

Kelas LB95

Assurance of Learning

Presentation

**Trustify: Rangkuman Inovatif Berbasis Artificial
Intelligence untuk Meningkatkan Literasi Digital**

Presented by **Kelompok 1**

Kelompok 1

Dosen: Dr. Prabowo Wahyu Sudarno, S.Kom.

Raynaldo Winson Amadeus Kusuma

2802419370

Maurel Chairinniswah Yasmine

2802475320

Sabrina Arfanindia Devi

2802448755

Albert Wesley Dion TJA

2802493090

Hastuardy Pandu Mahendra

2802423506

Latar Belakang

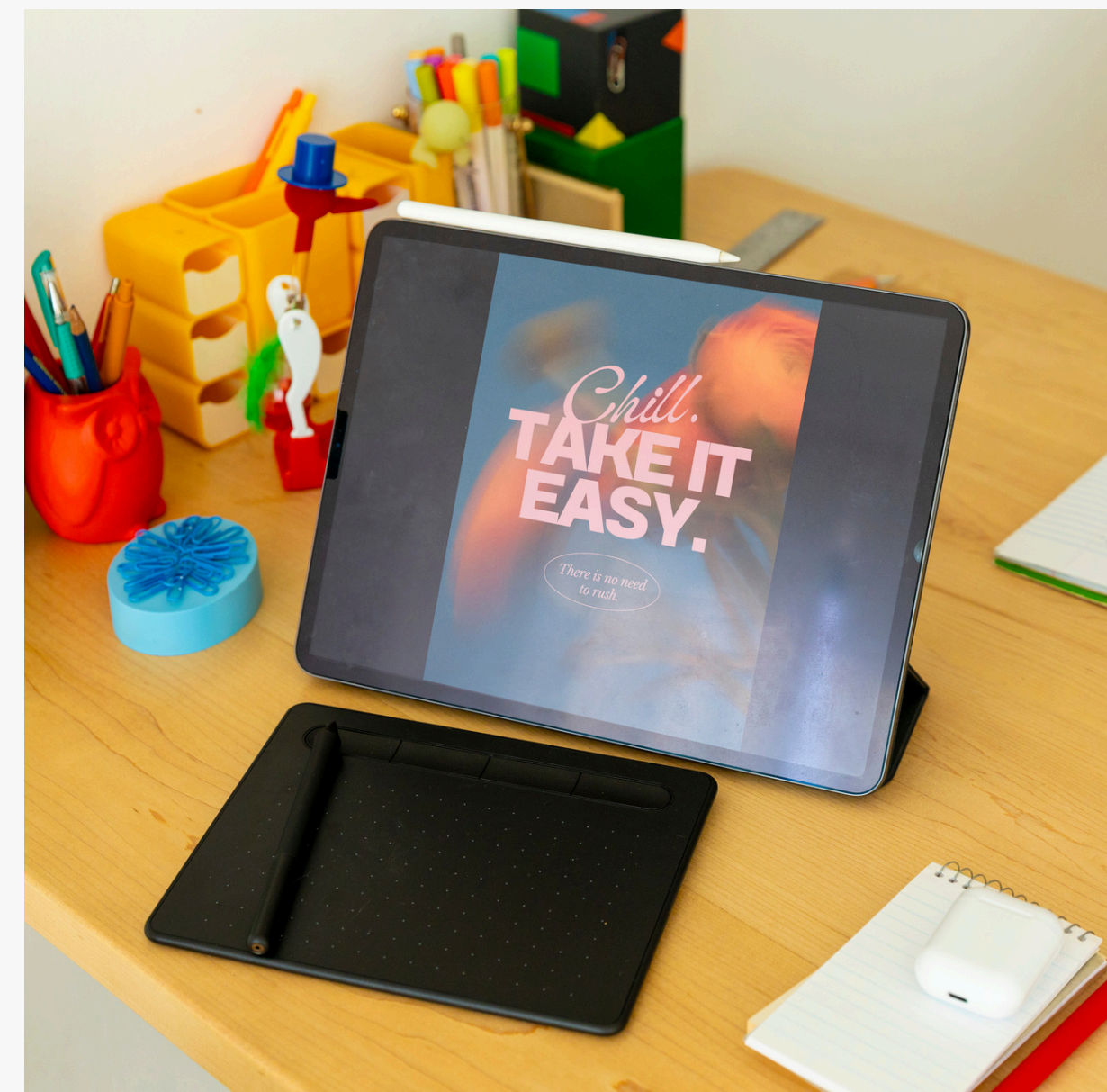
- Arus informasi digital yang sangat cepat menyebabkan information overload
- Pembaca sering melakukan skimming sehingga berisiko salah memahami isi berita
- Rendahnya literasi membaca dan pemahaman teks di Indonesia memperparah kondisi ini



