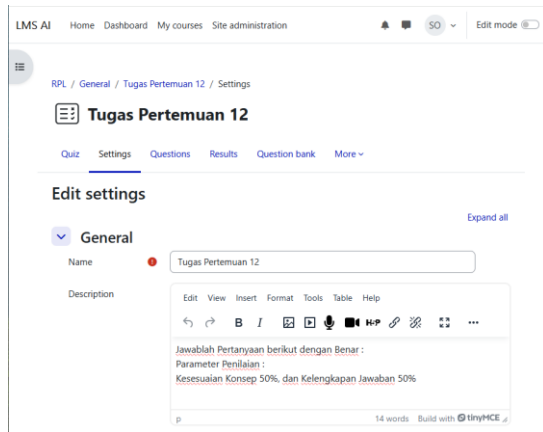
##### a. Desain Antarmuka Plugin Moodle

Pada tahap desain dan pengembangan, sistem direalisasikan dalam bentuk plugin penilaian esai pada LMS Moodle. Plugin ini dirancang agar terintegrasi dengan alur penilaian esai yang telah familiar bagi dosen.

Fitur utama plugin meliputi:

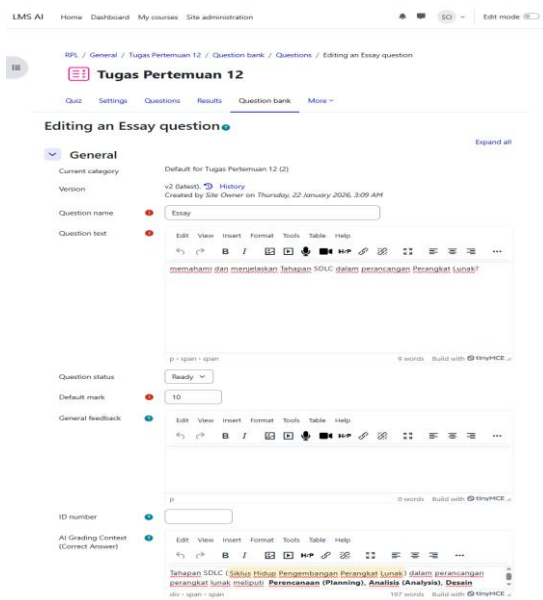
1. Input rubrik penilaian, soal esai dan materi pendukung, oleh dosen

Pada halaman ini dosen dapat input nama soal, dan Rubrik penilaian didalam deskripsi soal



Gambar 2. Input deskripsi sekaligus Rubrik Penilaian.

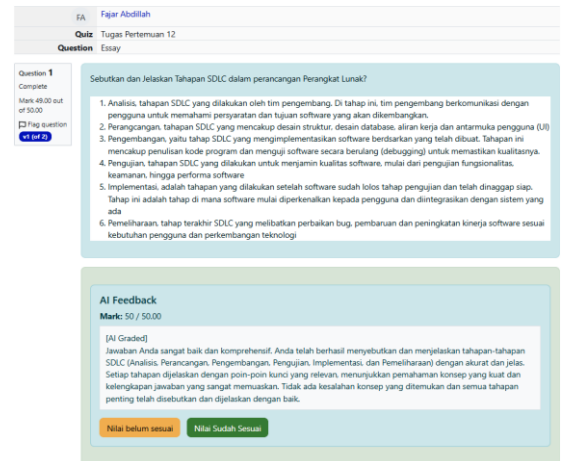
Pada bagian berikut dosen dapat menginput soal, nilai dan materi atau jawaban yang benar dari soal tersebut.



Gambar 3. Input deskripsi sekaligus Rubrik Penilaian.

- Antarmuka rekomendasi skor dan umpan balik dari Gemini AI dan validasi Human-in-the-Loop dengan opsi Approve dan Revise

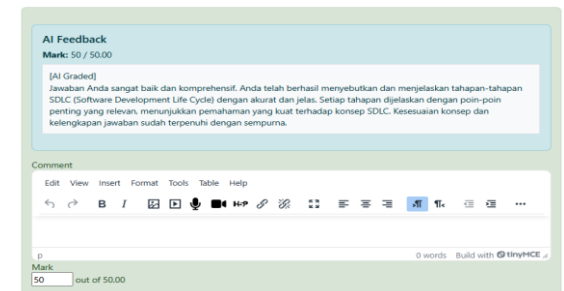
Pada bagian ini dosen masuk ke gradebook untuk menilai jawaban mahasiswa lalu menekan tombol "Grade With AI" maka AI akan memberikan nilai dan umpan balik. Dan dosen dapat melakukan verifikasi apakah nilai sudah sesuai atau belum.



