

Pembuatan Data Sintetis Untuk *Software Quality Assurance* (QA) Menggunakan *Generative AI Large Language Model*

¹Alim Purnomo Aji Surya M.R., ²Hilmy Aliy Andra Putra, ³Gugun Gunadi
Ilmu Komputer, Universitas Djuanda, Bogor

E-mail: ¹alimsuryam.r@gmail.com, ²hilmy.aliy@unida.ac.id, ³gugun.gunadi@unida.ac.id

ABSTRAK

Pengujian perangkat lunak (*Software Quality Assurance/QA*) modern dalam ekosistem pengembangan yang tangkas (*Agile*) membutuhkan volume data uji yang masif, dinamis, dan representatif. Namun, pemanfaatan data produksi riil secara langsung menghadapi rintangan operasional dan kendala regulasi privasi yang ketat, seperti Undang-Undang Pelindungan Data Pribadi (UU PDP) Nomor 27 Tahun 2022. Teknik anonimisasi konvensional terbukti tidak memadai karena seringkali merusak integritas relasional statistik, sementara model generatif *Deep Learning* tradisional seperti *Generative Adversarial Networks* (GANs) rentan terhadap anomali *mode collapse* pada data kategorikal berdimensi tinggi serta menuntut biaya komputasi yang mahal. Penelitian ini mengusulkan kerangka kerja komputasi inovatif menggunakan pendekatan *zero-shot inference* berbasis *Large Language Model* (LLM), secara spesifik menggunakan arsitektur Gemini 3, untuk membangkitkan data uji sintetis berformat tabular murni tanpa melalui proses pelatihan ulang (*fine-tuning*). Evaluasi kinerja dilakukan pada 1.000 baris kerangka data uji demografi dan finansial (berdasar parameter dataset *Glassdoor Gender Pay Gap*). Analisis fidelitas data diukur menggunakan metrik analitik *Cosine Similarity*, *Jensen-Shannon (JS) Divergence*, dan uji jarak *Kolmogorov-Smirnov* (KS). Hasil eksperimen menunjukkan performa presisi tingkat tinggi: LLM mereplikasi distribusi data kategorikal dengan rata-rata kemiripan *Cosine* mencapai 99,38% (mengeliminasi halusinasi leksikon) dan stabilitas proporsi populasi *JS-Divergence* sebesar 96,52%. Pada dimensi numerikal kontinu, model mempertahankan kepatuhan batas logis (*boundary constraints*) dengan tingkat akurasi jarak KS rata-rata sebesar 89,88%. Pendekatan ini berhasil mendobrak batasan model terdahulu, menghadirkan terobosan solusi pengadaan data uji industri yang efisien secara biaya, berskala besar, seketika (*real-time API*), dan menjamin 100% kepatuhan privasi data.

Kata kunci : *Data Sintetis, Generative AI, Large Language Model, Quality Assurance*

ABSTRACT

Modern software testing (*Quality Assurance/QA*) within agile development ecosystems requires massive, dynamic, and highly representative volumes of test data. However, the direct utilization of real production data faces significant operational hurdles and strict privacy regulations, such as Indonesia's Personal Data Protection Law (PDP) No. 27 of 2022. Conventional anonymization techniques have proven inadequate as they frequently destroy statistical relational integrity, while traditional generative Deep Learning models like *Generative Adversarial Networks* (GANs) are highly susceptible to *mode collapse* anomalies on high-dimensional categorical data and demand expensive computational costs. This study proposes an innovative computational framework employing a *zero-shot inference* approach based on a *Large Language Model* (LLM), specifically utilizing the Gemini 3 architecture, to generate pure tabular-formatted synthetic test data without the need for iterative *fine-tuning* processes. Performance evaluation was conducted on 1,000 rows of demographic and financial test data (parameterized from the Glassdoor Gender Pay Gap dataset). Data fidelity analysis was measured using *Cosine Similarity*, *Jensen-Shannon (JS) Divergence*, and the *Kolmogorov-Smirnov* (KS) distance metric. Experimental results demonstrate a high-level

precision performance: the LLM successfully replicated categorical data distributions with an average *Cosine* similarity reaching 99.38% (eliminating lexical hallucination) and achieved a *JS-Divergence* population proportion stability of 96.52%. In the continuous numerical dimension, the model maintained logical boundary compliance with an average KS distance accuracy rate of 89.88%. This approach successfully breached the limitations of predecessor models, presenting a breakthrough industrial test data procurement solution that is cost-efficient, massively scalable, instantaneous (*real-time API*), and guarantees 100% data privacy compliance.

