

**PENDEKATAN DATA-DRIVEN UNTUK MENGATASI  
MASALAH KUALITAS PADA VIZWIZ CAPTIONING**

**SKRIPSI**

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Informatika



disusun oleh

**RHAMA RANGGA DHIPUTRA**

**22.11.5048**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2026**

**PENDEKATAN DATA-DRIVEN UNTUK MENGATASI  
MASALAH KUALITAS PADA VIZWIZ CAPTIONING**

**SKRIPSI**

untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Informatika



disusun oleh

**RHAMA RANGGA DHIPUTRA**

**22.11.5048**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2026**

**HALAMAN PERSETUJUAN**

**SKRIPSI**

**PENDEKATAN DATA-DRIVEN UNTUK MENGATASI MASALAH  
KUALITAS PADA VIZWIZ CAPTIONING**

yang disusun dan diajukan oleh

**Rhama Rangga Dhiputra**

**22.11.5048**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 25 Mei 2026

Dosen Pembimbing,



Anna Baita S.Kom., M.Kom  
NIK. 190302290

**HALAMAN PENGESAHAN**  
**SKRIPSI**  
**PENDEKATAN DATA-DRIVEN UNTUK MENGATASI MASALAH**  
**KUALITAS PADA VIZWIZ CAPTIONING**

yang disusun dan diajukan oleh

**Rhama Rangga Dhiputra**

**22.11.5048**

Telah dipertahankan di depan Dewan Penguji  
pada tanggal 25 Mei 2026

**Susunan Dewan Penguji**

**Nama Penguji**

**Tanda Tangan**

Theopilus Bayu S, S.Kom., M.Eng  
NIK. 190302375



Subektiningsih, S.Kom., M.Kom  
NIK. 190302413



Anna Baita, S.Kom., M.Kom  
NIK. 190302290

Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 25 Mei 2026

**DEKAN FAKULTAS ILMU KOMPUTER**



Prof. Dr. Kusriani, M.Kom.  
NIK. 190302106

## HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Rhama Ranga Dhiputra  
NIM : 22.11.5048

Menyatakan bahwa Skripsi dengan judul berikut:

### **PENDEKATAN DATA-DRIVEN UNTUK MENGATASI MASALAH KUALITAS PADA VIZWIZ CAPTIONING**

Dosen Pembimbing : Anna Baita, S.Kom., M.Kom.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 25 Mei 2026

Yang Menyatakan,



DEAQR011917124

Rhama Ranga Dhiputra

## HALAMAN PERSEMBAHAN

Segala puji dan syukur senantiasa tercurah ke hadirat Tuhan Yang Maha Esa atas segala rahmat, kekuatan, dan petunjuk-Nya. Tanpa campur tangan-Nya, perjalanan panjang dan melelahkan dalam menyelesaikan karya tulis ini tidak akan pernah mencapai titik akhir. Karya ini adalah wujud syukur atas segala nikmat akal dan kesabaran yang telah dianugerahkan.

Karya sederhana ini, dengan segala kerendahan hati, penulis persembahkan sebagai wujud bakti, cinta, dan hormat yang tak terhingga kepada:

1. Bapak Budi Santoso dan Ibu Earlya Handayani tercinta. Dua sosok pahlawan tanpa tanda jasa yang peluhnya menjadi jalan bagi kesuksesan penulis, yang air matanya menjelma menjadi doa-doa tak terputus di hening malam, dan yang kasih sayangnya tidak pernah lekang oleh waktu. Terima kasih atas segala pengorbanan yang tak terhitung jumlahnya, bimbingan moral yang selalu menguatkan, serta kepercayaan penuh yang diberikan sejak langkah pertama penulis menempuh pendidikan. Skripsi ini adalah hadiah kecil yang belum mampu membalas sejuta kebaikan yang telah Bapak dan Ibu berikan, namun semoga kelak dapat menjadi secercah kebanggaan.
2. Ibu Anna Baita, S.Kom., M.Kom., selaku Dosen Pembimbing yang tidak hanya memberikan arahan akademis, tetapi juga kesabaran dan keikhlasan dalam membimbing. Serta kepada jajaran pimpinan, Bapak/Ibu Dosen, dan segenap civitas academica Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta. Terima kasih atas ilmu, pengalaman, dan ruang berkembang yang telah membentuk pola pikir penulis menjadi lebih dewasa dan kritis.
3. Adik-adikku tersayang, Satria Bagas Wibisono, Kania Yumna Nazaleya, dan Hilmiya Zulfa Hadziqah, serta seluruh keluarga besar yang selalu memberikan kehangatan, dukungan, dan motivasi. Terima kasih atas setiap

senyuman dan kata-kata penyemangat yang selalu menjadi pelipur lara di saat penulis menghadapi kebuntuan dan rasa lelah dalam merangkai baris demi baris kode dan kalimat dalam penelitian ini.

4. Sahabat-sahabat seperjuangan di bangku perkuliahan, rekan-rekan diskusi, serta semua pihak yang telah menjadi "support system" terbaik. Terima kasih atas solidaritas, tawa, keluh kesah bersama, dan malam-malam panjang yang kita lalui demi menyelesaikan tugas akhir ini. Kalian adalah saksi bisu dari perjuangan dan pendewasaan ini.
5. Terakhir, persembahkan ini saya tujukan untuk diri saya sendiri, Rhama Ranga Dhiputra. Terima kasih karena memilih untuk tidak menyerah. Terima kasih sudah bertahan melewati segala keraguan, rasa lelah, dan malam-malam tanpa tidur. Perjalanan menyusun kecerdasan buatan untuk aksesibilitas ini telah mengajarkan bahwa setiap tetes usaha tidak akan pernah mengkhianati hasil.

Semoga karya tulis ini tidak hanya menjadi pajangan di rak perpustakaan, melainkan dapat memberikan manfaat yang nyata bagi ilmu pengetahuan dan masyarakat luas.

## KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa, karena atas rahmat, hidayah, dan karunia-Nya, penulis dapat menyelesaikan penyusunan skripsi yang berjudul "Pendekatan Data-Driven untuk Mengatasi Masalah Kualitas pada VizWiz Captioning" dengan baik. Skripsi ini disusun untuk memenuhi salah satu syarat dalam mencapai derajat Sarjana Komputer pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta.

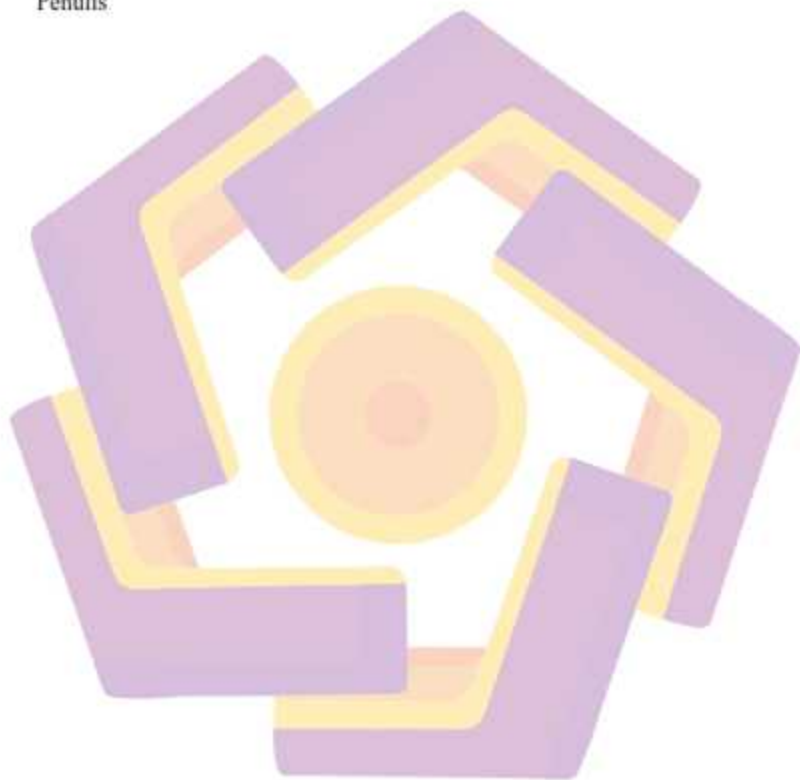
Penulis menyadari bahwa keberhasilan dalam penyusunan skripsi ini tidak lepas dari bantuan, bimbingan, arahan, serta dukungan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan rasa hormat dan ucapan terima kasih yang sebesar-besarnya kepada:

