

TESIS
**STUDI KOMPARATIF METODE PARAMETER-
EFFICIENT FINE-TUNING UNTUK ABSTRACTIVE
TEXT SUMMARIZATION**



Oleh:

Muhammad Syafiq Akmal

24206052002

PROGRAM STUDI INFORMATIKA
PROGRAM MAGISTER FAKULTAS SAINS DAN TEKNOLOGI
UIN SUNAN KALIJAGA
YOGYAKARTA

2026

PENGESAHAN TUGAS AKHIR



KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
FAKULTAS SAINS DAN TEKNOLOGI

Jl. Marsda Adisucipto Telp. (0274) 540971 Fax. (0274) 519739 Yogyakarta 55281

PENGESAHAN TUGAS AKHIR

Nomor : B-2073/Un.02/DST/PP.00.9/08/2026

Tugas Akhir dengan judul : Studi Komparatif Metode Parameter-Efficient Fine-Tuning untuk Abstractive Text Summarization

yang dipersiapkan dan disusun oleh:

Nama : MUHAMMAD SYAFIQ AKMAL, S.Kom.
Nomor Induk Mahasiswa : 24206052002
Telah diujikan pada : Selasa, 18 Agustus 2026
Nilai ujian Tugas Akhir : A-

dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta

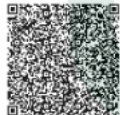
TIM UJIAN TUGAS AKHIR



Ketua Sidang

Dr. Agung Fatwanto, S.Si., M.Kom.
SIGNED

Valid ID: 6a8965beb8d2e



Penguji I

Dr. Agus Mulyanto, S.Si., M.Kom., ASEAN
Eng.
SIGNED

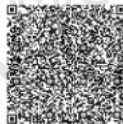
Valid ID: 6a893aef6b500f



Penguji II

Ir. Maria Ulfah Siregar, S.Kom., MIT., Ph.D.
SIGNED

Valid ID: 6a89162949584



Yogyakarta, 18 Agustus 2026
UIN Sunan Kalijaga
Dekan Fakultas Sains dan Teknologi

Prof. Dr. Dra. Hj. Khurul Wardati, M.Si.
SIGNED

Valid ID: 6a8b60b70f23f

SURAT PERNYATAAN KEASLIAN

SURAT PERNYATAAN KEASLIAN

Yang bertanda tangan di bawah ini:

Nama : Muhammad Syafiq Akmal
NIM : 24206052002
Jenjang : Magister
Program Studi : Informatika

menyatakan bahwa naskah tesis ini secara keseluruhan adalah hasil penelitian/karya saya sendiri, kecuali pada bagian-bagian yang dirujuk sumbernya.

Yogyakarta, 10 Agustus 2026

Saya yang menyatakan



Muhammad Syafiq Akmal
NIM: 24206052002

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

SURAT PERNYATAAN BEBAS PLAGIASI

SURAT PERNYATAAN BEBAS PLAGIASI

Yang bertanda tangan di bawah ini:

Nama : Muhammad Syafiq Akmal
NIM : 24206052002
Jenjang : Magister
Program Studi : Informatika

menyatakan bahwa naskah tesis ini secara keseluruhan benar-benar bebas dari plagiasi. Jika di kemudian hari terbukti melakukan plagiasi, maka saya siap ditindak sesuai ketentuan hukum yang berlaku

Yogyakarta, 10 Agustus 2026

Saya yang menyatakan



Muhammad Syafiq Akmal

NIM: 24206052002

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

NOTA DINAS PEMBIMBING

NOTA DINAS PEMBIMBING

Kepada Yth.,

Dekan Fakultas Sains dan teknologi
UIN Sunan Kalijaga
Yogyakarta

Assalamu 'alaikum wr. wb.

Disampaikan dengan hormat, setelah melakukan bimbingan, arahan, dan koreksi terhadap penulisan tesis yang berjudul:

**STUDI KOMPARATIF METODE PARAMETER-EFFICIENT FINE-TUNING
UNTUK ABSTRACTIVE TEXT SUMMARIZATION**

Yang ditulis oleh :

Nama : Muhammad Syafiq Akmal
NIM : 24206052002
Fakultas : Sains dan Teknologi
Jenjang : Program Studi Magister
Program Studi : Informatika

Saya berpendapat bahwa tesis tersebut sudah dapat diajukan kepada Program Studi Magister (S2) Informatika UIN Sunan Kalijaga untuk diujikan dalam rangka memperoleh gelar Magister Komputer.

Wassalamu'alaikum wr. wb.

Yogyakarta, 10 Agustus 2026

Pembimbing



Dr. Agung Farwanto, S.Si., M.Kom.
19770103 200501 003

SURAT PERSETUJUAN TUGAS AKHIR

Hal : Persetujuan Tugas Akhir

Lamp : -

Kepada:

Yth. Dekan Fakultas Sains dan Teknologi
UIN Sunan Kalijaga Yogyakarta
Di Yogyakarta

Assalamualaikum wr.wb.

Setelah membaca, meneliti, memberikan petunjuk dan mengoreksi serta mengadakan perbaikan seperlunya, maka saya selaku pembimbing berpendapat bahwa tesis saudara:

Nama : Muhammad Syafiq Akmal

NIM : 24206052002

Judul Tesis : Studi Komparatif Metode Parameter-Efficient Fine-Tuning
Untuk Abstractive Text Summarization

Sudah dapat diajukan kepada program studi Magister Informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta sebagai salah satu syarat untuk memperoleh gelar Magister Informatika.

Dengan ini saya mengharap agar tugas akhir tersebut di atas dapat segera dimunaqosyahkan. Atas perhatiannya saya ucapkan terimakasih.

Yogyakarta, 10 Agustus 2026

Pembimbing

Dr. Agung Fatwanto, S.Si., M.Kom.

