

# Penerapan *Large Language Models* Dalam Pembaruan Artikel Biografi Wikipedia

Najma Qalbi Dwiharani<sup>1\*</sup>, Yudi Wibisono<sup>2</sup>, Yaya Wihardi<sup>3</sup>

<sup>1,2,3</sup>Fakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam, Ilmu Komputer, Universitas Pendidikan Indonesia, Bandung, Indonesia

Email: <sup>1</sup>najmadwiharani@upi.edu, <sup>2</sup>yudi@upi.edu, <sup>3</sup>yayawihardi@upi.edu  
(\* Email Corresponding Author: najmadwiharani@upi.edu)

Received: July 24, 2025 | Revision: July 25, 2025 | Accepted: July 31, 2025

## Abstrak

Wikipedia merupakan sumber informasi daring yang sangat populer di Indonesia, namun pembaruan artikelnya masih sangat bergantung pada kontribusi penyunting. Pada kategori artikel biografi, pembaruan informasi secara berkala sangat penting karena adanya perkembangan karier dan peristiwa terkini dari tokoh yang bersangkutan. Penelitian ini bertujuan untuk mengeksplorasi penerapan *Large Language Models* (LLM) dalam menambahkan informasi baru ke artikel biografi Wikipedia Indonesia dengan referensi dari satu artikel berita daring. Model utama yang digunakan adalah Gemma 3 yang kemudian dibandingkan dengan model *baseline* Phi-3-mini. Penelitian ini juga menguji efektivitas lima strategi *prompting* yang berbeda, yaitu *simple prompt*, *system prompt (en)*, *system prompt (id)*, *one-shot*, dan *prompt chaining* untuk mengarahkan model dalam menghasilkan keluaran yang relevan dan sesuai dengan gaya Wikipedia. Proses *fine-tuning* dilakukan menggunakan data berbentuk kombinasi artikel Wikipedia sebelum diperbarui, artikel berita sebagai referensi, dan teks berisi informasi baru yang relevan untuk ditambahkan ke dalam artikel Wikipedia sebagai target keluaran. Evaluasi dilakukan dengan metrik ROUGE untuk mengukur kesamaan antara hasil keluaran model dan referensi dari penyunting Wikipedia. Hasil penelitian menunjukkan bahwa *fine-tuning* model Gemma 4B secara signifikan meningkatkan performa, khususnya pada strategi *prompt chaining* dengan rata-rata skor ROUGE-1 sebesar 0.3687. Dibandingkan dengan *baseline* Phi-3-mini, model Gemma memberikan hasil yang lebih konsisten dan relevan. Temuan ini menunjukkan bahwa pendekatan berbasis LLM dapat menjadi solusi potensial dalam membantu proses pembaruan artikel biografi Wikipedia.

**Kata Kunci:** Wikipedia, Biografi, Berita, Pembaruan Artikel, *Large Language Models*

## Abstract

Wikipedia is a widely popular online source of information in Indonesia, yet the updating of its articles still heavily relies on contributions from volunteer editors. In the case of biographical articles, regular updates are crucial due to the continuous developments in individuals' careers and public events. This study aims to explore the use of *Large Language Models* (LLMs) to assist in adding new information to Indonesian Wikipedia biography articles using a single online news article as a reference. The main model used is Gemma 3, which is compared with the baseline model Phi-3-mini. The study also evaluates the effectiveness of five different prompting strategies—*simple prompt*, *system prompt (en)*, *system prompt (id)*, *one-shot*, and *prompt chaining*—to guide the model in generating outputs that are relevant and aligned with Wikipedia's writing style. Fine-tuning was performed using data composed of a combination of the original (pre-update) Wikipedia article, a news article as a reference, and a target text containing new information deemed relevant to be added to the Wikipedia article. Evaluation was conducted using the ROUGE metric to measure the similarity between the model's output and the editor-written references. The results show that fine-tuning the Gemma 4B model significantly improves performance, particularly with the *prompt chaining* strategy, achieving an average ROUGE-1 score of 0.3687. Compared to the Phi-3-mini baseline, the Gemma model produces more consistent and relevant outputs. These findings suggest that LLM-based approaches, when guided by appropriate prompting strategies, can serve as a potential solution to support the automatic updating of biographical articles on Wikipedia.

**Keywords:** Wikipedia, Biography, Online News, Article Update, *Large Language Models*

## 1. PENDAHULUAN

Wikipedia adalah proyek ensiklopedia daring multibahasa yang bebas dan terbuka, yang dijalankan oleh Yayasan Wikimedia. Artikel-artikel di Wikipedia merupakan hasil kolaborasi oleh para penyunting dari seluruh dunia. Siapapun dapat menyunting, menambah, dan memperbaiki informasi yang ada pada halaman artikel Wikipedia [1]. Berdasarkan data statistik dari Wikimedia, jumlah pengunjung halaman Wikipedia Indonesia pada tahun 2022-2023 mencapai tiga miliar dengan lebih dari sekitar tiga juta pengunjung per harinya [2]. Hal ini menunjukkan bahwa Wikipedia merupakan situs yang cukup populer di Indonesia. Sebagai situs yang bebas dan terbuka, informasi yang ada pada Wikipedia bergantung pada jumlah dan keaktifan para penyuntingnya. Berdasarkan data pada Wikipedia Indonesia di bulan April 2024, jumlah artikel yang ada pada halaman Wikipedia Indonesia mencapai 691.364 artikel dengan total 3.833.833 halaman dan 25.584.539 suntingan. Terdapat 1.474.915 pengguna terdaftar, namun hanya 2.404 pengguna yang aktif pada bulan April 2024 dan 47 pengurus aktif Wikipedia Indonesia [3]. Keterbatasan jumlah pengguna aktif dan kurangnya penyunting yang terlibat secara rutin ini dapat mempengaruhi kualitas dan akurasi informasi pada Wikipedia Indonesia.

Pada bulan April 2024, terdapat 92.595 halaman artikel Wikipedia Indonesia yang termasuk dalam kategori biografi dengan 57.659 halamannya merupakan kategori orang yang masih hidup [4]. Artikel biografi merupakan artikel yang perlu perubahan secara rutin karena adanya perkembangan karir, pencapaian, dan peristiwa terkini yang dialami oleh orang yang bersangkutan. Artikel yang tidak diperbarui secara teratur dapat mengandung informasi yang kurang

lengkap atau sudah tidak lagi relevan dengan keadaan orang tersebut. Oleh karena itu, para penyunting Wikipedia perlu memaksimalkan pembaruan informasi dalam artikel-artikelnya, terutama pada artikel dengan kategori biografi. Artikel biografi Wikipedia Indonesia menggunakan banyak referensi penyuntingan yang bersumber dari berita. Hal ini menunjukkan bahwa berita memiliki peran penting dalam menyediakan data yang aktual dan relevan untuk artikel Wikipedia Indonesia.