1. Ibu Anna Baita, S.Kom., M.Kom., selaku Dosen Pembimbing Skripsi yang telah meluangkan waktu, tenaga, dan pikiran untuk memberikan bimbingan, arahan, saran, serta motivasi yang sangat berharga selama proses penyusunan skripsi ini.
2. Tim Dosen Penguji Skripsi, selaku Dewan Penguji yang telah memberikan kritik, koreksi, saran, dan masukan konstruktif demi perbaikan dan kesempurnaan naskah skripsi ini saat sidang dan masa revisi.
3. Ibu Prof. Dr. Kusriani, M.Kom., selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
4. Ibu Eli Pujastuti, M.Kom., selaku Ketua Program Studi Informatika Universitas Amikom Yogyakarta.
5. Seluruh Dosen dan Staf Fakultas Ilmu Komputer Universitas Amikom Yogyakarta, yang telah membekali penulis dengan ilmu pengetahuan serta memberikan bantuan administrasi selama masa perkuliahan.
6. Kedua Orang Tua dan Seluruh Keluarga Tercinta, yang senantiasa memberikan doa yang tulus, limpahan kasih sayang, dukungan moral, spiritual, serta material yang tiada henti kepada penulis dalam menyelesaikan studi ini.

Penulis berharap semoga skripsi ini dapat memberikan manfaat, wawasan, serta kegunaan yang baik bagi perkembangan ilmu pengetahuan, khususnya di bidang informatika dan teknologi aksesibilitas.

Yogyakarta, 25 Mei 2026

Penulis



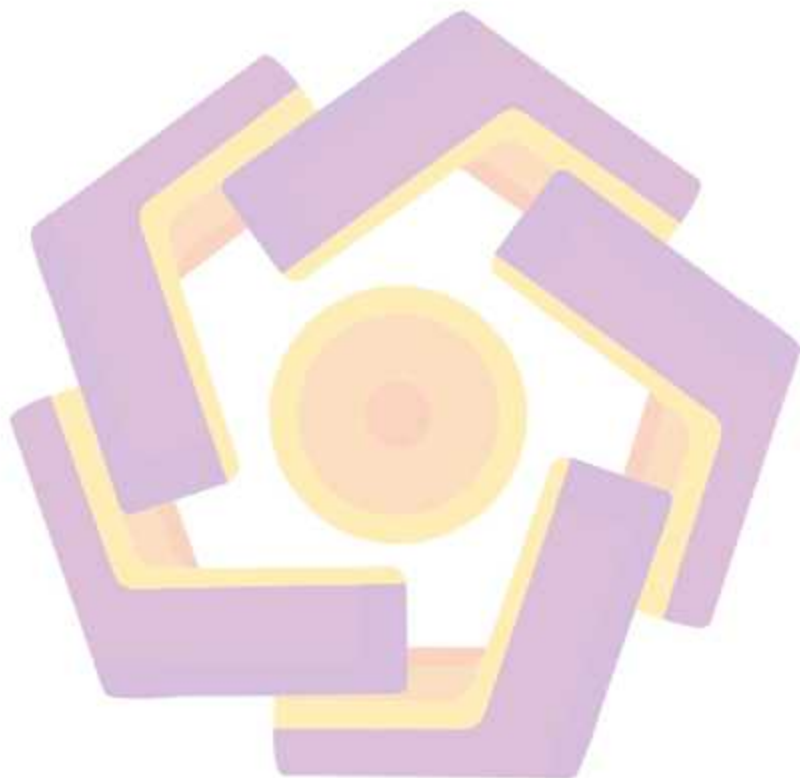
## DAFTAR ISI

|                                           |      |
|-------------------------------------------|------|
| HALAMAN JUDUL .....                       | i    |
| HALAMAN PERSETUJUAN.....                  | ii   |
| HALAMAN PENGESAHAN .....                  | iii  |
| HALAMAN PERNYATAAN KEASLIAN SKRIPSI ..... | iv   |
| HALAMAN PERSEMBAHAN .....                 | v    |
| KATA PENGANTAR .....                      | vii  |
| DAFTAR ISI.....                           | ix   |
| DAFTAR TABEL.....                         | xiii |
| DAFTAR GAMBAR .....                       | xiv  |
| DAFTAR LAMBANG DAN SINGKATAN .....        | xvi  |
| DAFTAR ISTILAH.....                       | xx   |
| INTISARI .....                            | xxx  |
| ABSTRACT.....                             | xxxi |
| BAB I PENDAHULUAN.....                    | 1    |
| 1.1 Latar Belakang .....                  | 1    |
| 1.2 Rumusan Masalah .....                 | 3    |
| 1.3 Batasan Masalah .....                 | 3    |
| 1.4 Tujuan Penelitian .....               | 5    |
| 1.5 Manfaat Penelitian .....              | 6    |
| 1.6 Sistematika Penulisan .....           | 8    |
| BAB II TINJAUAN PUSTAKA .....             | 9    |

|                                   |                                                                                 |    |
|-----------------------------------|---------------------------------------------------------------------------------|----|
| 2.1                               | Studi Literatur .....                                                           | 9  |
| 2.2                               | Dasar Teori.....                                                                | 15 |
| 2.2.1                             | Sistem Visi-Bahasa ( <i>Vision-Language</i> ) dan <i>Image Captioning</i> ..... | 15 |
| 2.2.2                             | Analisis Data Eksploratif (EDA) Multidimensi pada Citra .....                   | 18 |
| 2.2.3                             | Prapemrosesan Citra dan Teks ( <i>Preprocessing</i> ) .....                     | 22 |
| 2.2.4                             | Arsitektur Bootstrapping Language-Image Pre-training (BLIP) .....               | 27 |
| 2.2.5                             | Efisiensi Parameter melalui <i>Low-Rank Adaptation</i> (LoRA).....              | 31 |
| 2.2.6                             | Optimasi Pemelajaran Mesin dan Manajemen Memori.....                            | 34 |
| 2.2.7                             | Metrik Evaluasi Kinerja (Leksikal & Semantik) .....                             | 38 |
| BAB III METODE PENELITIAN .....   |                                                                                 | 42 |
| 3.1                               | Objek Penelitian .....                                                          | 42 |
| 3.2                               | Alur Penelitian .....                                                           | 42 |
| 3.2.1                             | Pengumpulan Data ( <i>Data Acquisition</i> ) .....                              | 45 |
| 3.2.2                             | Analisis Data Eksploratif ( <i>Exploratory Data Analysis/EDA</i> ) .....        | 46 |
| 3.2.3                             | Prapemrosesan Citra dan Teks ( <i>Preprocessing</i> ) .....                     | 49 |
| 3.2.4                             | Pemodelan ( <i>Modelling</i> ).....                                             | 51 |
| 3.2.5                             | Strategi Dekode Inferensi .....                                                 | 53 |
| 3.2.6                             | Evaluasi Kinerja.....                                                           | 54 |
| 3.2.7                             | Implementasi ( <i>Deployment</i> ).....                                         | 56 |
| 3.3                               | Alat dan Bahan .....                                                            | 58 |
| 3.3.1                             | Data Penelitian .....                                                           | 58 |
| 3.3.2                             | Alat dan Instrumen Penelitian.....                                              | 62 |
| BAB IV HASIL DAN PEMBAHASAN ..... |                                                                                 | 64 |
| 4.1                               | Hasil Pengumpulan dan Kurasi Data ( <i>Data-Centric Curation</i> ).....         | 64 |
| 4.1.1                             | Karakteristik Dataset Mentah VizWiz .....                                       | 64 |