NIP. 197701032005011003

ABSTRAK

Peringkasan teks abstraktif untuk Bahasa Indonesia masih menghadapi tantangan biaya komputasi yang tinggi ketika mengadaptasi model bahasa pralatih berskala besar melalui full fine-tuning, sementara penelitian terdahulu pada dataset XL-Sum Bahasa Indonesia baru menguji pendekatan full fine-tuning tanpa membandingkannya dengan metode Parameter-Efficient Fine-Tuning (PEFT) seperti LoRA, AdaLoRA, maupun QLoRA. Penelitian ini bertujuan membandingkan kualitas ringkasan dan efisiensi komputasi antara full fine-tuning dengan tiga metode PEFT tersebut pada backbone mT5-base untuk tugas peringkasan teks abstraktif Bahasa Indonesia. Eksperimen dilakukan pada subset Bahasa Indonesia dataset XL-Sum, dengan pembagian data latih, validasi, dan uji masing-masing sebesar 38.242, 4.780, dan 4.780 pasangan dokumen-ringkasan; kualitas ringkasan dievaluasi menggunakan ROUGE-1, ROUGE-2, ROUGE-L, BERTScore, dan METEOR disertai uji signifikansi statistik Wilcoxon dan bootstrap berpasangan, sementara efisiensi komputasi diukur melalui rasio parameter terlatih, penggunaan memori GPU puncak, waktu pelatihan, dan waktu inferensi. Hasil penelitian menunjukkan bahwa ketiga metode PEFT secara konsisten menghasilkan kualitas ringkasan yang lebih baik dibandingkan full fine-tuning pada seluruh metrik yang diukur, dengan QLoRA mencapai skor tertinggi pada sebagian besar metrik, sedangkan dari sisi efisiensi AdaLoRA paling efisien dari segi parameter dan waktu pelatihan, sementara QLoRA paling hemat memori GPU meski waktu pelatihannya paling lama. Temuan ini menunjukkan bahwa metode PEFT, khususnya LoRA dan QLoRA, dapat menjadi alternatif yang lebih efisien dan berkualitas dibandingkan full fine-tuning untuk peringkasan teks Bahasa Indonesia, meskipun pemilihan metode optimal tetap bergantung pada prioritas antara kualitas, efisiensi memori, atau kecepatan pelatihan, karena tidak ada satu metode yang unggul mutlak pada seluruh dimensi

Kata kunci: peringkasan teks abstraktif, Parameter-Efficient Fine-Tuning, LoRA, AdaLoRA, QLoRA, mT5, XL-Sum

ABSTRACT

Abstractive text summarization for Indonesian still faces high computational costs when adapting large pretrained language models through full fine-tuning, while prior studies on the Indonesian XL-Sum dataset have only tested full fine-tuning without comparing it to Parameter-Efficient Fine-Tuning (PEFT) methods such as LoRA, AdaLoRA, or QLoRA. This study aims to compare the summarization quality and computational efficiency of full fine-tuning against three PEFT methods, namely LoRA, AdaLoRA, and QLoRA, applied to the mT5-base backbone for Indonesian abstractive text summarization. Experiments were conducted on the Indonesian subset of the XL-Sum dataset, split into 38,242 training, 4,780 validation, and 4,780 test document-summary pairs; summarization quality was evaluated using ROUGE-1, ROUGE-2, ROUGE-L, BERTScore, and METEOR, supplemented with Wilcoxon signed-rank and paired bootstrap significance tests, while computational efficiency was measured through the trainable parameter ratio, peak GPU memory usage, training time, and inference time. The results show that all three PEFT methods consistently achieved better summarization quality than full fine-tuning across all measured metrics, with QLoRA reaching the highest scores on most metrics, while AdaLoRA was the most parameter- and training-time-efficient method and QLoRA was the most GPU-memory-efficient despite requiring the longest training time. These findings indicate that PEFT methods, particularly LoRA and QLoRA, can serve as more efficient and higher-quality alternatives to full fine-tuning for Indonesian text summarization, although the optimal method choice still depends on whether quality, memory efficiency, or training speed is prioritized, since no single method dominates across every dimension.

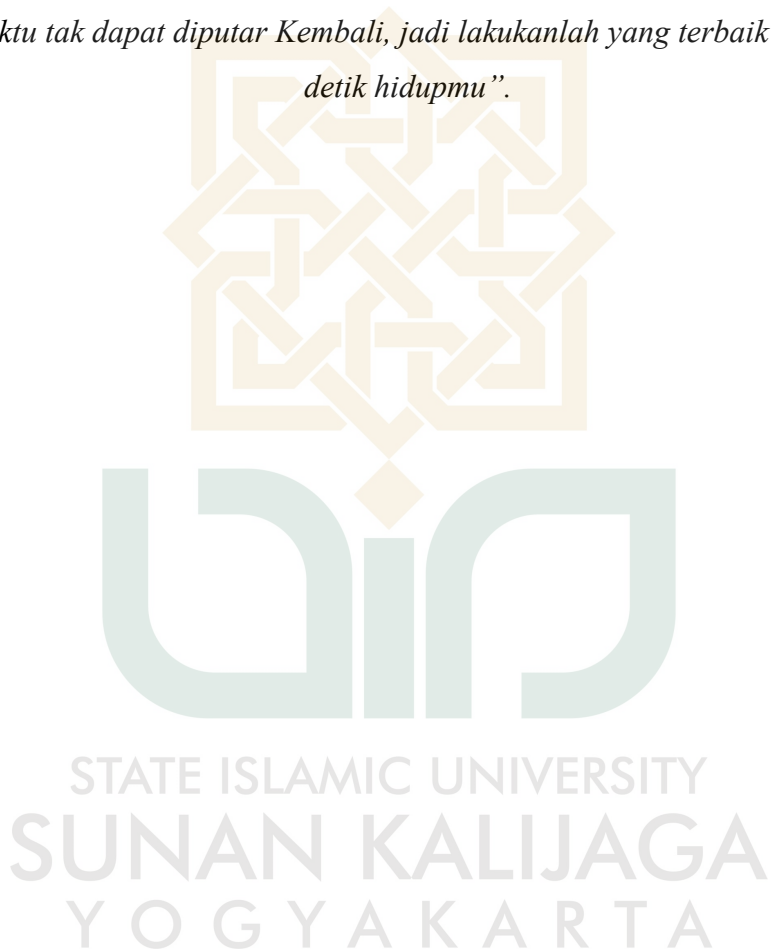
Keywords: abstractive text summarization, Parameter-Efficient Fine-Tuning, LoRA, AdaLoRA, QLoRA, mT5, XL-Sum

MOTTO

“Dan bersabarlah kamu sesungguhnya janji Allah adalah benar”

(Q.S. Ar-Rum: 60)

“Waktu tak dapat diputar Kembali, jadi lakukanlah yang terbaik disetiap detik hidupmu”.