Dalam proses pembaruan artikel Wikipedia, khususnya artikel biografi, penyunting perlu memantau perkembangan terbaru dari subjek dan mencari sumber informasi baru yang relevan. Salah satu langkah penting dalam proses ini adalah menambahkan informasi baru yang belum tercantum pada artikel. Penyunting harus membaca dan membandingkan isi berita dengan artikel Wikipedia yang ada kemudian menyusun kalimat tambahan yang sesuai dengan gaya penulisan Wikipedia serta menyisipkannya ke bagian yang relevan jika ada informasi baru yang layak untuk ditambahkan. Proses penambahan informasi ini bersifat berulang dan membutuhkan waktu. Tantangan semakin besar jika mempertimbangkan jumlah artikel biografi yang sangat banyak di Wikipedia, sedangkan jumlah penyunting aktif terbatas. Oleh karena itu, dibutuhkan pendekatan yang dapat membantu penyunting dalam mengidentifikasi dan menyusun informasi tambahan secara lebih efisien.

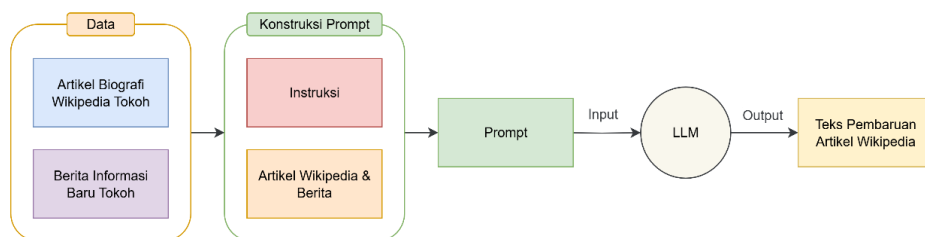
Perkembangan teknologi kecerdasan buatan atau *Artificial Intelligence* (AI) saat ini telah banyak dimanfaatkan untuk melakukan pekerjaan dalam berbagai bidang. Model bahasa besar atau *Large Language Models* (LLM) seperti GPT-3 (*Generative Pretrained Transformer 3*) telah membuat kemajuan yang signifikan dalam bidang *Natural Language Processing* (NLP) pada beberapa tahun terakhir. LLM merupakan kategori model AI yang dilatih dengan data teks dalam jumlah besar dan mampu menghasilkan teks seperti buatan manusia, menjawab pertanyaan, serta menyelesaikan permasalahan yang terkait bahasa dengan akurasi tinggi [5]. Penggunaan LLM dapat membantu dalam otomatisasi penyelesaian masalah terkait bahasa yang sebelumnya memakan waktu dan tenaga kerja, sehingga dapat meningkatkan efisiensi dari berbagai pekerjaan [6].

Sejumlah penelitian telah mengkaji pemanfaatan AI dan model bahasa untuk penulisan atau pembaruan artikel bergaya Wikipedia. Penelitian dengan pendekatan *multi-document abstractive summarization* untuk membangun artikel Wikipedia secara otomatis dari berbagai sumber referensi telah dilakukan sebelumnya dan dinilai mampu menghasilkan artikel utuh yang menyerupai gaya Wikipedia melalui model *transformer* [7]. Penelitian lain mengembangkan sistem WikiWrite, yang menyusun artikel Wikipedia baru dengan meniru struktur dari artikel Wikipedia yang mirip, kemudian menyusun ulang informasi dari web eksternal melalui teknik abstraksi dan *paraphrasing* [8]. Kedua penelitian tersebut dilakukan pada Wikipedia berbahasa Inggris. Penelitian terkait artikel biografi Wikipedia Indonesia juga telah dilakukan sebelumnya dengan menyusun paragraf biografi dengan pendekatan *extractive summarization* dari berita daring [9]. Meskipun sesuai konteks bahasa dan domain yang diteliti dalam penelitian ini, pendekatan pada penelitian tersebut masih bergantung pada pengambilan langsung kalimat dari berita dan belum memanfaatkan kemampuan LLM untuk menulis ulang informasi agar sesuai dengan gaya Wikipedia. Beberapa penelitian telah menunjukkan penerapan LLM dalam penyusunan artikel Wikipedia, baik secara otomatis [10] maupun sebagai bantuan dalam penulisan berdasarkan kerangka atau isi awal [11]. Berdasarkan penelitian-penelitian tersebut, LLM terbukti memiliki kemampuan yang cukup baik dalam menghasilkan teks dengan gaya penulisan ensiklopedik seperti Wikipedia.

Penelitian ini berfokus pada penambahan informasi baru berbasis berita ke dalam artikel biografi Wikipedia Indonesia yang sudah ada. Penelitian ini memanfaatkan kemampuan generatif dari LLM untuk menyusun kalimat tambahan yang sesuai dengan gaya penulisan Wikipedia berdasarkan kombinasi isi artikel lama dan satu berita daring sebagai referensi. Dengan pendekatan tersebut, penelitian ini diharapkan dapat berkontribusi pada penerapan LLM dalam pembaruan artikel Wikipedia, khususnya untuk artikel biografi bahasa Indonesia. Penelitian ini juga membuka potensi pengembangan alat bantu penyuntingan artikel berbasis model bahasa besar.

## 2. METODOLOGI PENELITIAN

### 2.1 Penerapan LLM untuk Pembaruan Artikel Wikipedia



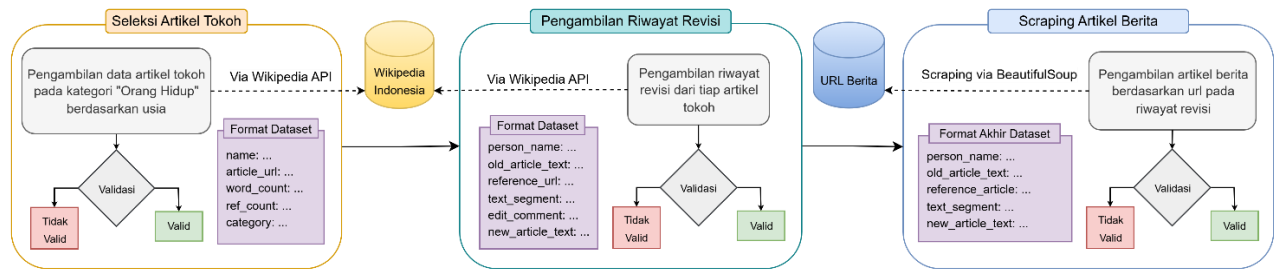
**Gambar 1.** Diagram Alur Penerapan LLM untuk Pembaruan Artikel Wikipedia Biografi

Solusi yang diusulkan dalam penelitian ini adalah penerapan *Large Language Model* (LLM) untuk membantu proses pembaruan artikel biografi di Wikipedia Indonesia secara otomatis. Sebagaimana ditunjukkan pada Gambar 1, proses dimulai dengan dua jenis data utama, yaitu artikel Wikipedia sebelum diperbarui dan artikel berita daring yang memuat informasi baru mengenai tokoh terkait. Kedua teks tersebut diproses dalam tahap konstruksi *prompt* yang

mencakup penyusunan instruksi eksplisit untuk mengarahkan model dalam menghasilkan keluaran teks pembaruan yang relevan, faktual, dan sesuai dengan gaya penulisan Wikipedia. *Prompt* yang telah dikonstruksi kemudian diberikan sebagai masukan (*input*) kepada LLM yang selanjutnya menghasilkan kalimat atau paragraf pembaruan sebagai keluaran (*output*).