SDG yang diambil

Tujuan dan Manfaat

- Mengembangkan aplikasi AI untuk merangkum berita berbahasa Indonesia secara otomatis
- Membantu pengguna memahami inti informasi tanpa kehilangan konteks penting
- Mengurangi risiko misinformasi akibat pemahaman yang dangkal
- Mendukung peningkatan literasi digital masyarakat di era informasi



METODE

1

Requirements

Kebutuhan aplikasi dianalisa untuk menentukan arsitektur model terbaik dan sumber dataset yang tepat untuk finetuning

(NLP-Summarization)

2

Memilih base model

Base model yang tepat ditentukan berdasarkan beberapa faktor pertimbangan. Arsitektur yang digunakan adalah model T5

3

Memilih dataset

Dataset yang bisa digunakan untuk finetuning harus mengandung teks asli berita beserta (human-generated) summary. Dataset dari Huggingface diutamakan untuk integrasi

6

Deployment

LORA hasil training di-backup menggunakan wandb.ai lalu disimpan dalam artifact ke Huggingface model repository

Proof-of-Concept dibuat menggunakan Gradio sebagai UI element dan di-hosting di infrastructure Huggingface. 8-bit quantization diterapkan untuk POC

5

Benchmarking

Dalam notebook yang sama, benchmarking menggunakan beberapa metric dilakukan pada hasil finetuning

(ROUGE & BERTScore)

4

Finetuning

Dengan memanfaatkan ecosystem Huggingface, finetuning dilakukan dengan strategi PEFT menggunakan teknik LoRA. Finetuning dilakukan dalam Google Colab


Model Selection


3-Coder ▾ 

Recommended Architecture for Indonesian Summarization:

T5 (Text-to-Text Transfer Transformer) - Primary Recommendation:

T5 is architecturally superior for summarization tasks due to several key advantages:

- **Unified text-to-text framework:** Treats all NLP tasks as sequence-to-sequence problems, making it naturally suited for summarization
- **Denoising pre-training objective:** Trains on corrupted text reconstruction, which develops strong generation capabilities
- **Superior long-form understanding:** Better handles document-level relationships compared to BERT2BERT
- **Available Indonesian variants:** Models like  "yahya/t5-base-indonesian-summarization" are

 Image Edit

 Web Dev

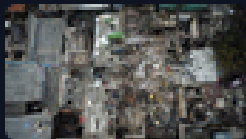
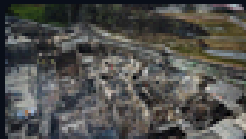
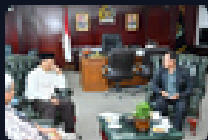
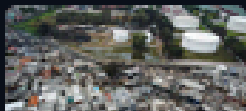
 Image Generation

 Video Generation

 Artifacts

Dataset

► The full dataset viewer is not available (click to read why). Only showing a preview of the rows.

id	source	title	image	url	content	date	
int64	string	string	string	string	string	string	
83	tempo	Depo Plumpang Terbakar,...	 https:// statik.tempo.co/.../1186060_720.jp...	https:// nasional.tempo.co...	TEMPO.CO, Jakarta - Anggota Komisi VII DPR...	2023-03-04 06:18:13+00	
84	tempo	Jokowi Perintahkan...	 https:// statik.tempo.co/.../1186059_720.jp...	https:// nasional.tempo.co...	TEMPO.CO, Jakarta - Presiden Joko Widodo...	2023-03-04 06:04:38+00	
85	tempo	HNW Mendukung Jamaah Umroh...	 https:// statik.tempo.co/.../1186020_720.jp...	https:// nasional.tempo.co...	INFO NASIONAL - Wakil Ketua MPR RI Dr. H. M...	2023-03-04 06:18:04+00	
86	tempo	Tim Dokkes		https://	TEMPO.CO, Jakarta - Tim	2023-03-04	

Dataset

<code>created_at</code> string	<code>updated_at</code> string	<code>summary</code> string
2023-03-04 07:03:39.039332	2023-03-04 07:03:39.039332	Anggota Komisi VII DPR RI...