PERSEMBAHAN

Dengan rasa syukur yang mendalam kepada Allah SWT, Yang telah memberikan rahmat, petunjuk, dan kekuatan sehingga karya ini dapat terselesaikan dengan baik.

Karya ini kupersembahkan untuk:

1. Orang tuaku tercinta, Moh. Syamsul Bally dan Hasnita, Terima kasih atas cinta, doa, dan pengorbanan tanpa henti yang selalu menjadi penyemangat dan pelita dalam setiap langkahku. Segala yang kuperoleh ini tak terlepas dari restu dan doa kalian.
2. Dosen pembimbing, Dr. Agung Fatwanto, S.Si., M.Kom., Terima kasih atas ilmu, arahan, dan bimbingan yang penuh kesabaran dan ketulusan. Nasihat dan dukungan Bapak menjadi bekal berharga dalam menyelesaikan karya ini.
3. Kepada Hanny Handayani Sucinta, yang telah memberikan dukungan, semangat, perhatian, dan motivasi selama proses penyusunan tesis ini. Terima kasih telah menjadi bagian dari perjalanan ini, baik dalam keadaan sulit maupun menyenangkan, serta selalu memberikan dorongan agar saya tetap kuat dan terus berusaha hingga tesis ini dapat diselesaikan.
4. Almamater tercinta, UIN Sunan Kalijaga Sebagai tempat aku belajar, bertumbuh, dan mengasah kemampuan, almamater ini adalah saksi perjuangan dan langkah-langkah kecilku menuju masa depan. Semoga karya ini menjadi awal kontribusiku untuk ilmu pengetahuan dan memberi manfaat bagi banyak orang. Aamiin.

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT atas limpahan rahmat, taufik, dan hidayah-Nya, sehingga penulis dapat menyelesaikan tesis yang berjudul "Studi Komparatif Metode Parameter-Efficient Fine-Tuning Untuk Abstractive Text Summarization" dengan baik. Tesis ini disusun sebagai salah satu syarat untuk memperoleh gelar Magister pada Program Studi Magister Informatika, UIN Sunan Kalijaga Yogyakarta.

Penulis menyadari bahwa penyusunan tesis ini tidak terlepas dari bantuan, bimbingan, dukungan, serta doa dari berbagai pihak. Oleh karena itu, dengan segala kerendahan hati, penulis menyampaikan terima kasih kepada:

1. Bapak Prof. Noorhaidi, S.Ag., M.A., M.Phil., Ph.D. selaku Rektor UIN Sunan Kalijaga Yogyakarta.
2. Bapak Prof. Dr. Dra. Hj. Khurul Wardati, M.Si. selaku Dekan Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta.
3. Bapak Dr. Ir. Sumarsono, S.T., M.Kom. selaku Ketua Program Studi Magister Informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta.
4. Dr. Agung Fatwanto, S.Si., M.Kom. selaku dosen pembimbing tesis yang telah banyak memberikan ilmu, arahan, bimbingan, motivasi, nasihat, dan saran dalam proses menyelesaikan tesis penulis.
5. Dr. Ir. Sumarsono, S.T., M.Kom. selaku dosen pembimbing akademik yang telah banyak memberikan ilmu, arahan, bimbingan, motivasi, nasihat, dan saran serta dukungan sejak awal perkuliahan.

6. Seluruh dosen program studi Magister Informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta yang telah ikhlas memberikan ilmu serta pengalaman pengetahuannya kepada penulis selama masa perkuliahan.
7. Kedua orang tua tercinta Bapak Moh. Syamsul Bally dan Hasnita serta Adik tersayang Muh. Naufal Akbar, inspirasi hidup penulis yang selalu memberikan dukungan moril, materil maupun non materil, perhatian, kasih sayang, dan senantiasa selalu melangitkan do'a untuk penulis.
8. Teman-teman seperjuangan kuliah di Fakultas Sains dan Teknologi Yogyakarta sekaligus responden dalam penelitian ini.

Semoga segala kebaikan yang diberikan dari semua pihak mendapatkan balasan dari Allah SWT. Penulis menyadari bahwa dalam penulisan tesis ini masih banyak kekurangan, sehingga kritik dan saran yang membangun dari pembaca akan sangat penulis hargai. Akhir kata, penulis berharap semoga tesis ini dapat bermanfaat.

Wassalamu'alaikum Warahmatullahi Wabarakatuh.

Yogyakarta, 10 Agustus 2026

Penulis,

Muhammad Syafiq Akmal
NIM 24206052002

DAFTAR ISI

PENGESAHAN TUGAS AKHIR	ii
SURAT PERNYATAAN KEASLIAN	iii
SURAT PERNYATAAN BEBAS PLAGIASI	iv
NOTA DINAS PEMBIMBING	v
SURAT PERSETUJUAN TUGAS AKHIR	vi
ABSTRAK	vii
ABSTRACT	viii
MOTTO	ix
PERSEMBAHAN	x
KATA PENGANTAR	xi
DAFTAR ISI	xiii
DAFTAR TABEL	xvi
DAFTAR GAMBAR	xvii
DAFTAR LAMPIRAN	xviii
BAB I PENDAHULUAN	1
A. Latar Belakang	1
B. Rumusan Masalah	5
C. Batasan Masalah	5
D. Tujuan Penelitian	6
E. Manfaat penelitian	6
BAB II KAJIAN PUSTAKA DAN LANDASAN TEORI	7
A. Kajian Pustaka	7
B. Landasan Teori	15
1. Abstractive Text Summarization	15
2. Model mT5	16
3. <i>Parameter Efficient Fine-Tuning (PEFT)</i>	18
4. Low-Rank Adaptation (LoRA)	18

