

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Tingkat literasi di Indonesia menunjukkan tren positif dalam beberapa tahun terakhir. Berdasarkan data dari Perpustakaan Nasional Republik Indonesia (Perpusnas), Indeks Pembangunan Literasi Masyarakat (IPLM) nasional pada tahun 2024 mencapai skor 73,52, melampaui target 71,4 dan meningkat dari skor 69,42 pada tahun sebelumnya. Selain itu, Tingkat Gemar Membaca (TGM) nasional juga mengalami peningkatan signifikan, dari 66,7 pada tahun 2023 menjadi 72,44 pada tahun 2024, yang dikategorikan sebagai tingkat sedang [1].

Penggunaan media pembelajaran seperti *flashcard* telah terbukti efektif dalam meningkatkan hasil belajar dan motivasi siswa di Indonesia. Penelitian oleh Hapsari dan Kristin (2024) menunjukkan bahwa penerapan media *flashcard* dalam pembelajaran Matematika di kelas II SD Negeri Tingkir Lor 02 mampu meningkatkan rata-rata hasil belajar siswa dari 66,15 pada siklus pertama menjadi 68,83 pada siklus kedua, dengan persentase ketuntasan meningkat dari 56% menjadi 89%. Motivasi belajar siswa juga mengalami peningkatan, dari 69% (kategori cukup) menjadi 77% (kategori baik).

Selain itu, penelitian oleh Gultom dan Mudiono (2024) menemukan bahwa penggunaan *flashcard* secara konsisten meningkatkan kemampuan membaca, menulis, dan pemahaman teks pada siswa kelas I di SDN 3 Kotalama. Pendekatan ini juga mendorong peningkatan motivasi belajar serta keterlibatan aktif siswa dalam proses pembelajaran. Di sisi lain, teknologi *Retrieval-Augmented Generation* (RAG) mulai diterapkan dalam berbagai bidang di Indonesia. Samudra et al. (2024) mengimplementasikan RAG dalam perancangan *chatbot* kesehatan pencernaan guna meningkatkan akurasi dan relevansi informasi yang disampaikan. Metode RAG mengintegrasikan model generatif dengan sistem pencarian berbasis dokumen untuk memberikan jawaban yang lebih andal dan berbasis bukti (*evidence-based*).

Penerapan RAG juga telah diuji dalam konteks layanan publik. Pujiono et al. (2024) mengembangkan *chatbot* dengan menggunakan RAG dan basis data vektor untuk menjawab pertanyaan terkait regulasi layanan publik. Model GPT-4 yang digunakan menunjukkan kinerja terbaik dengan nilai *cosine similarity* rata-rata sebesar 0,404, yang mengindikasikan potensi besar dalam meningkatkan responsivitas *chatbot* di sektor pelayanan publik.

Integrasi teknologi RAG dalam pembuatan *flashcard* dari dokumen PDF berpotensi besar untuk meningkatkan efisiensi dan efektivitas pembelajaran di Indonesia. Dengan memanfaatkan RAG, informasi relevan dari dokumen dapat diekstraksi dan disajikan dalam bentuk *flashcard* secara otomatis, sehingga mempermudah proses belajar mengajar serta berkontribusi terhadap peningkatan kualitas pendidikan.

### 1.2 Rumusan Masalah

Berikut adalah rumusan masalah yang dapat dikemukakan:

1. Bagaimana merancang dan mengembangkan sistem otomatis yang dapat mengekstraksi informasi penting dari dokumen PDF dan mengubahnya menjadi *flashcard* menggunakan pendekatan *Retrieval-Augmented Generation* (RAG)?
2. Apa saja tahapan optimal dalam proses ekstraksi, *retrieval*, dan *generation* yang dapat menghasilkan *flashcard* dengan informasi yang ringkas, relevan, dan terstruktur?
3. Bagaimana cara mengintegrasikan *model retrieval* dan *generative model* dalam arsitektur RAG agar dapat bekerja secara efisien dalam konteks dokumen berbahasa Indonesia?

### 1.3 Batasan Masalah

Berikut adalah batasan masalah pada penelitian ini:

1. Penelitian ini hanya mencakup proses konversi dokumen PDF menjadi *flashcard* secara otomatis menggunakan teknologi *Retrieval-Augmented Generation* (RAG), tanpa melibatkan proses evaluasi langsung oleh pengguna akhir
2. Jenis dokumen yang digunakan terbatas pada dokumen PDF berbahasa Inggris yang bersifat informatif atau edukatif, seperti artikel ilmiah populer, ringkasan materi pelajaran, atau modul pembelajaran.
3. Model *retriever* dan *generator* yang digunakan akan dibatasi pada model-model sumber terbuka (*open-source*) atau yang dapat dijalankan secara lokal, seperti FAISS untuk vektor dan LLM lokal
4. Output sistem dibatasi dalam bentuk teks *flashcard* sederhana (pertanyaan dan jawaban atau poin-poin utama), tanpa pengolahan visual tambahan seperti gambar atau audio.
5. Evaluasi sistem dilakukan berdasarkan kualitas hasil output (ketepatan, relevansi, dan keterbacaan) melalui penilaian manual, bukan berdasarkan efektivitas dalam meningkatkan hasil belajar pengguna.