LORA Training

```
1 lora_config = LoraConfig(  
2     r=8,                # Low rank for efficiency  
3     lora_alpha=32,       # Higher alpha for r=8  
4     lora_dropout=0.1,    # Moderate regularization  
5     bias="none",         # Preserve original bias behavior  
6     task_type="SEQ_2_SEQ_LM", # For T5 summarization  
7     target_modules=["q", "v", "k", "o", "DenseReluDense.wi", "DenseReluDense.wo"] # Key T5 layers  
8 )  
9  
10 model = get_peft_model(model, lora_config)
```

```
1 model.print_trainable_parameters()
```

```
trainable params: 3,244,032 || all params: 226,147,584 || trainable%: 1.4345
```

Benchmarking

```
1  # Print results
2  print(f"ROUGE-1: {rouge_scores['rouge1']:.4f}")
3  print(f"ROUGE-2: {rouge_scores['rouge2']:.4f}")
4  print(f"ROUGE-L: {rouge_scores['rougeL']:.4f}")
5  print(f"BERTScore F1: {sum(bertscore_scores['f1']) / len(bertscore_scores['f1']):.4f}")
```

```
ROUGE-1: 0.5161
ROUGE-2: 0.3154
ROUGE-L: 0.4207
BERTScore F1: 0.7955
```

Deployment

Indonesian Text Summarization

Enter Indonesian text to generate a summary using T5 model with PEFT adapters

Input Text

Enter Indonesian text here...