Gambar 4. Rekomendasi score dan umpan balik AI

- Antarmuka Revisi nilai dan umpan balik AI

Dosen dapat merevisi nilai dan umpan balik AI apabila nilai atau umpan balik tidak sebagaimana mestinya.



Gambar 5. Revisi nilai dan umpan balik AI

- Penyampaian nilai akhir dan umpan balik kepada mahasiswa.

Mahasiswa dapat melihat nilai dan umpan balik setelah dosen menyimpan hasil penilaian baik itu hasil AI yang di setuju dosen, atau hasil revisi dosen.



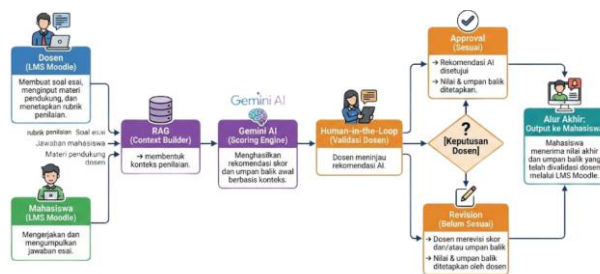
Gambar 6. Nilai akhir dan umpan balik kepada mahasiswa

Berdasarkan hasil perancangan dan implementasi, antarmuka plugin penilaian esai pada LMS Moodle berhasil dikembangkan secara sederhana dan fungsional, dengan alur

kerja yang selaras dengan proses penilaian esai yang telah familiar bagi dosen. Seluruh aktivitas utama, mulai dari input soal, materi, dan rubrik hingga validasi hasil penilaian berbasis AI, terintegrasi dalam satu antarmuka yang mudah digunakan sehingga tidak menuntut perubahan signifikan terhadap kebiasaan kerja dosen. Keberadaan fitur validasi *Human-in-the-Loop* yang ditampilkan secara jelas pada antarmuka menegaskan peran dosen sebagai pengambil keputusan akademik akhir, sementara AI diposisikan sebagai alat bantu penilaian.

#### b. Implementasi Prompt Engineering dan RAG

Hasil pengembangan menunjukkan bahwa mekanisme *prompt engineering* berbasis RAG dapat mengintegrasikan data soal esai, materi pendukung, rubrik penilaian, dan jawaban mahasiswa ke dalam satu konteks penilaian terstruktur. Konteks ini digunakan oleh Gemini AI untuk menghasilkan rekomendasi skor dan umpan balik awal yang relevan.



Gambar 6. Mekanisme *prompt engineering* berbasis RAG

Pendekatan ini membedakan sistem yang dikembangkan dari sistem AES konvensional, karena sumber konteks penilaian sepenuhnya ditentukan oleh dosen saat pembuatan soal esai.

#### C. Hasil Tahap Evaluate

##### a. Hasil Pengujian Fungsional (*Black-Box Testing*)

Pengujian fungsional sistem dilakukan menggunakan metode *black-box testing* untuk memastikan seluruh fungsi utama sistem berjalan sesuai dengan kebutuhan pengguna. Hasil pengujian menunjukkan bahwa seluruh fitur utama sistem dapat berjalan dengan baik dan sesuai spesifikasi.

Tabel 1. Daftar Skenario Black-Box Testing

| No | Skenario Uji           | Aktivitas Pengguna                 | Hasil yang Diharapkan                   | Hasil Aktual | Status   |
|----|------------------------|------------------------------------|---|--------------|----------|
| 1  | Input soal esai        | Dosen menginput soal esai          | Data tersimpan di sistem                | Sesuai       | Berhasil |
| 2  | Input materi pendukung | Dosen menginput materi pendukung   | Materi tersimpan & terhubung ke soal    | Sesuai       | Berhasil |
| 3  | Input rubrik penilaian | Dosen menginput rubrik & bobot     | Rubrik dapat digunakan untuk penilaian  | Sesuai       | Berhasil |
| 4  | Pengumpulan jawaban    | Mahasiswa mengirim jawaban esai    | Jawaban tersimpan di database           | Sesuai       | Berhasil |
| 5  | Konstruksi prompt RAG  | Sistem menggabungkan soal, materi, | Prompt terbentuk dengan konteks lengkap | Sesuai       | Berhasil |