Keyword : Generative AI, Large Language Model, Quality Assurance, Synthetic Data, Zero-Shot Inference

1. PENDAHULUAN

Dalam dunia pengembangan perangkat lunak, proses pengujian aplikasi melalui *Quality Assurance (QA)* memegang peranan vital untuk menjamin stabilitas, keandalan, dan kualitas sistem sebelum dirilis. Setiap aplikasi memerlukan verifikasi dan validasi yang ketat guna memastikan fungsi berjalan sesuai kebutuhan fungsional maupun non-fungsional, serta bebas dari kesalahan (*bug*) yang dapat merusak pengalaman pengguna. Kegagalan dalam proses *QA* berpotensi menimbulkan kerugian finansial, merusak reputasi merek, hingga menurunkan kepercayaan pengguna, karena *software bugs, errors, dan defects* dapat membawa konsekuensi yang berbahaya (Taley & Pathak, 2020). Komponen fundamental dalam *QA* adalah ketersediaan data uji (*test data*) yang berfungsi sebagai bahan bakar pengujian komprehensif, mulai dari functional testing yang memvalidasi logika bisnis hingga stress testing yang mencari titik kegagalan sistem. Tanpa data uji yang representatif dan beragam, pengujian tidak dapat mencerminkan kondisi nyata di lingkungan produksi. Namun, penggunaan data produksi asli sering kali terkendala karena mengandung informasi sensitif dan rahasia yang tunduk pada regulasi privasi ketat seperti *General Data Protection Regulation (GDPR)* di Eropa atau Undang-Undang Perlindungan Data Pribadi (UU PDP) di Indonesia. Pelanggaran terhadap batasan penggunaan informasi sensitif ini dapat mengakibatkan sanksi hukum berat, seperti rekor denda sebesar \$1.28 miliar yang dialami Meta pada tahun 2023 (Yuan et al. 2025). Teknik tradisional seperti anonimisasi sering kali merusak integritas relasional dan menyebabkan data kehilangan korelasi statistik yang penting. Selain itu, metode

pengumpulan data konvensional memiliki kelemahan terkait cakupan edge case yang tidak memadai serta biaya usaha yang tinggi (Rusum & Anasuri, 2023). Sebagai solusi, muncul teknologi data sintetis—data buatan sistem komputer yang meniru pola dan karakteristik data nyata tanpa mengandung informasi pribadi (Rusum & Anasuri, 2023). Data sintetis memberikan fleksibilitas bagi tim *QA* untuk menciptakan skenario pengujian ekstrem yang mungkin tidak ada dalam data aktual, sehingga meningkatkan cakupan pengujian secara drastis (Rusum & Anasuri, 2023). Perkembangan *Large Language Model (LLM)* dan *Generative AI* telah merevolusi cara data sintetis dihasilkan, dengan potensi peningkatan produktivitas pekerja antara 10% hingga 56% (Filippucci et al., 2024). Penggunaan alat bantu seperti *GitHub Copilot* bahkan terbukti mempercepat penyelesaian tugas pengembang hingga 55.8% (Peng et al., 2023). Berbeda dengan model generatif terdahulu seperti *Generative Adversarial Networks (GANs)* yang sulit dilatih dan rentan terhadap mode collapse, atau *Variational Autoencoders (VAEs)* yang hasilnya sering kurang presisi (Rusum & Anasuri, 2023), *LLM modern* mampu memahami konteks semantik secara mendalam. Penelitian ini mengisi kesenjangan pada metode evaluasi *QA* yang sebelumnya hanya berfokus pada metrik statistik tanpa menangkap kelayakan semantik (Rusum & Anasuri, 2023). Kerangka kerja yang diusulkan mengimplementasikan pendekatan *zero-shot inference* menggunakan *API gemini-3-flash-preview*. Inti dari framework ini bertumpu pada mekanisme *QA* kuantitatif berlapis yang mengevaluasi data sintetis menggunakan instrumen statistik multivariat seperti *Kolmogorov-Smirnov Test (KS-Test)*, *Jensen-Shannon Divergence (JS-Divergence)*, serta pengukuran jarak semantik