Generated Summary

Clear

Submit

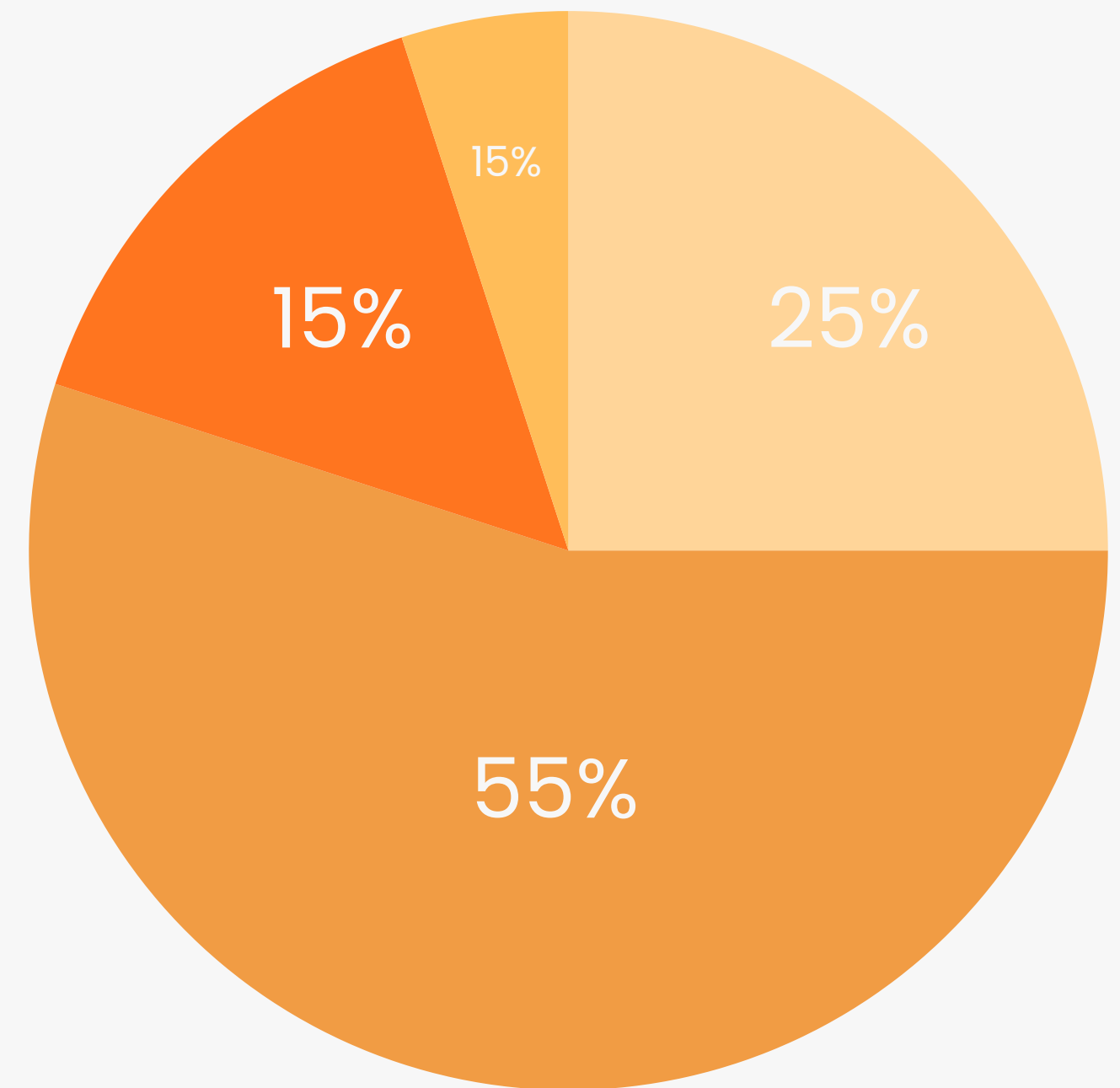
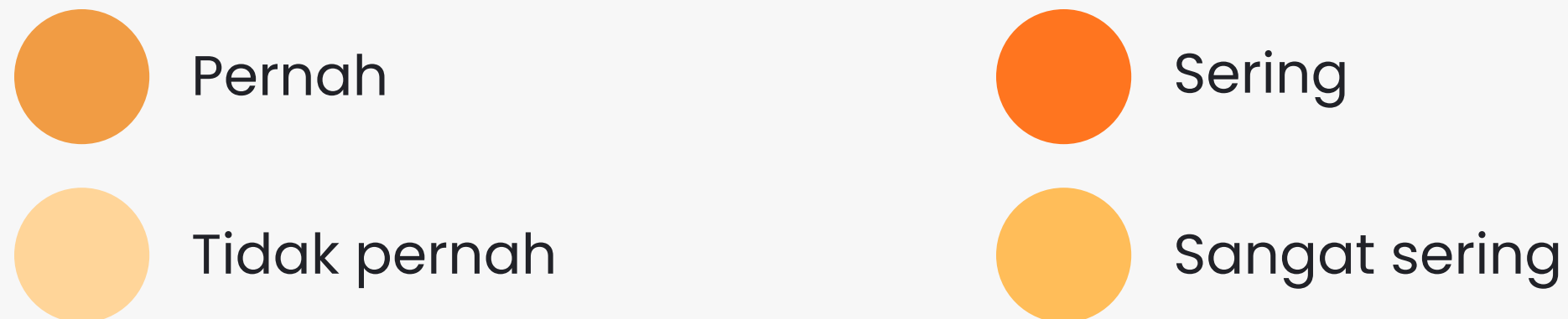
Model Loading

```
def load_model_once():  
    global model, tokenizer  
    if model is None:  
        hf_token = os.environ.get('HF_TOKEN')  
        login(token=hf_token)  
  
        # Load base model  
        base_model_name = "cahya/t5-base-indonesian-summarization-cased"  
        tokenizer = T5Tokenizer.from_pretrained(base_model_name)  
        base_model = T5ForConditionalGeneration.from_pretrained(  
            base_model_name,  
            load_in_8bit=True, # Quantize for CPU efficiency  
            device_map="auto"  
        )  
  
        model = PeftModel.from_pretrained(  
            base_model,  
            "reydeuss/trustify-t5-adapter",  
        )  
    return model, tokenizer
```

Hasil Survey

Survey dilakukan melalui *Google Forms* untuk mengetahui sejauh mana pengguna memahami bacaan dari teks yang sedang dibaca.