|                                                                                           |     |
|-------------------------------------------------------------------------------------------|-----|
| 4.1.2 Hasil Analisis Kualitas Citra (Dimensi 1).....                                      | 67  |
| 4.1.3 Hasil Analisis Kualitas Teks (Dimensi 2).....                                       | 73  |
| 4.1.4 Hasil Penggabungan dan Taksonomi Penyaringan Akhir (Dimensi 3) 80                   |     |
| 4.1.5 Distribusi Dataset Akhir.....                                                       | 89  |
| 4.2 Hasil Prapemrosesan Data Multimodal .....                                             | 90  |
| 4.2.1 Hasil Transformasi Citra .....                                                      | 90  |
| 4.2.2 Hasil Tokenisasi Teks .....                                                         | 93  |
| 4.3 Hasil Pelatihan Model (Pemodelan BLIP + LoRA) .....                                   | 94  |
| 4.3.1 Metrik Pelatihan ( <i>Loss &amp; Stability</i> ) .....                              | 94  |
| 4.3.2 Efisiensi Komputasi (Memory-Efficient) .....                                        | 97  |
| 4.4 Hasil Evaluasi Kinerja Model ( <i>Ablation Study</i> ).....                           | 98  |
| 4.4.1 Skenario Pengujian .....                                                            | 99  |
| 4.4.2 Hasil Evaluasi Leksikal dan Semantik.....                                           | 99  |
| 4.4.3 Analisis Kualitatif Model.....                                                      | 102 |
| 4.5 Pembahasan Utama: Eliminasi "Paradoks Kualitas" .....                                 | 104 |
| 4.5.1 Pembagian Tertil Kualitas ( <i>Tertile Split</i> ).....                             | 104 |
| 4.5.2 Hasil Uji Statistik Independen .....                                                | 105 |
| 4.5.3 Kesimpulan Eliminasi Paradoks .....                                                 | 107 |
| 4.6 Hasil Implementasi Sistem ( <i>Deployment</i> ) Aplikasi "Mata AI" .....              | 108 |
| 4.6.1 Arsitektur dan Antarmuka Aplikasi .....                                             | 108 |
| 4.6.2 Integrasi LLM dan Konsep <i>Trustworthy AI</i> .....                                | 110 |
| 4.6.3 Pengujian Kualitatif Dunia Nyata ( <i>Real-World Testing</i> ).....                 | 112 |
| 4.6.4 Pengujian Kondisi Ekstrem dan Batasan Sistem ( <i>Fail-Safe Analysis</i> )<br>..... | 116 |
| 4.7 Dokumentasi Implementasi dan Transparansi Sistem .....                                | 119 |

|                      |     |
|----------------------|-----|
| BAB V PENUTUP .....  | 120 |
| 5.1 Kesimpulan ..... | 120 |
| 5.2 Saran .....      | 121 |
| REFERENSI .....      | 123 |



## DAFTAR TABEL

|                                                                                                       |     |
|-------------------------------------------------------------------------------------------------------|-----|
| Table 2.1 Keaslian Penelitian .....                                                                   | 11  |
| Table 3.1 Distribusi dataset sebelum dan sesudah tahap kurasi multidimensi .....                      | 59  |
| Table 3.2 Rincian alasan pembuangan data (drop reason breakdown).....                                 | 60  |
| Table 3.3 Perbandingan metrik kualitas dataset sebelum dan sesudah kurasi .....                       | 61  |
| Table 3.4 Rincian instrumen perangkat keras dan perangkat lunak yang digunakan dalam penelitian ..... | 63  |
| Table 4.1 Definisi dan distribusi kategori taksonomi penyaringan dataset VizWiz .....                 | 83  |
| Table 4.2 Rangkuman hasil transformasi dimensi citra masukan.....                                     | 91  |
| Table 4.3 Contoh hasil konversi teks deskripsi menjadi tensor menggunakan BERT Tokenizer .....        | 93  |
| Table 4.4 Perbandingan metrik kinerja kuantitatif model BLIP pada berbagai skenario pelatihan. ....   | 100 |
| Table 4.5 Perbandingan Kinerja dan Uji Statistik: Model V1 (Raw) vs Model V2 (Final).....             | 105 |

## DAFTAR GAMBAR

|                                                                                                   |        |
|---------------------------------------------------------------------------------------------------|--------|
| Gambar 2.1 Arsitektur Multimodal Visi-Bahasa (BLIP/BLIP-2).....                                   | 16     |
| Gambar 2.2 Contoh degradasi visual pada himpunan data VizWiz .....                                | 17     |
| Gambar 2.3 Taksonomi isu kualitas visual VizWiz.....                                              | 21     |
| Gambar 2.4 Ilustrasi proses letterbox padding .....                                               | 24     |
| Gambar 2.5 Mekanisme sub-word tokenization pada prapemrosesan teks .....                          | 26     |
| Gambar 2.6 Ekstraksi patch spasial pada arsitektur Vision Transformer (ViT).....                  | 28     |
| Gambar 2.7 Mekanisme cross-attention antara ruang visual dan bahasa .....                         | 30     |
| Gambar 2.8 Dekomposisi matriks berpangkat rendah pada arsitektur LoRA .....                       | 33     |
| Gambar 2.9 Ilustrasi pencarian heuristik menggunakan Beam Search.....                             | 37     |
| <br>Gambar 3.1 Alur Penelitian .....                                                              | <br>44 |
| Gambar 3.2 Modifikasi arsitektur BLIP melalui injeksi adaptor LoRA.....                           | 52     |
| Gambar 3.3 Rancangan Antarmuka Purwarupa Sistem Aplikasi (Prototype) pada Tahap Implementasi..... | 57     |
| <br>Gambar 4.1 Distribusi variasi dimensi (lebar, tinggi, resolusi, dan rasio aspek) ..66         |        |
| Gambar 4.2 Distribusi dimensi citra pada dataset mentah VizWiz .....                              | 70     |
| Gambar 4.3 Validasi visual sampel citra berdasarkan tingkat kekaburan (blur tier) .....           | 72     |
| Gambar 4.4 Distribusi statistik anotasi teks pada dataset VizWiz.....                             | 74     |
| Gambar 4.5 Distribusi statistik anotasi teks pada dataset VizWiz.....                             | 76     |
| Gambar 4.6 Analisis deteksi kegagalan anotasi teks dan distribusinya.....                         | 79     |
| Gambar 4.7 Matriks korelasi Pearson antara kualitas citra dan anotasi teks .....                  | 81     |
| Gambar 4.8 Distribusi taksonomi penyaringan dan keputusan akhir dataset VizWiz .....              | 85     |
| Gambar 4.9 Distribusi alasan pembuangan data pada tahap penyaringan akhir.....                    | 88     |
| Gambar 4.10 Contoh Hasil Letterbox Padding .....                                                  | 92     |
| Gambar 4.11 Perbandingan kurva konvergensi loss: Model baseline vs. Model final .....             | 95     |

