

## **BAB 1**

### **PENDAHULUAN**

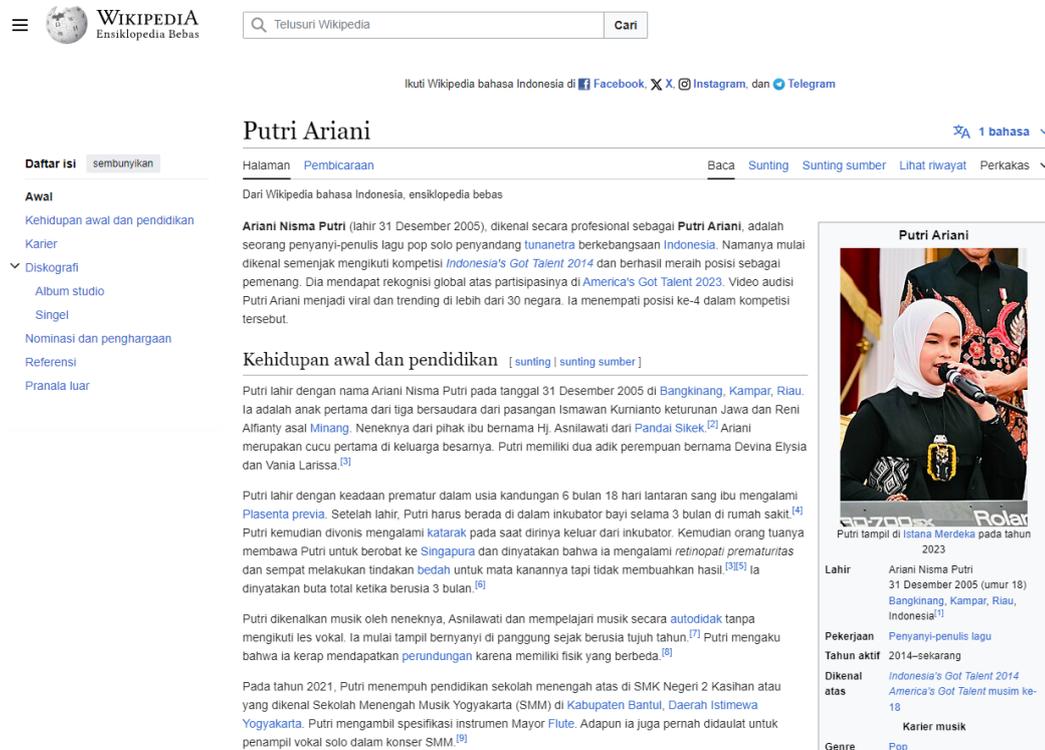
#### **1.1 Latar Belakang**

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 (Wikipedia contributors, 2024a). 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 (Wikimedia Foundation, 2024). 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 (Wikimedia Foundation, 2024). Terdapat 1.474.915 pengguna terdaftar, namun hanya 2.404 pengguna yang aktif pada bulan April 2024 dan 47 pengurus aktif Wikipedia Indonesia (Wikipedia contributors, 2024b). Keterbatasan jumlah pengguna aktif dan kurangnya penyunting yang terlibat secara rutin ini dapat mempengaruhi kualitas dan akurasi informasi pada Wikipedia Indonesia.

Artikel dengan kategori biografi pada Wikipedia Indonesia merupakan artikel yang menjelaskan tentang kehidupan seseorang. 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 (Wikimedia Foundation, 2024). Artikel biografi merupakan artikel yang perlu perubahan secara rutin karena adanya perkembangan karier, 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.



The image shows a screenshot of the Wikipedia article for Putri Ariani. The article is in Indonesian and is titled "Putri Ariani". It includes a navigation bar with tabs for "Halaman", "Pembicaraan", "Baca", "Sunting", "Sunting sumber", "Lihat riwayat", and "Perkakas". The main text of the article is visible, starting with "Ariani Nisma Putri (lahir 31 Desember 2005), dikenal secara profesional sebagai Putri Ariani, adalah seorang penyanyi-penulis lagu pop solo penyandang tunanetra berkebangsaan Indonesia. Namanya mulai dikenal semenjak mengikuti kompetisi Indonesia's Got Talent 2014 dan berhasil meraih posisi sebagai pemenang. Dia mendapat rekognisi global atas partisipasinya di America's Got Talent 2023. Video audisi Putri Ariani menjadi viral dan trending di lebih dari 30 negara. Ia menempati posisi ke-4 dalam kompetisi tersebut."