Jumlah Responden: 20


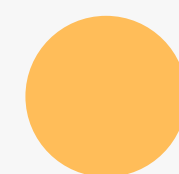


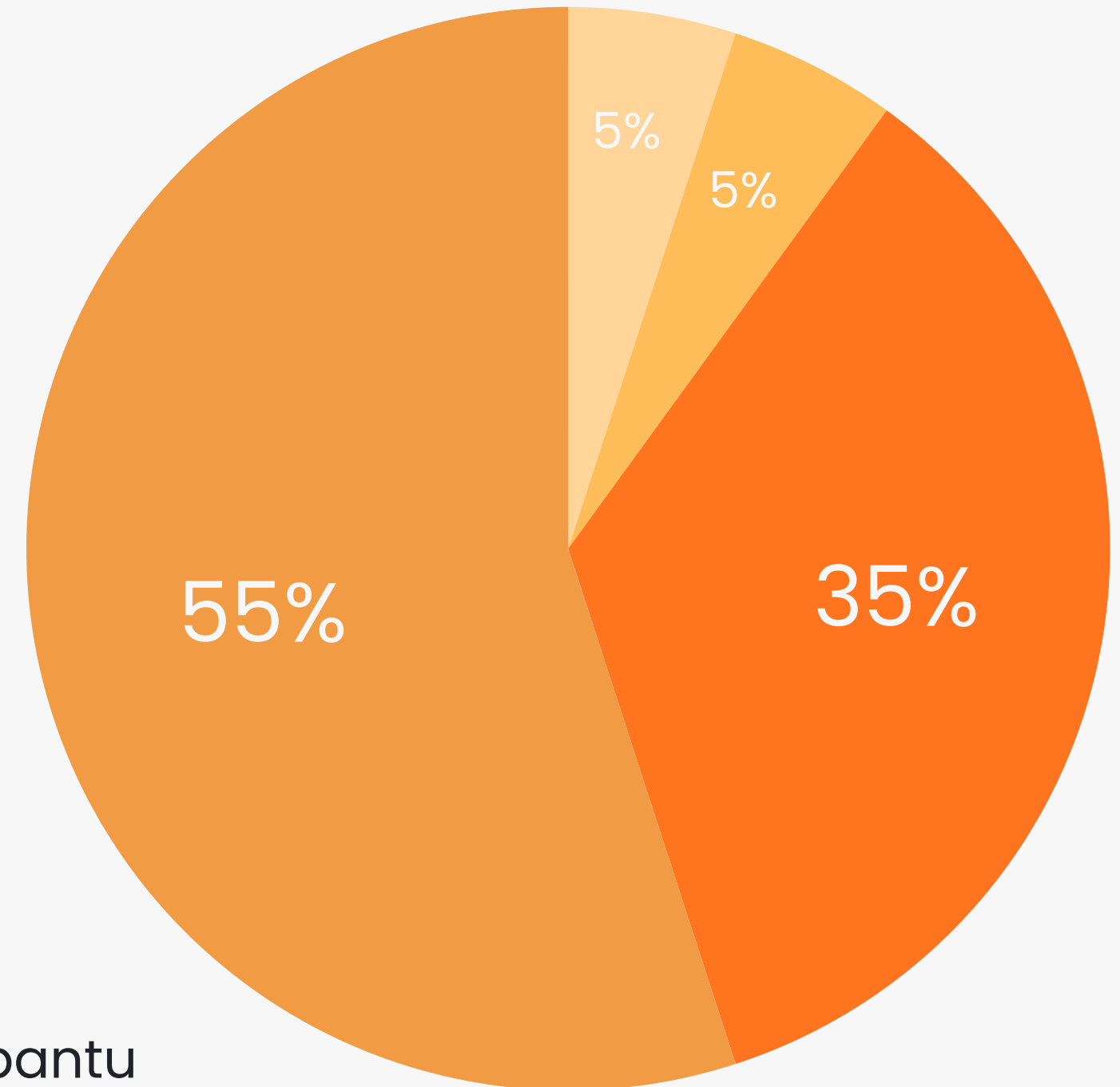
Hasil Survey

Pertanyaan yang diberikan yakni untuk mengukur sejauh mana ringkasan otomatis membantu pengguna dalam mempercepat pemahaman informasi dari teks yang dibaca.