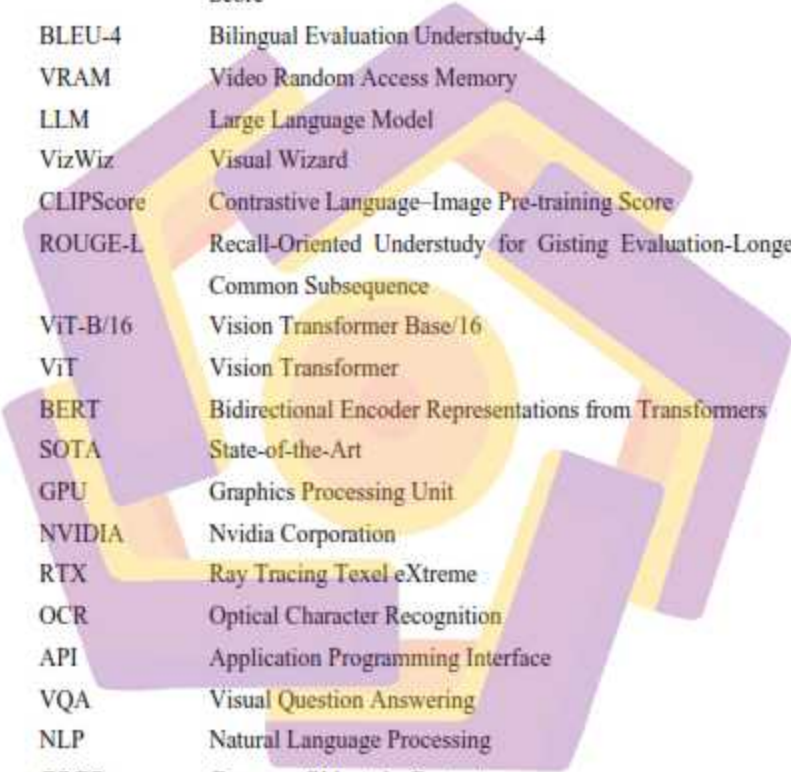
|                                                                                                                                   |     |
|-----------------------------------------------------------------------------------------------------------------------------------|-----|
| Gambar 4.12 Kurva kerugian (loss) skenario Full Fine-Tuning.....                                                                  | 96  |
| Gambar 4.13 Komparasi hasil deskripsi: Ground-truth, model baseline, dan model final .....                                        | 103 |
| Gambar 4.14 Perbandingan performa model berdasarkan strata kualitas citra....                                                     | 106 |
| Gambar 4.15 Tampilan antarmuka aplikasi "Mata AI" .....                                                                           | 109 |
| Gambar 4.16 Alur pemrosesan bahasa pada modul LLM .....                                                                           | 111 |
| Gambar 4.17 Antarmuka hasil analisis kualitatif objek pada kondisi ideal.....                                                     | 114 |
| Gambar 4.18 Antarmuka hasil analisis kualitatif objek dunia nyata pada kondisi motion blur.....                                   | 115 |
| Gambar 4.19 Tangkapan layar batasan operasional purwarupa "Mata AI" pada kondisi pengujian ekstrem tanpa deteksi objek utama..... | 118 |



## DAFTAR LAMBANG DAN SINGKATAN

|                  |                                                             |
|------------------|-------------------------------------------------------------|
| $v$              | Representasi vektor visual hasil ekstraksi fitur            |
| $T$              | Representasi vektor teks hasil pemrosesan bahasa            |
| $E_v$            | Fungsi penyandi (encoder) visual                            |
| $E_t$            | Penyandi tekstual                                           |
| $I$              | Citra masukan (input image)                                 |
| $T$              | Urutan teks masukan (input text sequence)                   |
| $w$              | Vektor teks pada ruang kesamaan kosinus (cosine similarity) |
| $F(u, v)$        | Nilai pada domain frekuensi                                 |
| $f(x, y)$        | Nilai intensitas piksel pada koordinat spasial              |
| $M, N$           | Dimensi matriks citra                                       |
| $u, v$           | Variabel domain frekuensi                                   |
| $x, y$           | Variabel koordinat spasial                                  |
| $\bar{f}$        | Rata-rata intensitas piksel pada citra                      |
| $B$              | Nilai kecerahan rata-rata citra                             |
| $L(x)$           | Nilai filter Lanczos pada posisi $x$                        |
| $\text{sinc}(x)$ | Fungsi sinus kardinal                                       |
| $a$              | Ukuran jendela filter Lanczos                               |
| $z$              | Nilai piksel yang telah dinormalisasi                       |
| $x$              | Nilai piksel masukan                                        |
| $\mu$            | Nilai rata-rata empiris distribusi data                     |
| $\sigma$         | Deviasi standar distribusi data                             |
| $Q$              | Matriks Query                                               |
| $K$              | Matriks Key                                                 |
| $V$              | Matriks Value                                               |
| $dk$             | Dimensi matriks Key (scaling factor)                        |
| $h$              | Vektor keluaran hasil transformasi linear                   |
| $W_0$            | Matriks bobot asli yang dibekukan (frozen weights)          |
| $\Delta W$       | Pembaruan bobot pada arsitektur LoRA                        |

|              |                                                              |
|--------------|--------------------------------------------------------------|
| $A$          | Matriks dekomposisi berpangkat rendah ( $r \times k$ )       |
| $r$          | Rank dekomposisi ( $r \ll d$ )                               |
| $\theta_t$   | Nilai bobot pada langkah $t$                                 |
| $\eta$       | Laju pembelajaran (learning rate)                            |
| $\eta_{min}$ | Laju pembelajaran minimum                                    |
| $\eta_{max}$ | Laju pembelajaran maksimum                                   |
| $\nabla L$   | Gradien fungsi kerugian                                      |
| $\lambda$    | Koefisien weight decay                                       |
| $T_{cur}$    | Jumlah langkah sejak restart terakhir                        |
| $T_{max}$    | Jumlah langkah maksimum dalam satu siklus                    |
| $N$          | Jumlah langkah akumulasi gradien                             |
| $P(Y X)$     | Probabilitas urutan keluaran $Y$ diberikan masukan $X$       |
| $T$          | Panjang total urutan keluaran                                |
| $C$          | Panjang teks prediksi (BLEU)                                 |
| $R$          | Panjang teks referensi (BLEU)                                |
| $w_n$        | Bobot seragam n-gram                                         |
| $P_n$        | Probabilitas presisi modifikasi n-gram ke- $n$               |
| $g_n(c)$     | Vektor TF-IDF n-gram untuk teks prediksi                     |
| $g_n(sf)$    | Vektor TF-IDF n-gram untuk teks referensi ke- $j$            |
| $m$          | Jumlah kalimat referensi manusia                             |
| $RLCS$       | Nilai recall berdasarkan LCS                                 |
| $PLCS$       | Nilai precision berdasarkan LCS                              |
| $\beta$      | Rasio parameter recall terhadap precision                    |
| $x$          | Vektor token teks referensi (BERTScore)                      |
| $y$          | Vektor token teks prediksi (BERTScore)                       |
| $ X $        | Jumlah token pada teks referensi                             |
| $E_I(I)$     | Vektor penyematan visual dari citra $I$ (CLIPScore)          |
| $E_C(C)$     | Vektor penyematan teks dari kalimat prediksi $C$ (CLIPScore) |
| $w$          | Konstanta penskalaan (CLIPScore)                             |
| FFT          | Fast Fourier Transform                                       |