5. AdaLoRA (<i>Adaptive Budget Allocation</i>)	19
6. QLoRA (<i>Quantized LoRA</i>).....	20
7. ROUGE.....	21
8. METEOR	22
9. BERTScore	23
10. Wilcoxon Signed-Rank Test.....	24
11. Paired Bootstrap	25
BAB III METODE PENELITIAN	27
A. Kerangka Penelitian.....	27
B. Alat dan Lingkungan Penelitian	29
C. Variabel dan Definisi Operasional Variabel Penelitian.....	31
D. Tahapan Penelitian.....	33
1. Instalasi Pustaka dan Pemeriksaan Lingkungan Komputasi.....	33
2. Pemuatan dan Tokenisasi Dataset.....	33
3. Pembagian Dataset (Data Split)	37
4. Konfigurasi dan Pemuatan Model sesuai Skenario Adaptasi ...	38
5. Pelatihan model dan Pencatatan Log	40
6. Evaluasi Model dan Data Uji.....	40
7. Uji Signifikansi Statistik	41
8. Dokumentasi dan Penyimpanan Hasil	42
BAB IV HASIL DAN PEMBAHASAN.....	43
A. Gambaran Umum Hasil Penelitian	43
B. Hasil Evaluasi Kualitas Ringkasan	44
C. Hasil Evaluasi Efisiensi Komputasi.....	55
D. Perbandingan dengan peneliti terdahulu.....	59
E. Analisis Trade-off Kualitas dan Efisiensi	62
BAB V PENUTUP	65
A. Kesimpulan	65
B. Saran	66

DAFTAR PUSTAKA.....	67
PERSETUJUAN TIM PENGUJI UJIAN TESIS.....	75
DATA RIWAYAT HIDUP	76
CURRICULUM VITAE (CV)	79



DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu.....	11
Tabel 3.1 Spesifikasi Lingkungan Penelitian	30
Tabel 3.2 contoh sampel hasil verifikasi tokenisasi	35
Tabel 3.3 Distribusi Pembagian Dataset.....	37
Tabel 3.4 Konfigurasi Hyperparameter per skenario Adaptasi	38
Tabel 4.1 Hasil Evaluasi Kualitas Ringkasan pada Data Uji	45
Tabel 4.2 Hasil Uji Signifikansi Statistik pada Subset 500 Sampel Acak	48
Tabel 4.3 Hasil Uji Signifikansi Statistik pada Seluruh 4.780 Data Uji ...	49
Tabel 4.4 Hasil Evaluasi Efisiensi Komputasi	56
Tabel 4.5 Perbandingan Hasil ROUGE dengan penelitian terdahulu	60



DAFTAR GAMBAR

Gambar 3.1 Kerangka Penelitian.....	27
Gambar 4.1 Perbandingan Skor ROUGE per Skenario Adaptasi	46
Gambar 4.2 Perbandingan BERTScore F1 dan METEOR per Skenario Adaptasi.....	47
Gambar 4.3 Peta Signifikansi Statistik pada Subset 500 Sampel Acak	51
Gambar 4.4 Peta Signifikansi Statistik pada Seluruh 4.780 Data Uji	52
Gambar 4.5 Perbandingan Nilai p antar-Metode PEFT: Subset 500 Sampel vs Seluruh Data Uji	53
Gambar 4.6 Profil Kualitas Ringkasan Ternormalisasi per Skenario Adaptasi.....	54
Gambar 4.7 Perbandingan Efisiensi Komputasi per Skenario Adaptasi ...	58

DAFTAR LAMPIRAN

Lampiran 1. Persetujuan Tim Penguji Tesis.....	75
Lampiran 2. Data riwayat hidup.....	78
Lampiran 3. Curriculum Vitae (CV)	79



BAB I

PENDAHULUAN

A. Latar Belakang

Pertumbuhan volume data teks digital dalam beberapa tahun terakhir berlangsung sangat pesat, sehingga kebutuhan akan peringkasan teks otomatis (*automatic text summarization*) meningkat signifikan untuk memahami inti dokumen panjang secara cepat dan efisien (H Shakil, Farooq and Kalita, 2024). Secara umum, peringkasan teks terbagi menjadi dua pendekatan, yaitu *extractive summarization*, yang menyusun ringkasan dengan memilih langsung kalimat-kalimat terpenting dari teks sumber tanpa mengubah redaksinya, dan *abstractive summarization*, yang menghasilkan kalimat baru yang berpotensi berbeda dari teks sumber (Giarelis, Mastrokostas and Karacapilidis, 2023). Ringkasan yang dihasilkan melalui pendekatan *abstractive* terbukti lebih menyerupai cara manusia meringkas dibandingkan pendekatan *extractive*, meskipun pendekatan ini masih menghadapi tantangan dalam membentuk representasi kontekstual yang memadai agar ringkasan yang dihasilkan benar-benar koheren (Hassan Shakil, Farooq and Kalita, 2024). Untuk mengatasi tantangan representasi kontekstual tersebut, perkembangan pesat model bahasa prelatih (*pretrained language models*) berbasis *deep learning* dalam beberapa tahun terakhir turut mendorong peningkatan signifikan pada berbagai tugas pemrosesan bahasa alami, termasuk kemampuan memahami dan menyintesis konteks pada peringkasan teks abstraktif (Giarelis, Mastrokostas and Karacapilidis, 2023).

Namun demikian, kemajuan tersebut sebagian besar dicapai melalui peningkatan skala model bahasa prelatih secara drastis, dari jutaan menjadi

miliaran parameter, sehingga di samping membuka kemampuan baru, ketersediaan daya komputasi yang semakin murah turut memunculkan tantangan baru dalam penerapannya (Patil and Gudivada, 2024). Mengadaptasi model berskala besar tersebut ke tugas tertentu secara konvensional dilakukan melalui *full fine-tuning*, yaitu memperbarui seluruh parameter model selama pelatihan, namun seiring bertambahnya ukuran model, pendekatan ini menjadi semakin mahal secara komputasi dan pada titik tertentu praktis tidak lagi memungkinkan untuk dilakukan (Ding *et al.*, 2023). *Full fine-tuning* pada model berskala puluhan hingga ratusan miliar parameter menuntut sumber daya GPU yang sangat besar, sehingga secara nyata membatasi partisipasi peneliti dengan infrastruktur komputasi terbatas dalam mengembangkan dan mengadaptasi model bahasa besar (Lv *et al.*, 2024). Keterbatasan inilah yang mendorong kebutuhan akan pendekatan adaptasi model yang jauh lebih hemat sumber daya komputasi dan memori, sebagai alternatif yang tetap kompetitif tanpa harus memperbarui seluruh parameter model (Sledzieski *et al.*, 2024).