Jumlah Responden: 20

-  Sangat membantu
-  Tidak membantu sama sekali

-  Membantu
-  Cukup membantu



Estimasi Biaya Pengembangan AI

1. Belanja Bahan (maks. 60%)

(Untuk kebutuhan pengembangan dan dokumentasi proyek)

No	Jenis Pengeluaran	Volume	Harga Satuan (Rp)	Total (Rp)
1	ATK (kertas, tinta, map, dll)	1 paket	400.000	400.000
2	Biaya kurasi & preprocessing dataset (technical)	1 akses	300.000	300.000
3	Lisensi software pendukung (AI/NLP tools)	2 lisensi	1.000.000	2.000.000
4	Cloud storage & compute (training model)	1 paket	1.600.000	1.600.000
Subtotal				4.300.000

- Tidak beli laptop / hardware, serta dataset yang dicari opensource-Free
- Fokus ke AI & software

2. Belanja Sewa (maks. 15%)

No	Jenis Pengeluaran	Volume	Harga Satuan (Rp)	Total (Rp)
1	Layanan deployment & hosting aplikasi AI (demo & API)	1 paket	700.000	700.000
2	Sewa tools evaluasi / testing AI	1 paket	300.000	300.000
Subtotal				1.000.000

Belanja sewa mencakup penggunaan layanan deployment dan hosting aplikasi AI berbasis Hugging Face untuk menampilkan demo dan API sistem Trustify. Selain itu, alokasi ini digunakan untuk tools evaluasi dan pengujian performa model, termasuk pengukuran kualitas ringkasan dan stabilitas sistem selama uji coba.