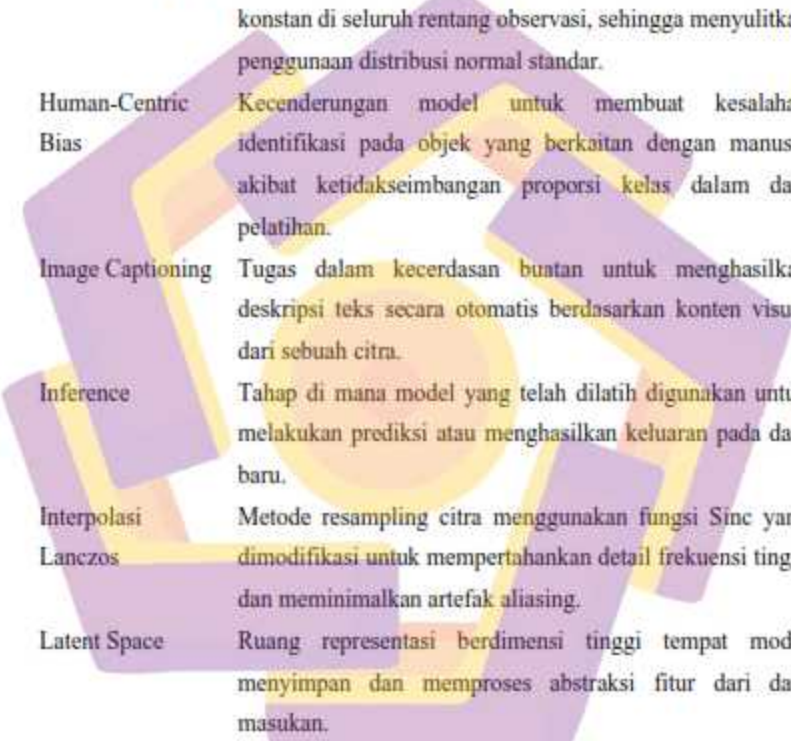
|           |                                                                              |
|-----------|------------------------------------------------------------------------------|
| PEFT      | Parameter-Efficient Fine-Tuning                                              |
| LoRA      | Low-Rank Adaptation                                                          |
| AdamW     | Adaptive Moment Estimation with Weight Decay                                 |
| CIDER     | Consensus-based Image Description Evaluation                                 |
| BERTScore | Bidirectional Encoder Representations from Transformers Score                |
| BLEU-4    | Bilingual Evaluation Understudy-4                                            |
| VRAM      | Video Random Access Memory                                                   |
| LLM       | Large Language Model                                                         |
| VizWiz    | Visual Wizard                                                                |
| CLIPScore | Contrastive Language-Image Pre-training Score                                |
| ROUGE-L   | Recall-Oriented Understudy for Gisting Evaluation-Longest Common Subsequence |
| ViT-B/16  | Vision Transformer Base/16                                                   |
| ViT       | Vision Transformer                                                           |
| BERT      | Bidirectional Encoder Representations from Transformers                      |
| SOTA      | State-of-the-Art                                                             |
| GPU       | Graphics Processing Unit                                                     |
| NVIDIA    | Nvidia Corporation                                                           |
| RTX       | Ray Tracing Texel eXtreme                                                    |
| OCR       | Optical Character Recognition                                                |
| API       | Application Programming Interface                                            |
| VQA       | Visual Question Answering                                                    |
| NLP       | Natural Language Processing                                                  |
| COCO      | Common Objects in Context                                                    |
| EDA       | Exploratory Data Analysis                                                    |
| RMS       | Root Mean Square                                                             |
| OOV       | Out of Vocabulary                                                            |
| AMP       | Automatic Mixed Precision                                                    |
| FP32      | Floating Point 32-bit                                                        |
| FP16      | Floating Point 16-bit                                                        |

|            |                                           |
|------------|-------------------------------------------|
| RAM        | Random Access Memory                      |
| BP         | Brevity Penalty                           |
| TF-IDF     | Term Frequency-Inverse Document Frequency |
| LCS        | Longest Common Subsequence                |
| CLIP       | Contrastive Language-Image Pre-training   |
| JSON       | JavaScript Object Notation                |
| RGB        | Red, Green, Blue                          |
| VLM        | Vision-Language Model                     |
| OOM        | Out-of-Memory                             |
| TTS        | Text-to-Speech                            |
| MS-COCO    | Microsoft Common Objects in Context       |
| UI         | User Interface                            |
| JSONL      | JavaScript Object Notation Lines          |
| LSTM       | Long Short-Term Memory                    |
| ResNet-152 | Residual Network-152                      |
| LiDAR      | Light Detection and Ranging               |

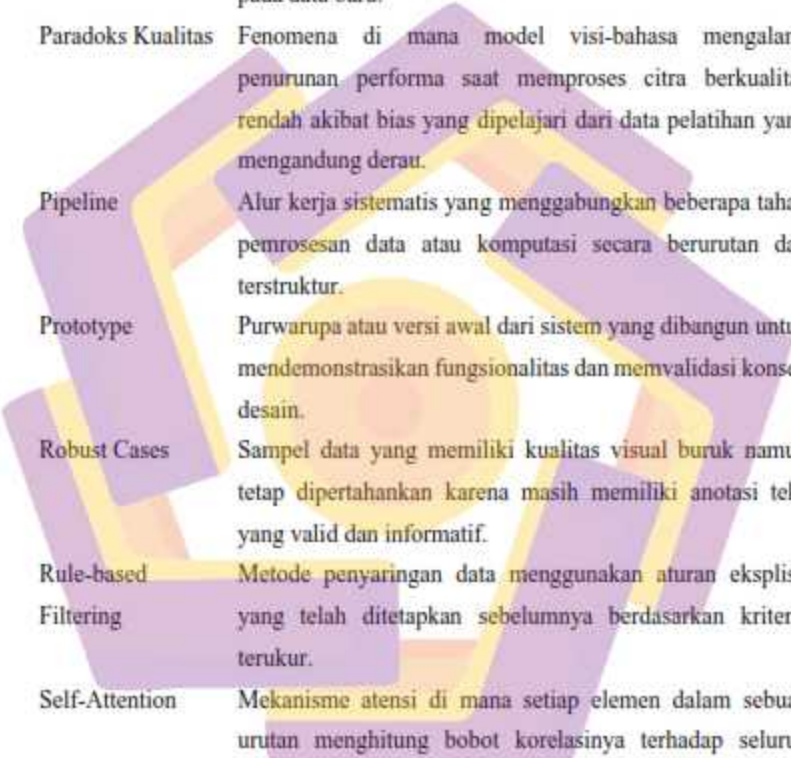
## DAFTAR ISTILAH

| Singkatan               | Istilah                                                                                                                                                                       |
|-------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Ablasi (Ablation Study) | Metode pengujian model dengan menghilangkan atau menonaktifkan komponen tertentu dari sistem untuk mengukur kontribusi masing-masing komponen terhadap performa keseluruhan.  |
| Attention Masking       | Mekanisme pada arsitektur transformator untuk mengabaikan area tertentu seperti padding dalam kalkulasi bobot kepentingan antar token.                                        |
| Baseline                | Model standar atau pembandingan yang digunakan sebelum dilakukan optimasi atau modifikasi arsitektur lebih lanjut.                                                            |
| Beam Search             | Algoritma pencarian heuristik yang mempertahankan sejumlah kandidat urutan kata terbaik secara bersamaan untuk menghasilkan teks yang lebih koheren.                          |
| Captioning              | Proses otomatis dalam menghasilkan deskripsi teks berdasarkan konten visual dari sebuah gambar.                                                                               |
| Catastrophic Forgetting | Fenomena di mana model kehilangan kemampuan generalisasi sebelumnya setelah dilatih pada tugas atau dataset baru yang spesifik.                                               |
| Cohen's d               | Ukuran efek statistik yang mengukur besaran perbedaan antara dua kelompok data secara kuantitatif.                                                                            |
| Cosine Annealing        | Metode penjadwalan laju pembelajaran yang menurunkan nilai laju secara bertahap mengikuti kurva fungsi kosinus untuk menstabilkan konvergensi.                                |
| Cosine Similarity       | Ukuran kesamaan antara dua vektor berdasarkan sudut kosinus di antara keduanya dalam ruang berdimensi tinggi.                                                                 |
| Cross-Attention         | Mekanisme atensi yang menghubungkan dua modalitas berbeda, seperti teks dan gambar, dengan memetakan Query dari satu modalitas terhadap Key dan Value dari modalitas lainnya. |

|                          |                                                                                                                                                                            |
|--------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Cultural Domain Gap      | Kesenjangan performa model yang terjadi ketika sistem digunakan pada konteks budaya yang berbeda dari distribusi data pelatihannya.                                        |
| Data-centric AI          | Paradigma pengembangan kecerdasan buatan yang memprioritaskan peningkatan kualitas dan kurasi data untuk mencapai performa model yang lebih baik.                          |
| Data Leakage             | Kondisi di mana informasi dari data pengujian bocor ke dalam proses pelatihan sehingga hasil evaluasi menjadi bias dan tidak representatif.                                |
| Data-Driven Thresholding | Pendekatan penetapan ambang batas secara dinamis berdasarkan distribusi empiris data aktual, bukan nilai tetap yang ditentukan secara manual.                              |
| Early Stopping           | Teknik regularisasi pelatihan yang menghentikan proses pelatihan lebih awal ketika performa pada data validasi tidak lagi membaik.                                         |
| Explainability           | Kemampuan sistem kecerdasan buatan untuk memberikan penjelasan yang dapat dipahami manusia atas keputusan atau prediksi yang dihasilkan.                                   |
| Fail-Safe                | Mekanisme keamanan sistem yang memastikan kegagalan terjadi secara terkontrol tanpa menghasilkan informasi yang berbahaya atau menyesatkan.                                |
| Fine-tuning              | Proses melatih ulang model pra-latih pada dataset spesifik agar memiliki kemampuan adaptasi yang lebih baik terhadap domain target.                                        |
| Frozen Weights           | Kondisi di mana bobot parameter model dikunci dan tidak diperbarui selama proses pelatihan lanjutan.                                                                       |
| Gradient Accumulation    | Teknik manajemen memori GPU yang mengakumulasi nilai gradien dari beberapa mini-batch sebelum melakukan pembaruan bobot, untuk mensimulasikan batch size yang lebih besar. |



|                     |                                                                                                                                                                 |
|---------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Ground-truth        | Data referensi atau label kebenaran yang dihasilkan oleh manusia sebagai standar evaluasi performa model.                                                       |
| Hallucination (AI)  | Kondisi di mana model kecerdasan buatan menghasilkan informasi yang tidak akurat atau tidak relevan dengan masukan yang diberikan.                              |
| Heteroskedastik     | Kondisi statistik di mana varians dari suatu variabel tidak konstan di seluruh rentang observasi, sehingga menyulitkan penggunaan distribusi normal standar.    |
| Human-Centric Bias  | Kecenderungan model untuk membuat kesalahan identifikasi pada objek yang berkaitan dengan manusia akibat ketidakseimbangan proporsi kelas dalam data pelatihan. |
| Image Captioning    | Tugas dalam kecerdasan buatan untuk menghasilkan deskripsi teks secara otomatis berdasarkan konten visual dari sebuah citra.                                    |
| Inference           | Tahap di mana model yang telah dilatih digunakan untuk melakukan prediksi atau menghasilkan keluaran pada data baru.                                            |
| Interpolasi Lanczos | Metode resampling citra menggunakan fungsi Sinc yang dimodifikasi untuk mempertahankan detail frekuensi tinggi dan meminimalkan artefak aliasing.               |
| Latent Space        | Ruang representasi berdimensi tinggi tempat model menyimpan dan memproses abstraksi fitur dari data masukan.                                                    |
| Letterbox Padding   | Teknik pengubahan dimensi citra dengan menambahkan piksel netral bernilai nol pada sisi yang kosong untuk menjaga rasio aspek asli tetap utuh.                  |
| Motion Blur         | Efek buram pada gambar yang terjadi akibat pergerakan objek atau kamera selama proses pengambilan gambar.                                                       |



|                      |                                                                                                                                                                                |
|----------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Multimodal           | Sistem atau arsitektur yang mampu memproses dan mengintegrasikan informasi dari lebih dari satu jenis modalitas data, seperti teks dan gambar.                                 |
| Overfitting          | Kondisi di mana model terlalu menyesuaikan diri dengan data pelatihan sehingga kehilangan kemampuan generalisasi pada data baru.                                               |
| Paradoks Kualitas    | Fenomena di mana model visi-bahasa mengalami penurunan performa saat memproses citra berkualitas rendah akibat bias yang dipelajari dari data pelatihan yang mengandung derau. |
| Pipeline             | Alur kerja sistematis yang menggabungkan beberapa tahap pemrosesan data atau komputasi secara berurutan dan terstruktur.                                                       |
| Prototype            | Purwarupa atau versi awal dari sistem yang dibangun untuk mendemonstrasikan fungsionalitas dan memvalidasi konsep desain.                                                      |
| Robust Cases         | Sampel data yang memiliki kualitas visual buruk namun tetap dipertahankan karena masih memiliki anotasi teks yang valid dan informatif.                                        |
| Rule-based Filtering | Metode penyaringan data menggunakan aturan eksplisit yang telah ditetapkan sebelumnya berdasarkan kriteria terukur.                                                            |
| Self-Attention       | Mekanisme atensi di mana setiap elemen dalam sebuah urutan menghitung bobot korelasinya terhadap seluruh elemen lain dalam urutan yang sama.                                   |
| Semantic Resilience  | Ketangguhan model dalam mempertahankan pemahaman makna konteks objek meskipun citra masukan mengalami degradasi visual yang parah.                                             |
| Shortcut Learning    | Kecenderungan model untuk menggunakan pola-pola sederhana atau tidak relevan dalam data pelatihan alih-alih mempelajari fitur semantik yang mendalam.                          |

|                         |                                                                                                                                                                              |
|-------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Sub-word                | Pendekatan tokenisasi yang memecah kata-kata tidak dikenal menjadi unit sub-kata yang lebih kecil agar dapat direpresentasikan oleh model bahasa.                            |
| Tokenization            |                                                                                                                                                                              |
| Transfer Learning       | Teknik pembelajaran mesin yang memanfaatkan pengetahuan dari model pra-latih pada satu domain untuk meningkatkan performa pada domain yang berbeda.                          |
| Trustworthy AI          | Konsep kecerdasan buatan yang dapat dipercaya karena mampu memberikan transparansi, penjelasan, dan keamanan atas keputusan yang dihasilkan.                                 |
| Underexposure           | Kondisi di mana citra terlalu gelap akibat kurangnya cahaya yang ditangkap sensor kamera selama proses perekaman.                                                            |
| Overexposure            | Kondisi di mana citra terlalu terang akibat cahaya berlebihan yang ditangkap sensor kamera, sehingga detail visual hilang.                                                   |
| User-generated Content  | Konten yang dibuat dan diunggah secara langsung oleh pengguna akhir, bukan oleh profesional atau dalam kondisi terkontrol.                                                   |
| Viewfinder              | Layar bidik pada kamera yang digunakan untuk melihat dan membingkai objek sebelum pengambilan gambar dilakukan.                                                              |
| Vision-Language Model   | Arsitektur kecerdasan buatan yang dirancang untuk memahami dan mengintegrasikan informasi dari modalitas visual dan bahasa secara bersamaan.                                 |
| Weight Decay            | Teknik regularisasi dalam pelatihan model yang memberikan penalti pada bobot berukuran besar untuk mencegah overfitting.                                                     |
| Ablasi (Ablation Study) | Metode pengujian model dengan menghilangkan atau menonaktifkan komponen tertentu dari sistem untuk mengukur kontribusi masing-masing komponen terhadap performa keseluruhan. |