|    |                        | rubrik, jawaban                          |                               |        |          |
|----|------------------------|--|-------------------------------|--------|----------|
| 6  | Koneksi API Gemini     | Sistem mengirim permintaan ke Gemini API | API merespons tanpa error     | Sesuai | Berhasil |
| 7  | Rekomendasi AI         | Gemini AI menghasilkan skor & feedback   | Skor & feedback tampil        | Sesuai | Berhasil |
| 8  | Validasi HIL (Approve) | Dosen menyetujui hasil AI                | Nilai disimpan sebagai final  | Sesuai | Berhasil |
| 9  | Validasi HIL (Revise)  | Dosen merevisi skor/feedback             | Nilai tersimpan sesuai revisi | Sesuai | Berhasil |
| 10 | Publikasi nilai        | Sistem menampilkan nilai ke mahasiswa    | Nilai & feedback tampil       | Sesuai | Berhasil |

#### b. Hasil Evaluasi Human-in-the-Loop

Evaluasi *Human-in-the-Loop* (HIL) dilakukan untuk menilai tingkat penerimaan dosen terhadap sistem penilaian esai otomatis berbasis Gemini AI, khususnya dalam konteks kendali manusia, akuntabilitas akademik, efisiensi proses penilaian, serta posisi AI sebagai alat bantu pengambilan keputusan. Evaluasi ini dilakukan menggunakan kuesioner berbasis skala Likert lima tingkat yang diisi oleh dosen setelah mencoba sistem secara langsung.

Keterangan kategori:

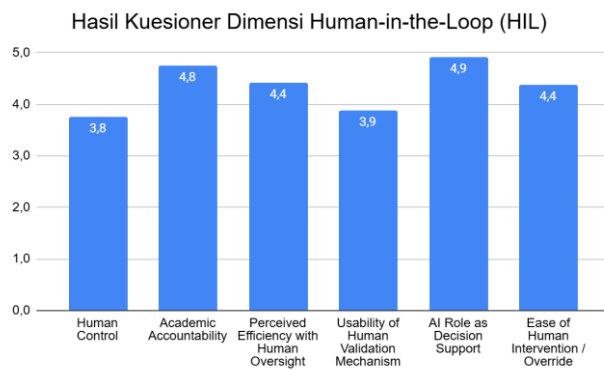
- 4.21 – 5.00 : Sangat Tinggi
- 3.41 – 4.20 : Tinggi
- 2.61 – 3.40 : Sedang
- 1.81 – 2.60 : Rendah
- 1.00 – 1.80 : Sangat Rendah

Sebanyak 24 responden dosen dari 14 perguruan tinggi swasta di wilayah Banten, DKI Jakarta, dan Jawa Barat berpartisipasi dalam evaluasi ini. Seluruh responden telah menggunakan sistem penilaian esai berbasis AI yang terintegrasi dengan LMS Moodle dan melibatkan mekanisme validasi dosen (*Human-in-the-Loop*) sebelum memberikan penilaian.

#### c. Hasil Evaluasi Dimensi Human-in-the-Loop

Dimensi *Human-in-the-Loop* mencakup enam indikator utama yang merepresentasikan aspek kendali, tanggung jawab, efisiensi, dan kemudahan interaksi dosen dalam sistem. Yang disajikan pada diagram berikut :

#### d. Pembahasan Evaluasi Human-in-the-Loop



Gambar 7. Hasil kuesioner dimensi Human-in-the-Loop (HIL)

Hasil evaluasi menunjukkan bahwa sistem penilaian esai berbasis Gemini AI memperoleh tingkat penerimaan yang sangat positif dari dosen. Indikator *AI Role as Decision Support* memperoleh skor rata-rata tertinggi (4,9), yang mengindikasikan bahwa dosen secara kuat memandang AI sebagai alat bantu penilaian, bukan sebagai pengganti peran manusia dalam pengambilan keputusan akademik. Temuan ini menegaskan bahwa desain sistem berhasil menjaga batas epistemik antara peran AI dan otoritas dosen.

Indikator *Academic Accountability* juga menunjukkan skor yang sangat tinggi (4,8), menandakan bahwa dosen tetap merasa bertanggung jawab penuh atas nilai akhir yang diberikan kepada mahasiswa. Hal ini menunjukkan bahwa penerapan mekanisme *Human-in-the-Loop* tidak mengurangi rasa tanggung jawab akademik dosen, tetapi justru memperkuat posisi dosen sebagai pengambil keputusan akhir.

Indikator *Perceived Efficiency with Human Oversight* dan *Ease of Human Intervention/Override* masing-masing memperoleh skor rata-rata 4,4. Hasil ini mengindikasikan bahwa keterlibatan dosen dalam proses validasi tidak dipersepsikan sebagai hambatan, melainkan sebagai bagian dari proses yang tetap efisien dan fleksibel. Dosen dapat dengan mudah menyetujui atau merevisi rekomendasi penilaian AI sesuai dengan pertimbangan akademik.

Indikator *Human Control* (3,8) dan *Usability of Human Validation Mechanism* (3,9) berada pada kategori tinggi namun relatif lebih rendah dibandingkan indikator lainnya. Temuan ini menunjukkan adanya peluang perbaikan pada aspek desain antarmuka dan pengalaman pengguna, khususnya untuk semakin memperjelas kontrol dosen dalam proses validasi. Namun demikian, skor tersebut tetap mengindikasikan bahwa kendali dosen atas proses penilaian telah terjaga dengan baik.

#### IV. KESIMPULAN DAN SARAN

##### A. Kesimpulan

Penelitian ini berhasil mengembangkan dan mengevaluasi sistem penilaian esai otomatis berbasis Gemini AI yang terintegrasi dengan Learning Management System Moodle melalui pendekatan Retrieval-Augmented Generation (RAG) dan Human-in-the-Loop (HIL). Sistem dirancang untuk mendukung proses penilaian esai dengan

memanfaatkan kecerdasan buatan sebagai alat bantu, tanpa menghilangkan peran dosen sebagai pengambil keputusan akademik akhir.

Hasil pengujian fungsional (*black-box testing*) menunjukkan bahwa seluruh fungsi utama sistem, mulai dari input soal, materi pendukung, rubrik penilaian, pengumpulan jawaban mahasiswa, pemanggilan API Gemini, hingga publikasi nilai akhir, dapat berjalan dengan baik dan sesuai dengan kebutuhan pengguna. Hal ini menunjukkan bahwa sistem secara teknis layak digunakan dalam lingkungan LMS Moodle.

Evaluasi *Human-in-the-Loop* melalui kuesioner dosen menunjukkan tingkat penerimaan yang tinggi hingga sangat tinggi terhadap sistem. Dosen menilai bahwa AI berperan secara tepat sebagai decision-support system, sementara kendali dan tanggung jawab akademik tetap berada pada dosen. Tingginya skor pada indikator tanggung jawab akademik dan peran AI sebagai alat bantu menegaskan bahwa integrasi AI dalam sistem penilaian ini tidak menggeser otoritas penilaian dari manusia ke mesin.

Selain itu, hasil validasi dosen terhadap rekomendasi penilaian AI menunjukkan bahwa sebagian besar rekomendasi dapat disetujui tanpa revisi, sementara sebagian kecil memerlukan koreksi. Temuan ini mengonfirmasi pentingnya mekanisme *Human-in-the-Loop* sebagai pengaman akademik untuk menjaga akurasi, fleksibilitas, dan akuntabilitas penilaian. Dengan demikian, sistem yang dikembangkan dinilai efektif sebagai solusi pendukung penilaian esai pada skala kelas besar, sekaligus tetap mematuhi prinsip tanggung jawab akademik.

##### B. Saran

Berdasarkan hasil penelitian, beberapa saran yang dapat diberikan untuk pengembangan selanjutnya adalah sebagai berikut:

1. Penyempurnaan antarmuka validasi dosen, khususnya untuk memperkuat persepsi kendali dan kemudahan penggunaan mekanisme *Human-in-the-Loop*.
2. Pengujian sistem dalam skala yang lebih luas, baik dari sisi jumlah mata kuliah maupun variasi bidang ilmu, untuk menguji konsistensi performa sistem.
3. Pengembangan mekanisme penjelasan hasil penilaian AI (explainable assessment) agar dosen dapat lebih mudah memahami dasar rekomendasi skor yang diberikan.
4. Studi lanjutan terkait aspek etika dan kebijakan penggunaan AI, khususnya mengenai transparansi, privasi data, dan regulasi institusional dalam penerapan AI untuk penilaian akademik.