## 2.2 Pengumpulan dan *Preprocessing* Data

Proses pengumpulan dan *preprocessing* data dalam penelitian ini dilakukan melalui beberapa tahap berurutan sebagaimana ditunjukkan pada Gambar 2. Tujuan dari tahapan ini adalah untuk membangun *dataset* yang utamanya terdiri atas pasangan artikel Wikipedia sebelum dan setelah revisi, berita daring yang menjadi referensi, serta informasi baru yang ditambahkan. *Dataset* ini akan digunakan untuk melatih dan menguji model dalam menghasilkan informasi baru untuk pembaruan artikel biografi.



**Gambar 2.** Tahapan Pengumpulan dan *Preprocessing* Data

Tahap pertama pengumpulan dan *preprocessing* data dimulai dengan seleksi artikel tokoh, yaitu dengan mengambil daftar tokoh dari kategori “Orang hidup” di Wikipedia Indonesia. Untuk menyaring artikel yang layak diproses, diterapkan sejumlah kriteria seperti panjang artikel, jumlah referensi, dan penyebutan identitas tokoh sebagai orang Indonesia. Setelah itu dilakukan validasi manual artikel tokoh untuk memastikan bahwa artikel benar-benar memenuhi kriteria dan sesuai dengan fokus penelitian.

Tahap kedua adalah pengambilan riwayat revisi artikel menggunakan Wikipedia API. Setiap pasangan revisi dibandingkan satu sama lain untuk mendeteksi penambahan informasi baru. Revisi-revisi yang terpilih adalah yang tidak bersifat minor (seperti perbaikan ejaan atau format) dan mencantumkan referensi berita daring sebagai sumber. Setiap pasangan revisi ini disimpan dalam format awal yang mencakup *person\_name*, *old\_article\_text*, *new\_article\_text*, *reference\_links*, dan *edit\_comment*. Setelah itu pasangan revisi dan referensinya diperiksa dan divalidasi secara manual untuk memastikan bahwa benar terjadi penambahan informasi berdasarkan berita.

Tahap ketiga adalah proses *scraping* artikel berita yang dilakukan untuk mengambil isi lengkap dari tautan referensi dalam revisi Wikipedia yang telah dikumpulkan. *Scraping* dilakukan menggunakan *library* BeautifulSoup. Proses ini menghasilkan teks penuh dari artikel berita sebagai bahan informasi yang akan digunakan oleh model. Kemudian dilakukan validasi manual data hasil *scraping* untuk memastikan bahwa isi berita benar-benar relevan dengan informasi yang ditambahkan ke artikel Wikipedia. Format akhir *dataset* yang digunakan dalam eksperimen terdiri dari lima komponen utama:

- person\_name*: nama tokoh dalam artikel Wikipedia.
- old\_article\_text*: isi artikel sebelum penambahan informasi baru.
- reference\_article*: isi artikel berita yang menjadi referensi.
- text\_segment*: Potongan teks baru yang ditambahkan dalam revisi.
- new\_article\_text*: isi artikel setelah penambahan informasi.

Tabel 1 menampilkan salah satu contoh data yang dikumpulkan dan diproses dalam penelitian ini. Contoh ini menunjukkan struktur data yang terdiri dari nama tokoh, artikel Wikipedia sebelum pembaruan, artikel berita yang dijadikan referensi, segmen teks yang berisi informasi baru yang ditambahkan, serta artikel Wikipedia setelah dilakukan pembaruan. Dalam penelitian ini, telah dikumpulkan sebanyak 192 sampel data. Sebanyak 172 data akan digunakan sebagai data pelatihan, sedangkan 20 data lainnya dipilih sebagai data uji untuk evaluasi.

**Tabel 1.** Contoh entri data

<i>person_name</i>	Putri Ariani
<i>old_article_text</i> (artikel wikipedia sebelum pembaruan)	Ariani Nisma Putri (lahir 31-12-2005), atau yang lebih dikenal sebagai Putri Ariani, adalah seorang penyanyi pop solo Disabilitas Netra wanita berkebangsaan Indonesia. Namanya mulai dikenal semenjak mengikuti kompetisi Indonesia's Got Talent 2014 dan berhasil meraih posisi sebagai pemenang.  == Kehidupan awal dan pendidikan ==  Putri lahir dengan nama Ariani Nisma Putri pada tanggal 31 Desember 2005 di Bangkinang, Kampar, Riau.

reference_article (artikel berita)	... YOGYAKARTA - Putri Ariani, memukau dunia dengan suara emasnya. Dengan suaranya yang merdu ketika menyanyikan lagunya sendiri, dia berhasil mendapatkan Golden Buzzer dari Juri American Got Talent. Gadis kelahiran 31 Desember 2025 ini ternyata adalah siswa kelas 11 di SMKN 2 Kasihan atau yang dikenal Sekolah Menengah Musik (SMM). Bocah ini adalah salah satu dari dua murid tuna netra di sekolah tersebut. ...
text_segment (teks pembaruan yang akan ditambahkan)	... Pada tahun 2021, Putri menempuh pendidikan sekolah menengah atas di SMK Negeri 2 Kasihan atau yang dikenal Sekolah Menengah Musik Yogyakarta. Putri mengambil spesifikasi Instrumen Mayor Flute. Adapun ia juga pernah didaulat untuk solo vokal dalam konser SMM. ...
new_article_text (artikel wikipedia yang telah ditambahkan teks pembaruan)	... Ariani Nisma Putri (lahir 31-12-2005), atau yang lebih dikenal sebagai Putri Ariani, adalah seorang penyanyi pop solo Disabilitas Netra wanita berkebangsaan Indonesia. Namanya mulai dikenal semenjak mengikuti kompetisi Indonesia's Got Talent 2014 dan berhasil meraih posisi sebagai pemenang. == Kehidupan awal dan pendidikan == Putri lahir dengan nama Ariani Nisma Putri pada tanggal 31 Desember 2005 di Bangkinang , Kampar , Riau . ... <b>Pada tahun 2021, Putri menempuh pendidikan sekolah menengah atas di SMK Negeri 2 Kasihan atau yang dikenal Sekolah Menengah Musik Yogyakarta. Putri mengambil spesifikasi Instrumen Mayor Flute . Adapun ia juga pernah didaulat untuk solo vokal dalam konser SMM.</b> ...