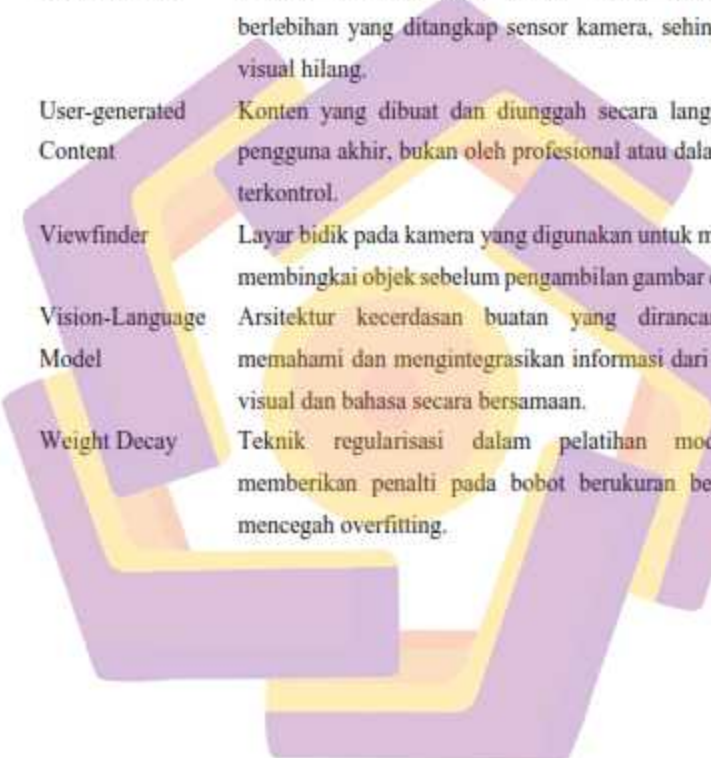
|                         |                                                                                                                                                                               |
|-------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Attention Masking       | Mekanisme pada arsitektur transformator untuk mengabaikan area tertentu seperti padding dalam kalkulasi bobot kepentingan antar token.                                        |
| Baseline                | Model standar atau pembandingan yang digunakan sebelum dilakukan optimasi atau modifikasi arsitektur lebih lanjut.                                                            |
| Beam Search             | Algoritma pencarian heuristik yang mempertahankan sejumlah kandidat urutan kata terbaik secara bersamaan untuk menghasilkan teks yang lebih koheren.                          |
| Captioning              | Proses otomatis dalam menghasilkan deskripsi teks berdasarkan konten visual dari sebuah gambar.                                                                               |
| Catastrophic Forgetting | Fenomena di mana model kehilangan kemampuan generalisasi sebelumnya setelah dilatih pada tugas atau dataset baru yang spesifik.                                               |
| Cohen's d               | Ukuran efek statistik yang mengukur besaran perbedaan antara dua kelompok data secara kuantitatif.                                                                            |
| Cosine Annealing        | Metode penjadwalan laju pembelajaran yang menurunkan nilai laju secara bertahap mengikuti kurva fungsi kosinus untuk menstabilkan konvergensi.                                |
| Cosine Similarity       | Ukuran kesamaan antara dua vektor berdasarkan sudut kosinus di antara keduanya dalam ruang berdimensi tinggi.                                                                 |
| Cross-Attention         | Mekanisme atensi yang menghubungkan dua modalitas berbeda, seperti teks dan gambar, dengan memetakan Query dari satu modalitas terhadap Key dan Value dari modalitas lainnya. |
| Cultural Domain Gap     | Kesenjangan performa model yang terjadi ketika sistem digunakan pada konteks budaya yang berbeda dari distribusi data pelatihannya.                                           |
| Data-centric AI         | Paradigma pengembangan kecerdasan buatan yang memprioritaskan peningkatan kualitas dan kurasi data untuk mencapai performa model yang lebih baik.                             |

|                          |                                                                                                                                                                            |
|--------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Data Leakage             | Kondisi di mana informasi dari data pengujian bocor ke dalam proses pelatihan sehingga hasil evaluasi menjadi bias dan tidak representatif.                                |
| Data-Driven Thresholding | Pendekatan penetapan ambang batas secara dinamis berdasarkan distribusi empiris data aktual, bukan nilai tetap yang ditentukan secara manual.                              |
| Early Stopping           | Teknik regularisasi pelatihan yang menghentikan proses pelatihan lebih awal ketika performa pada data validasi tidak lagi membaik.                                         |
| Explainability           | Kemampuan sistem kecerdasan buatan untuk memberikan penjelasan yang dapat dipahami manusia atas keputusan atau prediksi yang dihasilkan.                                   |
| Fail-Safe                | Mekanisme keamanan sistem yang memastikan kegagalan terjadi secara terkontrol tanpa menghasilkan informasi yang berbahaya atau menyesatkan.                                |
| Fine-tuning              | Proses melatih ulang model pra-latih pada dataset spesifik agar memiliki kemampuan adaptasi yang lebih baik terhadap domain target.                                        |
| Frozen Weights           | Kondisi di mana bobot parameter model dikunci dan tidak diperbarui selama proses pelatihan lanjutan.                                                                       |
| Gradient Accumulation    | Teknik manajemen memori GPU yang mengakumulasi nilai gradien dari beberapa mini-batch sebelum melakukan pembaruan bobot, untuk mensimulasikan batch size yang lebih besar. |
| Ground-truth             | Data referensi atau label kebenaran yang dihasilkan oleh manusia sebagai standar evaluasi performa model.                                                                  |
| Hallucination (AI)       | Kondisi di mana model kecerdasan buatan menghasilkan informasi yang tidak akurat atau tidak relevan dengan masukan yang diberikan.                                         |



|                     |                                                                                                                                                                 |
|---------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Heteroskedastik     | Kondisi statistik di mana varians dari suatu variabel tidak konstan di seluruh rentang observasi, sehingga menyulitkan penggunaan distribusi normal standar.    |
| Human-Centric Bias  | Kecenderungan model untuk membuat kesalahan identifikasi pada objek yang berkaitan dengan manusia akibat ketidakseimbangan proporsi kelas dalam data pelatihan. |
| Image Captioning    | Tugas dalam kecerdasan buatan untuk menghasilkan deskripsi teks secara otomatis berdasarkan konten visual dari sebuah citra.                                    |
| Inference           | Tahap di mana model yang telah dilatih digunakan untuk melakukan prediksi atau menghasilkan keluaran pada data baru.                                            |
| Interpolasi Lanczos | Metode resampling citra menggunakan fungsi Sinc yang dimodifikasi untuk mempertahankan detail frekuensi tinggi dan meminimalkan artefak aliasing.               |
| Latent Space        | Ruang representasi berdimensi tinggi tempat model menyimpan dan memproses abstraksi fitur dari data masukan.                                                    |
| Letterbox Padding   | Teknik pengubahan dimensi citra dengan menambahkan piksel netral bernilai nol pada sisi yang kosong untuk menjaga rasio aspek asli tetap utuh.                  |
| Motion Blur         | Efek buram pada gambar yang terjadi akibat pergerakan objek atau kamera selama proses pengambilan gambar.                                                       |
| Multimodal          | Sistem atau arsitektur yang mampu memproses dan mengintegrasikan informasi dari lebih dari satu jenis modalitas data, seperti teks dan gambar.                  |
| Overfitting         | Kondisi di mana model terlalu menyesuaikan diri dengan data pelatihan sehingga kehilangan kemampuan generalisasi pada data baru.                                |