Kebutuhan tersebut dijawab oleh *Parameter-Efficient Fine-Tuning* (PEFT), sebuah kelompok pendekatan yang membekukan mayoritas bobot model prelatih dan hanya melatih sebagian kecil parameter tambahan, sehingga secara signifikan menekan jumlah parameter yang perlu dilatih tanpa mengorbankan performa model secara berarti (Abou Baker, Rohrschneider and Handmann, 2024). Salah satu metode PEFT yang paling banyak digunakan adalah *Low-Rank Adaptation* (LoRA), yang membekukan bobot model prelatih dan menyisipkan matriks dekomposisi berperingkat rendah (*low-rank*) yang dapat dilatih ke dalam tiap lapisan arsitektur Transformer, sehingga secara drastis mengurangi jumlah parameter yang perlu diperbarui (Hu *et al.*, 2021). Sebagai pengembangan

lanjutan, AdaLoRA memperkenalkan alokasi anggaran peringkat (rank) secara adaptif antar-matriks bobot berdasarkan skor kepentingannya, sehingga distribusi parameter yang dilatih menjadi lebih efisien dibandingkan LoRA yang membagi anggaran secara merata (Zhang *et al.*, 2023). Varian lain, QLoRA, menggabungkan prinsip LoRA dengan kuantisasi bobot model prelatih ke presisi 4-bit, memungkinkan penyesuaian model dengan puluhan miliar parameter hanya menggunakan satu unit GPU tunggal tanpa mengorbankan performa dibandingkan *fine-tuning* presisi penuh (Dettmers *et al.*, 2023).

Hasan dkk. (2021) memperkenalkan XL-Sum, dataset peringkasan abstraktif multibahasa berskala besar yang mencakup 44 bahasa termasuk Bahasa Indonesia, sekaligus menetapkan baseline full fine-tuning multibahasa pada *backbone* mT5 sebagai acuan awal performa di seluruh bahasa yang dicakup (Hasan *et al.*, 2021). Astuti, Muljono dan Sutriawan (2024) menerapkan *fine-tuning* penuh pada model mBART khusus untuk subset Bahasa Indonesia dataset XL-Sum, mencapai ROUGE-1 sebesar 35,94, namun tanpa menyertakan metode PEFT sebagai pembanding (Astuti, Muljono and Sutriawan, 2024). Whitehouse dkk. (2024) membandingkan efektivitas LoRA terhadap *full fine-tuning* pada tugas peringkasan multibahasa, menemukan bahwa LoRA kompetitif pada kondisi data melimpah dan unggul pada skenario data terbatas maupun transfer lintas-bahasa, namun penelitian ini hanya menguji LoRA tanpa AdaLoRA maupun QLoRA dan tidak menyertakan Bahasa Indonesia (Whitehouse *et al.*, 2024). Di sisi lain, penelitian oleh Aharoni dkk. (2023) menyoroti tantangan konsistensi faktual pada ringkasan multibahasa di 45 bahasa dataset XLSum, sementara studi Uthus dkk. (2023) mengembangkan arsitektur mLongT5 yang lebih efisien untuk input

panjang pada tugas peringkasan dan tanya-jawab multibahasa, menunjukkan bahwa aspek adaptasi model secara efisien pada tugas peringkasan multibahasa masih menjadi area penelitian yang terus berkembang (Aharoni *et al.*, 2023),(Uthus *et al.*, 2023).

Kajian sistematis terhadap metode-metode PEFT menunjukkan bahwa sebagian besar studi pembandingan masih berfokus pada tugas dan bahasa bersumber daya tinggi seperti Bahasa Inggris, sehingga evaluasi menyeluruh yang mencakup beberapa varian PEFT sekaligus dalam satu kerangka pengujian masih jarang dilakukan (Xu *et al.*, 2023). Kesenjangan ini diperparah oleh temuan bahwa hanya sebagian kecil penelitian NLP berbasis model pralatih yang melaporkan pengujian signifikansi statistik atas perbedaan performa yang diklaim, sehingga keandalan perbandingan antar metode menjadi diragukan (Casola, Lauriola and Lavelli, 2022). Persoalan ini semakin terasa untuk Bahasa Indonesia, mengingat dari lebih dari 700 bahasa yang digunakan di Indonesia, mayoritasnya, termasuk Bahasa Indonesia sendiri dalam banyak tugas, masih kurang terwakili dalam sumber daya dan tolak ukur penelitian NLP (Aji *et al.*, 2022). Dengan demikian, studi komprehensif yang membandingkan beberapa metode PEFT secara bersamaan disertai validasi statistik yang memadai, khususnya untuk bahasa bersumber daya terbatas, masih menjadi arah penelitian yang perlu ditindaklanjuti (Z. Han *et al.*, 2024).

Untuk menjawab kesenjangan tersebut, penelitian ini secara sistematis membandingkan *full fine-tuning* dengan tiga varian PEFT, yaitu LoRA, AdaLoRA, dan QLoRA, pada arsitektur mT5-base yang mendukung Bahasa Indonesia sebagai salah satu dari bahasa yang dicakup pralatihannya (Xue *et al.*, 2021). Kualitas ringkasan yang dihasilkan setiap metode dievaluasi menggunakan metrik ROUGE, yang dilengkapi dengan

pengujian signifikansi statistik untuk memastikan bahwa perbedaan performa antar metode bukan sekadar kebetulan, melainkan perbedaan yang benar-benar bermakna secara statistik. Perbandingan dilakukan secara menyeluruh, mencakup aspek kualitas hasil ringkasan sekaligus efisiensi komputasi seperti kebutuhan memori dan waktu pelatihan, sehingga trade-off antar metode dapat terlihat secara jelas. Dengan berfokus secara khusus pada Bahasa Indonesia, penelitian ini diharapkan memberikan bukti empiris yang dapat menjadi rujukan bagi pemilihan metode adaptasi model bahasa besar yang efisien untuk tugas peringkasan teks berbahasa Indonesia.

B. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

- 1) Bagaimana perbandingan kualitas ringkasan (ROUGE-1, ROUGE-2, ROUGE-L, BERTScore, METEOR) antara model mT5-base yang diadaptasi menggunakan *Full Fine-Tuning*, LoRA, AdaLoRA, dan QLoRA pada subset Bahasa Indonesia dataset XL-Sum?
- 2) Bagaimana perbandingan efisiensi komputasi (rasio parameter terlatih, waktu pelatihan, konsumsi memori GPU puncak, waktu inferensi) di antara keempat metode tersebut?