### 2.3 Model Phi-3-mini dan Gemma 3

*Large Language Models* (LLM) adalah model bahasa yang memanfaatkan jaringan saraf berisi miliaran parameter, melalui proses pelatihan (*pre-training*) pada data teks tak berlabel dalam jumlah besar menggunakan pendekatan *machine learning* serta memiliki kemampuan untuk melakukan beragam penyelesaian masalah dalam bidang NLP [6]. Ciri utama LLM dibandingkan dengan model bahasa lainnya ada pada proses *pre-training* menggunakan data ekstensif yang bertujuan agar model dapat mempelajari representasi bahasa secara umum melalui *self-supervised learning* [12]. Sebagian besar LLM dibangun di atas arsitektur *transformer* [13]. LLM yang digunakan dalam penelitian ini dipilih berdasarkan beberapa kriteria utama: tersedia secara *open-source*, mendukung Bahasa Indonesia, tersedia dalam versi *instruction-tuned*, serta memiliki kemampuan pemrosesan konteks yang cukup panjang. Berdasarkan kriteria tersebut, dua model dipilih, yaitu Phi-3-mini [14] sebagai *baseline* dan Gemma 3 [15] sebagai model utama.

Phi-3-mini yang dikembangkan oleh Microsoft Research merupakan model berparameter 3,8 miliar yang dilatih pada 3,3 triliun *token*. Model ini dirancang dengan prinsip *data-optimal training*, yang menekankan kualitas data pelatihan. Phi-3-mini menggunakan arsitektur *transformer decoder-only* dengan *context window* 4.000 *token* dan *tokenizer* Llama-2. Meskipun kecil, model ini efisien dan mendukung format instruksi, menjadikannya tolok ukur yang relevan dalam eksperimen. Kapasitas *context window* 4.000 *token* ini masih cukup untuk memuat artikel Wikipedia dan satu berita, namun termasuk batas bawah dibandingkan model lain. Oleh karena itu, model ini digunakan sebagai tolok ukur minimum performa.

Gemma 3 dikembangkan oleh Google DeepMind dan dirancang untuk menghasilkan teks multibahasa yang faktual serta koheren. Gemma 3 tersedia dalam beberapa ukuran parameter, yaitu 1B, 4B, 12B, dan 2. Varian 1B mendukung *context window* hingga 32.000 *token* dan varian lainnya mendukung hingga 128.000 *token*. Arsitekturnya menggunakan *transformer decoder-only* dan mengadopsi sejumlah optimasi seperti *interleaved local-global attention*, *Grouped-Query Attention* (GQA), *Rotary Positional Embeddings* (RoPE), yang memungkinkan efisiensi komputasi pada *input* panjang. Gemma 3 dilatih menggunakan data multibahasa berskala besar dan telah melalui tahap *instruction tuning* menggunakan distilasi pengetahuan serta *Reinforcement Learning with Human Feedback* (RLHF). Gemma 3 dipilih sebagai model utama karena kemampuannya dalam menangani *input* yang panjang, sehingga memungkinkan pemrosesan gabungan artikel Wikipedia dan berita dalam satu *prompt* secara utuh. Hal ini memberikan fleksibilitas yang lebih besar dalam merancang strategi *prompting* dan pelatihan lanjutan (*fine-tuning*) untuk menghasilkan keluaran yang lebih relevan dan kontekstual. Dalam penelitian ini, dua varian Gemma 3 digunakan, yaitu Gemma 3 1B dan Gemma 3 4B.

## 2.4 Skenario Eksperimen

Skenario eksperimen diawali tahap konstruksi *prompt* dengan menentukan perancangan strategi *prompting* untuk mengevaluasi kemampuan model. *Prompt* dalam penelitian ini dirancang untuk mengarahkan model agar mampu mengidentifikasi informasi baru yang relevan tentang tokoh dari artikel berita dan belum tercantum dalam artikel Wikipedia yang diberikan. Selain itu, *prompt* disusun sedemikian rupa untuk menghasilkan keluaran yang sesuai dengan aturan dan gaya penulisan Wikipedia, serta memastikan bahwa informasi yang ditambahkan bersifat baru tanpa mengulang isi yang telah ada sebelumnya. Eksperimen dilakukan dengan menggunakan lima strategi *prompting* yang berbeda untuk mengamati bagaimana variasi format dan konteks pada *input* memengaruhi kualitas *output* model.

### a. *Simple Prompt*

Strategi ini memberikan instruksi langsung kepada model tanpa contoh sebelumnya (*zero-shot*). Model diminta membaca satu artikel Wikipedia dan satu artikel berita, lalu menghasilkan satu atau dua kalimat informasi baru dari berita yang belum ada dalam artikel Wikipedia. Format *prompt* ini ditunjukkan pada Tabel 2.

**Tabel 2.** Format *simple prompt*

<i>user</i>	<p>Berikut adalah artikel Wikipedia dan sebuah artikel berita. Tugas Anda adalah membaca keduanya, lalu menambahkan satu atau dua kalimat faktual yang berkaitan dengan tokoh dalam artikel Wikipedia. Informasi tersebut harus berasal dari berita dan belum terdapat dalam artikel Wikipedia. Jangan menyalin ulang artikel Wikipedia. Jangan mengubah atau menyempurnakan teks yang sudah ada. Tulislah hanya informasi baru yang layak ditambahkan ke artikel Wikipedia, menggunakan gaya penulisan netral khas Wikipedia, dalam Bahasa Indonesia.</p> <p>Artikel Wikipedia: {artikel wikipedia} Berita: {berita}</p>
-------------	---

### b. *System Prompt*

Strategi ini menambahkan arahan yang lebih eksplisit kepada model tentang hal-hal yang boleh dan tidak boleh dilakukan pada *system prompt*. Dua versi digunakan: satu dalam bahasa Inggris (Tabel 3) dan satu dalam bahasa Indonesia (Tabel 4) untuk membandingkan pengaruh bahasa instruksi terhadap keluaran model.