melalui *Cosine Similarity*. Integrasi ini diharapkan mampu mengatasi hambatan ketersediaan data uji yang aman sekaligus membuktikan secara empiris bahwa efisiensi LLM dapat menyaingi atau melampaui fidelitas model generatif tradisional dalam ekosistem Software Quality Assurance.

2. LANDASAN TEORI

2.1 Software Development

Software Development merupakan disiplin rekayasa yang berkaitan dengan semua aspek produksi perangkat lunak. Proses ini mencakup penerapan pendekatan sistematis, disiplin, dan terukur terhadap pengembangan, operasi, serta pemeliharaan perangkat lunak.

2.2 Software Quality Assurance (QA)

Dalam ekosistem modern, QA berfungsi sebagai mekanisme kontrol preventif untuk menjamin perangkat lunak memenuhi spesifikasi teknis dan bebas dari cacat fatal. Komponen vital dalam QA adalah data uji (test data), di mana ketersediaan data yang representatif menjadi prasyarat mutlak untuk validasi keandalan sistem.

2.3 Artificial Intelligence (AI) dan Generative AI

AI adalah bidang ilmu yang bertujuan menciptakan sistem komputer yang mampu meniru kemampuan berpikir manusia. AI modern beroperasi melalui tiga aktivitas kognitif: pembelajaran (learning), penalaran (reasoning), dan koreksi diri (self-correction). Generative AI merupakan cabang AI yang menggunakan jaringan saraf untuk mengidentifikasi pola guna menghasilkan konten baru dan orisinal. Teknologi ini memungkinkan pembuatan data sintetis yang aman dan bebas risiko privasi.

2.4 Large Language Model (LLM) dan Gemini

LLM adalah jenis sistem AI yang mampu menghasilkan teks mirip manusia berdasarkan pola dari data masif. Gemini, yang dikembangkan oleh Google DeepMind, merupakan keluarga model multimodal yang dirancang untuk penalaran lintas teks, gambar, audio, dan kode. Varian *gemini-3-flash-preview* dioptimalkan untuk latensi rendah dan throughput tinggi, menjadikannya pilihan efisien untuk tugas-tugas workhorse seperti generasi data sintetis skala besar.

2.5 Data Sintetis dan Data Testing

Data sintetis adalah data buatan yang meniru pola dan karakteristik data nyata tanpa

mengandung informasi sensitif. Penggunaannya memungkinkan tim QA melakukan uji fungsional dan performa tanpa melanggar regulasi privasi seperti GDPR atau UU PDP. Dalam era DevOps, otomatisasi pembuatan data uji melalui data sintetis dapat menghemat waktu, biaya, dan mempercepat siklus umpan balik.

2.6 Metrik Evaluasi Kualitas Data

Untuk mengukur fidelitas data sintetis, digunakan beberapa instrumen statistik:

(a) Kolmogorov-Smirnov Test (KS-Test):

Instrumen non-parametrik untuk mengukur kesamaan distribusi probabilitas antara dua himpunan data independen dengan menghitung jarak vertikal maksimum pada kurva ECDF. Dengan rumus pada persamaan (1):

$$D_{n,m} = \sup_x |F_{1,n}(x) - F_{2,m}(x)| \quad (1)$$

(b) Jensen-Shannon Divergence (JS-Divergence):

Metrik untuk mengukur kemiripan distribusi probabilitas yang bersifat simetris dan stabil, sangat krusial untuk mengevaluasi fitur kategorikal. Bisa dilihat pada persamaan (2) berikut:

$$JSD(P|Q) = \frac{1}{2} D_{KL}(P|M) + \frac{1}{2} D_{KL}(Q|M) \quad (2)$$

(c) Cosine Similarity:

Metrik untuk mengukur kesamaan semantik dan kedekatan relasional antar atribut dengan menghitung kosinus sudut antar vektor di ruang multidimensi. Dengan formula pada persamaan (3):

$$\frac{\Sigma AB}{|A| \times |B|} \quad (3)$$

3. METODOLOGI

3.1 Objek dan Lokasi Penelitian

Objek dalam penelitian ini adalah dataset kesehatan yang diperoleh dari platform daring, yaitu Kaggle dan data.world. Seluruh kegiatan penelitian ini dilakukan secara virtual karena

proses pengumpulan, pengolahan, dan analisis data dilakukan melalui akses online terhadap situs resmi tersebut. Dengan demikian, seluruh tahapan penelitian mulai dari pengunduhan data, pemeriksaan awal, hingga analisis akhir dieksekusi secara digital menggunakan lingkungan kerja Google Colab tanpa memerlukan lokasi fisik tertentu.

3.2 Dataset Penelitian

Dataset yang digunakan bersumber dari platform [kaggle.com](https://www.kaggle.com) dengan judul "Glassdoor- Analyze Gender Pay Gap" yang disusun oleh Neelima Jauhari. Data ini berfokus pada informasi gaji untuk berbagai jabatan berdasarkan jenis kelamin dan telah digunakan dalam penelitian sebelumnya oleh Sanghvi, K. pada tahun 2023. Dataset ini bersifat publik dan *open-access* untuk keperluan akademik. Data terakhir diperbarui pada tahun 2020 dengan total 1.000 entri data dan terdiri dari 9 kolom fitur. Rincian mengenai struktur kolom dataset ini dapat dilihat pada Tabel 1.

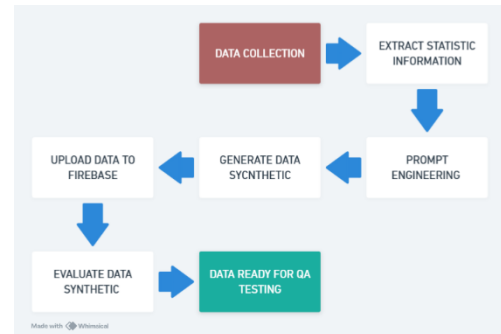
Tabel 1. Rincian Kolom

Kolom	Deskripsi	Tipe Data	Jumlah
<i>JobTitle</i>	Jabatan pekerjaan karyawan	object (Teks)	1000
<i>Gender</i>	Jenis kelamin karyawan (Female/Male)	object (Teks)	1000
<i>Age</i>	Usia karyawan	int64 (Angka)	1000
<i>PerfEval</i>	Skor evaluasi kinerja	int64 (Angka)	1000
<i>Education</i>	Tingkat pendidikan terakhir	object (Teks)	1000
<i>Dept</i>	Departemen tempat karyawan bekerja	object (Teks)	1000
<i>Seniority</i>	Tingkat senioritas (kemungkinan level)	int64 (Angka)	1000
<i>BasePay</i>	Gaji pokok	int64 (Angka)	1000
<i>Bonus</i>	Bonus tahunan	int64 (Angka)	1000

3.3 Tahapan Penelitian

Penelitian ini menerapkan metode eksperimen untuk membandingkan kinerja *Generative AI Large Language Model*, khususnya Gemini, dalam menghasilkan data sintesis sebagai data uji pada proses *software*

quality assurance. Alur kerja atau tahapan penelitian secara visual digambarkan pada Gambar 1.



Gambar 1. Tahapan Penelitian

Secara sistematis, tahapan penelitian ini dibagi menjadi tujuh langkah utama:

- Data Collection:** Dataset "Glassdoor- Analyze Gender Pay Gap" dikumpulkan dari Kaggle dan disimpan terlebih dahulu di dalam Firebase sebelum digunakan.
- Extract Statistic Information:** Proses pengambilan informasi statistik seperti *median*, *mean*, standar deviasi, serta nilai minimum dan maksimum dari data asli dilakukan agar data sintesis yang dihasilkan nantinya memiliki karakteristik yang serupa dengan aslinya.
- Prompt Engineering:** Penentuan instruksi strategis diberikan kepada model LLM, mencakup penetapan peran model, tujuan, konteks data, serta batasan yang harus dipatuhi untuk memastikan keluaran sesuai kebutuhan.
- Generate Data Synthetic:** Model LLM memanfaatkan pemahamannya terhadap pola bahasa dan struktur data untuk menciptakan baris data baru mengikuti format dataset asli tanpa mengandung informasi sensitif. Penjelasan konseptual mengenai proses di balik layar tahap ini dirangkum pada
- Evaluate Data Synthetic:** Dilakukan evaluasi untuk menilai sejauh mana data sintesis mendekati karakteristik data asli melalui pendekatan kesamaan statistik dan representasi vektor embedding. Upload Data To Firebase: Data sintesis yang telah tervalidasi diunggah ke Firebase untuk menyediakan akses cepat dan

sinkronisasi real-time bagi lingkungan pengembangan.

- (f) *Ready For Testing*: Dataset yang telah tersimpan siap digunakan untuk berbagai skenario pengujian perangkat lunak, mulai dari pengujian fungsional hingga simulasi beban.

4. HASIL DAN PEMBAHASAN

Ekstraksi Informasi Statistik Dataset Pada tahap awal penelitian, dilakukan proses ekstraksi informasi statistik terhadap dataset yang digunakan sebagai sumber data utama (*ground truth*). Dataset yang digunakan adalah data *Glassdoor - Analyze Gender Pay Gap* yang terdiri dari atribut profil karyawan dengan tipe data numerik dan kategorikal. Tahapan ini bertujuan untuk mengidentifikasi karakteristik statistik dari dataset sehingga pola distribusi data dapat dipahami secara menyeluruh sebelum proses pembuatan data sintetis dilakukan. Hasil statistik deskriptif dari data tersebut dapat dilihat pada Tabel 2 dan 3.

Tabel 2. Statistik Numerik

Column	mean	std	min	25%	50%	75%	max
Seniority	2.971	1.395028711	1	2	3	4	5
Base Pay	94472.653	25337.49327	34208	76850.25	93327.5	111558	179726
Age	41.393	14.2948555	18	29	41	54.25	65
PerfEval	3.037	1.423958798	1	2	3	4	5
Bonus	6467.161	2004.377365	1703	4849.5	8026.507	1126	11293

Tabel 3. Statistik Kategorik

Column	count	unique	top	freq
JobTitle	1000	10	Marketing Associate	118
Gender	1000	2	Male	532
Dept	1000	5	Operations	210
Education	1000	4	High School	265

Berdasarkan Tabel 2 dan 3, dataset secara proporsional terdiri dari atribut kategorikal (*JobTitle*, *Gender*, *Department*, dan *Education*) serta atribut numerik (*BasePay*, *Age*, *Bonus*, *Seniority*, *PerfEval*). Atribut

JobTitle memiliki 10 kategori unik dengan posisi *Marketing Associate* sebagai kategori dominan. Di sisi lain, atribut *Gender* sedikit didominasi oleh kelompok *Male* dengan total 532 data, yang secara akurat mencerminkan komposisi demografis awal dataset. Pada atribut numerik, *BasePay* mencatatkan nilai rata-rata sebesar 94.472 dengan standar deviasi 25.337 yang merepresentasikan sebaran variasi gaji riil di lapangan, sedangkan rentang *Age* berkisar pada usia produktif antara 18 hingga 65 tahun. Keseluruhan informasi statistik deskriptif ini tidak hanya berfungsi sebagai ringkasan, tetapi diekstraksi menjadi acuan instruksi (*prompt*) utama bagi *Large Language Model* (LLM). Melalui penyediaan parameter dan batasan yang presisi ini, LLM dipandu untuk tidak melakukan halusinasi informasi, sehingga mampu menghasilkan simulasi data sintetis yang mempertahankan karakteristik, varians, dan bentuk distribusi probabilitas dari data aslinya secara utuh.

Evaluasi Fitur Kategorikal Evaluasi pada fitur kategorikal (teks) dilakukan dengan membandingkan distribusi frekuensi kategori antara dataset asli dan dataset sintetis menggunakan pendekatan matematis ganda. Pengukuran *Jensen-Shannon Divergence* (JSD) digunakan untuk melihat penyimpangan probabilitas, *Cosine Similarity* digunakan guna menjaga kedekatan makna semantik teks. Selain itu, *Cosine Similarity* digunakan guna menjaga kedekatan makna semantik teks. Setiap atribut kategorikal direpresentasikan sebagai vektor frekuensi, lalu diukur kesamaan kosinus sudut antar vektornya yang dapat dilihat pada Tabel 4

Tabel 4. Perbandingan Kategorik

SYNTHETIC				
Column	count	unique	top	freq
JobTitle	1000	10	Marketing Associate	105
Gender	1000	2	Male	509
Dept	1000	5	Operations	255
Education	1000	4	Masters	288
REAL				
Column	count	unique	top	freq
JobTitle	1000	10	Marketing Associate	118

Gender	1000	2	Male	532
Dept	1000	5	Operations	210
Education	1000	4	High School	265

Berdasarkan Tabel 4, dapat dilihat bahwa seluruh atribut kategorikal memiliki nilai *JS Divergence* yang sangat baik dengan skor di atas 95%, di mana atribut *Gender* merupakan atribut yang distribusinya paling identik dengan data asli. Nilai *Cosine Similarity* pada seluruh atribut kategorikal berada di atas 99%, yang menegaskan bahwa distribusi frekuensi kategori antara dataset asli dan dataset sintetis memiliki makna semantik yang nyaris sempurna.

Evaluasi Fitur Numerik Evaluasi terhadap fitur numerik (*Seniority*, *BasePay*, *Age*, *PerfEval*, *Bonus*) dilakukan menggunakan *Kolmogorov-Smirnov Distance (KS-Distance)* dan *Jensen-Shannon Divergence*. Hukum fundamental dari *KS-Distance* mendikte pencarian simpangan absolut paling ekstrem (Supremum / D) antara dua kurva Fungsi Distribusi Kumulatif Empiris atau *Empirical Cumulative Distribution Function (ECDF)*. Hasil perhitungan kemiripan numerik beserta statistik deskriptif perbandingannya dirangkum dalam Tabel 5 dan 6.

Tabel 5. Evaluasi Fitur Numerik

Column	KS Distance%	JS Divergence%
Seniority	97.5	97.86
BasePay	88.6	83.70
Age	93.1	84.71
PerfEval	79.8	75.54
Bonus	90.4	82.77

Tabel 6. Statistik Deskriptif Numerik

SYNTHETIC							
Column	mean	std	min	25%	50%	75%	max
Seniority	3.025	1.39295 2015	1	2	3	4	5
BasePay	98658. 962	32796.1 9477	342 08	7400 0	9500 0	122 000	179 726
Age	41.091	12.6016 4753	18	30.75	41	52	65

REAL							
Column	mean	std	min	25%	50%	75%	max
Seniority	2.971	1.39502 8711	1	2	3	4	5
BasePay	94472. 653	25337.4 9327	342 08	7685 0.25	9332 7.5	111 558	179 726
Age	41.393	14.2948 555	18	29	41	54.2 5	65
PerfEval	3.037	1.42395 8798	1	2	3	4	5
Bonus	6467.1 61	2004.37 7365	170 3	4849. 5	6507	802 6	112 93

Atribut *Seniority* menunjukkan tingkat kemiripan paling tinggi (97,5%), mengindikasikan bahwa model generatif mampu mempertahankan pola distribusi ordinal dengan sangat baik. Atribut *Age*, *Bonus*, dan *BasePay* juga menunjukkan tingkat kesamaan yang tinggi secara berurutan sebesar 93,1%, 90,4%, dan 88,6%. Sebaliknya, atribut *PerfEval* menunjukkan tingkat kesamaan terendah yakni 79,8%. Penurunan ini disebabkan oleh variasi distribusi penilaian kinerja yang kompleks dan tidak beraturan dalam dataset asli. Visualisasi perbandingan kepadatan (*density*) fitur numerik ini dapat diamati pada Gambar 5 hingga Gambar 9, yang menunjukkan bahwa model secara impresif mempertahankan rentang batas (*boundary constraints*) dan letak pemusatan data (*central tendency*).

5. KESIMPULAN

Berdasarkan hasil perancangan, implementasi, dan evaluasi framework menggunakan *Large Language Model (LLM) Gemini 3* untuk membangkitkan data uji sintetis pada sistem *Software Quality Assurance (QA)*, dapat ditarik beberapa kesimpulan utama. Pertama, model *Gemini 3* terbukti sangat unggul dalam mereplikasi distribusi variabel berdimensi teks (kategorikal) dengan presisi tinggi. Melalui metrik *Cosine Similarity*, model ini

mencatatkan skor rata-rata kemiripan semantik sebesar 99,38%, serta skor rata-rata penyimpangan probabilitas sebesar 96,52% menggunakan *Jensen-Shannon (JS) Divergence*. Hal ini membuktikan bahwa arsitektur *LLM* secara konsisten mampu merangkai kosakata riil dengan proporsi populasi yang identik dengan kerangka data aslinya tanpa mengalami halusinasi data.

Kedua, pada variabel angka (numerik), model menunjukkan stabilitas ekstraksi pola yang solid. Pengukuran simpangan kumulatif menggunakan *Kolmogorov-Smirnov (KS) Distance* menghasilkan nilai rata-rata akurasi sebesar 89,88%. Mesin berhasil memetakan pola distribusi yang tidak beraturan, seperti pola asimetris pada variabel usia maupun fluktuasi pencilaan pada gaji dasar, dan tetap sanggup mempertahankan batas logis (*boundary constraints*) tanpa harus melalui proses pelatihan ulang (*fine-tuning*).

Ketiga, penggunaan pendekatan zero-shot dari *LLM* ini menawarkan efisiensi komputasi dan tingkat keamanan privasi yang jauh lebih baik jika dibandingkan dengan penelitian terdahulu yang mengandalkan arsitektur *Generative Adversarial Networks (GANs)*. Eksperimen ini mengeliminasi kebutuhan infrastruktur komputasi kelas berat dan waktu pelatihan yang panjang. Melalui integrasi *endpoint API*, data sintesis yang mempertahankan integritas statistik dapat diproduksi seketika (*real-time*). Hilangnya keterikatan langsung antara data sintesis dan identitas individu asli menjadikan dataset ini sepenuhnya aman untuk diterjunkan ke dalam siklus pengujian berkelanjutan (*Continuous Testing*) tanpa melanggar regulasi perlindungan data.

6. UCAPAN TERIMA KASIH

Berisi ucapan terima kasih terutama kepada pihak yang telah memberi pendanaan penelitian atau pengabdian Masyarakat.

DAFTAR PUSTAKA

Filippucci, F., Gal, P., Laengle, K., & Schief, M. (2025). Macroeconomic productivity gains from Artificial Intelligence in G7 economies. *OECD Artificial Intelligence Papers*. <https://www.oecd-ilibrary.org/en/publications/macroecomic-productivity-gains-from->

[artificial-intelligence-in-g7-economies_a5319ab5-en.html](https://doi.org/10.37817/ikraith-informatika.v9i3)

Peng, S., Kalliamvakou, E., Cihon, P., & Demirer, M. (2023). The impact of ai on developer productivity: Evidence from github copilot. *arXiv preprint arXiv:2302.06590*.

<https://doi.org/10.48550/arXiv.2302.06590>

Rusum, G. P., & Anasuri, S. (2023). Synthetic test data generation using generative models. *International Journal of Emerging Trends in Computer Science and Information Technology*, 4(4), 96–108.

Taley, D. S. (2020). Comprehensive study of software testing techniques and strategies: A review. *International Journal of Engineering Research & Technology (IJERT)*, 9(8). <https://doi.org/10.5281/zenodo.18951330>

Yuan, R., Liu, Y., & Cheng, L. (2024). A multi-faceted evaluation framework for assessing synthetic data generated by large language models. *arXiv preprint arXiv:2404.14445*. <https://doi.org/10.48550/arXiv.2404.14445>