C. Batasan Masalah

Agar penelitian lebih terarah, maka batasan masalah dalam penelitian ini adalah sebagai berikut:

- 1) Model pralatih dibatasi pada mT5-base, tanpa perbandingan dengan mBART50 atau mLongT5 meski keduanya dijadikan bahan kajian teoretis.
- 2) Dataset dibatasi pada subset Bahasa Indonesia dataset XL-Sum, dengan split 38.242/4.780/4.780 data latih/validasi/uji.

- 3) Metode adaptasi dibatasi pada *Full Fine-Tuning*, LoRA, AdaLoRA, dan QLoRA.
- 4) Evaluasi kualitas dibatasi pada metrik otomatis (ROUGE, BERTScore, METEOR), tanpa evaluasi manusia terhadap koherensi maupun faktualitas ringkasan.

D. Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

- 1) Mengukur dan membandingkan kualitas ringkasan serta efisiensi komputasi *Full Fine-Tuning*, LoRA, AdaLoRA, dan QLoRA pada mT5-base dengan dataset XL-Sum Bahasa Indonesia.
- 2) Menganalisis signifikansi statistik perbedaan kualitas ringkasan antar-metode menggunakan uji *Wilcoxon signed-rank* dan *paired bootstrap*.
- 3) Merumuskan rekomendasi metode adaptasi model yang paling sesuai untuk peringkasan teks Bahasa Indonesia dalam kondisi sumber daya komputasi terbatas.

E. Manfaat penelitian

Manfaat yang diharapkan dari penelitian ini adalah sebagai berikut:

- 1) Untuk Praktisi NLP: memberikan acuan praktis *trade-off* kualitas-efisiensi antar-metode adaptasi model untuk proyek dengan keterbatasan GPU.
- 2) Untuk Akademisi: memperkaya literatur studi komparatif PEFT untuk *abstractive summarization* Bahasa Indonesia, yang masih terbatas dibanding studi multibahasa umum.
- 3) Untuk Masyarakat/Industri: mendukung pengembangan sistem peringkasan teks otomatis Bahasa Indonesia yang lebih hemat biaya komputasi.

BAB V

PENUTUP

A. Kesimpulan

Penelitian ini menyimpulkan bahwa ketiga metode Parameter-Efficient Fine-Tuning (PEFT), yakni LoRA, AdaLoRA, dan QLoRA, secara konsisten menghasilkan kualitas ringkasan yang lebih baik dan signifikan secara statistik dibandingkan pendekatan Full Fine-Tuning pada seluruh metrik evaluasi. Hal ini dibuktikan oleh capaian QLoRA yang mencatat skor ROUGE-1 tertinggi sebesar 31,46, disusul LoRA (31,33) dan AdaLoRA (30,35), yang semuanya jauh melampaui skor Full Fine-Tuning (24,36). Meskipun secara kualitas PEFT sangat unggul, evaluasi efisiensi komputasinya justru menunjukkan pola trade-off yang jelas karena tidak ada satu pun metode yang memenangkan seluruh dimensi sumber daya. Sebagai bukti, AdaLoRA terbukti paling cepat dilatih dengan jumlah parameter terkecil, QLoRA paling hemat memori GPU meskipun waktu pelatihannya paling lambat, sementara Full Fine-Tuning justru tetap menjadi yang tercepat pada tahap inferensi.

Berdasarkan analisis trade-off tersebut, tidak ada satu metode tunggal yang unggul mutlak, sehingga perumusan rekomendasi adaptasi model harus disesuaikan sepenuhnya dengan prioritas dan limitasi sumber daya komputasi yang ada. Jika sistem dihadapkan pada keterbatasan memori (VRAM) yang sangat ketat, QLoRA menjadi pilihan utama karena mampu memangkas kebutuhan memori secara drastis meskipun waktu pelatihannya menjadi lebih lama. Namun, apabila siklus eksperimen menuntut kecepatan waktu training yang maksimal dengan modifikasi parameter terminimal, maka AdaLoRA adalah opsi yang paling direkomendasikan. Terakhir, jika

mencari titik keseimbangan terbaik (balanced) antara kualitas ringkasan yang tinggi, waktu pelatihan yang wajar, serta kecepatan inferensi yang mumpuni, maka model LoRA adalah metode yang paling ideal untuk digunakan.

B. Saran

penelitian ini merekomendasikan beberapa arahan strategis untuk pengembangan selanjutnya. Pertama, perlu dilakukan ablation study dengan memvariasikan ukuran nilai rank pada metode PEFT guna menguji konsistensi trade-off antara kualitas dan efisiensi di berbagai tingkat kompresi parameter. Selain itu, perluasan komparasi arsitektur backbone ke model lain seperti mBART50, serta pengujian pada domain dataset yang lebih kompleks di luar teks berita, sangat diperlukan untuk mengukur kemampuan generalisasi metode adaptasi ini. Mengingat evaluasi saat ini baru sebatas metrik otomatis, pelibatan anotator manusia juga sangat disarankan di masa mendatang untuk mengukur aspek krusial seperti koherensi dan konsistensi faktual ringkasan secara lebih akurat. Sebagai penutup, saya sangat menyarankan agar penelitian-penelitian selanjutnya mulai menstandarisasi pelaporan metrik efisiensi komputasi seperti jejak VRAM dan waktu pelatihan agar komparasi antar-studi dapat dievaluasi secara komprehensif, tidak hanya dari segi kualitas, tetapi juga ketahanan komputasinya.

DAFTAR PUSTAKA

- Abou Baker, N., Rohrschneider, D. and Handmann, U. (2024) 'Parameter-Efficient Fine-Tuning of Large Pretrained Models for Instance Segmentation Tasks', *Machine Learning and Knowledge Extraction*, 6(4), p. 133. Available at: <https://doi.org/10.3390/make6040133>.
- Aharoni, R. *et al.* (2023) 'Multilingual Summarization with Factual Consistency Evaluation', in *Findings of the Association for Computational Linguistics: ACL 2023*. Available at: <https://doi.org/10.18653/v1/2023.findings-acl.220>.
- Aji, A.F. *et al.* (2022) 'One Country, 700+ Languages: NLP Challenges for Underrepresented Languages and Dialects in Indonesia', in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Available at: <https://doi.org/10.18653/v1/2022.acl-long.500>.
- Akter, M., Bansal, N. and Karmaker, S.K. (2022) 'Revisiting Automatic Evaluation of Extractive Summarization Task: Can We Do Better than ROUGE?', in *Findings of the Association for Computational Linguistics: ACL 2022*. Available at: <https://doi.org/10.18653/v1/2022.findings-acl.122>.
- Astuti, R.H., Muljono, M. and Sutriawan, S. (2024) 'Indonesian News Text Summarization Using MBART Algorithm', *Scientific Journal of Informatics* [Preprint]. Available at: <https://doi.org/10.15294/sji.v11i1.49224>.
- Banerjee, S. and Lavie, A. (2005) 'METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments', *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation*

Measures for Machine Translation and/or Summarization [Preprint].