**Tabel 3.** Format *system prompt* dengan Bahasa Inggris

<i>system</i>	<p>You are a helpful assistant that analyzes an Indonesian Wikipedia article and a news article. Your task is to output only new factual information from the provided news that is not in the Wikipedia article and should be added to the Wikipedia article. Do not rewrite, rephrase, or enhance the original article. Do not include opinions, speculation, or promotional language. The output must be written in Bahasa Indonesia, in the style and tone of a Wikipedia article. Write the new content as a coherent sentences or paragraph suitable for direct inclusion in a Wikipedia article. Do not include any introduction, summary, or meta-comments. Start directly with the new content, as it would appear in a real Wikipedia article.</p>
<i>user</i>	<p>Berikut adalah artikel Wikipedia dan sebuah artikel berita. Tugas Anda adalah membaca keduanya, lalu menambahkan satu atau dua kalimat faktual yang berkaitan dengan tokoh dalam artikel Wikipedia. Informasi tersebut harus berasal dari berita dan belum terdapat dalam artikel Wikipedia. Jangan menyalin ulang artikel Wikipedia. Jangan mengubah atau menyempurnakan teks yang sudah ada. Tulislah hanya informasi baru yang layak ditambahkan ke artikel Wikipedia, menggunakan gaya penulisan netral khas Wikipedia, dalam Bahasa Indonesia.</p> <p>Artikel Wikipedia: {artikel wikipedia} Berita: {berita}</p>

**Tabel 4.** Format *system prompt* dengan Bahasa Indonesia

<i>system</i>	<p>Anda adalah asisten yang membantu menganalisis artikel Wikipedia berbahasa Indonesia dan sebuah artikel berita. Tugas Anda adalah mengeluarkan informasi faktual baru yang belum terdapat dalam artikel Wikipedia dari berita yang diberikan dan layak untuk ditambahkan ke artikel Wikipedia tersebut. Jangan menulis ulang, mengubah, atau menyempurnakan isi artikel Wikipedia yang sudah ada. Jangan menyertakan opini, spekulasi, atau bahasa yang bersifat promosi. Keluaran harus ditulis dalam Bahasa Indonesia, dengan gaya penulisan dan nada yang sesuai dengan artikel Wikipedia. Tulislah informasi baru dalam bentuk kalimat atau paragraf yang koheren dan dapat langsung dimasukkan ke dalam artikel Wikipedia. Jangan menambahkan pengantar, ringkasan, atau komentar meta apa pun. Langsung mulai dengan konten baru, seperti layaknya bagian dari artikel Wikipedia.</p>
---------------	--

---

<i>user</i>	<p>Berikut adalah artikel Wikipedia dan sebuah artikel berita. Tugas Anda adalah membaca keduanya, lalu menambahkan satu atau dua kalimat faktual yang berkaitan dengan tokoh dalam artikel Wikipedia. Informasi tersebut harus berasal dari berita dan belum terdapat dalam artikel Wikipedia. Jangan menyalin ulang artikel Wikipedia. Jangan mengubah atau menyempurnakan teks yang sudah ada. Tulislah hanya informasi baru yang layak ditambahkan ke artikel Wikipedia, menggunakan gaya penulisan netral khas Wikipedia, dalam Bahasa Indonesia.</p> <p>Artikel Wikipedia: {artikel wikipedia} Berita: {berita}</p>
-------------	---

---

c. *One-shot*

Strategi ini menambahkan pendekatan *in-context learning* dengan memberikan satu contoh lengkap pasangan *input-output* yang sesuai. Format *prompt one-shot* ditunjukkan pada Tabel 5.

**Tabel 5.** Format *prompt one-shot*

<i>system</i>	<p>You are a helpful assistant that analyzes an Indonesian Wikipedia article and a news article. Your task is to output only new factual information from the provided news that is not in the Wikipedia article and should be added to the Wikipedia article. Do not rewrite, rephrase, or enhance the original article. Do not include opinions, speculation, or promotional language. The output must be written in Bahasa Indonesia, in the style and tone of a Wikipedia article. Write the new content as a coherent sentences or paragraph suitable for direct inclusion in a Wikipedia article. Do not include any introduction, summary, or meta-comments. Start directly with the new content, as it would appear in a real Wikipedia article.</p>
<i>user</i>	<p>Berikut adalah artikel Wikipedia dan sebuah artikel berita. Tugas Anda adalah membaca keduanya, lalu menambahkan satu atau dua kalimat faktual yang berkaitan dengan tokoh dalam artikel Wikipedia. Informasi tersebut harus berasal dari berita dan belum terdapat dalam artikel Wikipedia. Jangan menyalin ulang artikel Wikipedia. Jangan mengubah atau menyempurnakan teks yang sudah ada. Tulislah hanya informasi baru yang layak ditambahkan ke artikel Wikipedia, menggunakan gaya penulisan netral khas Wikipedia, dalam Bahasa Indonesia.</p> <p>Artikel Wikipedia: {artikel wikipedia} Berita: {berita}</p>
<i>model</i>	<p>{contoh referensi pembaruan}</p>
<i>user</i>	<p>Berikut adalah artikel Wikipedia dan sebuah artikel berita. Tugas Anda adalah membaca keduanya, lalu menambahkan satu atau beberapa kalimat berisi informasi baru dari berita yang berkaitan dengan tokoh dalam artikel Wikipedia dan belum ada pada artikel. Jangan menyalin ulang artikel Wikipedia. Tulislah hanya informasi baru yang layak ditambahkan ke artikel Wikipedia, menggunakan gaya penulisan netral khas Wikipedia, dalam Bahasa Indonesia.</p> <p>Artikel Wikipedia: {artikel wikipedia} Berita: {berita}</p>

d. *Prompt Chaining*

Strategi ini memecah tugas menjadi dua langkah terpisah. Pada tahap pertama, model mengekstrak poin-poin informasi baru dari berita. Pada tahap kedua, model menyusun kalimat berbasis poin-poin tersebut untuk dimasukkan ke artikel Wikipedia. Format lengkap *prompt chaining* disajikan pada Tabel 6.

**Tabel 6.** Format *prompt chaining*

<i>user</i> (tahap 1)	<p>Berikut adalah artikel biografi tokoh Wikipedia dan sebuah artikel berita. Tugas Anda adalah mengekstraksi informasi faktual baru dari berita yang berkaitan dengan tokoh dalam artikel Wikipedia, dan belum disebutkan dalam artikel Wikipedia. Informasi tersebut dapat hanya terdiri dari satu atau beberapa poin saja, tuliskan hanya jika benar-benar baru dan relevan. Tuliskan hasil ekstraksi dalam bentuk poin-poin ringkas, dalam Bahasa Indonesia. Jangan menyalin ulang atau mengubah isi artikel Wikipedia.</p> <p>Artikel Wikipedia: {artikel wikipedia} Berita: {berita}</p> <p>Poin-poin informasi faktual baru:</p>
<i>model</i> (tahap 1)	<p>{poin-poin informasi faktual baru}</p>

*user* (tahap 2) Berikut adalah beberapa poin informasi faktual baru hasil ekstraksi dari sebuah artikel berita. Tugas Anda adalah menyusun satu atau beberapa kalimat baru berdasarkan poin-poin tersebut, yang dapat langsung ditambahkan ke artikel Wikipedia. Tulislah dalam Bahasa Indonesia, menggunakan gaya penulisan netral khas Wikipedia. Jangan menambahkan opini, spekulasi, atau komentar.  
Poin-poin informasi baru: {poin-poin informasi faktual baru}  
Teks yang akan ditambahkan pada artikel Wikipedia:

Setelah skenario *prompting* disusun, eksperimen dilakukan menggunakan model LLM yang dipilih. Eksperimen dilakukan dalam dua tahap utama. Eksperimen pertama menggunakan ketiga model *pre-trained*, yaitu Phi-3-mini, Gemma 3 1B, dan Gemma 3 4B, untuk menjalankan proses inferensi terhadap data uji. Masing-masing model diuji dengan lima strategi *prompting* yang telah dirancang. Eksperimen kedua dilakukan dengan melakukan *fine-tuning* metode LoRA [16] pada model Gemma 3 1B dan 4B menggunakan data pelatihan yang telah dikumpulkan. Model hasil *fine-tuning* kemudian digunakan untuk proses inferensi pada data uji yang sama dan dengan lima strategi *prompting* yang sama untuk mengetahui sejauh mana proses *fine-tuning* dapat meningkatkan kualitas keluaran dibandingkan model *pre-trained*.

