

IMPLEMENTASI FINE-TUNING DREAMBOOTH DAN REGULARISASI DROPOUT UNTUK GENERASI KARAKTER FANTASI BERGAYA ART NOUVEAU

Yunita Endah Sulistiyowati^{1*}, Sri Mulyono²

Universitas Islam Sultan Agung^{1,2}

yunitasulistiyowati@gmail.com

Received: 23-01-2025

Revised: 30-01-2025

Approved: 13-02-2025

ABSTRAK

Penelitian ini bertujuan untuk menghasilkan visualisasi karakter fantasi bergaya Art Nouveau melalui fine-tuning model Stable Diffusion menggunakan metode DreamBooth. Proses penelitian mencakup pengumpulan dataset karakter bergaya Art Nouveau, pelatihan model menggunakan Stable Diffusion v1.5, pengembangan sistem berbasis web dengan Streamlit, serta evaluasi hasil menggunakan CLIP Score dan analisis kualitatif. Fine-tuning dilakukan dengan metode DreamBooth untuk meningkatkan fidelitas visual dan personalisasi karakter berdasarkan deskripsi teks. Data pelatihan diolah melalui CLIP Text Encoder untuk menghasilkan embedding yang merepresentasikan teks, sedangkan model dilatih secara iteratif dengan loss calculation berbasis mean squared error (MSE) dan dropout regularization untuk mencegah overfitting. Evaluasi menunjukkan bahwa penggunaan prompt yang lebih spesifik serta negative prompt dapat meningkatkan CLIP Score rata-rata dari 0.65 menjadi 0.826 setelah fine-tuning, yang mencerminkan peningkatan akurasi visualisasi terhadap deskripsi teks. Aplikasi berbasis web dikembangkan untuk memudahkan pengguna dalam menghasilkan gambar dari teks secara real-time. Implementasi sistem ini memungkinkan eksplorasi visual karakter secara interaktif, dengan fitur validasi berbasis CLIP Score dan evaluasi estetika. Hasil penelitian menunjukkan bahwa model yang telah difine-tune mampu menghasilkan gambar dengan elemen khas Art Nouveau, seperti motif dekoratif, garis melengkung, dan palet warna pastel. Dengan demikian, penelitian ini memberikan kontribusi dalam pengembangan sistem generatif berbasis AI untuk industri kreatif, khususnya dalam visualisasi karakter fantasi yang estetis dan naratif.

Kata Kunci: Stable Diffusion, DreamBooth, Art Nouveau, Fine-Tuning

PENDAHULUAN

Kemajuan teknologi kecerdasan buatan (AI) telah membawa inovasi signifikan dalam berbagai bidang, termasuk text-to-image generation, yang memungkinkan penerjemahan deskripsi teks menjadi visualisasi gambar (Farwati et al. 2023). Salah satu pendekatan utama adalah diffusion models, yang secara bertahap mengurangi noise untuk membentuk gambar yang sesuai dengan input teks (Hendrawati, Wulandari, and Ginantra 2024). Model seperti Stable Diffusion dikenal karena kemampuannya menghasilkan gambar berkualitas tinggi, tetapi masih menghadapi tantangan dalam fidelitas visual dan personalisasi hasil. Model diffusion, seperti yang digunakan dalam penelitian "Custom Concept Text-to-Image Using Stable Diffusion Model," mampu menghasilkan gambar realistis dari prompt teks (Puspita 2023). Tantangan utama dalam meningkatkan kualitas generasi gambar terletak pada ketersediaan dataset besar, kebutuhan komputasi tinggi, dan nilai Frechet Inception Distance (FID) yang masih cukup tinggi (Pongtambing et al. 2023).

Dalam konteks ini, teknik fine-tuning seperti DreamBooth menjadi solusi efektif untuk meningkatkan kemampuan model dalam memahami atribut spesifik dari deskripsi karakter. DreamBooth memungkinkan model menghasilkan gambar yang lebih personal dengan mempertahankan fitur utama subjek. Selain itu, teknik ini terbukti efektif dalam bidang seni digital dan desain, dimana memanfaatkan

DreamBooth untuk personalisasi gambar dalam berbagai gaya. Penerapan regularization seperti Dropout juga berperan penting untuk mencegah overfitting selama pelatihan, sehingga model dapat menghasilkan visualisasi yang lebih general dan sesuai dengan deskripsi teks (Baldi, Yohannes, and ... 2023). Dreambooth berhasil diterapkan pada berbagai bidang, seperti pengenalan gambar medis untuk meningkatkan akurasi model CNN melalui augmentasi data dan generasi gambar untuk kebutuhan desain berbasis gaya tertentu, termasuk ikon siluet untuk antarmuka pengguna (Harahap et al. 2022). Selain itu, penelitian "DreamSalon" dan "DreamTuner" memperkenalkan kerangka kerja baru yang lebih efisien untuk mempertahankan identitas subjek dan menangani manipulasi detail dengan menjaga keseimbangan antara identitas subjek dan kemampuan generasi gambar.

Penelitian lain menunjukkan bahwa penggunaan tiga kondisi (gambar, teks, dan gaya) dalam model diffusion, seperti pada "Conditional Text Image Generation with Diffusion Models (CTIG-DM)," memungkinkan kontrol lebih besar terhadap atribut visual gambar, mencakup mode sintesis, augmentasi, pemulihan, dan imitasi. Pendekatan ini memperluas fleksibilitas generasi gambar, meskipun aplikasinya dalam skenario dunia nyata masih membutuhkan validasi lebih lanjut. Salah satu gaya visual yang relevan untuk visualisasi karakter dalam novel fantasi adalah Art Nouveau. Gaya ini dikenal dengan garis melengkung yang elegan, motif alam, dan warna pastel yang memberikan kesan estetis dan magis. Penelitian terdahulu menunjukkan bahwa menggabungkan Stable Diffusion dengan DreamBooth mampu menghasilkan karya seni yang mendetail dan estetis, termasuk dalam konteks gaya visual tertentu. Namun, tantangan fidelitas dan personalisasi tetap menjadi perhatian utama ketika menghasilkan karakter dengan atribut visual yang rumit, seperti pakaian, ekspresi wajah, dan pose.

Implementasi sistem berbasis web dalam penelitian ini dirancang untuk memudahkan pengguna menghasilkan gambar karakter bertema Art Nouveau hanya dengan memasukkan deskripsi teks sebagai input. Sistem ini tidak hanya bertujuan mendukung kebutuhan kreatif dalam seni digital, tetapi juga memberikan solusi praktis bagi pembaca dan kreator novel fantasi untuk memvisualisasikan karakter secara otomatis. Teknologi ini dapat mendukung industri kreatif dengan menghasilkan visualisasi yang relevan secara naratif sekaligus estetis. Rumusan masalah yang diangkat dalam penelitian ini adalah bagaimana deskripsi teks dapat diterjemahkan menjadi visualisasi gambar dengan gaya Art Nouveau menggunakan Stable Diffusion, serta bagaimana memastikan fidelitas dan personalisasi dalam hasil generasi gambar. Penelitian ini bertujuan untuk mengembangkan model yang mampu menghasilkan visualisasi karakter bergaya Art Nouveau dengan fidelitas tinggi melalui fine-tuning menggunakan DreamBooth dan regularization Dropout, sekaligus menyediakan platform berbasis web yang memungkinkan pengguna memasukkan deskripsi teks untuk menghasilkan gambar karakter sesuai kebutuhan.

METODE PENELITIAN

Penelitian ini bertujuan untuk menghasilkan visualisasi karakter fantasi bergaya Art Nouveau melalui fine-tuning model Stable Diffusion menggunakan metode DreamBooth. Tahapan penelitian meliputi pengumpulan data, pelatihan model, pengembangan sistem berbasis web, serta evaluasi hasil untuk memastikan kesesuaian visualisasi dengan deskripsi teks yang diberikan.



Gambar 1. Tahapan Penelitian

Penulis menggunakan Stable Diffusion v1.5 sebagai model dasar untuk text-to-image generation. Model ini di-fine-tune menggunakan metode DreamBooth untuk meningkatkan fidelitas visual dan personalisasi karakter berdasarkan deskripsi teks. Tahapan pelatihan dimulai dengan preprocessing data menggunakan CLIP Text Encoder untuk menghasilkan vektor embedding yang merepresentasikan teks [18]. Gambar referensi dari dataset kemudian diolah dalam model untuk mengaitkan deskripsi dengan karakter visual tertentu. Proses pelatihan melibatkan loss calculation menggunakan fungsi mean squared error (MSE), dan dropout regularization diterapkan untuk mencegah overfitting [19]. Model juga diuji secara iteratif untuk memastikan hasil yang optimal.

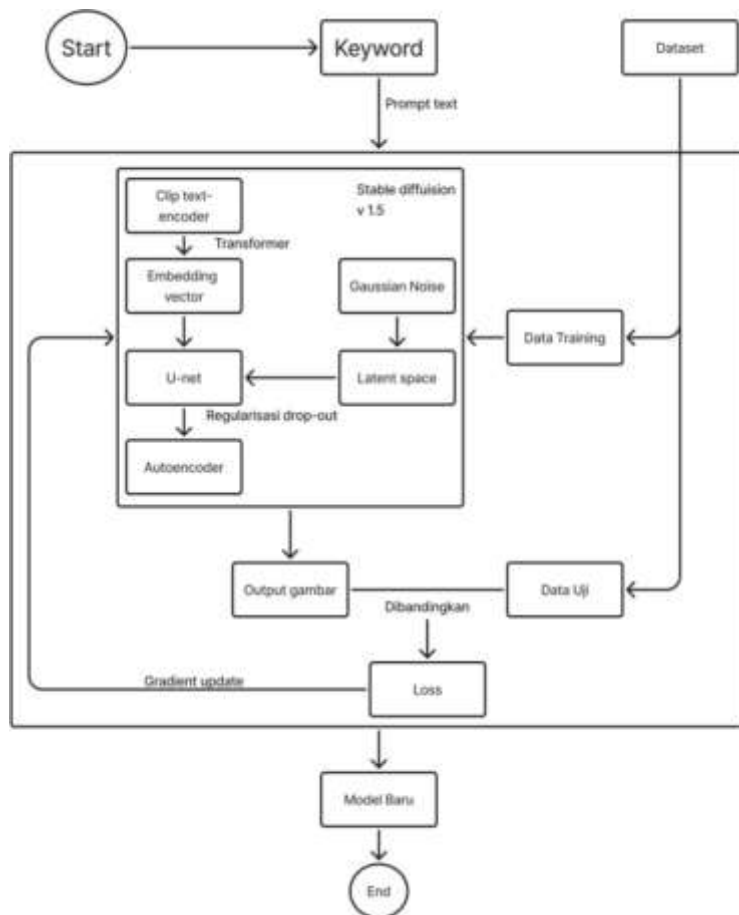
Studi literatur mencakup analisis jurnal dan artikel ilmiah terkait Stable Diffusion, DreamBooth, dan teknik regularization. Kajian ini bertujuan untuk memahami prinsip-prinsip dasar serta implementasi teknologi terkini yang mendukung pengembangan model generatif berbasis AI. Pengumpulan dan Pengolahan Dataset, Dataset terdiri dari gambar karakter bergaya Art Nouveau, yang dikumpulkan melalui platform seperti Civit AI. Gambar-gambar ini diseleksi berdasarkan atribut visual seperti pose, motif dekoratif, dan palet warna. Dataset kemudian diproses dengan langkah-langkah seperti pengelompokan gambar berdasarkan gaya dan validasi keberagaman atribut. Dataset ini dibagi menjadi data pelatihan (80%) dan data pengujian (20%) untuk mendukung proses fine-tuning model. Penggunaan Stable Diffusion 1.5, Stable Diffusion dipilih karena kemampuannya menghasilkan gambar berkualitas tinggi. Namun, tantangan seperti fidelitas visual untuk atribut kompleks dan personalisasi subjek memerlukan adaptasi khusus. DreamBooth digunakan untuk mengatasi keterbatasan ini dengan melatih model pada dataset spesifik yang relevan dengan gaya Art Nouveau.