### 1.4 Tujuan

Berikut adalah tujuan penelitian ini:

1. Mengembangkan sistem otomatis yang dapat mengubah dokumen *PDF* berbahasa Inggris menjadi kumpulan *flashcard* berbentuk teks menggunakan pendekatan *Retrieval-Augmented Generation* (RAG).
2. Menerapkan dan mengintegrasikan komponen utama RAG, yaitu *retriever* dan *generator*, untuk mengekstraksi informasi relevan dari dokumen dan menghasilkan *output flashcard* yang ringkas, informatif, dan terstruktur.
3. Mengevaluasi kinerja sistem RAG dalam menghasilkan *flashcard* berdasarkan aspek relevansi, keterbacaan, dan kesesuaian konten terhadap dokumen sumber.
4. Membangun *pipeline* pemrosesan dokumen PDF, mulai dari ekstraksi teks, *preprocessing*, *retrieval*, hingga *generation*, dengan fokus pada efisiensi dan kemudahan integrasi dalam konteks pembelajaran digital.

5. Menjelajahi potensi pemanfaatan teknologi RAG dalam bidang pendidikan, khususnya untuk mendukung literasi dan efektivitas pembelajaran melalui media digital seperti flashcard.

### 1.5 Manfaat

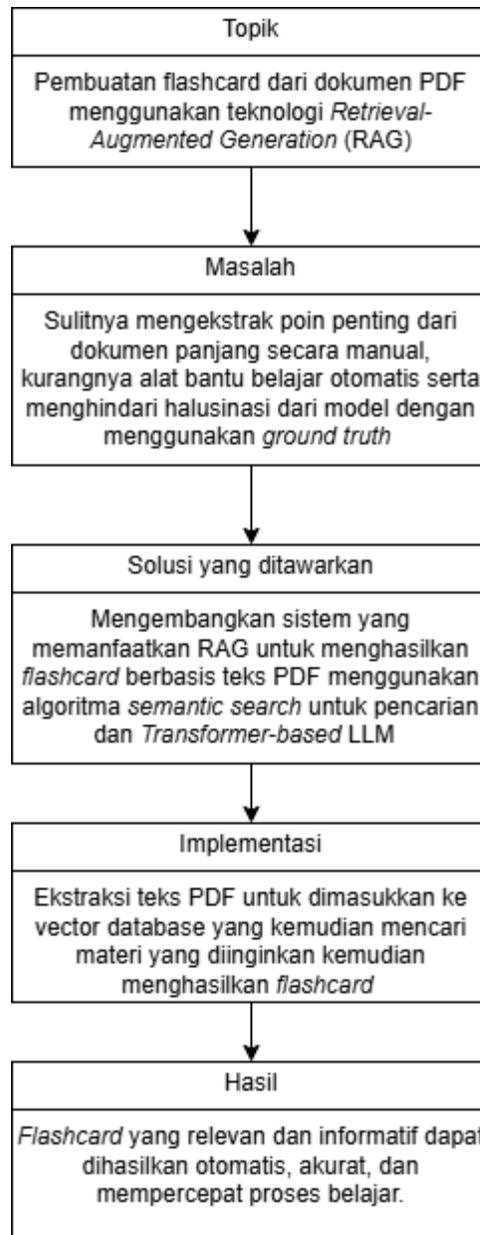
Berikut adalah manfaat dari penelitian ini:

1. Menjadi dasar pengembangan sistem otomatis berbasis RAG untuk mengekstrak informasi dari dokumen PDF berbahasa Inggris dan menyajikannya dalam format *flashcard* yang mudah dipahami.
2. Memberikan kontribusi ilmiah dalam pengembangan teknologi pembelajaran berbasis AI yang dapat menangani dan merangkum dokumen panjang berbahasa Inggris.
3. Meningkatkan pemahaman tentang integrasi antara *document retrieval* dan *generative language models* dalam konteks pendidikan dan literasi.
4. Memudahkan pendidik dan pelajar dalam membuat ringkasan materi pembelajaran dari sumber berbahasa Inggris secara otomatis dan efisien dalam bentuk *flashcard*.
5. Menjadi solusi bagi institusi pendidikan untuk memanfaatkan teknologi berbasis AI dalam menunjang proses pembelajaran, terutama dalam konteks globalisasi dan kebutuhan akan literatur internasional.

### 1.6 Kerangka Pemikiran

Dalam era digital saat ini, akses terhadap informasi pembelajaran semakin melimpah. Namun, volume informasi yang besar dalam dokumen seperti artikel, modul, dan jurnal ilmiah seringkali membuat pembelajar mengalami kesulitan dalam memahami dan mengingat inti materi secara efisien.

Salah satu metode pembelajaran aktif yang terbukti efektif dalam meningkatkan pemahaman dan daya ingat adalah penggunaan *flashcard*, yaitu kartu berisi pertanyaan dan jawaban ringkas yang dirancang untuk melatih daya ingat secara terstruktur dan berulang. Namun, pembuatan *flashcard* secara manual membutuhkan waktu dan usaha yang tidak sedikit. Oleh karena itu, dibutuhkan pendekatan otomatis berbasis AI untuk menghasilkan *flashcard* secara cepat dan relevan, terutama dari dokumen pembelajaran dalam format PDF. Teknologi *Retrieval-Augmented Generation* (RAG) hadir sebagai solusi inovatif untuk mengatasi tantangan tersebut.



*Gambar 1.1. Kerangka Pemikiran*