Beken Fikri, F., Oflazer, K. and Yanıkoğlu, B. (2021) ‘Semantic Similarity Based Evaluation for Abstractive News Summarization’, in *Proceedings of the 1st Workshop on Natural Language Generation, Evaluation, and Metrics (GEM 2021)*. Available at: <https://doi.org/10.18653/v1/2021.gem-1.3>.

Bestgen, Y. (2022) ‘Please, Don’t Forget the Difference and the Confidence Interval when Seeking for the State-of-the-Art Status’, in *Proceedings of the Thirteenth Language Resources and Evaluation Conference (LREC 2022)*. Available at: <https://doi.org/10.63317/2gavpzzwqkcx>.

Casola, S., Lauriola, I. and Lavelli, A. (2022) ‘Pre-trained transformers: an empirical comparison’, *Machine Learning with Applications* [Preprint]. Available at: <https://doi.org/10.1016/j.mlwa.2022.100334>.

Colombo, P. *et al.* (2021) ‘Automatic Text Evaluation through the Lens of Wasserstein Barycenters’, in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP 2021)*. Available at: <https://doi.org/10.18653/v1/2021.emnlp-main.817>.

Daraghmi, E., Atwe, L. and Jaber, A. (2025) ‘A Comparative Study of PEGASUS, BART, and T5 for Text Summarization Across Diverse Datasets’, *Future Internet* [Preprint]. Available at: <https://doi.org/10.3390/fi17090389>.

Denkowski, M. and Lavie, A. (2014) ‘Meteor Universal: Language Specific Translation Evaluation for Any Target Language’, in *Proceedings of the Ninth Workshop on Statistical Machine Translation*. Available at: <https://doi.org/10.3115/v1/w14-3348>.

Dettmers, T. *et al.* (2023) ‘QLORA: Efficient Finetuning of Quantized LLMs’, *Advances in Neural Information Processing Systems*, 36.

Deutsch, D., Dror, R. and Roth, D. (2021) ‘A Statistical Analysis of Summarization Evaluation Metrics Using Resampling Methods’, *Transactions of the Association for Computational Linguistics* [Preprint]. Available at: https://doi.org/10.1162/tacl_a_00417.

Deutsch, D., Dror, R. and Roth, D. (2022) ‘Re-Examining System-Level Correlations of Automatic Summarization Evaluation Metrics’, in *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Available at: <https://doi.org/10.18653/v1/2022.naacl-main.442>.

Ding, N. *et al.* (2023) ‘Parameter-efficient fine-tuning of large-scale pre-trained language models’, *Nature Machine Intelligence* [Preprint]. Available at: <https://doi.org/10.1038/s42256-023-00626-4>.

Giarelis, N., Mastrokostas, C. and Karacapilidis, N. (2023) ‘Abstractive vs. Extractive Summarization: An Experimental Review’, *Applied Sciences* [Preprint]. Available at: <https://doi.org/10.3390/app13137620>.

Han, R. *et al.* (2024) ‘Rethinking Efficient Multilingual Text Summarization Meta-Evaluation’, in *Findings of the Association for Computational Linguistics: ACL 2024*. Available at: <https://doi.org/10.18653/v1/2024.findings-acl.930>.

Han, Z. *et al.* (2024) ‘Parameter-Efficient Fine-Tuning for Large Models: A Comprehensive Survey’, *arXiv preprint* [Preprint]. Available at: <https://doi.org/10.48550/arxiv.2403.14608>.

Hasan, T. *et al.* (2021) ‘XL-Sum: Large-Scale Multilingual Abstractive Summarization for 44 Languages’, in *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*. Available at: <https://doi.org/10.18653/v1/2021.findings-acl.413>.

He, J., Berg-kirkpatrick, T. and Neubig, G. (2022) 'TOWARDS A UNIFIED VIEW OF PARAMETER-EFFICIENT TRANSFER LEARNING', pp. 1–15.

Hu, J.E. *et al.* (2021) 'LoRA: Low-Rank Adaptation of Large Language Models', *arXiv preprint arXiv:2106.09685* [Preprint]. Available at: <https://doi.org/10.48550/arXiv.2106.09685>.

Itsnaini, Q.A. *et al.* (2023) 'Abstractive Text Summarization using Pre-Trained Language Model "Text-to-Text Transfer Transformer (T5)"', *ILKOM Jurnal Ilmiah* [Preprint]. Available at: <https://doi.org/10.33096/ilkom.v15i1.1532.124-131>.

Kim, H.-S., Kang, J.-W. and Choi, S.-Y. (2025) 'ChatGPT vs. Human Journalists: Analyzing News Summaries Through BERTScore and Moderation Standards', *Electronics* [Preprint]. Available at: <https://doi.org/10.3390/electronics14112115>.

Koehn, P. (2004) 'Statistical Significance Tests for Machine Translation Evaluation', *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP)* [Preprint].

Koto, F., Lau, J.H. and Baldwin, T. (2021) 'Evaluating the Efficacy of Summarization Evaluation across Languages', in *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*. Available at: <https://doi.org/10.18653/v1/2021.findings-acl.71>.

Kumar, S., Solanki, A. and Jhanjhi, N.Z. (2024) 'ROUGE-SS: A New ROUGE Variant for the Evaluation of Text Summarization', *Recent Advances in Computer Science and Communications* [Preprint]. Available at: <https://doi.org/10.2174/0126662558304595240528111535>.

Lin, C.-Y. (2004) 'ROUGE: A Package for Automatic Evaluation of Summaries', *Text Summarization Branches Out: Proceedings of the ACL-*

04 Workshop [Preprint].

Lv, K. *et al.* (2024) 'Full Parameter Fine-tuning for Large Language Models with Limited Resources', in *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (ACL 2024)*. Association for Computational Linguistics. Available at: <https://doi.org/10.18653/v1/2024.acl-long.445>.