There is also a small image of Putri Ariani performing on stage, and a table of her personal information:

<b>Lahir</b>	Ariani Nisma Putri 31 Desember 2005 (umur 18) Bangkinang, Kampar, Riau, Indonesia <sup>[1]</sup>
<b>Pekerjaan</b>	Penyanyi-penulis lagu
<b>Tahun aktif</b>	2014–sekarang
<b>Dikenal atas</b>	Indonesia's Got Talent 2014 America's Got Talent musim ke-18
<b>Genre</b>	Pop

Gambar 1.1 Contoh Artikel Biografi Wikipedia

Salah contoh satu artikel Wikipedia Indonesia yang belum diperbarui adalah artikel biografi Putri Ariani pada Gambar 1.1, seorang penyanyi-penulis lagu pop solo yang juga penyandang tunanetra. Meskipun Putri Ariani merilis single kolaboratif dengan produser terkenal Alan Walker dan penyanyi Peder Elias pada bulan Januari 2024, informasi ini belum tercantum dalam artikel biografi Wikipedia Indonesia miliknya hingga bulan April 2024. Padahal informasi tersebut merupakan salah satu prestasi Putri Ariani yang telah diberitakan di berbagai situs berita daring seperti CNN Indonesia, Kompas, dan IDN Times. 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.

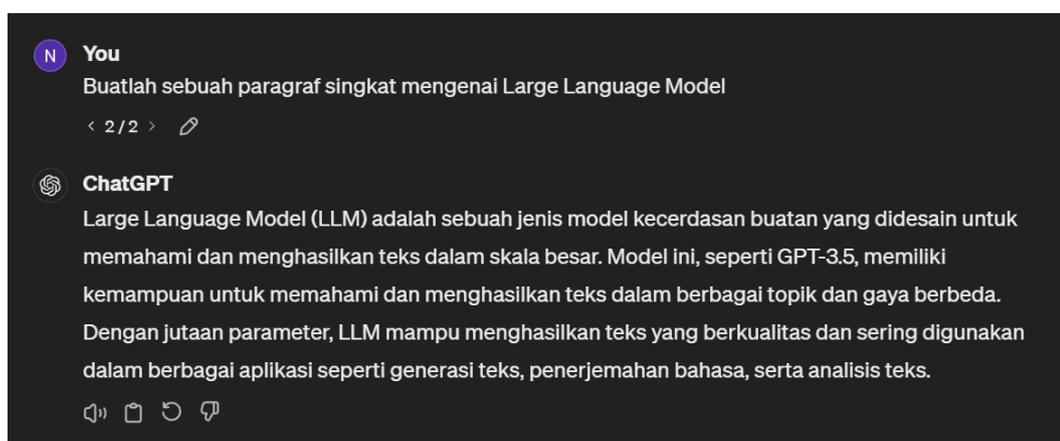
Pada proses pembaruan isi artikel, penyunting perlu mendapatkan sumber informasi terlebih dahulu dan membandingkannya dengan artikel yang sudah ada. Setelah itu, penyunting memberikan keputusan perlu atau tidaknya perbaikan atau

penambahan pada artikel. Jika diperlukan perbaikan atau penambahan maka penyunting akan melakukan perubahan terhadap artikel tersebut. Proses ini merupakan proses yang berulang dan cukup memakan waktu jika melihat perbandingan antara banyaknya artikel dan jumlah penyunting yang aktif, sehingga dibutuhkan otomatisasi dalam hal ini untuk memudahkan para penyunting.

Penelitian sebelumnya telah mengkaji mengenai pembaruan artikel biografi Wikipedia Indonesia menggunakan ringkasan ekstraktif dari berita daring (Chandraseta & Khodra, 2019). Pendekatan yang digunakan pada penelitian tersebut melibatkan pengumpulan artikel berita yang relevan, pembuatan ringkasan otomatis menggunakan model *Bi-directional Gated Recurrent Unit* (BiGRU), serta sejumlah langkah pemrosesan untuk memastikan informasi yang dihasilkan relevan, tidak berulang, dan sesuai dengan struktur artikel Wikipedia. Sistem yang dibangun dalam penelitian tersebut menunjukkan bahwa teknik pemrosesan bahasa alami dan peringkasan otomatis dapat dimanfaatkan untuk membantu pembaruan artikel Wikipedia secara lebih efisien. Selain itu, penelitian terkait pembuatan teks Wikipedia otomatis juga telah dilakukan dengan cara meringkas isi dari beberapa teks panjang (P. J. Liu dkk., 2018). Wikipedia yang digunakan pada penelitian tersebut adalah Wikipedia Bahasa Inggris.

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 (Kasneci dkk., 2023). 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 berhubungan dengan bahasa dengan akurasi tinggi (Kasneci dkk., 2023). Penggunaan LLM dapat membantu dalam otomatisasi penyelesaian masalah terkait bahasa yang sebelumnya memakan waktu dan tenaga kerja, sehingga dapat meningkatkan efisiensi dari berbagai pekerjaan (Raiaan dkk., 2024).

Salah satu contoh adaptasi LLM yang mudah diakses adalah ChatGPT yang bekerja dengan cara mengikuti instruksi yang diberikan, kemudian memberikan respons sesuai dengan instruksi tersebut. Pada Gambar 1.2, terdapat contoh teks yang dihasilkan oleh ChatGPT jika diberikan kalimat perintah “buatlah sebuah paragraf singkat mengenai Large Language Model”. Dari gambar tersebut dapat terlihat bahwa respon teks yang dihasilkan ChatGPT mirip dengan tulisan buatan manusia dan dapat dimengerti dengan baik.



Gambar 1.2 Contoh Teks Yang Dihasilkan LLM (ChatGPT)

Dalam penelitian ini akan dilakukan penerapan LLM dalam pembaruan artikel biografi Wikipedia Indonesia. Sumber informasi terbaru yang digunakan akan dibatasi pada situs berita daring. Hal ini dikarenakan berita merupakan salah satu sumber informasi tercepat yang mencakup fakta dan perkembangan terkini (A. S. Haris Sumadiria, 2005). Model LLM akan diterapkan dan diadaptasi dengan metode yang sesuai agar mampu memahami konteks artikel biografi tokoh serta gaya penulisan Wikipedia Indonesia. Dengan demikian, penelitian ini mengoptimalkan penerapan kinerja model LLM agar dapat menyelesaikan masalah yang lebih spesifik, yaitu pembaruan artikel biografi Wikipedia berdasarkan berita.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijabarkan sebelumnya, rumusan masalah penelitian ini adalah bagaimana penerapan *Large Language Models* (LLM) dalam pembaruan artikel biografi Wikipedia Indonesia dengan berita sebagai sumber informasi?

### 1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah diuraikan sebelumnya, maka tujuan dari penelitian yang dilakukan adalah menerapkan *Large Language Models* (LLM) dalam pembaruan artikel biografi Wikipedia Indonesia dengan berita sebagai sumber informasi.

### 1.4 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah:

1. Penelitian ini diharapkan dapat memberikan kontribusi pada pengembangan teknologi untuk pembaruan artikel Wikipedia Indonesia, khususnya pada artikel biografi tokoh Indonesia, dengan penerapan LLM.
2. Hasil penelitian diharapkan dapat membantu meningkatkan efisiensi dalam pembaruan artikel Wikipedia Indonesia.

### 1.5 Batasan Masalah

Penelitian ini memiliki beberapa batasan untuk menjaga ruang lingkup yang terfokus dan dapat diselesaikan. Batasan-batasan tersebut dijabarkan sebagai berikut:

1. Penelitian ini hanya berfokus pada penambahan informasi baru ke dalam artikel biografi Wikipedia, bukan penghapusan atau penyuntingan ulang terhadap informasi yang sudah ada. Informasi yang ditambahkan berasal dari satu artikel berita yang dijadikan sebagai sumber informasi.
2. Hanya artikel biografi tokoh Indonesia dalam Bahasa Indonesia yang menjadi objek pembaruan. Tokoh yang dimaksud adalah individu asal Indonesia, yang dibuktikan melalui isi artikel Wikipedia yang menyebutkan identitas kebangsaan atau asal Indonesia secara eksplisit.
3. Artikel berita yang digunakan sebagai sumber informasi baru merupakan berita daring berbahasa Indonesia yang dicantumkan pada bagian referensi dari Artikel Wikipedia Indonesia. Sumber berita digunakan sebagai landasan untuk menguji kemampuan model dalam mengekstrak dan menyisipkan informasi faktual yang relevan ke dalam struktur artikel Wikipedia.

4. Pembaruan hanya difokuskan pada penambahan informasi berbentuk teks, tanpa mencakup pembaruan pada tabel, *infobox*, atau elemen non-teks lainnya.
5. Model yang digunakan dalam eksperimen adalah model *pretrained instruct* Phi-3-mini (4K) dan Gemma 3 (1B dan 4B).
  - a. Phi-3-mini digunakan sebagai *baseline* karena memiliki lisensi terbuka, ukurannya cukup kecil, dan keterbatasan panjang konteks (*context window*) yaitu 4.000 *token*. Kapasitas *context window* ini cukup untuk memuat artikel Wikipedia dan satu berita, namun termasuk batas bawah dibandingkan model lain. Dengan keterbatasan tersebut, model ini digunakan untuk memberikan tolok ukur minimum terhadap kualitas *output* yang dihasilkan.
  - b. Gemma 3 dipilih sebagai model utama karena mendukung pemrosesan dalam Bahasa Indonesia, memiliki lisensi terbuka, dan menawarkan *context window* hingga 128.000 *token* serta memiliki berbagai ukuran parameter. Kemampuan ini memungkinkan pemrosesan *input* panjang berupa gabungan artikel Wikipedia dan berita dalam satu *prompt*, serta memberi ruang yang lebih luas untuk eksplorasi strategi *prompting* dan *fine-tuning*.