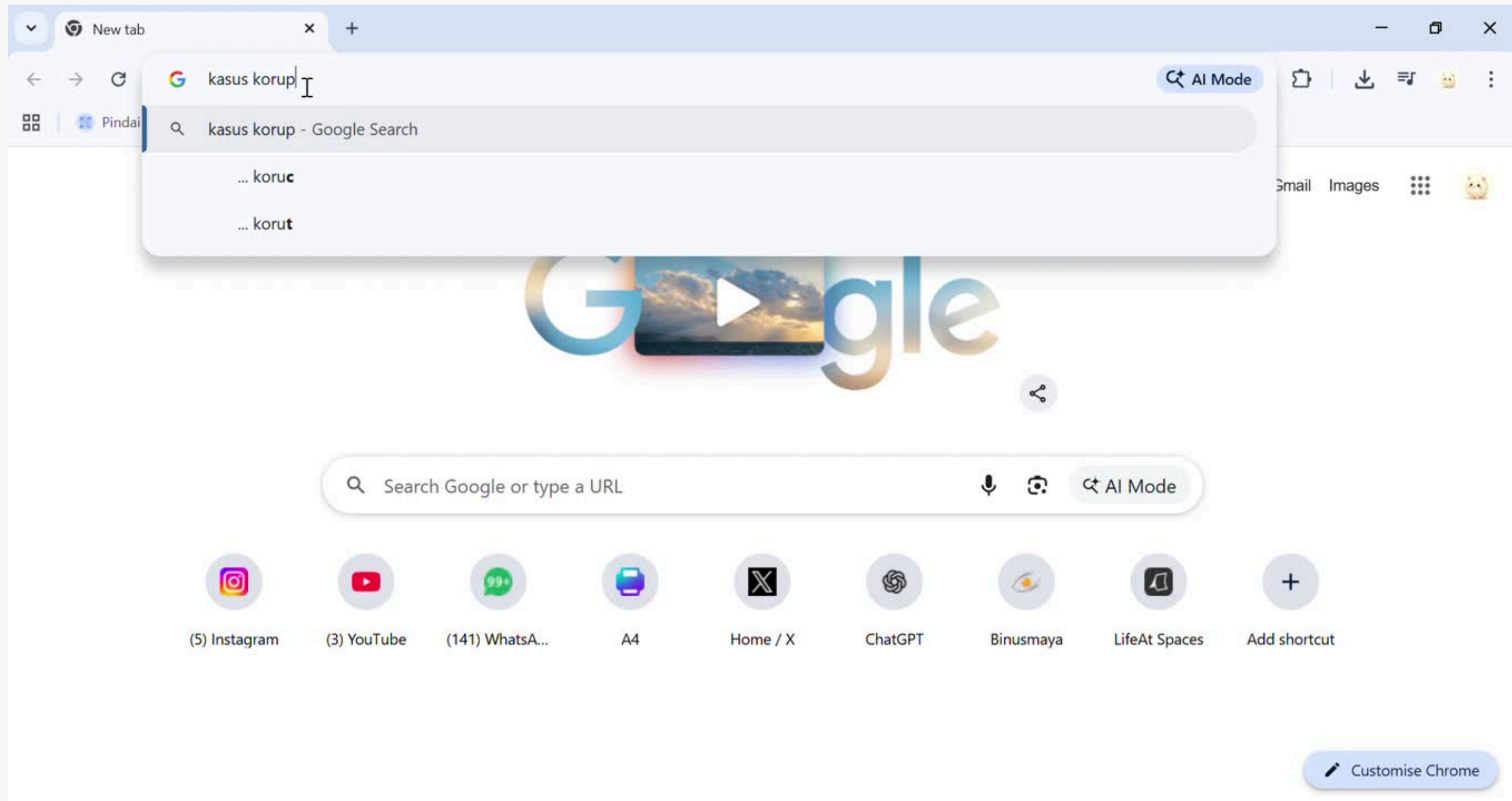
Biaya Lain-Lain

No	Jenis Pengeluaran	Volume	Harga Satuan (Rp)	Total (Rp)
1	Transport koordinasi tim	4	100.000	400.000
2	Transport dokumentasi kegiatan	2	150.000	300.000
Subtotal				700.000

No	Jenis Pengeluaran	Volume	Harga Satuan (Rp)	Total (Rp)
1	Jasa uji coba & evaluasi pengguna	1 paket	400.000	400.000
2	Percetakan laporan & dokumentasi	1 paket	300.000	300.000
3	Ads media sosial PKM (wajib)	1 paket	500.000	500.000
Subtotal				1.200.000

TOTAL AKHIR PROYEK: Rp7.200.000

Video Demo (Website) Penggunaan AI



Link Trustify: <https://huggingface.co/spaces/reydeuss/trustify>

Kesimpulan

Trustify merupakan solusi berbasis kecerdasan buatan yang dirancang untuk membantu meningkatkan pemahaman informasi di tengah tingginya arus berita digital. Dengan memanfaatkan teknologi *abstractive text summarization* berbasis Transformer, Trustify mampu menyajikan ringkasan berita yang ringkas, relevan, dan mudah dipahami. Hasil pengujian dan survei pengguna menunjukkan bahwa fitur ringkasan otomatis berpotensi mempercepat pemahaman isi informasi. Melalui pendekatan ini, Trustify berkontribusi dalam mendukung peningkatan literasi digital yang sejalan dengan SDG 4: Quality Education.

Saran dan Refleksi

- Pengembangan proyek ini juga berkolaborasi dengan AI, yakni chat GPT atau Gemini
- Optimalisasi performa sistem agar proses loading dan pemrosesan teks panjang menjadi lebih cepat.
- Pengembangan platform Trustify secara mandiri untuk mengurangi ketergantungan pada layanan pihak ketiga.
- Peningkatan aksesibilitas aplikasi agar dapat diakses langsung tanpa melalui tautan eksternal.
- Penyempurnaan desain antarmuka (UI/UX) untuk meningkatkan kenyamanan dan pengalaman pengguna.

Terima Kasih

Perhatiannyaa :)