|                       |                                                                                                                                                                                |
|-----------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Paradoks Kualitas     | Fenomena di mana model visi-bahasa mengalami penurunan performa saat memproses citra berkualitas rendah akibat bias yang dipelajari dari data pelatihan yang mengandung derau. |
| Pipeline              | Alur kerja sistematis yang menggabungkan beberapa tahap pemrosesan data atau komputasi secara berurutan dan terstruktur.                                                       |
| Prototype             | Purwarupa atau versi awal dari sistem yang dibangun untuk mendemonstrasikan fungsionalitas dan memvalidasi konsep desain.                                                      |
| Robust Cases          | Sampel data yang memiliki kualitas visual buruk namun tetap dipertahankan karena masih memiliki anotasi teks yang valid dan informatif.                                        |
| Rule-based Filtering  | Metode penyaringan data menggunakan aturan eksplisit yang telah ditetapkan sebelumnya berdasarkan kriteria terukur.                                                            |
| Self-Attention        | Mekanisme atensi di mana setiap elemen dalam sebuah urutan menghitung bobot korelasinya terhadap seluruh elemen lain dalam urutan yang sama.                                   |
| Semantic Resilience   | Ketangguhan model dalam mempertahankan pemahaman makna konteks objek meskipun citra masukan mengalami degradasi visual yang parah.                                             |
| Shortcut Learning     | Kecenderungan model untuk menggunakan pola-pola sederhana atau tidak relevan dalam data pelatihan alih-alih mempelajari fitur semantik yang mendalam.                          |
| Sub-word Tokenization | Pendekatan tokenisasi yang memecah kata-kata tidak dikenal menjadi unit sub-kata yang lebih kecil agar dapat direpresentasikan oleh model bahasa.                              |
| Transfer Learning     | Teknik pembelajaran mesin yang memanfaatkan pengetahuan dari model pra-latih pada satu domain untuk meningkatkan performa pada domain yang berbeda.                            |



|                        |                                                                                                                                              |
|------------------------|----------------------------------------------------------------------------------------------------------------------------------------------|
| Trustworthy AI         | Konsep kecerdasan buatan yang dapat dipercaya karena mampu memberikan transparansi, penjelasan, dan keamanan atas keputusan yang dihasilkan. |
| Underexposure          | Kondisi di mana citra terlalu gelap akibat kurangnya cahaya yang ditangkap sensor kamera selama proses perekaman.                            |
| Overexposure           | Kondisi di mana citra terlalu terang akibat cahaya berlebihan yang ditangkap sensor kamera, sehingga detail visual hilang.                   |
| User-generated Content | Konten yang dibuat dan diunggah secara langsung oleh pengguna akhir, bukan oleh profesional atau dalam kondisi terkontrol.                   |
| Viewfinder             | Layar bidik pada kamera yang digunakan untuk melihat dan membingkai objek sebelum pengambilan gambar dilakukan.                              |
| Vision-Language Model  | Arsitektur kecerdasan buatan yang dirancang untuk memahami dan mengintegrasikan informasi dari modalitas visual dan bahasa secara bersamaan. |
| Weight Decay           | Teknik regularisasi dalam pelatihan model yang memberikan penalti pada bobot berukuran besar untuk mencegah overfitting.                     |

## INTISARI

Model visi-bahasa modern saat ini masih mengalami kendala teknis yang signifikan saat memproses foto berkualitas rendah, berderau, atau kabur yang diambil secara autentik oleh individu tunanetra dalam skenario dunia nyata. Kesenjangan teknis ini memicu fenomena "Paradoks Kualitas" yang menyebabkan penurunan kinerja sistem pemandu visual secara drastis, sehingga berdampak negatif pada keamanan dan kemandirian navigasi pengguna. Penelitian ini bertujuan untuk mengeliminasi bias tersebut melalui pengembangan *pipeline* berbasis *data-centric* yang mengintegrasikan penyaringan kekaburan *Fast Fourier Transform* (FFT) dengan teknik *Parameter-Efficient Fine-Tuning* (PEFT) menggunakan *Low-Rank Adaptation* (LoRA). Langkah-langkah penyelesaian masalah mencakup kurasi data multidimensi, optimalisasi memori menggunakan pengoptimal AdamW pada kartu grafis tingkat konsumen, serta pengujian statistik mendalam melalui pembagian terbelah berbasis FFT. Hasil penelitian menunjukkan bahwa model yang dioptimalkan (V2-Final) berhasil mencapai skor CIDEr sebesar 0,4481, BERTScore F1 sebesar 0,8815, dan BLEU-4 sebesar 0,0511. Analisis statistik dengan nilai  $p = 0,622$  dan *Cohen's d* sebesar -0,038 membuktikan bahwa bias kualitas telah berhasil dihilangkan, menghasilkan performa yang seimbang antara citra buram dan tajam. Selain itu, sistem terbukti efisien secara komputasi dengan penggunaan VRAM puncak hanya sebesar 3,8 GB. Hasil penelitian ini dapat dimanfaatkan oleh pengembang teknologi bantu untuk menciptakan alat navigasi *real-time* yang tangguh bagi penyandang tunanetra. Penelitian selanjutnya disarankan untuk mengintegrasikan model ini dengan arsitektur *Large Language Model* (LLM) multimodal untuk interaksi percakapan yang lebih dinamis.

**Kata kunci:** aksesibilitas, BLIP, image captioning, kurasi data, LoRA.

## ABSTRACT

Modern vision-language models still encounter significant technical limitations when processing low-quality, noisy, or blurry photos authentically captured by visually impaired individuals in real-world scenarios. This technical gap triggers the "Quality Paradox" phenomenon, which drastically degrades the performance of visual guidance systems and negatively impacts the safety and navigational independence of the users. This study aims to eliminate this bias through the development of a data-centric pipeline that integrates Fast Fourier Transform (FFT) blur filtering with the Parameter-Efficient Fine-Tuning (PEFT) technique utilizing Low-Rank Adaptation (LoRA). The problem-solving steps include multi-dimensional data curation, memory optimization utilizing the AdamW optimizer on a consumer-grade graphics card, and rigorous statistical evaluation via FFT-based tertile splits. The results demonstrate that the optimized model (V2-Final) successfully achieves a CIDEr score of 0.4481, a BERTScore F1 of 0.8815, and a BLEU-4 of 0.0511. Statistical analysis yielding a p-value of 0.622 and a Cohen's d of -0.038 confirms that the quality bias has been successfully eradicated, resulting in a balanced performance across both severely blurry and crisp images. Additionally, the system proved to be computationally efficient, requiring a peak VRAM of only 3.8 GB. The findings of this research can be utilized by assistive technology developers to create robust, real-time navigational tools for the visually impaired. Future studies are recommended to integrate this model with multimodal Large Language Model (LLM) architectures to facilitate more dynamic conversational interactions.

**Keyword:** accessibility, BLIP, data curation, image captioning, LoRA.