#### DAFTAR PUSTAKA

- [1] Y. Wu and Q. Zheng, "Critical Thinking Writing Assessment in Middle School Language: Logic Chain Extraction and Expert Score Correlation Test Using BERT-CNN Hybrid Model," *Applied*

- Sciences*, vol. 15, no. 19, 2025, doi: 10.3390/app151910504.
- [2] E. D. Reilly, R. E. Stafford, K. M. Williams, and S. B. Corliss, "Evaluating the Validity and Applicability of Automated Essay Scoring in Two Massive Open Online Courses," *The International Review of Research in Open and Distributed Learning*, vol. 15, no. 5, 2014, doi: 10.19173/IRRODL.V15I5.1857.
  - [3] M. Beseiso, O. A. Alzubi, and H. Rashaideh, "A Novel Automated Essay Scoring Approach for Reliable Higher Educational Assessments," *Journal of Computing in Higher Education*, vol. 33, pp. 611–631, 2021, doi: 10.1007/S12528-021-09283-1.
  - [4] J. Burstein, J. Tetreault, and N. Madnani, "The E-rater® Automated Essay Scoring System," in *Handbook of Automated Essay Evaluation*, New York, NY, USA: Routledge, 2013, doi: 10.4324/9780203122761.ch4.
  - [5] D. Ramesh and S. K. Sanampudi, "A Multitask Learning System for Trait-based Automated Short Answer Scoring," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 10, 2023, doi: 10.14569/IJACSA.2023.0141048.
  - [6] M. Beseiso, O. A. Alzubi, and H. Rashaideh, "Transformer-based Models for Automated Essay Scoring: Addressing Coherence and Content Understanding," *Journal of Computing in Higher Education*, 2021.
  - [7] E. P. T. Tyndall, C. Gayheart, A. Some, J. Genz, T. Wagner, and B. Langhals, "Impact of Retrieval Augmented Generation and Large Language Model Complexity on Undergraduate Exams Created and Taken by AI Agents," *Data & Policy*, 2025, doi: 10.1017/dap.2025.10024.
  - [8] F. Pan, Q. Zhou, W. Guo, and H. Yang, "A Survey on Retrieval-Augmented Generation in Applications of Education and Teaching," 2025, doi: 10.1109/CSTE64638.2025.11092120.
  - [9] M. A. Adeoye, K. A. S. I. Wirawan, M. S. S. Pradnyani, and N. I. Septiarini, "Revolutionizing Education: Unleashing the Power of the ADDIE Model for Effective Teaching and Learning," *Jurnal Pendidikan Indonesia*, vol. 13, no. 1, 2024.
  - [10] D. Adriani, P. K. D. Lubis, and M. A. A. Triono, "Teaching Material Development of Educational Research Methodology with ADDIE Models," in *Proceedings of the EAI Conference*, 2020, doi: 10.4108/eai.4-12-2019.2293793.
  - [11] E. D. Reilly, R. E. Stafford, K. M. Williams, and S. B. Corliss, "Evaluating the Validity and Applicability of Automated Essay Scoring in Two Massive Open Online Courses," *The International Review of Research in Open and Distributed Learning*, vol. 15, no. 5, 2014.
  - [12] I. N. Zuhro, M. Sutomo, and Mashudi, "Desain Pembelajaran Pendidikan Agama Islam dengan Model ADDIE," *Ta'lim*, vol. 5, no. 2, 2022, doi: 10.52166/talim.v5i2.3085.
  - [13] M. Beseiso, O. A. Alzubi, and H. Rashaideh, "A Novel Automated Essay Scoring Approach for Reliable Higher Educational Assessments," *Journal of Computing in Higher Education*, 2021.
  - [14] J. Burstein, J. Tetreault, and N. Madnani, "The E-rater® Automated Essay Scoring System," New York, NY, USA: Routledge, 2013.
  - [15] D. Ramesh and S. K. Sanampudi, "Trait-based Automated Short Answer Scoring Using Multitask Learning," *International Journal of Advanced Computer Science and Applications*, 2023.
  - [16] E. P. T. Tyndall, C. Gayheart, A. Some, J. Genz, T. Wagner, and B. Langhals, "Impact of Retrieval Augmented Generation and Large Language Model Complexity on Undergraduate Exams Created and Taken by AI Agents," *Data & Policy*, 2025.
  - [17] F. Pan, Q. Zhou, W. Guo, and H. Yang, "A Survey on Retrieval-Augmented Generation in Applications of Education and Teaching," 2025.
  - [18] Y. Wu and Q. Zheng, "Critical Thinking Writing Assessment Using BERT-based Models," *Applied Sciences*, 2025.