Penjelasan skenario eksperimen ditunjukkan pada Tabel 7 yang merangkum kombinasi antara strategi *prompting* dan model LLM yang digunakan. Simbol centang (✓) pada tabel menunjukkan bahwa skenario tersebut dijalankan pada model terkait. Seluruh eksperimen dilakukan pada data uji yang sama untuk menjaga konsistensi dan memastikan perbandingan antar model dan strategi dapat dilakukan secara adil.

**Tabel 7.** Rancangan eksperimen berdasarkan model dan strategi *prompting*

Strategi <i>Prompting</i>	Phi-3-mini ( <i>pretrained</i> )	Gemma 3 1B ( <i>pretrained</i> )	Gemma 3 4B ( <i>pretrained</i> )	Gemma 3 1B ( <i>fine-tuned</i> )	Gemma 3 4B ( <i>fine-tuned</i> )
<i>Simple prompt</i>	✓	✓	✓	✓	✓
<i>System prompt</i> (en)	✓	✓	✓	✓	✓
<i>System prompt</i> (id)	✓	✓	✓	✓	✓
<i>One-shot</i>	✓	✓	✓	✓	✓
<i>Prompt chaining</i>	✓	✓	✓	✓	✓

## 2.5 Evaluasi

Hasil keluaran dari kedua eksperimen yang telah dilakukan akan dievaluasi. Penilaian dilakukan dengan membandingkan hasil keluaran model dengan referensi pembaruan yang benar-benar ditambahkan oleh penyunting pada artikel Wikipedia. Evaluasi menggunakan metrik ROUGE [17] yang mencakup ROUGE-1 (*unigram*), ROUGE-2 (*bigram*), dan ROUGE-L (*Longest Common Subsequence*) untuk menilai kesamaan isi secara tekstual. Hasil evaluasi digunakan untuk menilai efektivitas strategi *prompting* dan konfigurasi model dalam menghasilkan teks pembaruan artikel Wikipedia.

## 3. HASIL DAN PEMBAHASAN

Eksperimen dilakukan dengan memberikan masukan berupa artikel Wikipedia sebelum pembaruan dan artikel berita kepada model menggunakan skenario eksperimen yang telah dirancang. Eksperimen pertama dilakukan pada 20 data uji yang telah dikumpulkan sebelumnya. Tiga model *pre-trained* digunakan dalam tahap inferensi, yaitu Phi-3-mini (*baseline*), Gemma 3 1B, dan Gemma 3 4B. Inferensi dilakukan menggunakan platform Ollama dengan pengaturan parameter *temperature* = 1, *top\_k* = 64, dan *top\_p* = 0.95. Tabel 8 menyajikan nilai rata-rata hasil evaluasi ROUGE pada model Phi-3-mini pada berbagai strategi *prompting*. Hasil menunjukkan bahwa strategi *system prompt* dalam bahasa Inggris memberikan skor ROUGE tertinggi secara konsisten untuk Phi-3-mini. Hal ini sejalan dengan asumsi bahwa model Phi-3-mini lebih terlatih pada instruksi dalam bahasa Inggris.

**Tabel 8.** Rata-rata nilai ROUGE Phi-3-mini

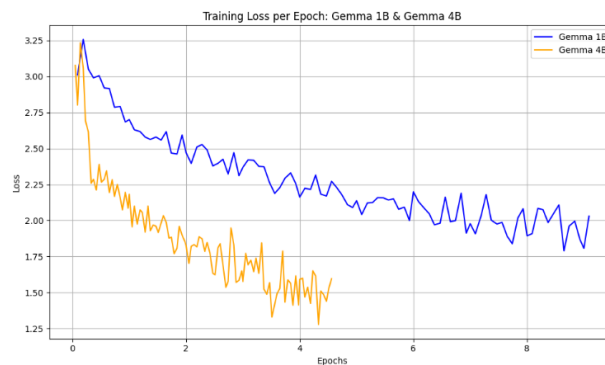
Strategi <i>Prompting</i>	ROUGE-1	ROUGE-2	ROUGE-L
<i>Simple prompt</i>	0.0939	0.0288	0.0675
<i>System prompt</i> (en)	<b>0.1746</b>	<b>0.0793</b>	<b>0.1257</b>
<i>System prompt</i> (id)	0.1083	0.0342	0.0735
<i>One-shot</i>	0.0744	0.0208	0.0503
<i>Prompt chaining</i>	0.1528	0.0377	0.1090

Selanjutnya dilakukan *fine-tuning* terhadap dua model Gemma 3, yaitu versi 1B dan 4B, menggunakan data pelatihan yang telah disiapkan dengan format *simple prompt*. Proses *fine-tuning* dilakukan dengan konfigurasi *hyperparameter* pada Tabel 9.

**Tabel 9.** Konfigurasi *hyperparameter fine-tuning*

Parameter	Gemma 3 1B	Gemma 3 4B
Model <i>Pretrained</i>	unsloth/gemma-3-1b-it	unsloth/gemma-3-4b-it
Max Context Length	30,000 <i>token</i>	100,000 <i>token</i>
Load in 8bit / 4bit	8bit	4bit
Full Finetuning	Tidak	Tidak
LoRA r	8	8
LoRA alpha	8	8
LoRA dropout	0.05	0.05
Epoch	9	4.5
Max steps	100	100
Learning rate	2e-4	2e-4
Batch size	2	1
Optimizer	adamw 8bit	adamw 8bit

*Fine-tuning* dilakukan menggunakan *Trainer* dari Hugging Face yang telah diadaptasi oleh *library* Unsloth untuk mendukung efisiensi pada model Gemma 3. Untuk model Gemma 3 1B, total waktu pelatihan sekitar 3003 detik atau sekitar 50 menit, dengan rata-rata *loss* akhir sebesar 2.27. Untuk model Gemma 3 4B, pelatihan berlangsung selama 2694 detik atau sekitar 45 menit, dengan *loss* akhir lebih rendah yaitu 1.87. Gambar 3 menunjukkan grafik penurunan *loss* berdasarkan *epoch* selama pelatihan kedua model. Garis biru menunjukkan Gemma 3 1B dan garis oranye menunjukkan Gemma 3 4B.