Muharam, A.F. *et al.* (2025) 'Enhancing Abstractive Multi-Document Summarization with Bert2Bert Model for Indonesian Language', *JISKA (Jurnal Informatika Sunan Kalijaga)* [Preprint]. Available at: <https://doi.org/10.14421/jiska.2025.10.1.110-121>.

Nasution, S., Ferdiana, R. and Hartanto, R. (2025) 'Towards Two-Step Fine-Tuned Abstractive Summarization for Low-Resource Language Using Transformer T5', *International Journal of Advanced Computer Science and Applications* [Preprint]. Available at: <https://doi.org/10.14569/ijacsa.2025.01602120>.

Patil, R. and Gudivada, V.N. (2024) 'A Review of Current Trends, Techniques, and Challenges in Large Language Models (LLMs)', *Applied Sciences*, 14(5), p. 2074. Available at: <https://doi.org/10.3390/app14052074>.

Peyrard, M. *et al.* (2021) 'Better than Average: Paired Evaluation of NLP Systems', in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*. Available at: <https://doi.org/10.18653/v1/2021.acl-long.179>.

Putri, I., Trisna, I.N.P. and Rusjayanthi, N.K.D. (2025) 'Abstractive Text Summarization to Generate Indonesian News Highlight Using Transformers Model', *Journal of Information Systems and Informatics*

[Preprint]. Available at: <https://doi.org/10.51519/journalisi.v7i2.1082>.

Rainio, O., Teuho, J. and Klén, R. (2024) 'Evaluation Metrics and Statistical Tests for Machine Learning', *Scientific Reports* [Preprint]. Available at: <https://doi.org/10.1038/s41598-024-56706-x>.

Roy, D., Fakhoury, S. and Arnaoudova, V. (2021) 'Reassessing Automatic Evaluation Metrics for Code Summarization Tasks', in *Proceedings of the 29th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*. Available at: <https://doi.org/10.1145/3468264.3468588>.

Saadany, H. and Orasan, C. (2021) 'BLEU, METEOR, BERTScore: Evaluation of Metrics Performance in Assessing Critical Translation Errors in Sentiment-oriented Text', in *Proceedings of the Translation and Interpreting Technology Online Conference (TRITON 2021)*. Available at: https://doi.org/10.26615/978-954-452-071-7_006.

Shakil, H, Farooq, A. and Kalita, J. (2024) 'Abstractive text summarization: State of the art, challenges, and improvements', *Neurocomputing* [Preprint]. Available at: <https://doi.org/10.1016/j.neucom.2024.128255>.

Shakil, Hassan, Farooq, A. and Kalita, J. (2024) 'Abstractive text summarization: State of the art, challenges, and improvements', *Neurocomputing* [Preprint]. Available at: <https://doi.org/10.1016/j.neucom.2024.128255>.

Sledzieski, S. *et al.* (2024) 'Democratizing protein language models with parameter-efficient fine-tuning', *Proceedings of the National Academy of Sciences* [Preprint]. Available at: <https://doi.org/10.1073/pnas.2405840121>.

Suleiman, D. and Awajan, A. (2020) 'Deep Learning Based Abstractive Text Summarization: Approaches, Datasets, Evaluation Measures, and Challenges', *Mathematical Problems in Engineering* [Preprint]. Available

at: <https://doi.org/10.1155/2020/9365340>.

Tang, G., Yousuf, O. and Jin, Z. (2024) 'Improving BERTScore for Machine Translation Evaluation Through Contrastive Learning', *IEEE Access* [Preprint]. Available at: <https://doi.org/10.1109/access.2024.3406993>.

Transformer, T., Raffel, C. and Liu, P.J. (2020) 'Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer', 21, pp. 1–67.

Uthus, D. *et al.* (2023) 'mLongT5: A Multilingual and Efficient Text-To-Text Transformer for Longer Sequences', in *Findings of the Association for Computational Linguistics: EMNLP 2023*. Available at: <https://doi.org/10.18653/v1/2023.findings-emnlp.628>.

Vaswani, A. (2017) 'Attention Is All You Need', (Nips).

Vu, T. *et al.* (2022) 'Overcoming Catastrophic Forgetting in Zero-Shot Cross-Lingual Generation', pp. 9279–9300.

Wainer, J. (2022) 'A Bayesian Bradley-Terry Model to Compare Multiple ML Algorithms on Multiple Data Sets', *arXiv (Cornell University)* [Preprint]. Available at: <https://doi.org/10.48550/arxiv.2208.04935>.

Wein, S. *et al.* (2023) 'Follow the Leader(board) with Confidence: Estimating p-values from a Single Test Set with Item and Response Variance', in *Findings of the Association for Computational Linguistics: ACL 2023*. Available at: <https://doi.org/10.18653/v1/2023.findings-acl.196>.

Whitehouse, C. *et al.* (2024) 'Low-Rank Adaptation for Multilingual Summarization: An Empirical Study', in *Findings of the Association for Computational Linguistics: NAACL 2024*. Available at: <https://doi.org/10.18653/v1/2024.findings-naacl.77>.

Wu, Y. *et al.* (2025) 'Can Large Language Models Serve as Evaluators for Code Summarization?', *IEEE Transactions on Software Engineering*

[Preprint]. Available at: <https://doi.org/10.1109/tse.2025.3595283>.

Xu, L. *et al.* (2023) 'Parameter-Efficient Fine-Tuning Methods for Pretrained Language Models: A Critical Review and Assessment', *arXiv preprint* [Preprint]. Available at: <https://doi.org/10.48550/arxiv.2312.12148>.

Xue, L. *et al.* (2021) 'mT5: A Massively Multilingual Pre-trained Text-to-Text Transformer', in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Available at: <https://doi.org/10.18653/v1/2021.naacl-main.41>.

Yuan, W., Neubig, G. and Liu, P. (2021) 'BARTScore: Evaluating Generated Text as Text Generation', *arXiv (Cornell University)* [Preprint]. Available at: <https://doi.org/10.48550/arxiv.2106.11520>.

Zhang, Q. *et al.* (2023) 'AdaLoRA: Adaptive Budget Allocation for Parameter-Efficient Fine-Tuning', *arXiv preprint arXiv:2303.10512* [Preprint]. Available at: <https://doi.org/10.48550/arXiv.2303.10512>.

Zhang, T. *et al.* (2020) 'BERTScore: Evaluating Text Generation with BERT', *arXiv (Cornell University)* [Preprint].

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA