

**Pengembangan *Legislative Activity Tracker* DPR RI  
menggunakan *Information Extraction* (NER dan  
*Entity Linking*) Multi-Sumber**

**Proposal Tugas Akhir**

Oleh

**Athhar Mahendra Umar  
18222080**



**PROGRAM STUDI SISTEM DAN TEKNOLOGI INFORMASI  
SEKOLAH TEKNIK ELEKTRO DAN INFORMATIKA  
INSTITUT TEKNOLOGI BANDUNG  
Desember 2025**

# LEMBAR PENGESAHAN

**Pengembangan *Legislative Activity Tracker* DPR RI  
menggunakan *Information Extraction (NER dan Entity Linking)*  
Multi-Sumber**

## **Proposal Tugas Akhir**

Oleh

**Athhar Mahendra Umar  
18222080**

Program Studi Sistem dan Teknologi Informasi  
Sekolah Teknik Elektro dan Informatika  
Institut Teknologi Bandung

Proposal Tugas Akhir ini telah disetujui dan disahkan  
di Bandung, pada tanggal 4 Desember 2025

Pembimbing

Ir. Windy Gambetta, MBA  
NIP. 196404301989031005

## DAFTAR ISI

<b>DAFTAR GAMBAR . . . . .</b>	<b>v</b>
<b>DAFTAR TABEL . . . . .</b>	<b>vi</b>
<b>I PENDAHULUAN . . . . .</b>	<b>1</b>
I.1 Latar Belakang . . . . .	1
I.2 Rumusan Masalah . . . . .	3
I.3 Tujuan . . . . .	3
I.4 Batasan Masalah . . . . .	3
I.5 Metodologi . . . . .	4
<b>II STUDI LITERATUR . . . . .</b>	<b>6</b>
II.1 Transparansi Parlemen . . . . .	6
II.1.1 Transparansi Prosedur . . . . .	6
II.1.2 Transparansi Kinerja . . . . .	7
II.1.3 Transparansi Perilaku . . . . .	7
II.2 Transparansi Parlemen Digital . . . . .	8
II.2.1 Open Government Data (OGD) dalam Konteks Legislatif . . . . .	8
II.2.2 Teknologi Sipil <i>Civic Tech</i> dan Pemantauan Parlemen . . . . .	9
II.2.3 Partisipasi Digital Publik dalam Pengawasan Legislatif . . . . .	9
II.3 Urgensi Transparansi Parlemen Digital . . . . .	9
II.3.1 Hubungan Transparansi Digital dengan Kepercayaan Publik dan Legitimasi . . . . .	10
II.3.2 Kebutuhan Data Terstruktur untuk Pengawasan Berkelanjutan . . . . .	11
II.4 Pendekatan Teknis Global dalam Pemantauan Kinerja Legislatif . . . . .	12
II.4.1 Studi dan Praktik di Eropa . . . . .	12
II.4.2 Studi dan Praktik di Amerika Utara . . . . .	13
II.5 Studi dan Praktik Transparansi Parlemen di Indonesia . . . . .	14
II.5.1 Transparansi Kelembagaan DPR RI . . . . .	14
II.5.2 Inisiatif Civic Tech: Studi Kasus <i>JariUngu</i> . . . . .	15
II.6 Tantangan Teknis dalam Pemrosesan Data Legislatif Indonesia . . . . .	16
II.7 Pendekatan Teknis Sistem <i>Legislative Activity Tracker</i> . . . . .	17
II.7.1 Pemrosesan Bahasa Alami <i>Natural Language Processing (NLP)</i> . . . . .	18
II.7.1.1 Arsitektur Model Neural untuk Pemrosesan Teks: <i>BERT</i> dan <i>ColBERT</i> . . . . .	19

II.7.1.2	Model Bahasa dan Sumber Daya Linguistik Indonesia . . . . .	20
II.7.1.3	Metodologi Pengenalan Entitas Bernama <i>Named Entity Recognition</i> . . . . .	21
II.7.1.4	<i>Information Extraction</i> . . . . .	22
II.7.1.5	<i>Information Retrieval</i> . . . . .	23
II.7.2	Akuisisi Data Web dan <i>Crawling</i> Terfokus . . . . .	24
II.7.2.1	<i>Focused Crawling</i> . . . . .	25
II.7.3	Pra pemrosesan Dokumen dan Analisis Tata Letak ( <i>Layout Analysis</i> ) . . . . .	26
II.7.3.1	<i>Optical Character Recognition (OCR)</i> . . . . .	27
II.7.4	<i>Event Log</i> . . . . .	28
II.7.5	Penautan Semantik ( <i>Semantic Linking</i> ) . . . . .	29
II.8	Analytic Hierarchy Process (AHP) . . . . .	30
II.8.1	Skala Saaty dalam Perbandingan Berpasangan . . . . .	31
<b>III</b>	<b>ANALISIS MASALAH . . . . .</b>	<b>32</b>
III.1	Analisis Kondisi Saat Ini . . . . .	32
III.1.1	Analisis Permasalahan . . . . .	32
III.1.2	Analisis Peluang . . . . .	34
III.2	Analisis Kebutuhan . . . . .	35
III.2.1	Identifikasi Masalah Pengguna . . . . .	35
III.2.2	Kebutuhan Fungsional . . . . .	37
III.2.3	Kebutuhan Nonfungsional . . . . .	38
III.3	Analisis Pemilihan Solusi . . . . .	39
III.3.1	Alternatif Solusi . . . . .	39
III.3.2	Analisis Penentuan Solusi . . . . .	40
<b>IV</b>	<b>DESAIN KONSEP SOLUSI . . . . .</b>	<b>42</b>
IV.1	Diagram Konseptual . . . . .	42
IV.1.1	Kondisi Sistem ( <i>As-Is</i> ) . . . . .	42
IV.1.2	Kondisi Sistem <i>Legislative Activity Tracker</i> . . . . .	43
IV.1.3	Perbandingan Sistem As-Is dan Sistem <i>Legislative Activity Tracker</i> . . . . .	45
IV.1.4	Diagram Arsitektur Tingkat Tinggi ( <i>High-Level Architecture</i> ) . . . . .	46
<b>V</b>	<b>RENCANA SELANJUTNYA . . . . .</b>	<b>50</b>
V.1	Rencana Implementasi . . . . .	50
V.1.1	Langkah-Langkah Implementasi . . . . .	50
V.1.2	Kebutuhan Alat dan Lingkungan Pengembangan . . . . .	51
V.2	Rencana Evaluasi . . . . .	52
V.2.1	Metode Pengujian . . . . .	52

V.2.2	Indikator Keberhasilan . . . . .	53
V.3	Analisis Risiko . . . . .	54
V.3.1	Identifikasi Risiko . . . . .	54
V.3.2	Strategi Mitigasi Risiko . . . . .	55
<b>LAMPIRAN 1: Detail Kebutuhan Fungsional . . . . .</b>		<b>63</b>
<b>LAMPIRAN 2: Penjelasan Detail Alternatif Solusi . . . . .</b>		<b>66</b>
<b>LAMPIRAN 3: Rincian Analisis Penentuan Solusi . . . . .</b>		<b>68</b>

## DAFTAR GAMBAR

III.1	Kondisi Transparansi Parlemen ( <i>as-is</i> ) . . . . .	33
III.2	Struktur ( <i>Analytical Hierarchy Process</i> ) . . . . .	41
IV.1	Kondisi Transparansi Parlemen ( <i>as-is</i> ) . . . . .	42
IV.2	Kondisi Transparansi Parlemen <i>Legislative Activity Tracker</i> . . . .	44
IV.3	Diagram Konseptual ( <i>Legislative Activity Tracker</i> ) . . . . .	47

## DAFTAR TABEL

II.1	Skala Saaty untuk Perbandingan Berpasangan . . . . .	31
III.1	Daftar Identifikasi Masalah (M) . . . . .	34
III.2	Daftar Identifikasi Peluang (P) . . . . .	35
III.3	Analisis Kebutuhan Pengguna . . . . .	37
III.4	Analisis Kebutuhan Non-Fungsional Sistem . . . . .	38
III.5	Perhitungan Bobot Akhir Alternatif Solusi . . . . .	41
IV.1	Perbandingan Sistem As-Is dan Sistem LAT Usulan . . . . .	46
V.1	Daftar Kebutuhan Alat dan Teknologi Pengembangan . . . . .	52
V.2	Kriteria dan Indikator Keberhasilan Sistem . . . . .	53
V.3	Identifikasi Risiko Pengembangan Sistem . . . . .	54
V.4	Strategi Mitigasi Risiko . . . . .	55
5	Analisis Kebutuhan Fungsional Sistem . . . . .	63
6	Alternatif Solusi dan Masalah yang Ingin Diselesaikan . . . . .	66
7	Kriteria Matriks Perbandingan Berpasangan (Skala Saaty) . . . . .	68
8	Matriks Perbandingan Berpasangan Antar Kriteria . . . . .	68
9	Normalisasi Nilai AHP Kriteria . . . . .	68
10	Rata-Rata Nilai Kriteria (Bobot Prioritas) . . . . .	69
11	Perbandingan Berpasangan Alternatif Terhadap Kriteria Efektivitas (K1) . . . . .	69
12	Normalisasi Matriks Alternatif Solusi (Kriteria Efektivitas) . . . . .	69
13	Rata-rata Nilai Alternatif Solusi (Kriteria Efektivitas) . . . . .	69

# BAB I

## PENDAHULUAN

### I.1 Latar Belakang

Transparansi parlemen merupakan domain tata kelola publik yang mensyaratkan keterbukaan dalam proses pengambilan keputusan legislatif, tidak sekadar melalui publikasi dokumen statis tetapi melalui penyediaan ekosistem informasi yang tersandarisasi. Secara konseptual, transparansi ini menuntut data kinerja, mulai dari pola pemungutan suara (*roll-call votes*) hingga notulensi rapat, untuk disajikan dalam format yang dapat diproses kembali (*reusable*) guna menjamin akuntabilitas fungsional (Begany 2024). Masalah mendasar dalam domain ini adalah kesenjangan (*gap*) antara ketersediaan informasi nominal dengan kegunaan praktisnya (*usability*).

Sering kali, data legislatif terperangkap dalam format tidak terstruktur tanpa interoperabilitas semantik, yang menghambat pelacakan hubungan kausal antara mandat politik dan tindakan legislatif nyata (F. Wang dkk. 2023). Ketidadaan mekanisme penelusuran (*audit trails*) yang valid ini menciptakan ambiguitas yang menjadikan transparansi hanya sebatas jargon administratif tanpa dampak pengawasan signifikan (Sheppard 2023). Inti permasalahan modern bukan lagi pada ketersediaan data, melainkan pada struktur dan kualitasnya untuk memungkinkan audit kinerja independen (Liang 2025).

Penyelesaian masalah transparansi ini krusial karena peran fundamentalnya dalam menjaga kepercayaan publik (*political trust*) sebagai mata uang utama demokrasi. Di era digital, warga menuntut bukti empiris yang dapat diverifikasi, bukan sekadar janji politik. Penelitian menunjukkan ketersediaan data transparan berkorelasi positif dengan legitimasi negara, sedangkan kekaburan informasi (*opacity*) diasosiasikan dengan persepsi korupsi yang tinggi (Chen dkk. 2023).

Tanpa transparansi fungsional, terjadi asimetri informasi yang melemahkan kontrak



sosial antara pemilih dan wakilnya (Ferry dkk. 2024). Lebih lanjut, transparansi berfungsi sebagai mekanisme pengendali internal (*internal control mechanism*) yang mendisiplinkan perilaku aktor politik melalui efek pengawasan (*surveillance effect*) (Bordignon dkk. 2020). Kegagalan menangani defisit ini bukan hanya masalah teknis, melainkan ancaman bagi keberlanjutan institusi demokrasi itu sendiri (Rienks 2023).

Dalam lanskap penelitian global, diskursus mengenai informatika parlemen (*parliamentary informatics*) dan transparansi digital telah bergeser dari sekadar digitalisasi dokumen menuju penerapan teknologi analitik canggih untuk memantau kinerja sektor publik. Sejumlah studi terbaru di Eropa dan Amerika Utara telah berhasil mendemonstrasikan penggunaan teknik *Process Mining* dan *Artificial Intelligence (AI)* untuk mengaudit prosedur administratif dan mendeteksi anomali dalam log aktivitas pemerintahan (Nai dkk. 2025; Himler dkk. 2024). Fokus utama dari penelitian-penelitian ini adalah membangun kerangka kerja teknis yang mampu mengekstraksi wawasan dari data yang sangat besar (*big data*), seperti memprediksi kepatuhan prosedur atau mengotomatisasi analisis teks hukum untuk meningkatkan responsivitas legislatif terhadap konstituen (Kreps dkk. 2023). Kesamaan mendasar antara penelitian-penelitian global ini dengan proyek yang diusulkan terletak pada penggunaan data terstruktur (*structured logs*) sebagai bahan baku utama untuk menciptakan akuntabilitas yang terukur dan berbasis bukti (*data-driven accountability*).

Dalam konteks Indonesia, literatur tata kelola data pemerintah telah berkembang namun masih terkonsentrasi pada *e-government* umum atau fase elektoral. Tantangan utama yang diidentifikasi adalah fragmentasi kelembagaan dan rendahnya kualitas infrastruktur data yang menghambat pemanfaatan optimal. Meski inisiatif teknologi sipil (*civic tech*) seperti JariUngu telah berhasil meningkatkan partisipasi politik, fokusnya masih terbatas pada visualisasi profil dan data pemilu, bukan pemantauan kinerja pasca-pemilu yang mendalam (Halimatusa'diyah dan Jannah 2025). *Research gap* yang nyata dalam literatur Indonesia adalah ketiadaan studi yang mengembangkan arsitektur *event log* untuk memantau aktivitas harian legislator DPR RI yang terhubung secara eksplisit dengan janji kampanye. Penelitian ini hadir untuk mengisi kekosongan tersebut dengan menawarkan solusi teknis interoperabilitas data yang belum dieksplorasi sebelumnya dalam konteks parlemen nasional.

Mengingat kompleksitas tantangan transparansi dan kekosongan instrumen pengawasan yang ada, terdapat kebutuhan mendesak untuk merumuskan kerangka teknis yang dapat memecahkan masalah fragmentasi data dan diskoneksi semantik dalam

ekosistem informasi parlemen saat ini. Diperlukan sebuah pendekatan baru yang menjawab pertanyaan fundamental mengenai konsistensi kinerja wakil rakyat

## **I.2 Rumusan Masalah**

Berdasarkan latar belakang dan *gap* yang ada, pertanyaan penelitian ini adalah:

1. Apa struktur data yang diperlukan untuk membangun dashboard yang mengintegrasikan aktivitas faktual legislator (apa yang mereka lakukan) dengan janji kampanye mereka (apa yang mereka katakan)?
2. Apa metrik dan indikator kinerja yang tepat untuk mengukur konsistensi antara komitmen kampanye dan aktivitas legislatif legislator yang ditampilkan dalam dashboard?
3. Apa mekanisme visualisasi informasi yang efektif untuk merepresentasikan perbandingan (komparasi) antara janji kampanye dan realisasi kinerja legislator pada antarmuka dashboard?

## **I.3 Tujuan**

Berdasarkan rumusan masalah yang telah dijelaskan, tujuan dari tugas akhir ini adalah sebagai berikut:

1. Merancang sistem yang mampu mengekstraksi dan mentransformasi dokumen aktivitas legislatif DPR RI yang tidak terstruktur menjadi event log yang terstandarisasi?
2. Merumuskan metrik dan logika komparasi untuk mengukur tingkat konsistensi antara aktivitas faktual harian legislator dengan janji kampanye mereka dalam sebuah mekanisme penelusuran (*audit trail*) yang terintegrasi.
3. Membangun antarmuka dashboard interaktif yang memvisualisasikan perbandingan antara janji kampanye dan realisasi kinerja legislator secara intuitif guna memudahkan publik melakukan pengawasan berbasis data.

## **I.4 Batasan Masalah**

Dalam pelaksanaan tugas akhir ini, batasan yang digunakan adalah:

1. Cakupan subjek adalah anggota DPR RI periode 2024-2029, dengan sumber data aktivitas dari dokumen publik resmi dan media kredibel, serta janji kampanye dari dokumen visi-misi terverifikasi.
2. Pengolahan data dibatasi pada dokumen dan berita berbahasa Indonesia menggunakan metode *Natural Language Processing* (NLP) untuk menghasilkan

struktur *event log* secara otomatis.

3. Sumber data utama yang digunakan untuk analisis aktivitas legislatif dibatasi pada dokumen Risalah Rapat Paripurna yang tersedia secara publik di situs resmi DPR RI, mengingat kelengkapan dan validitas formal dokumen tersebut sebagai representasi sikap resmi anggota dewan.
4. Penautan antara aktivitas legislatif dan janji kampanye didasarkan pada analisis kesamaan semantik teks, tanpa melakukan verifikasi kebenaran hukum atau penilaian etik terhadap konten.
5. Sistem yang dibangun merupakan prototipe akademik berupa *dashboard* visualisasi, tidak mencakup aspek skalabilitas infrastruktur produksi atau keamanan siber tingkat lanjut.

## **I.5 Metodologi**

Pelaksanaan tugas akhir ini dilakukan melalui serangkaian tahapan sistematis dan terstruktur untuk memastikan bahwa perumusan masalah didasarkan pada fakta empiris dan solusi yang ditawarkan didukung oleh landasan teori yang kuat serta metode pengambilan keputusan yang objektif. Tahapan penelitian, khususnya dalam penyusunan proposal ini, meliputi investigasi pengumpulan fakta untuk merumuskan masalah, studi literatur, serta penentuan solusi terbaik menggunakan metode kuantitatif.

Tahapan investigasi dan pengumpulan fakta di latar belakang dilakukan untuk merumuskan masalah secara akurat melalui langkah-langkah sebagai berikut:

1. Melakukan observasi langsung terhadap ekosistem informasi parlemen saat ini dengan mengakses situs resmi DPR RI, portal JDIH, dan sumber data pemilu (KPU) untuk memahami struktur dan ketersediaan data publik.
2. Melakukan simulasi pencarian manual untuk mengevaluasi pengalaman pengguna dalam menelusuri rekam jejak legislator, mulai dari mencari janji kampanye hingga membandingkannya dengan aktivitas di risalah rapat.
3. Mengidentifikasi hambatan utama (*pain points*) yang dihadapi publik, seperti fragmentasi sumber data, format dokumen yang tidak terstruktur (PDF), dan ketiadaan mekanisme penautan informasi secara otomatis.
4. Menganalisis temuan observasi untuk merumuskan masalah penelitian yang spesifik, yaitu kesenjangan antara ketersediaan data publik dengan kemudahan akses dan pemanfaatannya untuk pengawasan kinerja legislatif.

Selanjutnya, untuk mencari solusi terhadap masalah yang telah dirumuskan, dilakukan langkah-langkah pencarian, pengelompokan, dan penapisan literatur serta pe-

milihan solusi sebagai berikut:

1. **Pencarian Literatur:** Melakukan penelusuran referensi ilmiah melalui basis data akademik seperti Google Scholar dan ResearchGate. Kata kunci yang digunakan meliputi: "*legislative transparency system*", "*parliamentary monitoring NLP*", "*named entity recognition political text*", "*Analytic Hierarchy Process (AHP)*", dan "*web scraping government data*".
2. **Pengelompokan Topik:** Mengklasifikasikan literatur yang ditemukan ke dalam domain pengetahuan utama:
  - (a) Konsep transparansi parlemen dan akuntabilitas politik.
  - (b) Teknologi *Information Retrieval* (IR) dan metode *focused crawling*.
  - (c) Metode *Natural Language Processing* (NLP) untuk ekstraksi informasi.
  - (d) Metode pengambilan keputusan multikriteria (MCDM), khususnya *Analytic Hierarchy Process* (AHP).
3. **Penapisan Literatur:** Menyeleksi sumber informasi berdasarkan relevansi topik, kebaruan (publikasi 5–10 tahun terakhir), dan kredibilitas penerbit. Literatur mengenai implementasi AHP dalam pemilihan teknologi informasi menjadi prioritas untuk mendukung bab analisis masalah.
4. **Penentuan Solusi dengan AHP:** Setelah literatur dan kandidat solusi terkumpul, dilakukan proses pemilihan solusi terbaik menggunakan metode *Analytic Hierarchy Process* (AHP). Langkah ini mencakup penentuan kriteria evaluasi (seperti efisiensi, skalabilitas, biaya), pembobotan kriteria melalui perbandingan berpasangan, dan penilaian alternatif solusi untuk mendapatkan rekomendasi arsitektur sistem yang paling optimal.
5. **Dokumentasi dan Sintesis:** Mencatat temuan penting dari literatur dan hasil perhitungan AHP. Seluruh proses investigasi dan analisis ini didokumentasikan untuk menjamin validitas metodologis penelitian.

Hasil penggalian informasi dan sintesis teori dari langkah-langkah metodologi di atas selanjutnya akan diuraikan secara komprehensif pada bab berikutnya. Bab II Studi Literatur akan memaparkan landasan teoretis dan tinjauan pustaka yang menjadi fondasi akademik bagi pengembangan sistem *Legislative Activity Tracker* (LAT), termasuk pembahasan mendalam mengenai teknologi NLP, IR, dan kerangka kerja AHP yang digunakan.

## **BAB II**

### **STUDI LITERATUR**

#### **II.1 Transparansi Parlemen**

Transparansi parlemen didefinisikan sebagai prinsip tata kelola yang mewajibkan lembaga legislatif untuk membuka proses pengambilan keputusan, prosedur internal, dan hasil kinerja mereka kepada publik agar dapat dipantau dan dievaluasi secara objektif. Secara filosofis, transparansi ini berakar pada konsep kontrak sosial, di mana warga negara memberikan mandat kekuasaan kepada wakil rakyat dengan syarat adanya akuntabilitas atas penggunaan kekuasaan tersebut.

Berdasarkan analisis yang dilakukan Ferry dkk. (2024) mengenai akuntabilitas demokrasi, transparansi parlemen bukan sekadar tujuan akhir, melainkan prasyarat mutlak *sine qua non* bagi terciptanya kepercayaan publik *political trust* (Ferry dkk. 2024). Tanpa keterbukaan yang memadai, hubungan antara konstituen dan wakilnya akan mengalami asimetri informasi yang mengarah pada delegitimasi institusi perwakilan. Dalam praktiknya, transparansi ini mencakup tiga dimensi utama, yaitu transparansi prosedur (bagaimana hukum dibuat), transparansi kinerja (apa yang dihasilkan), dan transparansi perilaku (bagaimana integritas aktor politiknya), yang semuanya harus dapat diakses oleh publik untuk menjamin fungsi pengawasan eksternal berjalan efektif.

##### **II.1.1 Transparansi Prosedur**

Transparansi prosedur berkaitan dengan keterbukaan mekanisme dan aturan yang mengatur bagaimana keputusan legislatif diambil di dalam parlemen. Dimensi ini menekankan pentingnya akses publik terhadap informasi mengenai tahapan pembentukan undang undang, jadwal persidangan, agenda rapat komisi, serta tata tertib persidangan.

Berdasarkan studi Bordignon dkk. (2020) mengenai aturan dan akuntabilitas, transparansi prosedur berfungsi sebagai mekanisme pencegah manipulasi agenda *agenda setting manipulation*, di mana kejelasan aturan memastikan bahwa tidak ada keputusan strategis yang diambil secara tersembunyi atau melalui prosedur yang menyimpang (Bordignon dkk. 2020). Dalam konteks ini, publik harus dapat memverifikasi apakah wakil rakyat telah mematuhi tahapan legislasi yang baku, mulai dari pengusulan naskah akademik, pembahasan tingkat pertama, hingga pengambilan keputusan di rapat paripurna, sehingga legitimasi produk hukum yang dihasilkan dapat dipertanggungjawabkan secara formal.

### **II.1.2 Transparansi Kinerja**

Transparansi kinerja berfokus pada keterbukaan hasil nyata *output* dan dampak *outcome* dari aktivitas legislatif yang dilakukan oleh anggota parlemen, baik secara individu maupun kelembagaan. Dimensi ini mencakup penyediaan data mengenai produk legislasi yang disahkan, partisipasi dalam pemungutan suara *voting records*, kehadiran dalam rapat-rapat yang penting, serta kualitas intervensi substantif dalam perdebatan kebijakan.

Mengacu pada penelitian Chen dkk. (2023), transparansi kinerja merupakan indikator utama yang digunakan publik untuk menilai efektivitas wakil rakyat; ketersediaan data kinerja yang terukur memungkinkan konstituen untuk membandingkan antara janji politik yang disampaikan saat kampanye dengan realisasi tindakan di parlemen (Chen dkk. 2023). Tanpa transparansi pada dimensi ini, evaluasi publik hanya akan didasarkan pada persepsi atau popularitas semata, bukan pada bukti produktivitas kerja yang objektif, sehingga melemahkan mekanisme penghargaan dan sanksi *reward and punishment mechanism* dalam sistem demokrasi.

### **II.1.3 Transparansi Perilaku**

Transparansi perilaku menyoroti aspek integritas etis dan kepatuhan moral anggota parlemen dalam menjalankan mandat publiknya. Dimensi ini mensyaratkan keterbukaan informasi terkait profil keuangan, deklarasi konflik kepentingan, sumber pendanaan kampanye, serta catatan pelanggaran kode etik yang mungkin dilakukan oleh legislator.

Berdasarkan analisis Rienks (2023) mengenai skandal dan kompetensi politik, transparansi perilaku berfungsi sebagai sistem peringatan dini *early warning system* untuk mendeteksi potensi penyalahgunaan kekuasaan *abuse of power* atau korupsi po-

litik (Rienks 2023). Dengan membuka data perilaku ini, publik dapat memantau apakah keputusan yang diambil oleh seorang legislator didasari oleh kepentingan konstituen atau dipengaruhi oleh kepentingan pribadi dan kelompok tertentu, sehingga menjaga martabat institusi perwakilan dari degradasi moral dan krisis kepercayaan.

## **II.2 Transparansi Parlemen Digital**

Transparansi parlemen digital merupakan evolusi dari prinsip keterbukaan tradisional yang didorong oleh adopsi teknologi informasi dan komunikasi dalam lingkungan legislatif. Konsep ini tidak hanya mengubah media penyampaian informasi dari analog ke digital, tetapi juga mengubah paradigma interaksi antara parlemen dan publik menjadi lebih partisipatif dan berbasis data *data driven*.

Menurut Begany (2024), transformasi digital di sektor publik menuntut pergeseran fokus dari sekadar digitalisasi dokumen, yaitu memindahkan kertas ke berkas *PDF*, menjadi digitalisasi data, yaitu membuat informasi terstruktur dan dapat diolah mesin. Dalam ekosistem parlemen digital, transparansi diukur dari seberapa mudah data legislatif seperti *event log* rapat, teks rancangan undang-undang, dan rekam jejak *voting* dapat diakses, ditelusuri, dan diaudit kembali oleh algoritma komputer maupun masyarakat luas tanpa hambatan teknis, sehingga memungkinkan pengawasan yang bersifat *real time* dan berskala besar.

### **II.2.1 Open Government Data (OGD) dalam Konteks Legislatif**

*Open Government Data OGD* didefinisikan sebagai data pemerintah yang dapat diakses secara bebas, digunakan ulang, dan didistribusikan kembali oleh siapa saja tanpa batasan hak cipta atau kontrol paten. Dalam konteks legislatif, penerapan prinsip *OGD* bertujuan untuk memecahkan masalah asimetri informasi dengan menyediakan data mentah *raw data* yang berkualitas tinggi kepada publik.

Berdasarkan tinjauan literatur sistematis yang dilakukan Kempeneer, Pirannejad, dan Wolswinkel (2023), data parlemen yang memenuhi standar *OGD* harus memiliki karakteristik lengkap, primer atau bersumber langsung dari pencatat, tepat waktu, dapat diakses, dapat diproses mesin *machine processable*, non diskriminatif, non proprietary, dan bebas lisensi (Kempeneer, Pirannejad, dan Wolswinkel 2023). Penerapan standar ini krusial untuk memungkinkan pengembangan aplikasi pihak ketiga *civic tech* yang dapat mengolah data parlemen yang kompleks menjadi informasi yang mudah dipahami, sekaligus memfasilitasi audit otomatis terhadap kinerja lembaga perwakilan.

### **II.2.2 Teknologi Sipil *Civic Tech* dan Pemantauan Parlemen**

Teknologi sipil *Civic Tech* merujuk pada penggunaan teknologi digital untuk meningkatkan partisipasi warga, memperkuat infrastruktur demokrasi, dan mendorong efektivitas pemerintahan. Dalam domain legislatif, *Civic Tech* terwujud melalui platform organisasi pemantau parlemen *Parliamentary Monitoring Organizations (PMO)* yang berfungsi sebagai perantara informasi antara parlemen dan warga. Platform ini mengagregasi data parlemen yang kompleks dan menyajikannya dalam bentuk visualisasi yang lebih mudah dipahami publik.

Penelitian Halimatusa'diyah dan Jannah (2025) menyoroti bahwa *Civic Tech* berperan penting dalam mengisi kekosongan saluran komunikasi politik dan mengubah partisipasi pasif menjadi aktivisme digital yang terinformasi (Halimatusa'diyah dan Jannah 2025). Berbeda dengan portal resmi pemerintah yang sering bersifat kaku, inisiatif *Civic Tech* berfokus pada pengalaman pengguna *user experience* untuk memudahkan warga melakukan audit sosial terhadap kinerja wakil rakyat mereka.

### **II.2.3 Partisipasi Digital Publik dalam Pengawasan Legislatif**

Partisipasi publik digital *e participation* didefinisikan sebagai keterlibatan warga dalam proses pemerintahan dan pengambilan kebijakan melalui sarana teknologi informasi dan komunikasi. Dalam konteks pengawasan legislatif, *e participation* melampaui sekadar akses informasi menuju interaksi dua arah, di mana warga dapat memberikan umpan balik, kritik, dan penilaian terhadap kinerja wakil rakyat.

Studi yang dilakukan Kreps dkk. (2023) menunjukkan bahwa alat komunikasi digital yang dirancang dengan baik dapat meningkatkan responsivitas legislatif secara signifikan karena legislator merasa diawasi secara *real time* oleh konstituennya (Kreps dkk. 2023). Namun, efektivitas partisipasi ini sangat bergantung pada kualitas informasi yang disajikan oleh platform *Civic Tech*. Informasi yang dangkal hanya menghasilkan partisipasi simbolis, sedangkan informasi yang terstruktur dan berbasis bukti, misalnya perbandingan antara janji dan realisasi, mampu mendorong partisipasi substantif yang memperkuat akuntabilitas vertikal.

## **II.3 Urgensi Transparansi Parlemen Digital**

Ketersediaan data yang terbuka dan terstandarisasi telah menjadi variabel penting dalam membangun ekosistem demokrasi yang responsif. Penelitian oleh Chen dkk. (2023) menunjukkan bahwa tingkat kepercayaan masyarakat terhadap lembaga po-



litik dapat meningkat hingga 28 persen di lingkungan yang menerapkan *open government data* secara konsisten. Di sisi lain, laporan empiris Begany (2024) menemukan bahwa adopsi praktik digitalisasi transparansi, melalui portal data terbuka parlemen, berkorelasi positif dengan peningkatan tingkat partisipasi publik dalam pengawasan legislatif, dengan jumlah kontribusi umpan balik warga melonjak dua kali lipat selama dua tahun pertama inisiatif tersebut dijalankan.

Dampak sistemik dari kurangnya transparansi digital dapat diukur melalui persepsi korupsi dan pola partisipasi politik yang stagnan. Ferry dkk. (2024) melaporkan bahwa institusi yang gagal menyediakan akses akuntabilitas digital mengalami peningkatan 19 persen dalam indeks persepsi korupsi menurut survei publik nasional. Sementara itu, Bordignon dkk. (2020) menggarisbawahi bahwa adanya mekanisme pengawasan digital mendorong kedisiplinan perilaku anggota parlemen, terlihat dari penurunan kasus pelanggaran prosedural sebesar 14 persen setelah diterapkannya sistem pemantauan terkomputerisasi. Tidak hanya berdampak bagi masyarakat umum, transparansi digital juga memberikan landasan bagi organisasi masyarakat sipil dan media untuk memverifikasi rekam jejak legislator secara independen, sehingga meningkatkan efektivitas kontrak sosial antara wakil rakyat dengan para pemilihnya.

Jika institusi politik gagal mengadopsi standar keterbukaan digital ini, risiko jangka panjangnya adalah menurunnya legitimasi dan fragmentasi kepercayaan publik, yang dalam jangka waktu tertentu berpotensi melemahkan stabilitas demokrasi Rienks (2023). Oleh sebab itu, urgensi transparansi digital tidak hanya berakar pada norma etik dan hukum, tetapi juga pada hasil nyata berupa perbaikan kualitas hubungan antara warga dan parlemen yang terukur secara empiris.

### **II.3.1 Hubungan Transparansi Digital dengan Kepercayaan Publik dan Legitimasi**

Transparansi digital parlemen berkaitan erat dengan pembentukan dan pemeliharaan kepercayaan publik terhadap lembaga perwakilan. Seiring meningkatnya literasi digital warga, ekspektasi terhadap keterbukaan informasi juga bergeser dari sekadar akses dokumen menuju ketersediaan data yang dapat ditelusuri dan diverifikasi secara mandiri. Studi kuantitatif menunjukkan bahwa keberadaan kanal transparansi digital yang aktif dapat meningkatkan persepsi kepercayaan publik terhadap lembaga politik secara signifikan dibandingkan lembaga yang hanya mengandalkan mekanisme pelaporan konvensional (Chen dkk. 2023). Dalam konteks ini, transparansi tidak hanya berfungsi sebagai kewajiban administratif, tetapi juga sebagai mekanis-

me komunikasi simbolik yang menandakan kesediaan lembaga untuk diawasi.

Di sisi lain, legitimasi parlemen sebagai institusi demokratis juga sangat dipengaruhi oleh sejauh mana proses dan hasil kerjanya dapat diakses secara terbuka melalui platform digital. Ferry dkk. (2024) menunjukkan bahwa warga cenderung menganggap lembaga politik lebih sah, atau *legitimate*, ketika mereka memiliki kesempatan yang jelas untuk memantau dan mengevaluasi kinerja wakil rakyat secara berkelanjutan. Ketika informasi mengenai agenda sidang, hasil pemungutan suara, serta partisipasi anggota legislatif tersedia secara transparan dan mudah diakses, warga merasakan adanya saluran akuntabilitas yang nyata sehingga memperkuat keyakinan bahwa sistem demokrasi bekerja sesuai prinsip representasi. Sebaliknya, ketiadaan transparansi digital yang memadai membuka ruang bagi tumbuhnya kecurigaan, spekulasi negatif, dan erosi kepercayaan yang pada akhirnya dapat melemahkan legitimasi lembaga perwakilan itu sendiri (Rienks 2023).

### **II.3.2 Kebutuhan Data Terstruktur untuk Pengawasan Berkelanjutan**

Transparansi parlemen digital hanya dapat berfungsi secara nyata jika informasi yang disajikan tidak berhenti pada tingkat dokumen, tetapi diolah menjadi data yang terstruktur, konsisten, dan mudah dikaitkan lintas waktu. Studi literatur mengenai desain transparansi digital menunjukkan bahwa tanpa struktur data yang jelas, upaya membuka informasi justru berisiko menciptakan apa yang disebut sebagai “transparansi semu”, yaitu kondisi ketika volume informasi yang tersedia sangat besar tetapi sulit dipakai ulang untuk tujuan pengawasan (Wirtz, Weyerer, dan Rosch 2022). Data yang terstruktur memungkinkan aktivitas legislatif direkam sebagai jejak yang berkelanjutan sehingga pola perilaku, perubahan sikap, dan konsistensi posisi terhadap suatu isu dapat diamati secara longitudinal, bukan hanya sebagai potret sesaat (Kempeneer, Pirannejad, dan Wolswinkel 2023).

Sebaliknya, ketiadaan struktur data yang memadai membuat pengawasan publik bergantung pada pembacaan manual yang terputus putus dan sangat rentan bias. Begany (2024) menunjukkan bahwa inisiatif data terbuka yang tidak menyediakan format terstruktur dan dapat diproses mesin cenderung kurang dimanfaatkan oleh pengguna lanjutan seperti peneliti dan organisasi masyarakat sipil, sehingga potensi akuntabilitasnya tidak tercapai secara optimal. Ketika data aktivitas dan keputusan legislator disusun secara sistematis dan dapat dipantau dari waktu ke waktu, mekanisme akuntabilitas menjadi lebih kuat karena setiap perubahan posisi, ketidakhadiran, atau penyimpangan dari pola sebelumnya dapat segera terdeteksi dan dipertanyakan (Hong 2024).

Dengan demikian, kebutuhan akan data yang terstruktur dan dapat diawasi secara berkelanjutan bukan hanya persoalan teknis, tetapi merupakan prasyarat agar prinsip transparansi digital benar benar dapat diterjemahkan menjadi pengawasan yang efektif dan berkesinambungan (Wirtz, Weyerer, dan Rosch 2022).

## **II.4 Pendekatan Teknis Global dalam Pemantauan Kinerja Legislatif**

Dalam lanskap penelitian global, pendekatan pemantauan kinerja legislatif mulai bergeser dari pembacaan dokumen secara manual menuju penggunaan metode komputasional yang memanfaatkan data dalam skala besar. Sejalan dengan berkembangnya korpus digital dan inisiatif *open data*, berbagai studi menggabungkan teknik pemrosesan bahasa alami, analisis peristiwa berbasis teks, dan kerangka kerja data terbuka untuk mengekstraksi pola perilaku legislatif yang sebelumnya sulit diobservasi. Liu, Mitra, dan North (2023) menunjukkan bahwa paradigma ekstraksi peristiwa berbasis teks memungkinkan rangkaian kejadian politik direkonstruksi sebagai narasi terstruktur yang siap dianalisis secara kuantitatif. Pendekatan ini memberikan landasan bagi penelitian lanjutan di Eropa dan Amerika Utara yang memanfaatkan data parlemen terstruktur sebagai bahan baku untuk membangun akuntabilitas yang lebih terukur.

### **II.4.1 Studi dan Praktik di Eropa**

Di kawasan Eropa, pengembangan korpus sidang parlemen yang terstandarisasi menjadi salah satu fondasi utama bagi informatika parlemen modern. Erjavec dkk. (2023) mengembangkan *ParlaMint*, yaitu kumpulan besar transkrip sidang parlemen multibahasa yang dianotasi secara linguistik dan disusun dalam format terstruktur. Korpus ini memungkinkan peneliti melakukan analisis yang konsisten lintas negara, misalnya memetakan pola intervensi anggota parlemen, intensitas pembahasan isu tertentu, dan perubahan retorika politik dari waktu ke waktu. Dengan adanya struktur data yang seragam, proses ekstraksi entitas, relasi, dan peristiwa dari teks sidang dapat dilakukan secara lebih sistematis dan dapat direplikasi.

Selain korpus, kerangka kerja analisis data parlemen juga mulai berkembang. Berkovitz dkk. (2023) mengusulkan sebuah kerangka terbuka untuk menganalisis data parlemen publik yang menggabungkan prinsip *big data*, visualisasi, dan keterbukaan data. Kerangka ini dirancang untuk memudahkan integrasi berbagai sumber data legislatif seperti metadata rancangan undang undang, rekam jejak *voting*, dan daftar keanggotaan komisi ke dalam satu sistem yang dapat diolah dan divisualisasikan.

Dari sudut pandang penelitian ini, praktik di Eropa memberikan contoh konkret tentang bagaimana data sidang yang semula berupa teks naratif dapat diubah menjadi korpus terstruktur yang mendukung analisis perilaku legislatif. Namun, fokus utamanya masih pada pemetaan aktivitas dan pola wacana, belum secara spesifik menghubungkan aktivitas tersebut dengan janji kampanye individual sebagaimana yang diupayakan dalam proyek ini.

#### II.4.2 Studi dan Praktik di Amerika Utara

Di Amerika Utara, ekosistem pemantauan legislatif berkembang di atas kombinasi antara portal resmi pemerintah dan inisiatif teknologi sipil. Di tingkat federal Amerika Serikat, Library of Congress (no date) menyediakan *Congress.gov* sebagai portal resmi legislatif yang menyajikan data rancangan undang-undang, riwayat proses legislasi, serta informasi anggota Kongres dalam format yang dapat diakses publik. Panduan yang disusun oleh Library of Congress (no date) menjelaskan bahwa data *Congress.gov* dapat diunduh dan digunakan secara *offsite*, termasuk dalam bentuk *bulk data*, sehingga memungkinkan pengembang dan peneliti membangun aplikasi pemantauan berbasis data legislatif yang lebih canggih.

Di atas infrastruktur data resmi tersebut, platform independen seperti *GovTrack.us* memanfaatkan data *Congress.gov* untuk menyajikan rekam jejak anggota Kongres dalam bentuk yang lebih mudah dipahami publik. *GovTrack.us* (no date) menjelaskan bahwa *GovTrack* mengolah data *voting*, keanggotaan komite, dan riwayat legislasi untuk menghasilkan metrik seperti produktivitas, kedekatan ideologis, dan pola *co sponsorship*. Pendekatan ini menunjukkan bagaimana data legislatif yang disediakan pemerintah dapat ditransformasikan menjadi *dashboard* pemantauan yang lebih bersahabat bagi warga.

Di sisi lain, survei yang dilakukan oleh Liu, Mitra, dan North (2023) tentang ekstraksi narasi berbasis peristiwa memperlihatkan bagaimana teknik *event based analysis* digunakan untuk menyusun kembali rangkaian peristiwa politik dari teks berita, yang secara konseptual sejalan dengan upaya membangun log aktivitas legislatif dari dokumen sidang.

Dari perspektif penelitian ini, praktik di Amerika Utara memberikan inspirasi kuat mengenai pentingnya portal data resmi yang menyediakan data terstruktur dan dapat diakses, serta peran kunci platform perantara seperti *GovTrack* dalam menerjemahkan data tersebut menjadi informasi yang bermakna bagi publik. Namun, seperti halnya di Eropa, fokus utama ekosistem ini masih berpusat pada pemantauan aktivitas

dan *voting*, sementara aspek penautan eksplisit antara janji kampanye individu dengan aktivitas legislatif harian belum menjadi perhatian utama. Ruang kosong inilah yang kemudian menjadi titik masuk bagi kontribusi penelitian ini.

## **II.5 Studi dan Praktik Transparansi Parlemen di Indonesia**

Dalam konteks Indonesia, upaya mewujudkan transparansi parlemen berjalan melalui dua jalur utama, yaitu mekanisme resmi kelembagaan dan inisiatif teknologi sipil *civic tech*. Literatur mengenai tata kelola data pemerintah menunjukkan bahwa kebijakan dan infrastruktur digital di Indonesia masih didominasi pendekatan *e government* umum dan layanan administrasi, sementara pemanfaatan data untuk pemantauan kinerja legislatif pasca pemilu relatif tertinggal (Breuer dkk. 2023).

Di satu sisi, DPR RI dan lembaga terkait telah mengembangkan berbagai sistem informasi untuk menyajikan dokumen legislasi dan proses pembahasan. Di sisi lain, platform independen seperti *JariUngu* berusaha mengisi celah keterbukaan dengan menyajikan profil dan informasi representasi politik kepada publik (JariUngu, no date). Kedua jalur ini penting untuk dipahami guna melihat sejauh mana ekosistem yang ada saat ini sudah mendukung akuntabilitas kinerja anggota dewan.

### **II.5.1 Transparansi Kelembagaan DPR RI**

Secara kelembagaan, DPR RI telah mengembangkan berbagai instrumen formal untuk mendukung keterbukaan proses legislasi. Salah satu kerangka penting adalah Sistem Informasi Legislasi *SILEG* yang diatur melalui pedoman Sekretaris Jenderal DPR RI dan dirancang untuk mendokumentasikan tahapan proses pembentukan undang-undang, mulai dari perencanaan, pembahasan, hingga pengesahan (Sekretaris Jenderal DPR RI 2020).

Melalui sistem ini, publik dapat mengakses dokumen seperti naskah rancangan undang-undang, jadwal rapat, dan ringkasan proses legislasi. Selain itu, Jaringan Dokumentasi dan Informasi Hukum *JDIH* dan situs resmi DPR menyediakan produk hukum final serta informasi umum mengenai alat kelengkapan dewan dan keanggotaan.

Namun, temuan Breuer dkk. (2023) menyoroti bahwa pada tataran praktik, tata kelola data di Indonesia masih menghadapi fragmentasi kelembagaan dan keterbatasan kualitas infrastruktur. Kondisi ini tercermin dalam ekosistem informasi DPR, di mana data legislatif tersebar di berbagai portal dan sering kali tersaji dalam format

dokumen statis yang tidak terstruktur. Ketiadaan skema data terpadu membuat publik atau peneliti harus melakukan penelusuran manual jika ingin melacak aktivitas seorang anggota dewan lintas rapat, komisi, dan periode waktu.

Belum terdapat satu struktur data yang secara eksplisit merepresentasikan siapa melakukan apa, kapan, dan dalam konteks apa, sehingga sulit membangun *event log* yang komprehensif dari sumber resmi yang ada. Lebih jauh, tidak ada mekanisme yang menghubungkan secara langsung antara data aktivitas legislatif ini dengan dokumen janji kampanye, sehingga akuntabilitas terhadap mandat elektoral belum terwujud dalam bentuk data yang dapat dianalisis secara sistematis.

### II.5.2 Inisiatif Civic Tech: Studi Kasus *JariUngu*

Di luar jalur resmi kelembagaan, Indonesia juga menyaksikan berkembangnya inisiatif teknologi sipil yang berupaya menjembatani kesenjangan informasi antara warga dan wakil rakyat. Salah satu contoh menonjol adalah *JariUngu*, sebuah platform *web* yang berfungsi sebagai direktori informasi politik dan legislasi. Berdasarkan penjelasan pada laman *Tentang Kami*, *JariUngu* menyajikan berbagai fitur seperti profil anggota DPR dan DPRD, data calon legislatif, ringkasan regulasi, serta fasilitas pencarian berdasarkan nama atau daerah pemilihan (JariUngu, no date).

Dengan cara ini, *JariUngu* membantu pemilih untuk mengenali siapa wakil mereka, apa latar belakangnya, dan regulasi apa saja yang terkait. Studi mengenai aktivisme digital di Indonesia menunjukkan bahwa platform seperti ini berkontribusi dalam menurunkan hambatan informasi dan mendorong partisipasi politik yang lebih terinformasi (Halimatusa'diyah dan Jannah 2025).

Meskipun demikian, dari perspektif pemantauan kinerja legislatif pasca pemilu, pendekatan yang diambil *JariUngu* masih memiliki keterbatasan struktural. Fokus utama platform ini adalah pada penyediaan informasi profil dan data elektoral, sedangkan pencatatan aktivitas harian legislator seperti intervensi dalam rapat, pola kehadiran, atau posisi terhadap isu tertentu belum dimodelkan sebagai rangkaian peristiwa terstruktur. Data yang tersedia lebih mendekati katalog informasi statis daripada *event log* yang memungkinkan analisis jejak aktivitas secara kronologis. Selain itu, belum terdapat mekanisme teknis yang secara eksplisit menautkan janji kampanye individual dengan rekam tindakan legislatif yang terdokumentasi, sehingga publik masih harus menyimpulkan hubungan tersebut secara manual.

Dengan demikian, meskipun *JariUngu* menunjukkan bahwa terdapat kebutuhan dan

manfaat nyata dari platform pemantauan legislatif, masih terdapat ruang signifikan untuk pengembangan arsitektur data yang secara khusus dirancang untuk menghubungkan janji politik dan kinerja aktual anggota DPR RI, yang menjadi fokus utama penelitian ini.

## II.6 Tantangan Teknis dalam Pemrosesan Data Legislatif Indonesia

Meskipun terdapat perkembangan signifikan dalam informatika parlemen di tingkat global, penerapan teknologi serupa di Indonesia menghadapi sejumlah tantangan teknis spesifik yang membedakan konteks DPR RI dari praktik terbaik internasional.

Tantangan utama terletak pada fragmentasi dan heterogenitas format data. Berbeda dengan inisiatif *ParlaMint* di Eropa yang menyediakan korpus parlemen multibahasa dalam format *XML TEI* yang terstandarisasi (Erjavec dkk. 2023), atau *Congress.gov* di Amerika Serikat yang menyediakan akses data massal *bulk data* siap pakai, data legislatif Indonesia tersebar di berbagai subdomain terpisah seperti *berkas.dpr.go.id*, *dpr.go.id*, dan *jdih.dpr.go.id* tanpa skema data terpadu. Dokumen dokumen ini tersaji dalam format campuran yang tidak konsisten, mulai dari *PDF* hasil pemindaian *scanned PDF* untuk risalah lama, *PDF* berbasis teks untuk laporan singkat, hingga tabel *HTML* untuk jadwal kegiatan. Kondisi ini menjadikan tahap penemuan dokumen *Information Retrieval* sebagai langkah kritis yang harus dilakukan sebelum ekstraksi informasi dapat dijalankan, suatu kompleksitas yang jarang ditemui dalam penelitian yang menggunakan korpus siap pakai. Fragmentasi kelembagaan dalam pengelolaan data ini menghambat pemanfaatan data untuk analisis komputasional yang presisi.

Selain masalah format, tantangan metodologis terbesar dalam pemantauan kinerja DPR RI adalah ketiadaan catatan pemungutan suara individu *individual roll call votes*. Berbeda dengan sistem di Kongres Amerika Serikat di mana posisi setiap anggota tercatat secara eksplisit dalam setiap *voting*, mayoritas pengambilan keputusan di DPR RI dilakukan berdasarkan musyawarah mufakat atau diwakili oleh pandangan fraksi. Catatan resmi sering kali hanya menyatakan posisi agregat, misalnya tujuh fraksi menyetujui dan dua fraksi menolak, tanpa merinci siapa saja anggota yang hadir dan mendukung keputusan tersebut secara individu.

Keterbatasan data dasar ini menuntut pengembangan model inferensi sikap politik *stance inference* yang tidak bergantung pada *ground truth* langsung, melainkan diturunkan dari kombinasi keanggotaan fraksi, kehadiran dalam rapat, dan posisi resmi fraksi terkait. Pendekatan inferensial ini merupakan adaptasi metodologis yang

diperlukan untuk mengatasi kekosongan data *voting* individual yang juga tidak tersedia di platform pemantau eksternal seperti *JariUngu*.

Terakhir, terdapat kesenjangan sumber daya linguistik untuk domain legislatif bahasa Indonesia. Inisiatif seperti *ParlaMint* telah menyediakan korpus yang dianotasi secara linguistik untuk mengenali entitas bernama *Named Entity Recognition (NER)*, struktur kalimat, dan metadata pembicara secara otomatis. Di Indonesia, meskipun telah tersedia model bahasa umum seperti *IndoBERT* (Koto dkk. 2020) atau *benchmark IndoNLU* (Wilie dkk. 2020), belum ada korpus risalah DPR yang secara spesifik dianotasi untuk mengenali entitas legislatif seperti nama anggota, nomor rancangan undang undang, nama alat kelengkapan dewan, atau referensi pasal hukum yang kompleks.

Model *Natural Language Processing (NLP)* umum sering kali gagal menangkap struktur wacana dan terminologi khas parlemen tanpa penyesuaian khusus *domain adaptation*. Oleh karena itu, penelitian ini tidak dapat sekadar menerapkan perangkat lunak *NLP* yang sudah ada, tetapi perlu membangun komponen ekstraksi kustom *custom parsers* dan pola pencocokan *pattern matching* yang mampu menangani kompleksitas linguistik dalam dokumen DPR RI. Dengan mempertimbangkan tantangan tantangan teknis ini, yaitu fragmentasi data, inferensi sikap politik, dan ketiadaan korpus anotasi legislatif, penelitian ini mengusulkan pendekatan arsitektur teknis yang disesuaikan dengan realitas data DPR RI, yang akan dijabarkan pada bagian berikutnya.

## **II.7 Pendekatan Teknis Sistem *Legislative Activity Tracker***

Pengembangan sistem pemantauan legislatif *Legislative Activity Tracker* menuntut pendekatan teknis yang mampu menjembatani kesenjangan antara data tidak terstruktur yang tersedia di domain publik dengan kebutuhan akan informasi yang terstruktur dan dapat diaudit. Mengingat karakteristik data DPR RI yang heterogen, tersebar, dan didominasi oleh format naratif, solusi konvensional berbasis basis data relasional semata tidak lagi memadai. Oleh karena itu, penelitian ini mengadopsi pendekatan komputasional berbasis kecerdasan buatan, khususnya dalam ranah pemrosesan teks dan pencarian informasi, untuk mentransformasi tumpukan dokumen mentah menjadi pengetahuan yang bermakna.

Pendekatan teknis yang diusulkan dirancang secara berlapis *layered architecture*, dimulai dari akuisisi dan pemahaman konten dokumen menggunakan teknik pemrosesan bahasa, diikuti dengan ekstraksi struktur peristiwa yang merepresentasikan ak-



tivitas legislatif, hingga mekanisme penautan logis untuk menghubungkan aktivitas tersebut dengan konteks mandat politiknya. Strategi ini dipilih untuk memastikan bahwa sistem tidak hanya berfungsi sebagai arsip digital pasif, melainkan sebagai instrumen analitis aktif yang mampu memodelkan realitas institusional DPR RI, di mana dinamika fraksi dan partisipasi individu saling terkait, ke dalam format data yang siap diolah oleh mesin dan diakses oleh publik.

### II.7.1 Pemrosesan Bahasa Alami *Natural Language Processing (NLP)*

Pemrosesan Bahasa Alami *Natural Language Processing (NLP)* merupakan domain komputasional yang memanfaatkan teknik pembelajaran mendalam *deep learning* untuk memodelkan struktur linguistik kompleks ke dalam representasi vektor matematis. Perkembangan mutakhir dalam *NLP* didorong oleh adopsi arsitektur *Transformer* yang diperkenalkan oleh Vaswani dan kolega, yang menggantikan model sekuensial tradisional seperti *Recurrent Neural Networks (RNN)* dengan mekanisme atensi diri *self attention mechanism*. Secara matematis, mekanisme ini menghitung bobot hubungan antar token dalam sebuah sekuens input  $X$  melalui transformasi linear menjadi tiga matriks, yaitu *Query (Q)*, *Key (K)*, dan *Value (V)*. Skor atensi dihitung menggunakan fungsi

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right) V, \quad (II.1)$$

di mana  $d_k$  adalah dimensi dari vektor *key*. Mekanisme ini memungkinkan model untuk memproses seluruh sekuens secara paralel dan menangkap ketergantungan jarak jauh *long range dependencies* dengan efisiensi komputasi yang jauh lebih tinggi dibandingkan *RNN* (Islam dkk. 2024).

Dalam implementasi praktis, model berbasis *Transformer* seperti *BERT Bidirectional Encoder Representations from Transformers* telah menjadi standar de facto untuk berbagai tugas hilir *NLP*. T. Lin dkk. (2022) menjelaskan bahwa arsitektur *encoder decoder* pada *Transformer* memungkinkan model untuk mempelajari representasi kontekstual dua arah *bidirectional*, yang secara signifikan meningkatkan performa dalam tugas pemahaman bahasa.

Studi kuantitatif oleh Lu dkk. (2025) menunjukkan bahwa penggunaan model pra latih seperti *SciBERT* yang dikombinasikan dengan lapisan *BiLSTM Bidirectional Long Short Term Memory* mampu mencapai skor F1 sebesar 86.27 persen dalam tugas pengenalan entitas pada teks ilmiah, mengungguli model dasar *BERT* standar dengan skor F1 sebesar 84.91 persen dan *BERT+CRF* dengan skor F1 sebesar

83.15 persen. Peningkatan performa ini diatribusikan pada kemampuan model untuk menghasilkan *word embeddings* yang dinamis, di mana setiap kata direpresentasikan sebagai vektor padat *dense vector* dalam ruang multidimensi  $V$ , sehingga kata-kata dengan kedekatan semantik memiliki jarak *Euclidean* atau *cosine similarity* yang minimal (Rani dkk. 2021).

Efektivitas model *NLP* modern juga sangat bergantung pada teknik pembelajaran transfer *transfer learning*, di mana model yang telah dilatih pada korpus berukuran besar *pre training* diadaptasi untuk tugas spesifik melalui proses *fine tuning*. Alanazi dkk. (2022) melaporkan bahwa teknik augmentasi data dan *fine tuning* pada model pembelajaran mendalam dapat meningkatkan skor F1 antara satu persen sampai delapan persen dalam tugas klasifikasi teks medis dibandingkan model tanpa augmentasi.

Namun, kompleksitas komputasi model *Transformer* yang berskala kuadratik terhadap panjang sekuens input  $O(n^2)$  tetap menjadi tantangan utama, terutama untuk pemrosesan dokumen panjang. Oleh karena itu, varian efisien seperti *Sparse Transformers* atau teknik kompresi model seperti *pruning* dan *quantization* terus dikembangkan untuk menyeimbangkan antara akurasi prediktif dan efisiensi memori (Islam dkk. 2024).

#### **II.7.1.1 Arsitektur Model Neural untuk Pemrosesan Teks: *BERT* dan *ColBERT***

Dalam lanskap pemrosesan bahasa alami modern, model berbasis *Transformer* telah menjadi standar de facto untuk menangkap semantik teks yang kompleks. Arsitektur yang paling berpengaruh adalah *BERT Bidirectional Encoder Representations from Transformers*, yang diperkenalkan sebagai model yang memanfaatkan *multi layer bidirectional Transformer encoders*. Tidak seperti model sekuensial tradisional yang memproses teks secara searah, *BERT* dilatih menggunakan objektif *Masked Language Modeling (MLM)* yang memungkinkan model untuk mempelajari representasi kontekstual mendalam dari kedua arah, yaitu dari kiri ke kanan dan dari kanan ke kiri, secara simultan.

Studi oleh T. Lin dkk. (2022) menjelaskan bahwa arsitektur ini memungkinkan *BERT* untuk menghasilkan *embeddings* yang sensitif terhadap konteks, di mana representasi vektor dari sebuah kata berubah tergantung pada kata-kata di sekitarnya. Varian standar *BERT Base* memiliki dua belas lapisan *encoder*, dimensi tersembunyi *hidden size* sebesar tujuh ratus enam puluh delapan, dan total seratus sepuluh

juta parameter, yang secara empiris terbukti mencapai performa *state of the art* dalam berbagai tugas pemahaman bahasa (T. Lin dkk. 2022). Namun, keterbatasan utama *BERT* terletak pada biaya komputasi yang tinggi untuk pencarian informasi *information retrieval* skala besar, karena mekanisme *full self attention* memiliki kompleksitas kuadratik terhadap panjang input.

Untuk mengatasi inefisiensi tersebut dalam konteks pencarian informasi, arsitektur *ColBERT Contextualized Late Interaction over BERT* dikembangkan sebagai solusi hibrida yang mempertahankan kemampuan semantik *BERT* namun dengan efisiensi pencarian yang jauh lebih tinggi. X. Wang dkk. (2022) menjelaskan bahwa *ColBERT* memperkenalkan paradigma *late interaction*, di mana *query* dan dokumen dikodekan secara terpisah menjadi representasi vektor padat *dense vectors* menggunakan *BERT*, dan interaksi relevansi dihitung hanya pada tahap akhir menggunakan operasi *MaxSim*, yaitu maksimum kesamaan antar vektor token.

Secara kuantitatif, pendekatan ini memungkinkan *ColBERT* untuk melakukan pencarian dokumen dengan latensi yang rendah dalam skala milidetik sambil mempertahankan akurasi yang kompetitif dengan model interaksi penuh yang jauh lebih lambat. Evaluasi eksperimental menunjukkan bahwa pendekatan *dense retrieval* seperti *ColBERT* mampu meningkatkan efektivitas pencarian *passage retrieval* hingga seratus tiga puluh enam persen dibandingkan metode *sparse retrieval* tradisional dalam skenario tertentu (Y. Li dkk. 2023). Keunggulan ini menjadikan *ColBERT* pilihan arsitektural yang ideal untuk sistem yang membutuhkan keseimbangan antara pemahaman semantik mendalam dan kecepatan eksekusi pada korpus dokumen yang besar.

#### II.7.1.2 Model Bahasa dan Sumber Daya Linguistik Indonesia

Penerapan teknik *NLP* pada bahasa dengan sumber daya terbatas *low resource languages* seperti bahasa Indonesia menghadapi tantangan unik dibandingkan dengan bahasa Inggris yang memiliki ekosistem data melimpah. Untuk mengatasi hal ini, pengembangan sumber daya linguistik standar menjadi krusial.

Salah satu inisiatif fundamental adalah *IndoLEM Indonesian Language Evaluation Montage*, sebuah kerangka kerja *benchmark* komprehensif yang dirancang untuk mengevaluasi performa model bahasa Indonesia pada berbagai tugas *NLP*, termasuk morfologi, sintaksis, dan semantik. Nabiilah dkk. (2023) menjelaskan bahwa ketersediaan *benchmark* terstandarisasi seperti *IndoLEM* memungkinkan peneliti untuk mengukur efektivitas model secara objektif dan membandingkan arsitektur

yang berbeda pada dataset yang seragam, yang mencakup tugas tugas seperti penandaan kelas kata *POS tagging* dan pengenalan entitas bernama *Named Entity Recognition (NER)*.

Sejalan dengan kebutuhan tersebut, model bahasa pra latih *pre trained language models* khusus bahasa Indonesia telah dikembangkan untuk menangkap karakteristik linguistik lokal. *IndoBERT*, sebuah varian *BERT* yang dilatih pada korpus teks bahasa Indonesia berskala besar, termasuk *Wikipedia* Indonesia, artikel berita, dan *web corpus*, telah menjadi standar de facto untuk tugas tugas pemahaman bahasa Indonesia.

Studi oleh Ekakristi dkk. (2025) menunjukkan bahwa *IndoBERT* secara konsisten mengungguli model multibahasa seperti *mBERT* dalam tugas tugas yang memerlukan pemahaman nuansa lokal. Secara kuantitatif, *IndoBERT* mencapai skor rata rata F1 sebesar delapan puluh delapan koma empat puluh tiga persen untuk tugas klasifikasi dan delapan puluh satu koma sembilan puluh dua persen untuk tugas *sequence labeling* seperti *NER* pada dataset *IndoNLU*, menunjukkan superioritasnya dibandingkan model yang tidak dikhususkan untuk bahasa Indonesia.

Lebih lanjut, penelitian Ibrohim dkk. (2023) menyoroti bahwa efektivitas model seperti *IndoBERT* dapat ditingkatkan lebih lanjut melalui teknik *intermediate task transfer learning*, di mana model dilatih pada tugas perantara yang relevan sebelum diterapkan pada tugas akhir, yang terbukti meningkatkan robustitas model terhadap variasi data yang tidak terduga.

### II.7.1.3 Metodologi Pengenalan Entitas Bernama *Named Entity Recognition*

Pengenalan Entitas Bernama *Named Entity Recognition (NER)* secara teknis diformulasikan sebagai masalah pelabelan sekuens *sequence labeling problem*, di mana tujuannya adalah memetakan sekuens token input  $X = (x_1, x_2, \dots, x_n)$  menjadi sekuens label  $Y = (y_1, y_2, \dots, y_n)$  yang merepresentasikan kategori entitas. Metodologi konvensional untuk tugas ini sering kali menggunakan pendekatan *Conditional Random Fields (CRF)* untuk memodelkan ketergantungan antar label yang berdekatan.

Namun, perkembangan terkini telah bergeser ke arah arsitektur pembelajaran mendalam hibrida seperti *Bi LSTM CRF*, yang menggabungkan kemampuan *Bidirectional Long Short Term Memory (Bi LSTM)* untuk menangkap fitur kontekstual jarak jauh dengan lapisan *CRF* di bagian akhir untuk mengoptimalkan prediksi label secara

ra global (Goyal dkk. 2024). Arsitektur ini secara konsisten menunjukkan performa unggul karena kemampuannya memodelkan dependensi transisi label, misalnya label I ORG seharusnya mengikuti B ORG, secara eksplisit.

Dalam era *Large Language Models*, pendekatan berbasis *Transformer* telah merevolusi *NER* dengan menggantikan lapisan *Bi LSTM* dengan *encoder* berbasis atensi diri *self attention*. Jehangir dkk. (2023) menjelaskan bahwa model seperti *BERT NER* memanfaatkan representasi kontekstual dinamis untuk mengatasi ambiguitas makna kata, yang merupakan kelemahan utama dari pendekatan berbasis *embedding* statis seperti *Word2Vec*.

Meskipun demikian, tantangan signifikan tetap ada dalam menangani entitas bersarang *nested entities*, di mana satu entitas terkandung di dalam entitas lain, misalnya frasa Universitas Indonesia yang merupakan organisasi, namun kata Indonesia di dalamnya juga merupakan nama lokasi. Metode pelabelan sekuens standar seperti skema BIO sering gagal menangani struktur ini.

Untuk mengatasi masalah tersebut, Ma dkk. (2023) mengusulkan mekanisme *multi level topic aware* yang memungkinkan model untuk mengenali entitas pada berbagai tingkat granularitas secara simultan, serta pendekatan berbasis *span based classification* yang memperlakukan setiap rentang teks sebagai kandidat entitas independen, bukan sekadar memberikan label pada setiap token. Selain itu, integrasi logika pengetahuan domain *knowledge aware deep logic learning* telah terbukti meningkatkan interpretabilitas model *NER*, memungkinkan sistem untuk tidak hanya memprediksi entitas tetapi juga menjelaskan keputusan berdasarkan aturan logika yang dipelajari (J. Lin dkk. 2025).

#### **II.7.1.4 Information Extraction**

Ekstraksi Informasi *Information Extraction (IE)* adalah proses otomatisasi untuk mentransformasi data teks tidak terstruktur menjadi format data terstruktur, seperti basis data relasional atau graf pengetahuan. Secara teknis, *IE* terdiri dari dua sub-tugas utama yang saling berkaitan, yaitu Ekstraksi Relasi *Relation Extraction (RE)* dan Ekstraksi Peristiwa *Event Extraction (EE)*. Ekstraksi Relasi bertujuan untuk mengidentifikasi hubungan semantik biner atau n-ari antara entitas yang telah dikenali, misalnya hubungan *subject verb object*. Metode terkini dalam *RE* memanfaatkan struktur sintaksis kalimat, seperti pohon dependensi *dependency trees*, untuk mengidentifikasi frasa kunci yang menjadi penghubung antar entitas.

Liu dkk. (2025) mengusulkan penggunaan algoritma deteksi komunitas pada pohon dependensi yang dikombinasikan dengan *Graph Convolutional Networks (GCN)*, yang terbukti meningkatkan akurasi ekstraksi relasi dengan memodelkan konektivitas topologis antar kata secara lebih efektif dibandingkan model berbasis sekuensial murni.

Di sisi lain, Ekstraksi Peristiwa *Event Extraction* merupakan tugas yang lebih kompleks karena melibatkan identifikasi pemicu peristiwa *event trigger*, tipe peristiwa, serta argumen pendukungnya, seperti aktor, waktu, dan lokasi, yang sering kali tersebar di seluruh dokumen *document level extraction*. Pendekatan tradisional yang memproses kalimat per kalimat sering kali gagal menangkap konteks peristiwa yang lintas kalimat.

Untuk mengatasi hal ini, Xu dkk. (2025) mengembangkan metode *Learning the Argument of An Entity with Event Prompts (LAAP)*, yang menggunakan teknik *prompting* dengan *placeholders* untuk memandu model bahasa besar *Large Language Models (LLMs)* dalam mengenali tipe peristiwa dan argumennya secara simultan di tingkat dokumen. Selain itu, pergeseran paradigma menuju pembelajaran gabungan *joint learning* semakin dominan, di mana tugas pengenalan entitas dan ekstraksi relasi dilakukan secara bersamaan dalam satu kerangka kerja terpadu *end to end* untuk menghindari propagasi kesalahan *error propagation* dari metode pipa bertahap *pipeline methods*.

#### **II.7.1.5 Information Retrieval**

Temu Kembali Informasi *Information Retrieval (IR)* adalah domain penelitian yang berfokus pada pengembangan algoritma untuk menemukan dokumen relevan dari koleksi data berskala besar berdasarkan kueri pengguna. Secara historis, model *IR* didominasi oleh pendekatan *sparse retrieval* berbasis leksikal seperti *BM25*, yang menghitung skor relevansi berdasarkan frekuensi kata kunci yang tumpang tindih antara kueri dan dokumen.

Namun, keterbatasan metode ini dalam menangkap makna semantik, yang sering disebut sebagai masalah *vocabulary mismatch*, telah mendorong transisi menuju pendekatan *dense retrieval* berbasis model *neural*. Guo dkk. (2020) dalam tinjauan komprehensifnya menjelaskan bahwa model perankingan *neural neural ranking models* memanfaatkan jaringan saraf tiruan untuk mempelajari representasi vektor dari kueri dan dokumen, memungkinkan sistem untuk mencocokkan relevansi berdasarkan kedekatan semantik laten, bukan sekadar pencocokan kata literal.

Arsitektur seperti *Dense Passage Retrieval (DPR)* menggunakan *bi encoder* untuk memetakan kueri dan dokumen ke dalam ruang vektor yang sama secara independen, yang memungkinkan pencarian berbasis *nearest neighbor search* atau *Maximum Inner Product Search (MIPS)* yang sangat efisien pada skala miliaran dokumen.

Meskipun demikian, pendekatan *dense retrieval* murni terkadang kehilangan detail leksikal spesifik, seperti nama entitas atau nomor hukum yang presisi. Untuk mengatasi hal ini, model hibrida dan arsitektur *late interaction* seperti *ColBERT* dikembangkan, yang mempertahankan representasi tingkat token hingga tahap pencocokan akhir, memberikan keseimbangan antara pemahaman semantik mendalam dan presisi pencocokan istilah (X. Wang dkk. 2022).

Selain itu, integrasi model bahasa besar *Large Language Models* dalam kerangka kerja *IR*, misalnya untuk ekspansi kueri *query expansion* atau augmentasi data pelatihan *data augmentation*, terbukti meningkatkan kinerja model *retrieval* secara signifikan. Silva dkk. (2024) menunjukkan bahwa penggunaan *LLM* untuk menghasilkan kueri sintesis dari deskripsi dataset *query generation* dan menggunakannya untuk melatih ulang model *dense retriever* dapat meningkatkan metrik *NDCG at five* hingga enam puluh sembilan persen dibandingkan model pra latih standar, khususnya dalam skenario di mana data latih berlabel sangat terbatas seperti *zero shot* atau *few shot settings*.

## II.7.2 Akuisisi Data Web dan *Crawling* Terfokus

Akuisisi data dari sumber web heterogen merupakan tahap fundamental dalam pembangunan korpus legislatif digital. Metode standar yang digunakan adalah *web crawling* atau *web scraping*, yaitu proses otomatisasi pengambilan konten dari *World Wide Web* melalui agen perangkat lunak (*bots*) yang menavigasi struktur *hyperlink* secara sistematis.

Dalam konteks pengumpulan data spesifik domain yang tersebar di berbagai subdomain, pendekatan *General Purpose Crawling* sering kali tidak efisien karena cenderung mengunduh sejumlah besar halaman yang tidak relevan. Sebagai solusi, teknik *Focused Crawling* dikembangkan untuk memprioritaskan pengunduhan halaman yang memiliki relevansi topik tinggi dengan tujuan pencarian.

Dhanith dkk. (2024) mengusulkan mekanisme *weakly supervised learning* berbasis *GRU (Gated Recurrent Unit)* untuk meningkatkan modul komputasi relevansi pada *focused crawler*, yang memungkinkan agen untuk menyaring konten tidak penting

dan hanya mengekstrak bagian halaman yang mengandung informasi target, bahkan dalam lingkungan web dinamis dengan label data yang minim.

Tantangan utama dalam akuisisi data web modern adalah penanganan konten yang tidak terstruktur dan variasi format penyajian.

Nguyen dkk. (2025) mendemonstrasikan kerangka kerja otomatis yang menggabungkan teknik *crawling* dengan pemrosesan bahasa alami (*Natural Language Processing NLP*) untuk mengkategorikan informasi dari situs web secara *real time*. Pendekatan ini tidak hanya mengambil teks mentah (*raw text*), tetapi juga memanfaatkan struktur *DOM* (*Document Object Model*) dan *metadata HTML* untuk mempertahankan konteks hierarkis dokumen. Selain itu, untuk menangani skala data yang besar, strategi penjadwalan terdistribusi (*distributed scheduling strategy*) menjadi krusial.

F. Wang dkk. (2023) memperkenalkan strategi penjadwalan berbasis *rasa lapar* (*hunger based strategy*) untuk *crawler* terdistribusi, yang secara dinamis menyeimbangkan beban kerja antar node pengunduh untuk memaksimalkan *throughput* dan cakupan halaman (*coverage*) tanpa membebani server target secara berlebihan. Penerapan teknik ini menjamin bahwa proses akuisisi dokumen legislatif dapat berjalan secara efisien, terukur (*scalable*), dan mampu beradaptasi dengan perubahan struktur situs web sumber.

#### II.7.2.1 *Focused Crawling*

*Focused Crawling* merupakan paradigma dalam akuisisi data web yang dirancang untuk meningkatkan efisiensi pengumpulan halaman dengan membatasi eksplorasi hanya pada halaman yang relevan dengan topik atau domain tertentu. Berbeda dengan *general purpose web crawler* yang melakukan penjelajahan lebar (*breadth first search*) tanpa diskriminasi konten, *focused crawler* dilengkapi dengan mekanisme penilaian relevansi (*relevance scoring*) yang memprediksi apakah sebuah *URL* kandidat mengarah ke konten yang sesuai dengan tujuan *crawling*. Secara algoritmik, *focused crawler* beroperasi dengan memelihara antrian prioritas (*priority queue*) yang mengurutkan *URL* berdasarkan skor relevansi yang dihitung oleh fungsi klasifikasi, yang dapat berbasis pada analisis konten halaman induk, struktur *URL*, atau teks jangkar (*anchor text*).

Dhanith dkk. (2024) menjelaskan bahwa pendekatan pembelajaran terawasi lemah (*weakly supervised learning*) dapat diterapkan untuk melatih modul prediktor rele-



vansi menggunakan *dataset* berlabel tidak sempurna, di mana model *Gated Recurrent Unit (GRU)* digunakan untuk menangkap pola sekuensial dalam representasi fitur halaman web guna meningkatkan akurasi deteksi konten relevan.

Untuk mengatasi keterbatasan pendekatan berbasis heuristik statis, teknik pembelajaran penguatan (*Reinforcement Learning RL*) telah diintegrasikan ke dalam strategi *crawling* untuk memungkinkan agen belajar secara adaptif dari interaksi dengan lingkungan web. Dalam kerangka kerja ini, *crawler* dimodelkan sebagai agen *RL* yang menerima status (*state*) berupa representasi halaman saat ini, memilih aksi berupa *URL* berikutnya yang akan dikunjungi, dan menerima *reward* berdasarkan relevansi halaman yang berhasil diunduh.

Nguyen dkk. (2025) menunjukkan bahwa kombinasi teknik *crawling* dengan klasifikasi teks berbasis *NLP* secara *real time* memungkinkan kategorisasi konten segera setelah diunduh, sehingga umpan balik relevansi dapat digunakan untuk menyesuaikan strategi eksplorasi secara dinamis. Selain itu, metode optimasi bobot kata kunci (*keyword weight optimization*) telah dikembangkan untuk meningkatkan presisi *focused crawler*, di mana bobot *term* dalam vektor *query* diperbarui secara iteratif menggunakan strategi gradien berdasarkan *feedback* dari halaman yang telah diunduh, memastikan bahwa *crawler* semakin terfokus pada *subdomain* yang paling relevan seiring berjalannya waktu (Abbasi, Rashidi, dan Ahmadi 2021).

### II.7.3 Pra pemrosesan Dokumen dan Analisis Tata Letak (*Layout Analysis*)

Kualitas data yang diekstrak dari korpus legislatif sangat bergantung pada efektivitas tahap pra pemrosesan, terutama mengingat heterogenitas format dokumen yang mencakup *PDF* berbasis teks (*born digital*) dan hasil pindaian (*scanned images*) dengan tingkat degradasi visual yang bervariasi. Strategi *OCR* selektif diterapkan untuk menyeimbangkan akurasi dan efisiensi komputasi. Untuk dokumen pindaian, penerapan teknik *binarization* adaptif dan koreksi kemiringan (*skew correction*) terbukti krusial sebelum proses pengenalan karakter dilakukan. Ignasius dkk. (2023) menunjukkan bahwa kombinasi *denoising* dan *skew correction* sebagai langkah pra *OCR* dapat secara signifikan menurunkan *Character Error Rate (CER)* pada dokumen berlatar belakang kompleks.

Lebih lanjut, untuk menangani artefak visual pada risalah lama, algoritma penghapusan derau (*noise removal*) berbasis optimalisasi gabungan (*joint optimization*) digunakan untuk memisahkan karakter tulisan dari noda latar belakang tanpa merusak integritas morfologi huruf. Pendekatan ini berdampak langsung pada peningkatan

akurasi pengenalan mesin yang beroperasi di atas lapisan teks hasil ekstraksi citra.

Tantangan utama dalam memproses dokumen legislatif adalah kompleksitas tata letak (*layout*) yang sering kali memuat elemen non linier seperti tabel, *header*, *footer*, dan kolom ganda. Metode analisis tata letak dokumen (*Document Layout Analysis DLA*) modern telah beralih dari pendekatan berbasis aturan heuristik menuju arsitektur *deep learning*. Ignasius dkk. (2023) memperkenalkan kerangka kerja berbasis *YOLO (You Only Look Once)* yang dimodifikasi dengan mekanisme konvolusi dinamis untuk mendeteksi elemen tata letak multi skala secara simultan. Pendekatan ini memungkinkan sistem untuk membedakan antara blok teks naratif, tabel data presensi, dan judul sesi dengan presisi tinggi (*mAP* di atas 90 persen), bahkan pada dokumen dengan struktur non Manhattan yang rumit.

Selain itu, penggunaan model transformer multimodal seperti *LayoutLMv3* memungkinkan integrasi fitur visual dan tekstual untuk memahami semantik spasial dokumen. Ignasius dkk. (2023) mengusulkan kerangka kerja fusi dinamis (*Dynamic Fusion Network*) yang secara efektif menggabungkan fitur visual global dan detail tekstual lokal. Pendekatan ini memastikan bahwa urutan pembacaan logis (*logical reading order*) dapat direkonstruksi dengan benar meskipun elemen fisik dokumen tersebar secara acak. Kemampuan ini esensial untuk mengubah halaman risalah rapat yang statis menjadi aliran teks terstruktur yang siap diproses oleh komponen *NLP* di tahap selanjutnya.

### **II.7.3.1 Optical Character Recognition (OCR)**

*Optical Character Recognition (OCR)* adalah teknologi komputasional yang mengonversi gambar dokumen yang mengandung teks, baik hasil pindaian (*scanned*), foto, maupun tangkapan layar, menjadi data teks yang dapat diedit dan dicari secara mesin. Secara arsitektural, sistem *OCR* modern terdiri dari beberapa tahapan pemrosesan, yaitu pra pemrosesan gambar (*image preprocessing*), segmentasi karakter (*character segmentation*), pengenalan karakter (*character recognition*), dan pasca pemrosesan (*post processing*) untuk koreksi kesalahan.

Pra pemrosesan mencakup operasi seperti *binarization* (konversi gambar berwarna menjadi hitam putih), *deskewing* (koreksi kemiringan dokumen), dan penghilangan derau (*noise removal*) untuk meningkatkan kontras antara karakter dan latar belakang. F. Wang dkk. (2023) menjelaskan bahwa arsitektur berbasis *deep learning*, khususnya model *Convolutional Recurrent Neural Networks (CRNN)*, telah mendominasi tugas pengenalan teks dalam citra (*scene text recognition*) dan mengungguli

metode tradisional berbasis fitur tangan (*hand crafted features*) dengan margin akurasi yang signifikan.

Meskipun *OCR* telah mencapai akurasi tinggi pada teks cetak bersih, tantangan tetap ada pada dokumen dengan kualitas rendah, teks berlatar belakang kompleks, atau distorsi geometris. Untuk menangani kesalahan *OCR* (*OCR errors*), teknik pasca pemrosesan berbasis pemrosesan bahasa alami (*Natural Language Processing NLP*) telah dikembangkan. Ignasius dkk. (2023) mengusulkan algoritma koreksi pasca *OCR* (*OCR post correction*) yang memanfaatkan model *sequence to sequence* (*Seq2Seq*) untuk mendeteksi dan memperbaiki kesalahan pengenalan karakter secara kontekstual.

Pendekatan ini mengintegrasikan pengetahuan linguistik, seperti model bahasa *n gram* dan representasi *BERT*, untuk memprediksi kata yang paling mungkin benar berdasarkan konteks kalimat, sehingga meningkatkan *robustness* sistem terhadap gambar teks yang telah dimanipulasi atau mengandung derau *adversarial*. Selain itu, untuk ekstraksi data tabular dari dokumen pindaian yang kompleks, Ignasius dkk. (2023) mendemonstrasikan kerangka kerja hibrida yang menggabungkan deteksi tabel berbasis *DETR* (*DEtection TRansformer*) dengan model pengenalan struktur tabel *EDD* (*Encoder Dual Decoder*). Kerangka ini mencapai skor *Tree Edit Distance based Similarity* (*TEDS*) sebesar 0,699, yang menunjukkan kemampuan tinggi dalam merekonstruksi struktur dan konten tabel secara simultan dari gambar beresolusi rendah.

#### II.7.4 *Event Log*

*Event log* adalah struktur data fundamental dalam sistem informasi proses (*process aware information systems*) yang merekam jejak aktivitas sebagai sekuens peristiwa (*event*) yang terjadi selama eksekusi proses bisnis atau sistem komputasional. Secara formal, sebuah *event log L* didefinisikan sebagai koleksi dari *traces* atau *cases*, di mana setiap *case* merepresentasikan satu instansi eksekusi proses dan terdiri dari sekuens peristiwa yang terurut secara kronologis. Setiap peristiwa minimal memiliki atribut *activity name* (jenis aktivitas yang dilakukan), *timestamp* (waktu kejadian), dan secara opsional *resource* (pelaku atau sistem yang melakukan aktivitas tersebut) serta *data attributes* tambahan yang merekam konteks spesifik. Landauer dkk. (2023) menjelaskan bahwa struktur ini memungkinkan analisis retrospektif terhadap perilaku sistem, deteksi anomali, dan rekonstruksi alur kerja (*workflow*) yang aktual.

Dalam praktiknya, *event log* diperoleh dari sistem *monitoring* seperti basis data

transaksional, *server log*, atau aplikasi *enterprise* yang secara otomatis mencatat setiap interaksi pengguna dan operasi sistem. Tantangan utama dalam pengelolaan *event log* adalah ketidaksempurnaan data (*log imperfection*), yang mencakup *missing events* (peristiwa yang tidak tercatat), *incorrect timestamps*, dan *interleaved logs* dari proses paralel yang tidak memiliki *identifier* sesi yang jelas.

Untuk mengatasi hal ini, teknik pra pemrosesan seperti *log parsing* menjadi esensial. Alzu'bi dkk. (2025) mengusulkan pendekatan klasifikasi berbasis semantik menggunakan *BERT* untuk menganalisis pesan log yang tidak terstruktur dan mengekstrak pola semantik dari teks deskriptif log. Pendekatan ini memungkinkan sistem untuk secara otomatis mengkategorikan ribuan baris log ke dalam tipe peristiwa yang konsisten, yang kemudian dapat digunakan untuk analisis lanjutan seperti deteksi intrusi atau audit keamanan.

Selain itu, teknik *process mining*, yaitu ekstraksi model proses dari *event log*, telah menjadi domain penelitian aktif. Bántay dkk. (2023) mendemonstrasikan penggunaan *frequent pattern mining* untuk mempartisi log besar menjadi *sub log* yang merepresentasikan proses paralel yang berbeda. Pendekatan ini memungkinkan eksplorasi dan analisis proses yang lebih granular dengan kompleksitas komputasi yang lebih rendah. Kemampuan untuk mentransformasi log mentah menjadi representasi terstruktur ini menjadikan *event log* sebagai fondasi bagi sistem audit otomatis dan optimisasi proses operasional.

### II.7.5 Penautan Semantik (*Semantic Linking*)

Penautan semantik (*Semantic Linking*) adalah proses komputasional untuk menghubungkan entitas tekstual atau fragmen data yang terpisah ke dalam satu referensi makna yang kohesif, sering kali melintasi batas dokumen atau *dataset* yang berbeda. Dalam arsitektur sistem informasi, tugas ini sering dioperasionalkan sebagai *Entity Linking (EL)* atau *Semantic Matching*, di mana tujuannya adalah memetakan penyebutan entitas (*entity mention*) dalam teks sumber ke entitas target dalam basis pengetahuan (*knowledge base*) atau dokumen referensi lain. Ruas dkk. (2022) menjelaskan bahwa tantangan utama dalam *EL* adalah ambiguitas nama dan ketiadaan entitas dalam basis data (*NIL entities*). Untuk mengatasi hal ini, pendekatan berbasis atensi (*attention mechanism*) digunakan untuk membobot fitur kontekstual di sekitar penyebutan entitas, sehingga model dapat membedakan makna entitas berdasarkan konteks kalimatnya dan memprediksi tautan yang benar bahkan untuk entitas yang jarang muncul (*long tail entities*).

Selain pencocokan entitas eksplisit, penautan semantik modern juga memanfaatkan representasi vektor multidimensi untuk mengukur kesamaan makna antara dua segmen teks yang mungkin tidak memiliki tumpang tindih kata kunci (*vocabulary mismatch*). Huang dkk. (2022) memperkenalkan kerangka kerja pencocokan saraf (*neural matching network*) yang menggabungkan fitur leksikal (*TF IDF*), fitur kontekstual statis (*Word2Vec*), dan fitur dinamis (*ELMo* dan *BERT*) untuk membangun representasi semantik yang komprehensif. Dengan menghitung matriks kesamaan kosinus pada ruang vektor gabungan ini, sistem dapat mengidentifikasi hubungan semantik laten antara dokumen dengan akurasi yang jauh lebih tinggi dibandingkan metode pencocokan kata kunci tradisional.

Lebih lanjut, untuk menangkap hubungan implisit yang kompleks, J. Li dkk. (2025) mengusulkan ekstraksi Jaringan Tautan Semantik Kata (*Word Semantic Link Network WSLN*) yang dilengkapi dengan aturan penalaran (*reasoning rules*). Metode ini memungkinkan sistem untuk menurunkan hubungan baru yang tidak tertulis secara eksplisit dalam teks, serta memfasilitasi penemuan pola keterkaitan yang lebih dalam antara konsep yang tersebar dalam korpus data yang besar.

## II.8 Analytic Hierarchy Process (AHP)

*Analytic Hierarchy Process* (AHP) adalah metode pengambilan keputusan multikriteria (*Multi-Criteria Decision Making / MCDM*) yang dikembangkan oleh Thomas L. Saaty pada tahun 1970-an untuk menangani masalah keputusan yang kompleks dengan cara menguraikan masalah menjadi struktur hierarkis dan melakukan perbandingan berpasangan antar elemen (Leal 2020; Kumar, Kushwaha, Pandey, dkk. 2023). Metode ini telah digunakan secara luas dalam berbagai bidang seperti pemilihan teknologi, manajemen risiko, dan perencanaan strategis karena kemampuannya mengubah penilaian kualitatif menjadi nilai kuantitatif yang terukur (Kriswardhana dkk. 2025).

AHP bekerja melalui tiga tahapan utama: (1) pembentukan hierarki masalah yang terdiri dari tujuan, kriteria, dan alternatif; (2) penentuan bobot prioritas melalui perbandingan berpasangan menggunakan skala numerik; dan (3) agregasi bobot untuk menentukan peringkat alternatif terbaik. Keunggulan AHP terletak pada kemampuannya menangani faktor-faktor subjektif sekaligus menjaga konsistensi logis melalui perhitungan rasio konsistensi (*Consistency Ratio, CR*) (Guillén-Mena dkk. 2023).

### II.8.1 Skala Saaty dalam Perbandingan Berpasangan

Inti dari metode AHP adalah penggunaan *Saaty's Fundamental Scale* untuk melakukan perbandingan berpasangan antar kriteria atau alternatif. Skala ini menggunakan nilai numerik dari 1 hingga 9, di mana nilai 1 menunjukkan dua elemen memiliki kepentingan yang sama, sedangkan nilai 9 menunjukkan satu elemen jauh lebih penting dibandingkan elemen lainnya (Leal 2020). Tabel berikut merangkum interpretasi skala Saaty:

Tabel II.1 Skala Saaty untuk Perbandingan Berpasangan

Nilai	Interpretasi
1	Sama pentingnya ( <i>Equal importance</i> )
3	Sedikit lebih penting ( <i>Moderate importance</i> )
5	Lebih penting ( <i>Strong importance</i> )
7	Sangat lebih penting ( <i>Very strong importance</i> )
9	Mutlak lebih penting ( <i>Extreme importance</i> )
2, 4, 6, 8	Nilai antara dua penilaian berurutan

Setelah matriks perbandingan berpasangan disusun, AHP menghitung vektor prioritas (*priority vector*) menggunakan metode *eigenvalue* atau normalisasi rata-rata geometrik. Konsistensi penilaian dievaluasi melalui *Consistency Index* (CI) dan *Consistency Ratio* (CR), di mana nilai  $CR < 0.1$  dianggap dapat diterima (Frish dkk. 2025). Dalam konteks penelitian ini, AHP digunakan untuk memilih arsitektur solusi terbaik bagi sistem *Legislative Activity Tracker* dengan mempertimbangkan kriteria seperti efektivitas, skalabilitas, dampak transparansi, kemudahan implementasi, dan kualitas data.

## **BAB III**

### **ANALISIS MASALAH**

#### **III.1 Analisis Kondisi Saat Ini**

Analisis kondisi saat ini bertujuan untuk mengidentifikasi permasalahan dalam proses yang berlangsung serta mengevaluasi peluang yang muncul darinya. Analisis ini akan menjadi landasan dalam merumuskan kebutuhan sistem.

##### **III.1.1 Analisis Permasalahan**

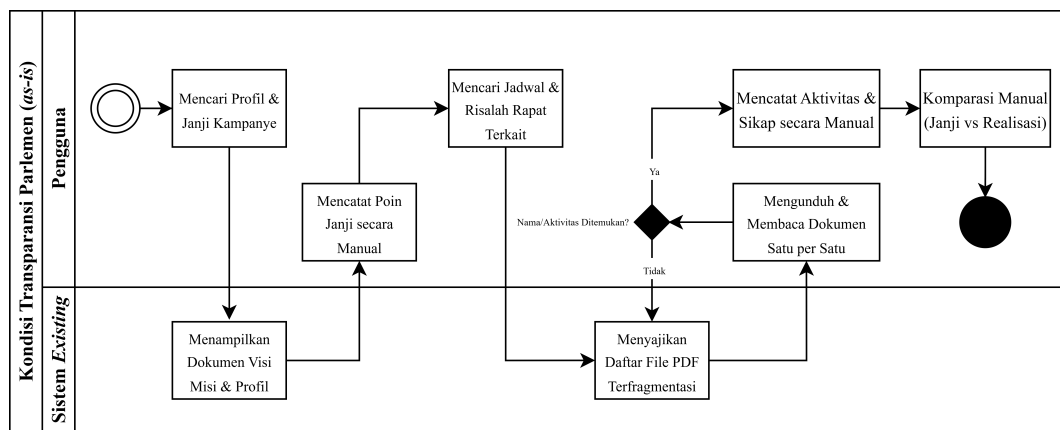
Kondisi transparansi parlemen di Indonesia saat ini masih terjebak pada paradigma ketersediaan dokumen (*availability*) semata, belum mencapai tahap aksesibilitas data (*accessibility*) yang mendukung akuntabilitas fungsional. Berdasarkan kerangka Transparansi Kinerja yang dibahas pada Sub bab II.1.2, transparansi yang efektif menuntut adanya data terukur mengenai *output* legislasi dan partisipasi anggota dewan agar publik dapat membandingkan janji politik dengan realisasi kerja.

Namun, realitas di lapangan menunjukkan bahwa informasi yang disajikan oleh kanal resmi DPR RI (seperti [dpr.go.id](http://dpr.go.id) dan [parlementaria](http://parlementaria.id)) lebih bersifat seremonial dan naratif, berfokus pada berita kegiatan daripada substansi perdebatan kebijakan. Absennya data kinerja yang granular ini menyebabkan terputusnya rantai informasi antara mandat konstituen dan tindakan wakil rakyat, yang sebagaimana dijelaskan dalam Sub bab II.3.1, berkontribusi langsung pada rendahnya kepercayaan publik (*political trust*) terhadap institusi legislatif.

Secara arsitektural, ekosistem informasi DPR RI saat ini terfragmentasi ke dalam beberapa *silo data* yang tidak saling terhubung. Mengacu pada tantangan teknis yang diuraikan dalam Sub bab II.5, data legislatif tersebar di berbagai subdomain terpisah tanpa skema integrasi yang jelas: profil anggota berada di situs utama, naskah akademik dan progres RUU di Sistem Informasi Legislasi (*SILEG*), sementara produk

hukum final di Jaringan Dokumentasi dan Informasi Hukum (JDIH).

Masalah ini diperparah oleh pelanggaran prinsip *Open Government Data (OGD)* yang dibahas pada Sub bab II.2.1, di mana data tidak disediakan dalam format yang *machine readable* maupun memiliki interoperabilitas semantik. Tidak adanya pengenalan unik (*unique identifier*) yang konsisten menghubungkan seorang legislator di satu sistem dengan aktivitasnya di sistem lain membuat upaya pengawasan komprehensif menjadi sangat sulit dilakukan secara manual, apalagi secara otomatis.



Gambar III.1 Kondisi Transparansi Parlemen (*as-is*)

Pada level teknis yang lebih mendalam, kualitas data mentah menjadi hambatan utama bagi penerapan analisis komputasional. Sebagian besar rekam jejak aktivitas legislatif tersimpan dalam format dokumen tidak terstruktur, seperti Risalah Rapat berupa *PDF* naratif yang panjang atau bahkan hasil pindaian (*scanned documents*) untuk arsip lama.

Sesuai dengan tinjauan Sub bab II.6.3 mengenai tantangan pra pemrosesan dokumen, format semacam ini menyulitkan mesin untuk membedakan antara elemen struktural (seperti daftar hadir atau keputusan fraksi) dan konten substansial (argumen perdebatan). Tanpa melalui proses *Information Extraction (IE)* dan *OCR* yang canggih sebagaimana dibahas pada Sub bab II.6.1.4 dan II.6.3.1, kekayaan informasi yang terkandung dalam ribuan halaman risalah ini tetap terkubur sebagai *dark data* yang tidak dapat diolah menjadi wawasan analitis.

Puncaknya, kelemahan sistem saat ini adalah ketiadaan mekanisme log peristiwa (*event log*) yang sistematis untuk memetakan aktivitas harian legislator. Platform pemantauan eksternal seperti JariUngu, yang dibahas pada Sub bab II.4.2, telah berupaya mengisi celah informasi ini, namun masih terbatas pada penyajian profil statis



dan agregasi berita, bukan pelacakan aktivitas berbasis urutan waktu (*temporal tracking*).

Belum ada sistem yang mampu memodelkan hubungan kausal secara eksplisit: Anggota X berjanji A saat kampanye, kemudian hadir di Rapat B, dan Fraksinya mengambil Sikap C yang mendukung atau menolak Janji A. Ketiadaan model data relasional ini, yang seharusnya dapat dijembatani melalui teknik *Semantic Linking* (Sub bab II.6.5), menjadikan evaluasi konsistensi politik anggota dewan, yang merupakan tujuan utama dari transparansi kinerja, mustahil dilakukan secara objektif dan berskala besar dengan infrastruktur yang ada saat ini.

Berdasarkan analisis kondisi faktual di atas, dapat disimpulkan bahwa hambatan transparansi DPR RI bersifat multidimensi, mencakup aspek tata kelola data yang terfragmentasi hingga keterbatasan infrastruktur teknis dalam pengolahan dokumen. Guna memfasilitasi perumusan solusi yang terarah, permasalahan-permasalahan tersebut diidentifikasi dan dikelompokkan ke dalam empat poin masalah fundamental (M-01 hingga M-04) yang akan menjadi acuan utama pengembangan sistem, sebagaimana dirangkum dalam Tabel III.1 berikut.

Tabel III.1 Daftar Identifikasi Masalah (M)

Kode	Deskripsi Masalah
M-01	Data aktivitas legislatif tersebar (terfragmentasi) di berbagai kanal (situs DPR, SILEG, JDIH) dan tersaji dalam format dokumen tidak terstruktur (PDF/Scan) yang sulit diolah mesin.
M-02	Ketiadaan mekanisme pencatatan historis ( <i>event log</i> ) yang merinci aktivitas anggota secara individual (siapa melakukan apa, di mana, dan kapan).
M-03	Diskoneksi semantik antara mandat politik (janji kampanye) dengan realisasi kinerja faktual, menyebabkan publik sulit melakukan evaluasi konsistensi legislator.
M-04	Absennya data pemungutan suara individual ( <i>roll-call votes</i> ) yang memaksa publik menebak posisi anggota hanya berdasarkan asumsi keanggotaan fraksi.

### III.1.2 Analisis Peluang

Di balik kompleksitas permasalahan yang teridentifikasi, perkembangan teknologi pemrosesan data modern menawarkan peluang strategis untuk merevolusi cara publik mengawasi parlemen. Ketersediaan dokumen digital, meskipun terfragmentasi,

sesungguhnya merupakan bahan baku yang kaya (*data lake*) jika dikelola dengan arsitektur yang tepat. Penerapan metode komputasional seperti *Information Retrieval* (IR) dan *Information Extraction* (IE) memungkinkan transformasi dokumen statis menjadi *event log* dinamis secara otomatis, memutus ketergantungan pada kurasi manual yang lambat. Lebih jauh, integrasi teknik analisis semantik membuka jalan bagi terwujudnya pengawasan berbasis bukti (*evidence-based oversight*), di mana konsistensi antara janji kampanye dan realisasi kinerja dapat diukur secara kuantitatif. Peluang-peluang pengembangan sistem yang muncul dari kesenjangan kondisi saat ini dirangkum dalam Tabel III.2 berikut.

Tabel III.2 Daftar Identifikasi Peluang (P)

Kode	Deskripsi Peluang
P-01	Integrasi data terpusat ( <i>centralized dashboard</i> ) yang menyediakan akses satu pintu terhadap rekam jejak legislator secara komprehensif.
P-02	Penerapan teknologi <i>NLP</i> dan <i>IE</i> untuk mengotomatisasi ekstraksi struktur peristiwa dari dokumen mentah menjadi data siap pakai ( <i>event log</i> ).
P-03	Penyediaan analisis komparatif berbasis data ( <i>data driven</i> ) untuk memvisualisasikan keselarasan antara janji kampanye dan realisasi kinerja.
P-04	Peningkatan transparansi fungsional melalui penyediaan data terbuka ( <i>open data</i> ) yang dapat diaudit, diverifikasi, dan digunakan ulang oleh publik.

## III.2 Analisis Kebutuhan

Analisis kebutuhan bertujuan untuk memetakan kesenjangan antara kondisi saat ini dengan solusi yang diharapkan, serta mendefinisikan spesifikasi fungsional yang harus dipenuhi oleh sistem *Legislative Activity Tracker* (LAT). Tahap ini diawali dengan identifikasi masalah dan peluang secara eksplisit, yang kemudian diturunkan menjadi kebutuhan pengguna.

### III.2.1 Identifikasi Masalah Pengguna

Pengguna sistem ini terdiri atas masyarakat umum (pemilih) dan peneliti/analisis kebijakan. Masing-masing pengguna memiliki kebutuhan serta tantangan yang berbeda dalam proses pencarian, pemantauan, maupun analisis kinerja legislatif. Sistem yang dirancang perlu mengakomodasi permasalahan yang dihadapi oleh setiap pengguna, yang diuraikan sebagai berikut:

#### 1. Masyarakat Umum (Pemilih)

- (a) Kesulitan menemukan profil dan rekam jejak aktivitas anggota DPR dalam satu ekosistem untuk melakukan evaluasi kinerja wakil rakyat.
- (b) Kesulitan menemukan informasi aktivitas yang sesuai dengan preferensi (nama anggota, fraksi, dapil, topik RUU) karena belum banyak platform yang menyediakan layanan pencarian aktivitas secara tergranular.
- (c) Kesulitan membandingkan janji kampanye dengan realisasi kinerja secara langsung karena dokumen visi-misi dan risalah rapat tersimpan di sumber yang terpisah.
- (d) Tidak tersedia sistem notifikasi atau ringkasan aktivitas terbaru, yang membuat pemilih sering tertinggal informasi mengenai keputusan krusial yang diambil wakilnya.

## 2. **Peneliti dan Analis Kebijakan**

- (a) Kesulitan mengumpulkan data aktivitas legislatif secara massal atau menjangkau dokumen historis yang tersebar di berbagai subdomain situs DPR.
- (b) Kurang tersedianya sistem untuk mengelola *dataset* legislatif, melakukan *filter* berdasarkan parameter spesifik, dan mengekspor data secara terpusat dalam satu platform.
- (c) Kurang adanya fitur untuk melakukan analisis tren atau menampilkan statistik kinerja fraksi/anggota secara sistematis.
- (d) Kesulitan mendapatkan konteks keterkaitan antar-dokumen (misalnya hubungan antara dokumen RUU dengan risalah rapat terkait) karena tidak ada sistem penautan referensi yang efisien.

Kebutuhan pengguna ini lebih merangkum dari kebutuhan dari analisis permasalahan dan peluang yang sudah dijelaskan sebelumnya serta merangkum bagaimana sistem dapat membantu pengguna. Tabel III.3 menunjukkan pemetaan kebutuhan pengguna yang telah diidentifikasi untuk mengatasi permasalahan pada proses interaksi antara pengajar, pelajar, dan penyedia tempat belajar.

Tabel III.3 Analisis Kebutuhan Pengguna

ID	Kebutuhan	Masalah Terkait	Peluang
K-01	Pengguna dapat mencari dan melihat profil anggota DPR beserta riwayat aktivitas legislatifnya (rapat, intervensi, kehadiran) dalam satu linimasa terpadu.	M-01, M-02	P-01, P-02
K-02	Pengguna dapat melihat perbandingan langsung ( <i>side by side</i> ) antara poin janji kampanye dengan sikap fraksi atau aktivitas anggota pada isu yang relevan.	M-03	P-03
K-03	Pengguna dapat melakukan filter aktivitas berdasarkan parameter spesifik: Topik RUU, Nama Komisi, Fraksi, atau Rentang Waktu.	M-01	P-01
K-04	Pengguna (khususnya peneliti) dapat mengakses atau mengunduh data aktivitas dalam format terstruktur ( <i>event log</i> ) untuk keperluan analisis lanjutan.	M-01, M-02	P-04
K-05	Pengguna dapat memverifikasi klaim data melalui tautan langsung ke dokumen sumber asli (Risalah Rapat, Berita Resmi) yang disediakan sistem.	M-03, M-04	P-04

### III.2.2 Kebutuhan Fungsional

Kebutuhan fungsional mendefinisikan fungsi-fungsi spesifik yang harus dijalankan oleh sistem untuk mendukung tujuan transparansi dan pengawasan legislatif. Rincian lengkap mengenai proses *input*, proses, dan *output* setiap fungsi dapat dilihat pada Lampiran. Berikut adalah daftar kebutuhan fungsional sistem:

1. **FR-01:** Sistem mampu mengakuisisi dan memperbarui dokumen legislatif secara otomatis dari berbagai sumber resmi (*web crawling*).
2. **FR-02:** Sistem mampu mengekstrak struktur peristiwa (*event log*) dari dokumen tidak terstruktur menggunakan teknik NLP dan IE.
3. **FR-03:** Sistem mampu menghitung dan menautkan kesamaan semantik antara janji kampanye dengan aktivitas legislatif (*semantic linking*).
4. **FR-04:** Sistem mampu menampilkan *dashboard* pencarian profil anggota dan riwayat aktivitasnya dengan fitur *filter*.

5. **FR-05:** Sistem mampu memvisualisasikan perbandingan komparatif (*side-by-side*) antara janji kampanye dan realisasi kinerja.

Penjelasan lebih rinci mengenai spesifikasi setiap kebutuhan fungsional di atas, termasuk deskripsi tujuan, data masukan, tahapan proses, serta format keluaran yang dihasilkan oleh sistem, disajikan secara lengkap dalam Tabel L-1 Spesifikasi Kebutuhan Fungsional yang terdapat pada Lampiran. Tabel tersebut menjabarkan logika teknis operasional sistem dalam memenuhi setiap persyaratan yang telah ditetapkan.

### III.2.3 Kebutuhan Nonfungsional

Selain kebutuhan fungsional, sistem juga harus memenuhi sejumlah batasan dan persyaratan kualitas agar dapat beroperasi dengan andal dan memberikan pengalaman pengguna yang baik. Kebutuhan non-fungsional sistem *Legislative Activity Tracker* (LAT) dijabarkan dalam Tabel III.5 berikut.

Tabel III.4 Analisis Kebutuhan Non-Fungsional Sistem

ID	Kebutuhan Non-Fungsional	Deskripsi
NFR-01	Kinerja ( <i>Performance</i> )	Sistem harus mampu menampilkan hasil pencarian profil atau aktivitas dalam waktu kurang dari 3 detik pada kondisi koneksi internet standar, guna menjamin kenyamanan pengguna.
NFR-02	Ketersediaan ( <i>Availability</i> )	Sistem harus dapat diakses oleh pengguna 24 jam sehari, 7 hari seminggu, dengan target <i>uptime</i> minimal 99% selama jam operasional normal.
NFR-03	Skalabilitas ( <i>Scalability</i> )	Arsitektur sistem, khususnya basis data <i>event log</i> , harus dirancang untuk menampung penambahan volume data dokumen harian secara terus-menerus tanpa degradasi performa yang signifikan.
NFR-04	Integritas Data ( <i>Data Integrity</i> )	Sistem harus menjamin bahwa setiap entri data aktivitas memiliki tautan referensi ( <i>provenance link</i> ) yang valid ke dokumen sumber asli, sehingga data yang disajikan dapat diverifikasi kebenarannya.

*Bersambung ke halaman berikutnya*

Tabel III.4 – Lanjutan dari halaman sebelumnya

ID	Kebutuhan Non-Fungsional	Deskripsi
NFR-05	Ketahanan ( <i>Robustness</i> )	Modul <i>crawler</i> dan <i>parser</i> sistem harus memiliki mekanisme penanganan kesalahan ( <i>error handling</i> ) yang baik, sehingga sistem tidak berhenti total saat menemui format dokumen PDF yang rusak atau tidak standar.
NFR-06	Kompatibilitas ( <i>Compatibility</i> )	Antarmuka <i>dashboard</i> sistem harus bersifat responsif ( <i>web-responsive</i> ) dan dapat diakses dengan tampilan yang baik melalui berbagai peramban web modern ( <i>Chrome, Firefox, Edge</i> ) di perangkat <i>desktop</i> maupun <i>mobile</i> .

### III.3 Analisis Pemilihan Solusi

Analisis pemilihan solusi bertujuan untuk mengidentifikasi kebutuhan dari setiap permasalahan yang dihadapi pengguna, kemudian akan dirumuskan sejumlah alternatif solusi yang relevan. Setelah itu, solusi yang paling relevan akan dipilih untuk dirancang.

#### III.3.1 Alternatif Solusi

Alternatif solusi merumuskan berbagai strategi pendekatan yang mungkin diambil untuk membangun sistem pemantauan legislatif. Berikut adalah ringkasan daftar alternatif solusi dari Lampiran 2. Penjelasan Detail Alternatif Solusi yang dipertimbangkan:

1. S01: Pengembangan Portal Direktori Manual (*Manual Curation*)  
Membangun basis data profil anggota dan aktivitas yang diinput secara manual oleh tim admin/relawan berdasarkan pemantauan berita dan dokumen fisik, menyerupai model awal platform *civic tech* konvensional.
  - (a) Masalah yang diselesaikan: M-03 (Mencoba mengatasi diskoneksi informasi dengan menyajikan profil dan janji dalam satu halaman).
  - (b) Keterbatasan: Gagal mengatasi M-01 dan M-02 secara *scalable* karena ketergantungan pada tenaga manusia yang lambat dan rentan *human error*.
2. S02: Sistem Temu Kembali Dokumen Terpusat (*IR-Only System*)

Membangun mesin pencari (*search engine*) yang mengindeks seluruh dokumen PDF dari berbagai situs DPR (*crawling*) agar dapat dicari berdasarkan kata kunci, tanpa melakukan ekstraksi struktur data mendalam.

- (a) Masalah yang diselesaikan: M-01 (Mengatasi masalah data tersebar dengan mengumpulkannya di satu repositori terpusat).
- (b) Keterbatasan: Gagal mengatasi M-02 dan M-04. *Output* sistem masih berupa dokumen mentah (PDF), sehingga pengguna tetap harus membaca manual.

3. S03: Sistem Ekstraksi Informasi Statis (*IE-Only on Local Corpus*)

Fokus hanya pada pengembangan algoritma NLP/IE untuk mengekstrak data dari kumpulan dokumen yang sudah diunduh secara lokal (*offline dataset*), tanpa modul *crawling* atau pembaruan otomatis.

- (a) Masalah yang diselesaikan: M-02 (Menjawab kebutuhan struktur data *event log*) dan M-04 (Membantu inferensi posisi fraksi).
- (b) Keterbatasan: Gagal mengatasi M-01 secara berkelanjutan karena tidak memiliki mekanisme *discovery* data baru dari web DPR yang dinamis.

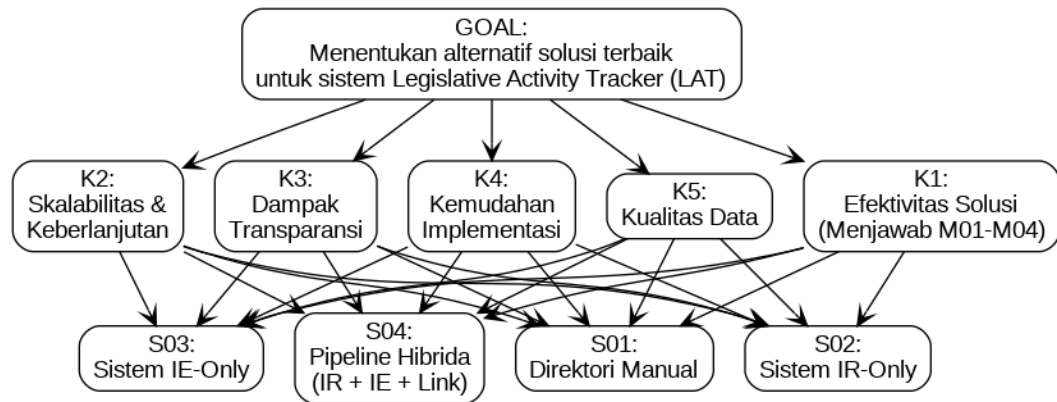
4. S04: Arsitektur *Pipeline* Hibrida (*Integrated IR + IE Pipeline*)

Mengintegrasikan modul *Information Retrieval* (untuk akuisisi dan manajemen korpus otomatis) dengan modul *Information Extraction* (untuk *parsing* NLP dan pembentukan *event log*), dilengkapi lapisan *Semantic Linking* untuk menghubungkan aktivitas dengan janji kampanye.

- (a) Masalah yang diselesaikan: M-01, M-02, M-03, dan M-04 secara komprehensif. Solusi ini mencakup penanganan hulu (data mentah) hingga hilir (visualisasi *dashboard*).
- (b) Keterbatasan: Memiliki kompleksitas implementasi yang paling tinggi dibandingkan opsi lain karena memerlukan pengembangan dua modul utama sekaligus (*IR* dan *IE*). Selain itu, akurasi ekstraksi data bergantung pada performa model *NLP* yang mungkin menghasilkan kesalahan (*false positives/negatives*) pada dokumen dengan struktur yang sangat tidak baku.

### III.3.2 Analisis Penentuan Solusi

Untuk menentukan solusi terbaik dari alternatif yang telah diusulkan, dilakukan analisis kuantitatif menggunakan metode *Analytical Hierarchy Process* (AHP). Proses ini diawali dengan pembentukan struktur hierarki keputusan (sebagaimana pada Gambar III.2), kemudian dilanjutkan dengan penilaian perbandingan berpasangan antar kriteria dan antar alternatif.



Gambar III.2 Struktur (*Analytical Hierarchy Process*)

Penilaian perbandingan dilakukan berdasarkan Skala Saaty, yang memberikan nilai numerik untuk merepresentasikan tingkat kepentingan relatif antara dua elemen. Rincian skala penilaian yang digunakan disajikan pada Lampiran 3. Rincian Analisis Penentuan Solusi berikut.

Tabel III.5 Perhitungan Bobot Akhir Alternatif Solusi

Alternatif	Efektivitas (0.348)	Skalabilitas (0.136)	Dampak (0.348)	Kemudahan (0.036)	Kualitas (0.136)	Bobot Global (Skor Akhir)
S01 (Manual)	0.016	0.007	0.070	0.018	0.034	0.145
S02 (IR)	0.035	0.048	0.035	0.011	0.014	0.143
S03 (IE)	0.072	0.007	0.035	0.004	0.041	0.159
S04 (Hibrida)	0.223	0.075	0.209	0.002	0.048	0.557

Berdasarkan hasil analisis dan perhitungan bobot prioritas kriteria (Tabel III.5), didapatkan bahwa Efektivitas Solusi (K1) dan Dampak Transparansi (K3) memiliki bobot yang paling tinggi dibandingkan dengan kriteria yang lain. Hal ini mengindikasikan bahwa kemampuan solusi dalam menjawab seluruh permasalahan secara komprehensif dan memberikan manfaat maksimal bagi publik menjadi prioritas utama.

Terdapat empat usulan solusi yang dievaluasi, antara lain Pengembangan Portal Direktori Manual (S01), Sistem Temu Kembali Dokumen (S02), Sistem Ekstraksi Informasi Statis (S03), dan Arsitektur *Pipeline* Hibrida (S04). Berdasarkan perhitungan bobot global yang disajikan pada Lampiran 3. Rincian Analisis Penentuan Solusi, dapat ditentukan bahwa Arsitektur *Pipeline* Hibrida (S04) dengan skor akhir 0.557 menjadi solusi yang akan dipilih untuk proyek kali ini. Meskipun solusi ini memiliki tingkat kesulitan implementasi tertinggi, kemampuannya dalam memberikan efektivitas dan dampak yang paling signifikan menjadikannya pilihan yang paling optimal.



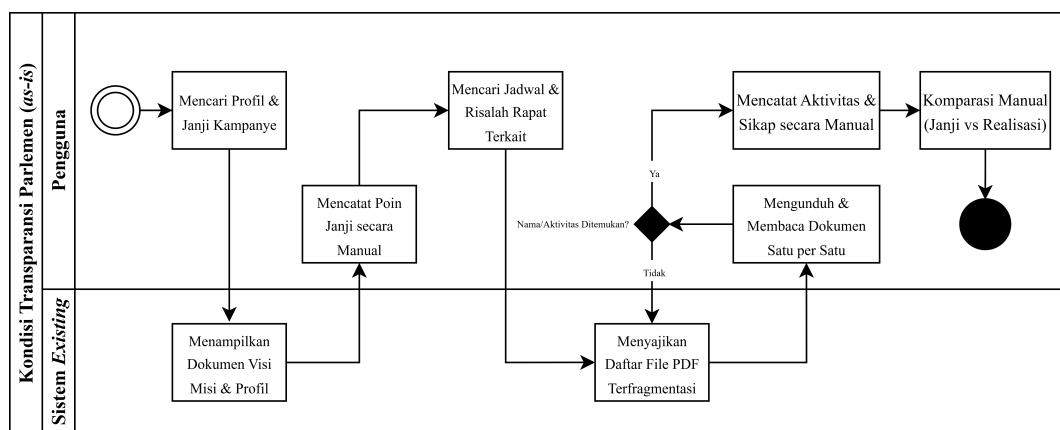
## BAB IV

### DESAIN KONSEP SOLUSI

#### IV.1 Diagram Konseptual

##### IV.1.1 Kondisi Sistem (*As-Is*)

Kondisi sistem transparansi parlemen saat ini telah dianalisis secara mendalam pada Sub-bab III.1.1 Analisis Permasalahan, di mana teridentifikasi sejumlah hambatan struktural yang menghambat pengawasan publik yang efektif. Untuk memberikan konteks perbandingan yang jelas, Gambar IV.1 berikut merangkum kembali alur kerja (*workflow*) yang harus dilalui oleh pengguna baik masyarakat umum maupun peneliti ketika berupaya mengevaluasi konsistensi kinerja seorang legislator terhadap janji kampanyenya dengan menggunakan ekosistem data DPR RI yang ada saat ini.



Gambar IV.1 Kondisi Transparansi Parlemen (*as-is*)

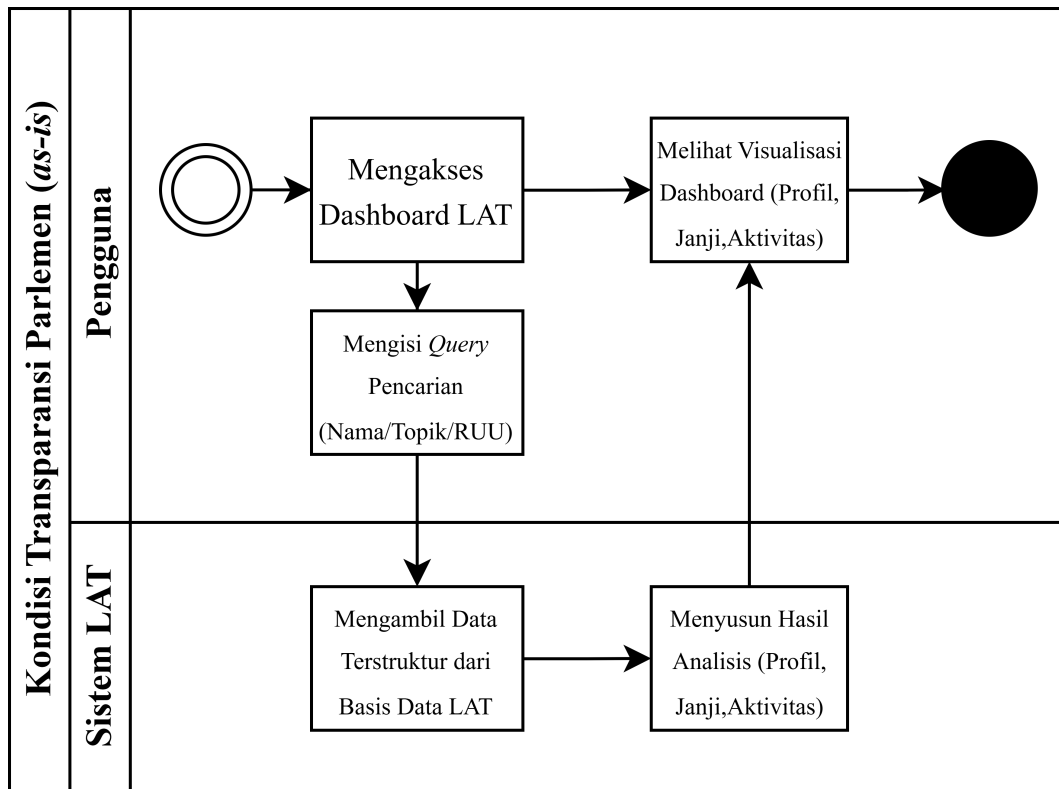
Sebagaimana terlihat dalam diagram, seluruh proses pemantauan masih bersifat manual dan terfragmentasi. Pengguna harus memulai dengan mencari profil dan dokumen janji kampanye dari sumber terpisah (misalnya situs KPU atau portal infor-

masi pemilu), kemudian beralih ke situs resmi DPR untuk mencari jadwal dan risalah rapat yang relevan. Setelah menemukan dokumen yang sesuai, pengguna harus mengunduh file PDF satu per satu, membaca secara manual untuk menemukan nama atau topik yang dicari, dan mencatat informasi aktivitas serta sikap legislator secara terpisah. Tahap terakhir adalah melakukan komparasi manual antara catatan janji kampanye dengan catatan aktivitas faktual untuk menarik kesimpulan mengenai konsistensi kinerja.

Alur kerja ini mencerminkan secara langsung keempat masalah utama yang telah diidentifikasi: fragmentasi data (M-01) yang memaksa pengguna berpindah-pindah platform, ketiadaan *event log* terstruktur (M-02) yang mengharuskan pembacaan dokumen naratif panjang, diskoneksi semantik (M-03) yang membuat penautan janji-realisisasi harus dilakukan secara subjektif, dan inferensi sikap berbasis asumsi (M-04) karena tidak tersedianya data *voting* atau transkrip pidato individual yang eksplisit. Kondisi ini mengakibatkan beban kerja kognitif yang sangat tinggi bagi pengguna, menjadikan pengawasan kinerja parlemen sebagai aktivitas yang hanya dapat dilakukan oleh pihak dengan sumber daya waktu dan keahlian penelusuran data yang memadai.

#### **IV.1.2 Kondisi Sistem *Legislative Activity Tracker***

Sebagai respon terhadap berbagai hambatan yang diuraikan pada kondisi *as-is*, sistem *Legislative Activity Tracker* (LAT) dirancang untuk mereduksi langkah manual pengguna secara drastis dan memindahkan beban pengolahan informasi ke dalam mekanisme otomatis. Gambar IV.2 menggambarkan rancangan alur aktivitas (*activity diagram*) kondisi *Legislative Activity Tracker*, ketika solusi LAT telah diimplementasikan. Diagram ini menggunakan dua *swimlane*, yaitu Pengguna dan Sistem LAT, untuk menonjolkan pergeseran peran antara manusia dan mesin.



Gambar IV.2 Kondisi Transparansi Parlemen *Legislative Activity Tracker*

Pada rancangan *Legislative Activity Tracker* ini, peran pengguna direduksi menjadi dua langkah utama: mengakses *dashboard* dan memasukkan parameter pencarian, kemudian menginterpretasikan hasil yang disajikan. Seluruh pekerjaan yang sebelumnya dilakukan secara manual mulai dari mencari dokumen di berbagai situs, mengunduh PDF, membaca satu per satu, hingga mencatat dan mengomparasi informasi dipindahkan ke dalam mekanisme otomatis di sisi Sistem LAT. Modul akuisisi data dan pemrosesan NLP yang telah dijelaskan pada Bab III bekerja di belakang layar untuk menyediakan basis data terstruktur yang siap dikueri, sementara modul analisis semantik menyusun hubungan eksplisit antara janji kampanye dan aktivitas legislatif.

Dengan demikian, alur *Legislative Activity Tracker* tidak hanya memperpendek jumlah langkah yang harus ditempuh pengguna, tetapi juga mengubah sifat tugas dari kerja administratif berulang menjadi kerja analitis berbasis bukti. Pengguna dapat langsung fokus pada penilaian substansi kinerja legislator, alih-alih tersandera oleh proses teknis penelusuran dan pembacaan dokumen.

#### IV.1.3 Perbandingan Sistem As-Is dan Sistem *Legislative Activity Tracker*

Berdasarkan diagram aktivitas yang telah dipaparkan pada Gambar IV.1 dan Gambar IV.2, terlihat perbedaan fundamental dalam pendekatan pemantauan kinerja legislatif antara kondisi saat ini (*as-is*) dan kondisi yang diusulkan (*Legislative Activity Tracker*). Pada kondisi *as-is*, pengguna memegang kendali penuh atas seluruh rantai proses: mulai dari penemuan dokumen (*discovery*), akuisisi data, pemrosesan (membaca dan mengekstraksi), hingga analisis (komparasi dan evaluasi). Beban kerja kognitif sepenuhnya berada di pundak pengguna, menjadikan proses pengawasan sangat rentan terhadap kelelahan, bias subjektif, dan kesalahan manusia (*human error*). Selain itu, fragmentasi sumber data yang parah memaksa pengguna untuk bertindak sebagai “integrator manual” yang harus menyambungkan informasi dari portal KPU, situs DPR, dan media secara mandiri.

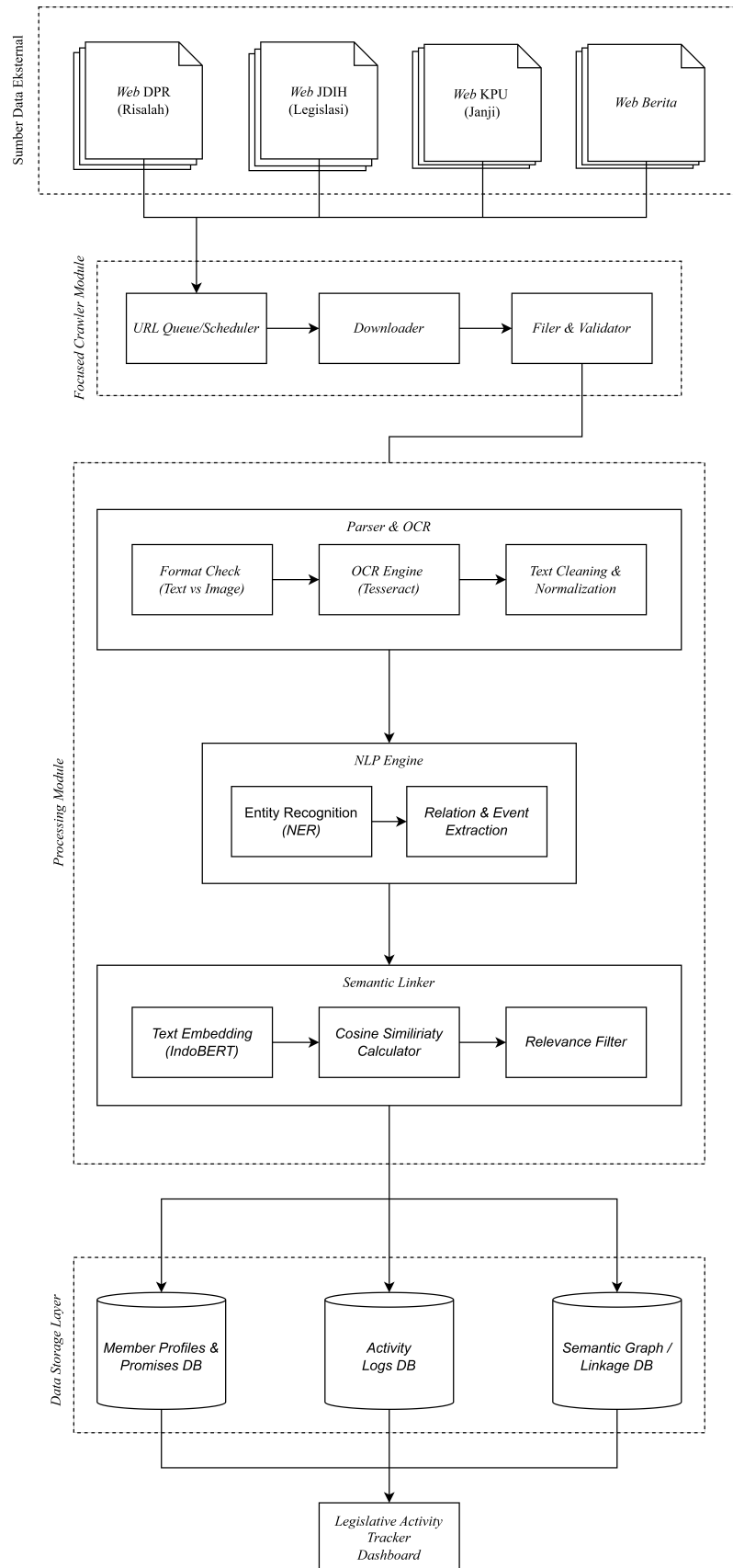
Sebaliknya, pada kondisi *Legislative Activity Tracker*, Sistem LAT mengambil alih seluruh beban teknis dan administratif tersebut melalui mekanisme otomatisasi di belakang layar (*backend automation*). Peran pengguna bertransformasi dari seorang “pengumpul data” menjadi “analisis informasi”. Pengguna tidak perlu lagi mengetahui kompleksitas lokasi penyimpanan dokumen atau menghabiskan waktu untuk membaca ratusan halaman risalah rapat; mereka cukup berinteraksi dengan antarmuka *dashboard* yang menyajikan hasil olahan data dalam bentuk visualisasi siap pakai. Dengan mendelegasikan tugas pencarian, ekstraksi, dan penautan semantik kepada algoritma sistem, proses pemantauan menjadi jauh lebih efisien, terukur, dan dapat diakses oleh publik secara luas tanpa memerlukan keahlian riset khusus.

Tabel IV.1 Perbandingan Sistem As-Is dan Sistem LAT Usulan

Aspek Perbandingan	Sistem Saat Ini ( <i>As-Is</i> )	Sistem Usulan <i>Legislative Activity Tracker</i>
Titik Akses Data	Terfragmentasi; pengguna harus mengakses banyak portal (KPU, DPR, JDIH) secara terpisah.	Terpusat ( <i>Centralized</i> ); pengguna mengakses satu pintu melalui <i>dashboard</i> LAT yang mengintegrasikan seluruh sumber data.
Metode Pengolahan	Manual; pengguna membaca dokumen PDF satu per satu untuk menemukan informasi relevan.	Otomatis; sistem menggunakan <i>Web Crawler</i> dan <i>NLP Engine</i> untuk mengekstrak data secara massal.
Format Informasi	Dokumen naratif mentah ( <i>Unstructured PDF</i> ) yang sulit diolah kembali.	Data peristiwa terstruktur ( <i>Structured Event Log</i> ) yang siap dikueri, difilter, dan dianalisis.
Mekanisme Evaluasi	Komparasi manual dan subjektif berdasarkan ingatan atau catatan pribadi pengguna.	Komparasi otomatis berbasis data dengan skor kesesuaian semantik ( <i>Semantic Similarity Score</i> ).
Efisiensi Waktu	Rendah; membutuhkan waktu berjam-jam hingga berhari-hari untuk satu topik atau anggota.	Tinggi; hasil pencarian dan analisis tersaji secara instan ( <i>real-time</i> ) dalam hitungan detik.
Skalabilitas	Terbatas; sulit dilakukan untuk memantau seluruh anggota dewan secara bersamaan.	Tinggi; sistem dapat memproses dan memantau aktivitas ratusan anggota dewan secara paralel.

#### IV.1.4 Diagram Arsitektur Tingkat Tinggi (*High-Level Architecture*)

Untuk merealisasikan alur kerja otomatis yang telah dijelaskan sebelumnya, sistem LAT dibangun di atas arsitektur *pipeline* hibrida yang mengintegrasikan teknik *Information Retrieval* (IR) dan *Natural Language Processing* (NLP). Gambar IV.3 memperlihatkan model konseptual sistem yang terdiri dari tiga lapisan utama: lapisan Sumber Data, lapisan Pemrosesan Inti (*Core Processing*), dan lapisan Presentasi. Diagram ini memetakan aliran data dari dokumen mentah hingga menjadi informasi visual di *dashboard* pengguna.



Gambar IV.3 Diagram Konseptual (*Legislative Activity Tracker*)

Arsitektur di atas dirancang secara spesifik untuk menjawab tantangan teknis yang telah diidentifikasi pada Bab III. Berikut adalah penjelasan fungsi setiap komponen utama dalam arsitektur tersebut:

1. Modul Akuisisi Data (*Data Acquisition Module*): Modul ini berfungsi sebagai gerbang awal sistem yang menangani masalah fragmentasi data (M-01). Menggunakan teknik *focused crawling*, modul ini secara otomatis memindai dan mengunduh dokumen terbaru (seperti risalah rapat, laporan singkat, dan dokumen legislasi) dari berbagai subdomain situs DPR RI ([dpr.go.id](http://dpr.go.id), [berkas.dpr](http://berkas.dpr), [sileg](http://sileg), [jdih](http://jdih)) serta dokumen profil dan visi-misi dari portal pemilu. Komponen ini memastikan sistem selalu memiliki data terkini tanpa intervensi manual.
2. Modul Pemrosesan Cerdas (*Intelligent Processing Module*): Modul ini merupakan inti kecerdasan sistem yang mengatasi masalah data tidak terstruktur (M-02) dan ketiadaan data *voting* (M-04).
  - (a) *Document Parser*: Mengubah format dokumen PDF atau pindaian menjadi teks mentah yang dapat dibaca mesin.
  - (b) *NLP Engine*: Melakukan ekstraksi entitas bernama (*Named Entity Recognition*) untuk mengenali aktor (nama anggota, fraksi) dan objek (RUU, topik). Selanjutnya, komponen *Event Extraction* mengidentifikasi tindakan spesifik (misalnya: “menyampaikan pandangan”, “menginterupsi”, “menyetujui”) dari narasi risalah rapat untuk membentuk *event log* terstruktur.
  - (c) *Semantic Linker*: Menjawab masalah diskoneksi informasi (M-03) dengan menghitung skor kesamaan semantik antara teks janji kampanye dan teks aktivitas legislatif. Komponen ini menciptakan tautan logis (*logical links*) yang memungkinkan pengguna melihat hubungan langsung antara janji dan realisasi.
3. Modul Penyimpanan Terstruktur (*Structured Storage*): Hasil pemrosesan disimpan dalam basis data relasional yang menyimpan profil anggota, *event log* aktivitas yang telah diekstraksi, serta metadata dokumen sumber. Struktur data ini memungkinkan kueri yang kompleks dan cepat, mendukung kebutuhan skalabilitas sistem.
4. Modul Presentasi (*Presentation Layer*): Lapisan ini menyediakan antarmuka berbasis web (*dashboard*) bagi pengguna akhir. Modul ini memvisualisasikan data terstruktur menjadi profil anggota, linimasa aktivitas interaktif, dan grafik komparasi janji-realisasi, serta menyediakan fitur pencarian dan filter data. Antarmuka ini dirancang untuk menyembunyikan kompleksitas teknis

di belakang layar, sehingga memberikan pengalaman pengguna yang sederhana dan intuitif.



## **BAB V**

### **RENCANA SELANJUTNYA**

Bab ini menguraikan langkah-langkah strategis yang akan ditempuh untuk merealisasikan desain konsep *Legislative Activity Tracker* (LAT) menjadi sistem yang fungsional. Rencana ini mencakup tahapan implementasi teknis mulai dari akuisisi data hingga pengembangan antarmuka pengguna, metode evaluasi untuk mengukur kinerja sistem, serta analisis risiko guna mengantisipasi potensi hambatan selama proses pengembangan.

#### **V.1 Rencana Implementasi**

Rencana implementasi menjabarkan peta jalan teknis (*technical roadmap*) yang terstruktur untuk membangun komponen-komponen sistem LAT. Proses ini dirancang secara bertahap untuk memastikan setiap modul mulai dari *backend* pemrosesan data hingga *frontend* visualisasi dapat terintegrasi dengan baik dan memenuhi kebutuhan fungsional yang telah ditetapkan.

##### **V.1.1 Langkah-Langkah Implementasi**

Proses pengembangan sistem LAT dibagi menjadi empat fase utama yang dilakukan secara sekuensial, dengan fokus pada transformasi data mentah menjadi informasi yang dapat diakses publik.

###### **1. Fase 1: Akuisisi Data (*Data Acquisition*)**

Fase ini berfokus pada pembangunan infrastruktur pengumpulan data otomatis. Langkah-langkah kuncinya meliputi:

- (a) Pengembangan skrip *web crawler* khusus untuk memindai dan mengunduh dokumen risalah rapat, agenda, dan legislasi dari subdomain situs DPR RI (`dpr.go.id`, `berkas.dpr.go.id`).
- (b) Pengembangan skrip untuk mengumpulkan data profil anggota dan dokumen visi-misi dari portal KPU atau sumber data pemilu terverifikasi.

- (c) Penyiapan repositori penyimpanan data mentah (*raw data storage*) untuk menampung fail PDF dan HTML hasil unduhan.

## 2. Fase 2: Pemrosesan dan Ekstraksi Teks (*NLP Backend*)

Fase ini merupakan inti dari kecerdasan sistem, di mana dokumen tidak terstruktur diolah menjadi data terstruktur.

- (a) Implementasi modul *PDF Parser* dan OCR (*Optical Character Recognition*) untuk mengekstrak teks dari dokumen digital maupun hasil pindai.
- (b) Pelatihan dan penyesuaian (*fine-tuning*) model *Named Entity Recognition* (NER) untuk mendeteksi entitas spesifik seperti Nama Anggota, Fraksi, Komisi, dan Topik RUU.
- (c) Pengembangan algoritma *Event Extraction* untuk mengidentifikasi aktivitas legislator (misalnya: bertanya, mengusulkan, menyetujui) dari narasi risalah.
- (d) Implementasi modul *Semantic Linker* menggunakan model bahasa (seperti IndoBERT) untuk menghitung skor relevansi antara teks janji kampanye dan teks aktivitas.

## 3. Fase 3: Pengembangan Basis Data dan API

Fase ini bertujuan menyediakan akses data yang efisien bagi aplikasi pengguna.

- (a) Perancangan skema basis data relasional untuk menyimpan tabel Anggota, Janji, *Event Log*, dan relasinya.
- (b) Pengembangan *RESTful API* sebagai jembatan komunikasi data, menyediakan *endpoint* untuk pencarian profil, riwayat aktivitas, dan statistik komparasi.

## 4. Fase 4: Pengembangan Antarmuka Pengguna (*Frontend*)

Fase terakhir adalah pembangunan *dashboard* visualisasi.

- (a) Desain antarmuka web (*UI/UX Design*) yang intuitif untuk fitur pencarian dan profil anggota.
- (b) Implementasi komponen visualisasi interaktif, seperti linimasa aktivitas (*timeline*) dan grafik batang untuk skor kesesuaian janji-realisisasi.
- (c) Integrasi *frontend* dengan API untuk menampilkan data secara dinamis.

### V.1.2 Kebutuhan Alat dan Lingkungan Pengembangan

Pengembangan sistem LAT membutuhkan serangkaian perangkat lunak, pustaka (*libraries*), dan lingkungan infrastruktur yang spesifik untuk mendukung proses *crawling*, analisis teks berbasis AI, hingga penyajian data web. Rincian kebutuhan alat dan teknologi yang akan digunakan disajikan pada Tabel V.1.

Tabel V.1 Daftar Kebutuhan Alat dan Teknologi Pengembangan

Kategori	Nama Alat/Teknologi	Fungsi dan Kegunaan
Bahasa Pemrograman	Python	Pengembangan modul <i>backend</i> , skrip <i>crawler</i> , dan pemrosesan NLP.
<i>Web Framework</i>	Flask / FastAPI	Membangun <i>RESTful API</i> untuk melayani permintaan data dari <i>frontend</i> .
<i>Frontend Library</i>	React.js / Vue.js	Membangun antarmuka pengguna ( <i>dashboard</i> ) yang interaktif dan responsif.
Basis Data	PostgreSQL	Menyimpan data terstruktur ( <i>Event Log</i> , Profil, Janji) dengan relasi yang kompleks.
<i>Scraping Tools</i>	Scrapy / Selenium	Kerangka kerja untuk melakukan akuisisi data otomatis dari situs web dinamis.
Pustaka NLP / AI	HuggingFace Transformers, PyTorch, IndoBERT	Menyediakan model pra-latih dan kerangka kerja untuk tugas NER, klasifikasi teks, dan analisis semantik.
OCR	Tesseract / pdfPlumber	Mengekstrak teks dari dokumen PDF berbasis gambar atau teks.
Lingkungan Pengembangan	VS Code, Google Colab	Editor kode dan lingkungan komputasi awan untuk pelatihan model AI.

## V.2 Rencana Evaluasi

Evaluasi merupakan tahap krusial untuk memastikan bahwa sistem LAT yang dikembangkan tidak hanya berfungsi secara teknis, tetapi juga mampu menghasilkan luaran informasi yang akurat dan dapat dipercaya. Rencana evaluasi ini mencakup metodologi pengujian yang akan diterapkan pada setiap komponen sistem, serta penetapan indikator atau kriteria keberhasilan kuantitatif sebagai tolok ukur pencapaian tujuan penelitian.

### V.2.1 Metode Pengujian

Pengujian sistem akan dilakukan melalui tiga pendekatan utama untuk memverifikasi kinerja modul kecerdasan buatan, fungsionalitas aplikasi, dan integrasi sistem secara keseluruhan.

#### 1. Pengujian Akurasi Model (*NLP Model Testing*)

Pengujian ini bertujuan mengukur kinerja modul ekstraksi informasi.

(a) Menggunakan *Ground Truth Dataset*, yaitu sekumpulan dokumen risa-

lah rapat yang telah dianotasi secara manual oleh manusia sebagai kunci jawaban.

- (b) Menghitung metrik evaluasi standar seperti *Precision*, *Recall*, dan *F1-Score* untuk tugas pengenalan entitas (NER) dan klasifikasi aktivitas.

## 2. Pengujian Fungsional (*Black Box Testing*)

Pengujian ini berfokus pada validasi fitur dari sudut pandang pengguna akhir.

- (a) Menguji skenario penggunaan utama, seperti pencarian profil anggota, pemfilteran aktivitas berdasarkan topik RUU, dan navigasi antarmuka.
- (b) Memastikan bahwa *dashboard* menampilkan data yang sesuai dengan kueri yang dimasukkan tanpa kesalahan logika.

## 3. Pengujian Integrasi (*End-to-End Testing*)

Pengujian ini memverifikasi aliran data dari hulu ke hilir.

- (a) Memastikan dokumen baru yang diambil oleh *crawler* berhasil diproses oleh modul NLP, disimpan ke basis data, dan muncul di *dashboard* secara otomatis tanpa intervensi manual.

### V.2.2 Indikator Keberhasilan

Keberhasilan pengembangan sistem LAT diukur berdasarkan pencapaian target kinerja kuantitatif pada aspek akurasi data, performa sistem, dan cakupan informasi. Rincian kriteria keberhasilan tersebut disajikan pada Tabel V.2.

Tabel V.2 Kriteria dan Indikator Keberhasilan Sistem

Aspek Pengujian	Parameter	Target Capaian
Akurasi NLP	Skor F1 untuk Ekstraksi Entitas Utama (Nama Anggota, Partai)	> 75%
	Skor F1 untuk Klasifikasi Jenis Aktivitas	> 70%
Kualitas Penautan	Relevansi Skor Kesamaan Semantik ( <i>Semantic Similarity</i> )	Nilai korelasi positif yang signifikan pada pasangan janji-aktivitas yang valid secara manual (> 0.7)
Performa Sistem	Waktu Respons Pencarian ( <i>Query Response Time</i> )	< 3 detik untuk kueri standar
Cakupan Data	Tingkat Keberhasilan Akuisisi Dokumen ( <i>Crawling Success Rate</i> )	Mampu mengunduh minimal 80% dokumen publik yang tersedia di situs target

### V.3 Analisis Risiko

Dalam pengembangan sistem perangkat lunak yang melibatkan integrasi data eksternal dan teknologi kecerdasan buatan, potensi risiko teknis maupun non-teknis tidak dapat dihindari. Analisis risiko ini bertujuan untuk mengidentifikasi berbagai hambatan yang mungkin muncul selama siklus pengembangan sistem LAT, serta merumuskan strategi mitigasi yang konkret untuk meminimalkan dampak negatif terhadap keberhasilan proyek.

#### V.3.1 Identifikasi Risiko

Risiko pengembangan dikategorikan berdasarkan sumber masalahnya, yaitu risiko teknis, risiko data, dan risiko sumber daya. Tabel V.3 menjabarkan potensi masalah yang telah diidentifikasi beserta tingkat dampaknya terhadap proyek.

Tabel V.3 Identifikasi Risiko Pengembangan Sistem

Kategori Risiko	Deskripsi Risiko	Dampak Potensial
Risiko Teknis	Perubahan struktur HTML pada situs sumber data ( <code>dpr.go.id</code> ) secara tiba-tiba.	Kegagalan modul <i>crawler</i> dalam mengunduh dokumen baru, menyebabkan data sistem tidak mutakhir.
	Kualitas dokumen sumber PDF yang sangat buruk (hasil pindai miring, buram, atau tulisan tangan).	Kegagalan modul OCR dan NLP dalam mengekstrak teks, mengakibatkan data aktivitas hilang atau tidak lengkap.
Risiko Data	Ambiguitas nama anggota dewan (banyak nama yang mirip atau sama).	Kesalahan atribusi aktivitas ke profil anggota yang salah, menurunkan akurasi data profil.
	Teks janji kampanye yang terlalu umum atau abstrak.	Kesulitan modul semantik dalam menemukan tautan yang relevan dengan aktivitas legislatif yang spesifik.
Risiko Sumber Daya	Keterbatasan sumber daya komputasi (GPU/RAM) untuk menjalankan model NLP yang berat.	Proses pemrosesan data menjadi sangat lambat atau kegagalan <i>deployment</i> model pada server produksi.

### V.3.2 Strategi Mitigasi Risiko

Berdasarkan risiko yang telah diidentifikasi, disusun strategi mitigasi untuk mencegah atau menangani masalah tersebut jika terjadi. Tabel V.4 merinci langkah-langkah penanganan untuk setiap risiko.

Tabel V.4 Strategi Mitigasi Risiko

Risiko Teridentifikasi	Strategi Mitigasi
Perubahan struktur HTML situs sumber	Merancang arsitektur <i>crawler</i> yang modular sehingga penyesuaian selektor HTML dapat dilakukan dengan cepat tanpa mengubah logika utama, serta menerapkan pemantauan ( <i>monitoring</i> ) log galat secara berkala.
Kualitas dokumen PDF buruk	Mengintegrasikan layanan OCR berbasis awan ( <i>cloud-based OCR</i> ) yang lebih canggih untuk dokumen sulit, atau menerapkan mekanisme penandaan manual ( <i>human-in-the-loop</i> ) untuk dokumen yang gagal diproses mesin.
Ambiguitas nama anggota	Menerapkan teknik <i>Entity Disambiguation</i> dengan memanfaatkan atribut tambahan seperti Partai Asal dan Daerah Pemilihan (Dapil) sebagai konteks pembeda.
Janji kampanye abstrak	Menggunakan model <i>Sentence Transformer</i> yang dilatih pada korpus hukum/politik Indonesia untuk menangkap makna kontekstual yang lebih luas, bukan sekadar pencocokan kata kunci.
Keterbatasan sumber daya komputasi	Menggunakan varian model NLP yang lebih ringan (teknik <i>Knowledge Distillation</i> ) atau menerapkan mekanisme pemrosesan data secara bertahap ( <i>batch processing</i> ) pada jam sepi trafik.

## DAFTAR PUSTAKA

- Abbasi, Alireza, Sedigheh Rashidi, dan Abdolreza Ahmadi. 2021. "Keyword Weight Optimization Using Gradient Strategies in Focused Crawling". *Pattern Recognition Letters*, <https://www.sciencedirect.com/science/article/abs/pii/S0167865520304335>.
- Alanazi, A.H., dkk. 2022. "Machine Learning and Deep Learning-based Natural Language Processing for Auto-Vetting Medical Referrals". *International Journal of Medical Informatics*, <https://www.sciencedirect.com/science/article/pii/S2352914822001071>.
- Alzu'bi, A., dkk. 2025. "Cyberattack Event Logs Classification Using Deep Learning with Semantic Feature Analysis". *Computers & Security*, <https://www.sciencedirect.com/science/article/abs/pii/S0167404824005285>.
- Bántay, Lilla, dkk. 2023. "Frequent Pattern Mining-Based Log File Partition for Process Mining". *Engineering Applications of Artificial Intelligence*, <https://www.sciencedirect.com/science/article/pii/S0952197623004050>.
- Begany, Gregory M. 2024. "Open Government Data Initiatives as Agents of Digital Transformation in the Public Sector: Exploring the Extent of Use among Early Adopters". *Government Information Quarterly*, <https://www.sciencedirect.com/science/article/abs/pii/S0740624X24000479>.
- Berkovitz, S., dkk. 2023. "An Open Framework for Analyzing Public Parliaments Data". *Journal of Big Data*, <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-023-00831-3>.
- Bordignon, Massimo, dkk. 2020. "Rules Transparency and Political Accountability". *Journal of Public Economics*, <https://www.sciencedirect.com/science/article/abs/pii/S0047272700001092>.

- Breuer, A., dkk. 2023. "Integrated Policymaking: Institutional Designs for Digital Governance in Indonesia". *World Development*, <https://www.sciencedirect.com/science/article/pii/S0305750X23001353>.
- Chen, M., dkk. 2023. "Determinants of Open Government Data Usage: Integrating Trust Theory and Social Cognitive Theory". *Government Information Quarterly*, <https://www.sciencedirect.com/science/article/abs/pii/S0740624X23000576>.
- Dhanith, P.R.J., dkk. 2024. "Weakly Supervised Learning for an Effective Focused Web Crawler". *Engineering Applications of Artificial Intelligence*, <https://www.sciencedirect.com/science/article/abs/pii/S0952197624001027>.
- Ekakristi, A.S., dkk. 2025. "Intermediate-Task Transfer Learning for Indonesian NLP Tasks". *Natural Language Processing Journal*, <https://www.sciencedirect.com/science/article/pii/S2949719125000378>.
- Erjavec, T., dkk. 2023. "The B Corpora of Parliamentary Proceedings". *Language Resources and Evaluation*, <https://link.springer.com/article/10.1007/s10579-021-09574-0>.
- Ferry, L., dkk. 2024. "Democracy, Accountability, Accounting and Trust: A Critical Reflection". *Critical Perspectives on Accounting*, <https://www.sciencedirect.com/science/article/pii/S1045235424000376>.
- Frish, Sergey, dkk. 2025. "Enhancing Consistency of AHP-Based Expert Judgments Through an Iterative Feedback Mechanism". *MethodsX* 14:103101. <https://doi.org/10.1016/j.mex.2025.103101>. <https://www.sciencedirect.com/science/article/pii/S2215016125001876>.
- GovTrack.us. No date. "About". Diakses terakhir. <https://substack.govtrack.us/about>.
- Goyal, N., dkk. 2024. "Named Entity Recognition and Relationship Extraction for Biomedical Texts: A Survey". *Neurocomputing*, <https://www.sciencedirect.com/science/article/abs/pii/S0925231224019428>.
- Guillén-Mena, Vanessa, dkk. 2023. "Lessons Learned from a Study Based on the AHP Method for Prioritizing Sustainable Neighborhood Criteria". *MethodsX* 11:102436. <https://doi.org/10.1016/j.mex.2023.102436>. <https://www.sciencedirect.com/science/article/pii/S2215016123004363>.



- Guo, Jiafeng, dkk. 2020. "A Deep Look into Neural Ranking Models for Information Retrieval". *Information Processing & Management*, <https://www.sciencedirect.com/science/article/abs/pii/S0306457319302390>.
- Halimatusa'diyah, Iim, dan A. N. Jannah. 2025. "Understanding Hidden Layers in Political Participation: Women's Representation in Indonesia's Election Management Bodies". *Women's Studies International Forum*, <https://www.sciencedirect.com/science/article/abs/pii/S0277539525000901>.
- Himler, P., dkk. 2024. "Anomaly Detection in Log-Event Sequences: A Federated Deep Learning Approach". *Machine Learning with Applications*, <https://www.sciencedirect.com/science/article/pii/S2666827024000306>.
- Hong, Seok-Joon. 2024. "Political Determinants of Government Transparency: Scope, Time, and Quality of Open Government Data". *Policy Studies Journal*, <https://onlinelibrary.wiley.com/doi/full/10.1111/polp.12607>.
- Huang, Zhao, dkk. 2022. "A Semantic Matching Approach Addressing Multidimensional Service Representation". *Expert Systems with Applications*, <https://www.sciencedirect.com/science/article/abs/pii/S0957417422015597>.
- Ibrohim, M.O., dkk. 2023. "Hate Speech and Abusive Language Detection in Indonesian Social Media". *Information Processing in Agriculture*, <https://www.sciencedirect.com/science/article/pii/S2405844023058553>.
- Ignasius, A., dkk. 2023. "Image Pre-Processing Effect on OCR's Performance for Scene Images". *Procedia Computer Science*, <https://www.sciencedirect.com/science/article/pii/S1877050923017660>.
- Islam, S., dkk. 2024. "A Comprehensive Survey on Applications of Transformers for Deep Learning Tasks". *Expert Systems with Applications*, <https://www.sciencedirect.com/science/article/abs/pii/S0957417423031688>.
- JariUngu. No date. "Tentang Kami". [https://jariungu.com/tentang\\_kami.php](https://jariungu.com/tentang_kami.php).
- Jehangir, B., dkk. 2023. "A Survey on Named Entity Recognition: Datasets, Tools, and Methodologies". *Natural Language Processing Journal*, <https://www.sciencedirect.com/science/article/pii/S2949719123000146>.

- Kempeneer, Sam, Ali Pirannejad, dan Jeroen Wolswinkel. 2023. "Open Government Data from a Legal Perspective: An AI-Driven Systematic Literature Review". *Government Information Quarterly*, <https://www.sciencedirect.com/science/article/pii/S0740624X23000230>.
- Koto, Fajri, Afshin Rahimi, Jey Han Lau, dan Timothy Baldwin. 2020. "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP". Dalam *Proceedings of COLING 2020*. <https://aclanthology.org/2020.coling-main.66/>.
- Kreps, S., dkk. 2023. "Can AI Communication Tools Increase Legislative Responsiveness?" *Government Information Quarterly*, <https://www.sciencedirect.com/science/article/abs/pii/S0740624X23000291>.
- Kriswardhana, Wahyu, dkk. 2025. "Analytic Hierarchy Process in Transportation Decision-Making: A Comprehensive Bibliometric and Systematic Review". *Expert Systems with Applications* 255:124583. <https://doi.org/10.1016/j.eswa.2024.124583>. <https://www.sciencedirect.com/science/article/pii/S0957417424023583>.
- Kumar, Anoop, Surendra Kushwaha, Dipendra Pandey, dkk. 2023. "Analytical Hierarchy Process for Sustainable Agriculture: An Overview". *MethodsX* 10:101954. <https://doi.org/10.1016/j.mex.2022.101954>. <https://www.sciencedirect.com/science/article/pii/S2215016122003260>.
- Landauer, Max, dkk. 2023. "Deep Learning for Anomaly Detection in Log Data: A Survey". *Machine Learning with Applications*, <https://www.sciencedirect.com/science/article/pii/S2666827023000233>.
- Leal, José Eugenio. 2020. "AHP-express: A Simplified Version of the Analytical Hierarchy Process Method". *MethodsX* 7:100748. <https://doi.org/10.1016/j.mex.2019.11.021>. <https://www.sciencedirect.com/science/article/pii/S2215016119303243>.
- Li, Jianhua, dkk. 2025. "Extracting Semantic Link Network of Words from Text for Semantics-Based Applications". *Expert Systems with Applications*, <https://www.sciencedirect.com/science/article/abs/pii/S0957417424026356>.
- Li, Y., dkk. 2023. "Generative Retrieval for Conversational Question Answering". *Information Processing & Management*, <https://www.sciencedirect.com/science/article/abs/pii/S0306457323002121>.

- Liang, Y. 2025. "Determinants of Open Government Data Continuance Usage and Value Creation: A Self-Regulation Framework Analysis". *Government Information Quarterly*, <https://www.sciencedirect.com/science/article/abs/pii/S0740624X25000164>.
- Library of Congress. No date. "Congress.gov". Portal resmi legislasi federal Amerika. <https://dl.acm.org/doi/full/10.1145/3584741>.
- . No date. "Using Congress.gov Data Offsite". Panduan akses data dan bulk data. <https://www.congress.gov/help/using-data-offsite>.
- Lin, J., dkk. 2025. "Interpretable Named Entity Recognition via Integrating Deep Learning with Logical Rules". *Expert Systems with Applications*, <https://www.sciencedirect.com/science/article/abs/pii/S0957417425029525>.
- Lin, T., dkk. 2022. "A Survey of Transformers". *AI Open*, <https://www.sciencedirect.com/science/article/pii/S2666651022000146>.
- Liu, Jingyi, Tanushree Mitra, dan Chris North. 2023. "A Survey on Event-Based News Narrative Extraction". *ACM Computing Surveys*, <https://dl.acm.org/doi/10.1145/3584741>.
- Liu, S., dkk. 2025. "Improved Relation Extraction Through Key Phrase Identification via Dependency Trees". *Computer Speech & Language*, <https://www.sciencedirect.com/science/article/abs/pii/S0885230824000895>.
- Lu, Chao, dkk. 2025. "Measuring Research Difficulty in Academic Papers: A Case Study in NLP". *Journal of Informetrics*, <https://www.sciencedirect.com/science/article/pii/S2694610625000396>.
- Ma, Qianli, dkk. 2023. "Sequence Labeling with MLTA: Multi-Level Topic-Aware Mechanism". *Information Sciences*, <https://www.sciencedirect.com/science/article/abs/pii/S0020025523005030>.
- Nabiilah, G.Z., dkk. 2023. "BERT Base Model for Toxic Comment Analysis on Indonesian Social Media Dataset". *Procedia Computer Science*, <https://www.sciencedirect.com/science/article/pii/S1877050922022669>.
- Nai, R., dkk. 2025. "Leveraging Process Mining and Event Log Enrichment in Legal Text Analysis". *Information Systems*, <https://www.sciencedirect.com/science/article/pii/S2212473X25000173>.

- Nguyen, H.A.T., dkk. 2025. “Developing an Automated Framework for Eco-Label Information Categorization Using Web Crawling and Natural Language Processing Techniques”. *Expert Systems with Applications*, <https://www.sciencedirect.com/science/article/abs/pii/S0957417425013107>.
- Rani, R., dkk. 2021. “A Weighted Word Embedding Based Approach for Extractive Text Summarization”. *Expert Systems with Applications*, <https://www.sciencedirect.com/science/article/abs/pii/S0957417421012264>.
- Rienks, H. 2023. “Corruption, Scandals and Incompetence: Do Voters Care?” *European Journal of Political Economy*, <https://www.sciencedirect.com/science/article/pii/S017626802300085X>.
- Ruas, Pedro, dkk. 2022. “NILINKER: Attention-Based Approach to NIL Entity Linking”. *Journal of Biomedical Informatics*, <https://www.sciencedirect.com/science/article/pii/S1532046422001526>.
- Sekretaris Jenderal DPR RI. 2020. *Peraturan Sekretaris Jenderal DPR RI No. 13 Tahun 2020: Pedoman Pelaksanaan SILEG*. Deskripsi SILEG pada draf.
- Sheppard, G. 2023. “Transparency Trade-offs in the Operation of National Public Private Partnership Units”. *Journal of Accounting and Public Policy*, <https://www.sciencedirect.com/science/article/abs/pii/S0278425423000601>.
- Silva, L., dkk. 2024. “Improving Dense Retrieval Models with LLM Augmented Data for Dataset Search”. *Computer Speech & Language*, <https://www.sciencedirect.com/science/article/abs/pii/S0950705124003757>.
- Wang, Fang, dkk. 2023. “Paths to Open Government Data Reuse: A Three-Dimensional Framework of Information Need, Data and Government Preparation”. *Government Information Quarterly*, <https://www.sciencedirect.com/science/article/abs/pii/S0378720623001271>.
- Wang, Xiao, dkk. 2022. “Improving Zero-Shot Retrieval Using Dense External Expansion”. *Information Processing & Management*, <https://www.sciencedirect.com/science/article/pii/S0306457322001364>.
- Wilie, B., dkk. 2020. “IndoNLU Benchmark and Resources for Evaluating Indonesian Natural Language Understanding”. Dalam *Proceedings of COLING 2020*. <https://aclanthology.org/2020.coling-main.557/>.

- Wirtz, Bernd W., Jan C. Weyerer, dan Michael Rosch. 2022. "Open Government Data: A Systematic Literature Review of Empirical Research". *Electronic Markets*, <https://link.springer.com/article/10.1007/s12525-022-00582-8>.
- Xu, J., dkk. 2025. "LAAP: Learning the Argument of An Entity with Event Prompts for Document-Level Event Extraction". *Neurocomputing*, <https://www.sciencedirect.com/science/article/abs/pii/S0925231224013559>.

## LAMPIRAN 1: Detail Kebutuhan Fungsional

Tabel 5 Analisis Kebutuhan Fungsional Sistem

ID	Functional Requirement	Deskripsi
FR-01	Sistem mampu mengakuisisi dan memperbarui dokumen legislatif secara otomatis dari berbagai sumber resmi ( <i>web crawling</i> ).	<p><b>Tujuan:</b> Mengumpulkan data mentah (dokumen) dari sumber yang terfragmentasi ke dalam satu repositori terpusat untuk diproses lebih lanjut.</p> <p><b>Masukan:</b></p> <ul style="list-style-type: none"><li>• Daftar URL target (dpr.go.id, berkas.dpr, dll).</li><li>• Parameter <i>crawling</i> (frekuensi, tipe file).</li></ul> <p><b>Proses:</b></p> <ul style="list-style-type: none"><li>• Sistem menjalankan <i>crawler</i> sesuai jadwal.</li><li>• Sistem mengunduh dokumen baru.</li><li>• Sistem menyimpan metadata sumber.</li></ul> <p><b>Keluaran:</b></p> <ul style="list-style-type: none"><li>• Koleksi dokumen PDF/HTML mentah di penyimpanan lokal sistem.</li></ul>

Tabel 5 – Lanjutan dari halaman sebelumnya

ID	Functional Requirement	Deskripsi
FR-02	Sistem mampu mengekstrak struktur peristiwa ( <i>event log</i> ) dari dokumen tidak terstruktur menggunakan teknik NLP dan IE.	<p><b>Tujuan:</b> Mengubah teks naratif (risalah/laporan) menjadi data terstruktur yang mencatat “siapa”, “apa”, “kapan”, dan “di mana”.</p> <p><b>Masukan:</b></p> <ul style="list-style-type: none"> <li>Dokumen PDF hasil akuisisi (FR-01).</li> <li>Model NER (<i>Named Entity Recognition</i>) dan aturan ekstraksi.</li> </ul> <p><b>Proses:</b></p> <ul style="list-style-type: none"> <li>Sistem melakukan <i>parsing</i> teks dan OCR (jika perlu).</li> <li>Sistem mengenali entitas (Anggota, Fraksi, RUU).</li> <li>Sistem mendeteksi aksi/peristiwa.</li> </ul> <p><b>Keluaran:</b></p> <ul style="list-style-type: none"> <li>Tabel basis data <i>Event Log (Legislative Activities)</i>.</li> </ul>
FR-03	Sistem mampu menghitung dan menautkan kesamaan semantik antara janji kampanye dengan aktivitas legislatif ( <i>semantic linking</i> ).	<p><b>Tujuan:</b> Menghubungkan dua <i>dataset</i> terpisah (Janji vs Aktivitas) berdasarkan kemiripan topik/konteks untuk analisis konsistensi.</p> <p><b>Masukan:</b></p> <ul style="list-style-type: none"> <li>Data Janji Kampanye (Visi-Misi).</li> <li>Data <i>Event Log</i> (FR-02).</li> </ul> <p><b>Proses:</b></p> <ul style="list-style-type: none"> <li>Sistem menghitung skor kesamaan (<i>similarity score</i>) antar teks.</li> <li>Sistem menetapkan <i>link</i> jika skor melebihi ambang batas.</li> </ul> <p><b>Keluaran:</b></p> <ul style="list-style-type: none"> <li>Tabel Relasi Janji-Aktivitas (<i>Linked Data</i>).</li> </ul>

Tabel 5 – Lanjutan dari halaman sebelumnya

ID	Functional Requirement	Deskripsi
FR-04	Sistem mampu menampilkan <i>dashboard</i> pencarian profil anggota dan riwayat aktivitasnya dengan fitur <i>filter</i> .	<p><b>Tujuan:</b> Memudahkan pengguna melihat rekam jejak legislator secara granular dan terorganisir.</p> <p><b>Masukan:</b></p> <ul style="list-style-type: none"> <li>• Kata kunci pencarian (Nama Anggota/Fraksi/Topik).</li> <li>• Parameter <i>filter</i> (Tahun, Komisi, Jenis Rapat).</li> </ul> <p><b>Proses:</b></p> <ul style="list-style-type: none"> <li>• Sistem menerima <i>query</i> dari pengguna.</li> <li>• Sistem mengambil data dari basis data <i>Event Log</i>.</li> </ul> <p><b>Keluaran:</b></p> <ul style="list-style-type: none"> <li>• Halaman Profil Anggota dengan linimasa aktivitas yang relevan.</li> </ul>
FR-05	Sistem mampu memvisualisasikan perbandingan komparatif ( <i>side-by-side</i> ) antara janji kampanye dan realisasi kinerja.	<p><b>Tujuan:</b> Menyajikan bukti empiris konsistensi politik secara intuitif kepada pengguna.</p> <p><b>Masukan:</b></p> <ul style="list-style-type: none"> <li>• ID Anggota yang dipilih.</li> <li>• Topik isu yang ingin dibandingkan.</li> </ul> <p><b>Proses:</b></p> <ul style="list-style-type: none"> <li>• Sistem mengambil data Janji dan Aktivitas yang sudah tertaut (FR-03).</li> <li>• Sistem menyandingkan keduanya dalam tampilan UI.</li> </ul> <p><b>Keluaran:</b></p> <ul style="list-style-type: none"> <li>• Visualisasi/Tabel Komparasi “Janji vs Realisasi”.</li> </ul>



## LAMPIRAN 2: Penjelasan Detail Alternatif Solusi

Tabel 6 Alternatif Solusi dan Masalah yang Ingin Diselesaikan

Kode Solusi	Alternatif Solusi	Masalah yang Ingin Diselesaikan
S01	<p><b>Pengembangan Portal Direktori Manual (<i>Manual Curation</i>)</b></p> <p>Membangun basis data profil anggota dan aktivitas yang diinput secara manual oleh tim admin/relawan berdasarkan pemantauan berita dan dokumen fisik, menyerupai model awal platform <i>civic tech</i> konvensional.</p>	<p><b>Masalah yang diselesaikan:</b></p> <p>M-03: Mencoba mengatasi diskoneksi informasi dengan menyajikan profil dan janji dalam satu halaman.</p> <p><b>Keterbatasan:</b></p> <p>Gagal mengatasi M-01 (fragmentasi data) dan M-02 (<i>event log</i>) secara <i>scalable</i> karena ketergantungan pada tenaga manusia yang lambat dan rentan <i>human error</i>. Tidak ada otomatisasi untuk volume data besar.</p>
S02	<p><b>Sistem Temu Kembali Dokumen Terpusat (<i>IR-only System</i>)</b></p> <p>Membangun mesin pencari (<i>search engine</i>) yang mengindeks seluruh dokumen PDF dari berbagai situs DPR (<i>crawling</i>) agar dapat dicari berdasarkan kata kunci, tanpa melakukan ekstraksi struktur data mendalam.</p>	<p><b>Masalah yang diselesaikan:</b></p> <p>M-01: Mengatasi masalah data tersebar dengan mengumpulkannya di satu repositori terpusat.</p> <p><b>Keterbatasan:</b></p> <p>Gagal mengatasi M-02 (<i>event log</i> terstruktur) dan M-04 (<i>voting data</i>). <i>Output</i> sistem masih berupa dokumen mentah (PDF), sehingga pengguna tetap harus membaca manual untuk menemukan “siapa melakukan apa”. Tidak ada analisis komparatif otomatis.</p>

*Bersambung ke halaman berikutnya*

Tabel 6 – Lanjutan dari halaman sebelumnya

Kode Solusi	Alternatif Solusi	Masalah yang Ingin Diselesaikan
S03	<p><b>Sistem Ekstraksi Informasi Statis (<i>IE-only on Local Corpus</i>)</b></p> <p>Fokus hanya pada pengembangan algoritma NLP/IE untuk mengekstrak data dari kumpulan dokumen yang sudah diunduh secara lokal (<i>offline dataset</i>), tanpa modul <i>crawling</i> atau pembaruan otomatis dari sumber web.</p>	<p><b>Masalah yang diselesaikan:</b></p> <p>M-02: Menjawab kebutuhan struktur data (<i>event log</i>) melalui <i>parsing</i> dokumen.</p> <p>M-04: Membantu inferensi posisi fraksi dari teks.</p> <p><b>Keterbatasan:</b></p> <p>Gagal mengatasi M-01 secara berkelanjutan karena tidak memiliki mekanisme <i>discovery</i> data baru dari web DPR yang dinamis. Sistem akan cepat usang (<i>obsolete</i>) tanpa <i>pipeline</i> akuisisi data.</p>
S04	<p><b>Arsitektur Pipeline Hibrida (<i>Integrated IR + IE Pipeline</i>)</b></p> <p>Mengintegrasikan modul <i>Information Retrieval</i> (untuk akuisisi dan manajemen korpus otomatis) dengan modul <i>Information Extraction</i> (untuk <i>parsing</i> NLP dan pembentukan <i>event log</i>), dilengkapi lapisan <i>Semantic Linking</i> untuk menghubungkan aktivitas dengan janji kampanye.</p>	<p><b>Masalah yang diselesaikan:</b></p> <ul style="list-style-type: none"> <li>• M-01: Menangani fragmentasi melalui <i>Focused Crawling</i> terpusat.</li> <li>• M-02: Menghasilkan <i>event log</i> terstruktur per anggota/fraksi melalui NLP.</li> <li>• M-03: Menghubungkan janji dan realisasi melalui analisis kesamaan semantik.</li> <li>• M-04: Memodelkan inferensi sikap politik fraksi secara sistematis.</li> </ul> <p>Solusi ini mencakup penanganan hulu (data mentah) hingga hilir (visualisasi <i>dashboard</i>).</p>

### LAMPIRAN 3: Rincian Analisis Penentuan Solusi

Tabel 7 Kriteria Matriks Perbandingan Berpasangan (Skala Saaty)

Nilai	Interpretasi
1	Kedua elemen sama pentingnya ( <i>Equal Importance</i> )
3	Salah satu elemen sedikit lebih penting ( <i>Moderate Importance</i> )
5	Salah satu elemen lebih penting ( <i>Strong Importance</i> )
7	Salah satu elemen jauh lebih penting ( <i>Very Strong Importance</i> )
9	Salah satu elemen mutlak lebih penting ( <i>Extreme Importance</i> )
2, 4, 6, 8	Nilai-nilai antara dua tingkat pertimbangan yang berdekatan

Tabel 8 Matriks Perbandingan Berpasangan Antar Kriteria

Kriteria	Efektivitas (K1)	Skalabilitas (K2)	Dampak (K3)	Kemudahan (K4)	Kualitas Data (K5)
Efektivitas (K1)	1	3	1	7	3
Skalabilitas (K2)	1/3	1	1/3	5	1
Dampak (K3)	1	3	1	7	3
Kemudahan (K4)	1/7	1/5	1/7	1	1/5
Kualitas Data (K5)	1/3	1	1/3	5	1
<b>Jumlah</b>	<b>2.81</b>	<b>8.20</b>	<b>2.81</b>	<b>25.00</b>	<b>8.20</b>

Tabel 9 Normalisasi Nilai AHP Kriteria

Kriteria	Efektivitas	Skalabilitas	Dampak	Kemudahan	Kualitas	Bobot Prioritas
Efektivitas	0.36	0.37	0.36	0.28	0.37	0.348
Skalabilitas	0.12	0.12	0.12	0.20	0.12	0.136
Dampak	0.36	0.37	0.36	0.28	0.37	0.348
Kemudahan	0.05	0.02	0.05	0.04	0.02	0.036
Kualitas	0.12	0.12	0.12	0.20	0.12	0.136

Tabel 10 Rata-Rata Nilai Kriteria (Bobot Prioritas)

Kriteria	Rata-rata Baris / Bobot Prioritas
Efektivitas	0.348
Dampak	0.348
Skalabilitas	0.136
Kualitas Data	0.136
Kemudahan	0.036

Tabel 11 Perbandingan Berpasangan Alternatif Terhadap Kriteria Efektivitas (K1)

Alternatif	S01 (Manual)	S02 (IR)	S03 (IE)	S04 (Hybrid)
S01	1	1/3	1/5	1/9
S02	3	1	1/3	1/7
S03	5	3	1	1/5
S04	9	7	5	1
<b>Jumlah</b>	<b>18.00</b>	<b>11.33</b>	<b>6.53</b>	<b>1.45</b>

Tabel 12 Normalisasi Matriks Alternatif Solusi (Kriteria Efektivitas)

Alternatif	S01	S02	S03	S04	Bobot Prioritas
S01	0.055	0.029	0.030	0.076	0.047
S02	0.166	0.088	0.050	0.096	0.100
S03	0.277	0.264	0.153	0.137	0.207
S04	0.500	0.617	0.765	0.689	0.642

Tabel 13 Rata-rata Nilai Alternatif Solusi (Kriteria Efektivitas)

Alternatif	Rata-rata (Bobot Lokal)
S01	0.047
S02	0.100
S03	0.207
S04	0.642