**Gambar 3** Grafik *Loss Finetuning* Gemma 1B dan 4B

Setelah dilakukan *fine-tuning* terhadap model Gemma 3 1B dan 4B, model diuji dengan dilakukan inferensi menggunakan data uji yang sama dengan saat inferensi sebelum *fine-tuning*. Inferensi dilakukan menggunakan *library Transformers* dari Hugging Face dengan pengaturan parameter yang sama seperti inferensi sebelum *fine-tuning*, yaitu *temperature* = 1, *top\_k* = 64, dan *top\_p* = 0.95. Tabel 10 menyajikan nilai rata-rata ROUGE untuk model Gemma 1B dan Gemma 4B dalam kondisi *base* (belum dilatih ulang) maupun *finetuned*, berdasarkan lima strategi *prompting*.

**Tabel 10.** Rata-rata nilai ROUGE Gemma 3

Strategi <i>Prompting</i>	Model	ROUGE-1	ROUGE-2	ROUGE-L	
Simple prompt	Gemma 1B	Base	<b>0.1493</b>	<b>0.0579</b>	<b>0.1055</b>
		Finetuned	0.0499	0.0057	0.0416
	Gemma 4B	Base	<b>0.3281</b>	<b>0.1585</b>	<b>0.2548</b>
		Finetuned	0.1047	0.0432	0.0802
System prompt (en)	Gemma 1B	Base	<b>0.2653</b>	<b>0.1209</b>	<b>0.1737</b>
		Finetuned	0.2085	0.0766	0.1539
	Gemma 4B	Base	0.2555	0.1111	0.1937
		Finetuned	<b>0.2795</b>	<b>0.1253</b>	<b>0.2181</b>
System prompt (id)	Gemma 1B	Base	<b>0.2628</b>	<b>0.1301</b>	<b>0.1884</b>
		Finetuned	0.1902	0.0664	0.1402
	Gemma 4B	Base	0.2443	0.0926	0.1864
		Finetuned	<b>0.3096</b>	<b>0.1559</b>	<b>0.2618</b>
One-shot	Gemma 1B	Base	<b>0.2519</b>	<b>0.1249</b>	0.1833
		Finetuned	0.2505	0.1062	<b>0.1935</b>
	Gemma 4B	Base	0.2705	0.1268	0.2099
		Finetuned	<b>0.3552</b>	<b>0.1685</b>	<b>0.2952</b>
Prompt chaining	Gemma 1B	Base	0.1804	0.0605	0.1242

	<i>Finetuned</i>	<b>0.2023</b>	<b>0.0618</b>	<b>0.1415</b>
Gemma 4B	<i>Base</i>	0.3447	0.1725	0.2551
	<i>Finetuned</i>	<b>0.3687</b>	<b>0.1920</b>	<b>0.2887</b>

Jika dibandingkan antar model, Gemma 4B melampaui Gemma 1B pada semua strategi, baik dalam kondisi *base* maupun *finetuned*. Perbedaan ini dapat dikaitkan dengan kapasitas parameter yang lebih besar pada Gemma 4B, yang memungkinkan pemrosesan konteks dan representasi informasi yang lebih kompleks dan akurat. Strategi *simple prompt* digunakan pada saat *fine-tuning*, namun menghasilkan nilai ROUGE yang justru lebih rendah dibandingkan model sebelum *finetune*. Bahkan nilai rata-rata ROUGE pada model Gemma 1B setelah *finetune* lebih rendah dari *baseline*. Permasalahan utama yang ditemukan adalah banyaknya kasus di mana model tidak menghasilkan keluaran selama inferensi. Hal ini mengindikasikan bahwa model kemungkinan mengalami *overfitting* terhadap pola *prompt* tersebut selama proses pelatihan, sehingga justru gagal merespons saat diberi masukan yang sama. Selain itu, jumlah data pelatihan yang terbatas (hanya 172 data) membuat model kurang mampu melakukan generalisasi, yang memperparah kegagalan dalam menghasilkan teks. Akibatnya, meskipun model telah melalui proses *fine-tuning*, performanya menurun karena banyak prediksi yang kosong, sehingga berdampak langsung pada rendahnya skor rata-rata ROUGE.

Selain strategi *simple prompt*, terlihat bahwa model Gemma 4B yang telah di-*finetune* menunjukkan peningkatan kinerja pada seluruh strategi *prompting* dibandingkan dengan model *base*. Hal ini memperlihatkan bahwa proses *finetuning* berhasil meningkatkan kemampuan model dalam menghasilkan keluaran yang lebih relevan dengan referensi, sebagaimana tercermin dari peningkatan rata-rata skor ROUGE. Strategi *prompting* yang memberikan hasil terbaik adalah *prompt chaining*. Tabel 11 menampilkan contoh keluaran dari berbagai model LLM dengan menerapkan strategi *prompt chaining*. Dari hasil yang ditampilkan, terlihat bahwa model *baseline* Phi-3-mini dan Gemma 1B versi *base* masih memberikan keluaran dalam bentuk poin-poin. Hal ini menunjukkan keterbatasan model dalam memahami instruksi pada langkah kedua. Model Gemma 1B yang telah di-*finetune* memberikan keluaran kalimat yang sesuai dengan struktur dan gaya penulisan Wikipedia, namun informasi yang diberikan tidak lengkap dan terdapat kesalahan faktual pada tanggal lahir. Model Gemma 4B versi *base* menghasilkan teks yang lebih panjang dan kaya informasi, namun masih menyertakan informasi berlebihan yang kurang sesuai dengan gaya ringkas dan netral Wikipedia. Hasil keluaran model Gemma 4B versi *finetuned* adalah yang paling mendekati referensi baik dari segi struktur dan penulisan, meskipun informasi tahun yang dihasilkan bersifat asumptif karena tidak tercantum secara eksplisit dalam artikel berita yang menjadi sumber informasi.

**Tabel 11.** Contoh keluaran model dengan strategi *prompt chaining*

	<b>Output Teks Pembaruan</b>
<b>Referensi</b>	Rio dan Atiqah telah dikaruniai seorang putri bernama Salma Jihane Putri Dewanto.
Phi-3-mini ( <i>baseline</i> )	1. Rio Dewanto adalah anak kucing pertama yang diakui dari Salma Jihane Putri Dewanto, puteri dalam umur di mana seorang aktris Atiqah Hasiholan berpacaran dengan mereka lewat 3,5 tahun. ...
Gemma 1B ( <i>base</i> )	Berikut beberapa kalimat baru yang layak ditambahkan ke artikel Wikipedia tentang Rio Dewanto dan Atiqah Hasiholan, berdasarkan poin-poin yang diberikan, dengan gaya penulisan netral khas Wikipedia: * Rio Dewanto dan Atiqah Hasiholan adalah pasangan pertama yang memiliki anak. ...
Gemma 1B ( <i>finetuned</i> )	Anak pertama dari pasangan Rio Dewanto dan Atiqah Hasiholan ini lahir pada Sabtu, 11 Maret 2023.
Gemma 4B ( <i>base</i> )	Pada 23 Juni 2016, Rio Dewanto dan Atiqah Hasiholan menyambut kelahiran putri pertama mereka, Salma Jihane Putri Dewanto, lahir pukul 08:30 WIB di RSIA Kemang Medical Care, Jakarta Selatan, dengan berat badan 3,38 kg dan panjang 51 cm. Rio Dewanto menyatakan bahwa nama Salma dipilih oleh Atiqah Hasiholan dan ia sendiri tidak mengetahui arti nama tersebut, namun ia ingin nama anaknya tetap mencantumkan unsur nama belakangnya, "Dewanto".
Gemma 4B ( <i>finetuned</i> )	Rio Dewanto dan Atiqah Hasiholan dikaruniai seorang putri bernama Salma Jihane Putri Dewanto pada Jumat, 23 Juni 2023.

Secara keseluruhan, hasil evaluasi menunjukkan bahwa model yang telah di-*finetune*, terutama Gemma 4B, memberikan keluaran yang paling sesuai baik dari sisi gaya penulisan maupun relevansi informasi. Temuan ini memperkuat potensi LLM sebagai alat bantu dalam proses pembaruan artikel Wikipedia.

## 4. KESIMPULAN

Berdasarkan hasil penelitian ini, dapat disimpulkan bahwa penggunaan model bahasa besar (LLM) untuk tugas penambahan informasi pada artikel biografi Wikipedia berbasis berita dapat memberikan hasil yang cukup baik, khususnya jika dilakukan dengan proses *fine-tuning* dan pemilihan strategi *prompting* yang tepat. Model utama yang digunakan dalam penelitian ini, yaitu Gemma 3, menunjukkan performa yang secara signifikan lebih tinggi dibandingkan

model *baseline* Phi-3-mini pada semua strategi *prompting*. Nilai ROUGE yang dihasilkan oleh Gemma 4B setelah *fine-tuning*, terutama dengan strategi *prompt chaining*, merupakan yang tertinggi di antara semua kombinasi, menandakan bahwa model mampu memahami instruksi kompleks dan menghasilkan keluaran yang lebih relevan serta informatif. Meskipun demikian, penelitian ini juga mengungkap beberapa tantangan, salah satunya adalah fenomena di mana model yang telah di-*finetune* menggunakan *simple prompt* justru menunjukkan penurunan skor ROUGE. Hal ini disebabkan oleh kesamaan format *prompt* pada saat pelatihan dan inferensi, yang menyebabkan model tidak menghasilkan keluaran pada sebagian data. Temuan ini menunjukkan bahwa penting untuk melakukan eksplorasi terhadap variasi *prompt* dalam proses pelatihan agar model tidak mengalami stagnasi saat inferensi. Selain itu, keterbatasan jumlah data pelatihan juga menjadi faktor yang membatasi potensi peningkatan performa model secara optimal. Namun, dengan konfigurasi pelatihan yang disesuaikan dan pemanfaatan LoRA *fine-tuning*, hasil yang diperoleh tetap menunjukkan bahwa *fine-tuning* skala kecil pada LLM seperti Gemma masih mampu meningkatkan kualitas keluaran model secara bermakna. Hasil penelitian ini menunjukkan potensi untuk pengembangan lanjutan dalam otomatisasi pembaruan konten Wikipedia menggunakan pendekatan LLM yang lebih optimal.

## REFERENCES

- [1] Wikipedia contributors, "Wikipedia." Accessed: Apr. 04, 2024. [Online]. Available: <https://id.wikipedia.org/wiki/Wikipedia>
- [2] Wikimedia Foundation, "Wikimedia Statistics." Accessed: Apr. 04, 2024. [Online]. Available: <https://stats.wikimedia.org/>
- [3] Wikipedia contributors, "Wikipedia." Accessed: Apr. 04, 2024. [Online]. Available: <https://id.wikipedia.org/wiki/Wikipedia:Pengumuman/Artikel>
- [4] Wikimedia Foundation, "Wikimedia." Accessed: Apr. 04, 2024. [Online]. Available: <https://petscan.wmcloud.org/>
- [5] E. Kasneci *et al.*, "ChatGPT for good? On opportunities and challenges of large language models for education," Apr. 01, 2023, *Elsevier Ltd.* doi: 10.1016/j.lindif.2023.102274.
- [6] M. A. K. Raiaan *et al.*, "A Review on Large Language Models: Architectures, Applications, Taxonomies, Open Issues and Challenges," *IEEE Access*, vol. 12, pp. 26839–26874, 2024, doi: 10.1109/ACCESS.2024.3365742.
- [7] P. J. Liu *et al.*, "Generating Wikipedia by Summarizing Long Sequences," Jan. 2018, [Online]. Available: <http://arxiv.org/abs/1801.10198>
- [8] S. Banerjee and P. Mitra, "WikiWrite: Generating Wikipedia Articles Automatically." [Online]. Available: <https://en.wikipedia.org/wiki/Wikipedia:Notability>
- [9] R. Chandraseta and M. L. Khodra, "Composing Indonesian Paragraph for Biography Domain using Extracive Summarization," 2019.
- [10] R. Yang *et al.*, "Evaluating Large Language Models on Wikipedia-Style Survey Generation", doi: 10.48550/arXiv.2308.10410.
- [11] Y. Shao, Y. Jiang, T. A. Kanell, P. Xu, O. Khattab, and M. S. Lam, "Assisting in Writing Wikipedia-like Articles From Scratch with Large Language Models," Feb. 2024, [Online]. Available: <http://arxiv.org/abs/2402.14207>
- [12] Y. Liu *et al.*, "Understanding LLMs: A Comprehensive Overview from Training to Inference," Jan. 2024, [Online]. Available: <http://arxiv.org/abs/2401.02038>
- [13] A. Vaswani *et al.*, "Attention Is All You Need," 2017.
- [14] M. Abdin *et al.*, "Phi-3 Technical Report: A Highly Capable Language Model Locally on Your Phone," Aug. 2024, [Online]. Available: <http://arxiv.org/abs/2404.14219>
- [15] Gemma Team *et al.*, "Gemma 3 Technical Report," Mar. 2025, [Online]. Available: <http://arxiv.org/abs/2503.19786>
- [16] E. J. Hu *et al.*, "LoRA: Low-Rank Adaptation of Large Language Models," Jun. 2021, [Online]. Available: <http://arxiv.org/abs/2106.09685>
- [17] C.-Y. Lin, "ROUGE: A Package for Automatic Evaluation of Summaries."