Fine-Tuning dengan DreamBooth

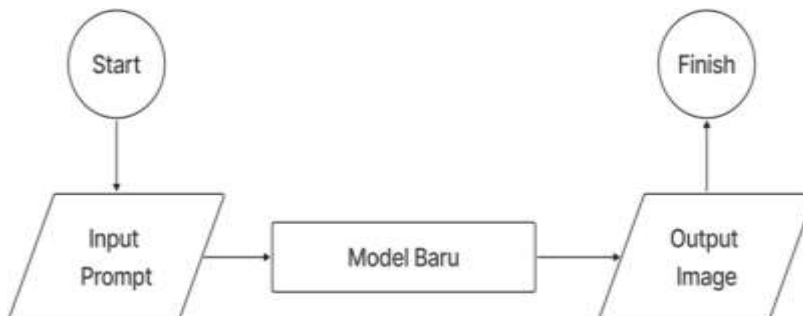
Proses pelatihan melibatkan beberapa tahap utama:

- Preprocessing Teks dan Gambar: Deskripsi teks diproses menjadi vektor embedding menggunakan CLIP Text Encoder.

- Pelatihan Model: Model dilatih menggunakan gambar referensi untuk menghubungkan teks dengan elemen visual karakter. Proses denoising menggunakan U-Net Architecture dilakukan secara iteratif untuk menghasilkan gambar dengan fidelitas tinggi [20].
- Validasi Model: Hasil pelatihan diuji menggunakan dataset validasi, dengan parameter seperti guidance scale dan inference steps disesuaikan untuk mencapai performa optimal.



Gambar 2. Alur Training Sistem



Gambar 3. Alur Kerja Sistem

Aplikasi berbasis web dikembangkan menggunakan Streamlit. Sistem ini memungkinkan pengguna untuk memasukkan deskripsi karakter secara langsung dan mendapatkan visualisasi gambar sesuai dengan parameter yang ditetapkan. Input berupa deskripsi teks diproses untuk menghasilkan representasi vektor, yang kemudian digunakan oleh model untuk menghasilkan gambar.



Gambar 4. Tampilan Awal



Gambar 5. Tampilan saat Sistem Dijalankan

Sistem dirancang untuk memenuhi kebutuhan perangkat keras dan perangkat lunak sebagai berikut:

- Perangkat Keras: Komputer dengan spesifikasi minimum RAM 8 GB, prosesor dual-core, serta akses internet stabil.
- Perangkat Lunak: Google Colab untuk pelatihan model, serta pustaka seperti PyTorch, Transformers, dan Diffusers untuk implementasi text-to-image generation.
- Fungsionalitas Sistem: Sistem mendukung generasi gambar dari deskripsi teks dalam gaya Art Nouveau, dengan validasi hasil melalui metrik seperti CLIP Score.

Evaluasi dan Validasi Hasil, Hasil model dievaluasi secara kuantitatif menggunakan CLIP Score untuk mengukur kesesuaian gambar dengan gambar referensi. Evaluasi kualitatif dilakukan untuk memastikan elemen gaya Art Nouveau seperti garis melengkung, motif bunga, dan palet warna pastel tercermin dalam gambar yang dihasilkan. Dengan pendekatan ini, penelitian diharapkan mampu menghasilkan sistem yang inovatif untuk mendukung industri kreatif, khususnya dalam visualisasi karakter fantasi yang estetis dan naratif.

HASIL PENELITIAN DAN PEMBAHASAN




Penelitian ini dimulai dengan penyesuaian format menjadi 512x512 piksel untuk memastikan keseragaman input. Dataset ini dirancang untuk mencakup berbagai variasi pose, ekspresi, dan elemen dekoratif yang relevan. Fokus utama adalah memberikan data yang cukup untuk memungkinkan model memahami karakteristik visual gaya Art Nouveau secara optimal. Proses evaluasi dilakukan secara iteratif untuk memastikan bahwa hasil yang dihasilkan mencerminkan elemen-elemen penting dari gaya tersebut. Hasil penelitian menunjukkan bahwa deskripsi teks (prompt) memiliki pengaruh besar terhadap kualitas gambar yang dihasilkan. Seperti yang ditunjukkan pada table 1, prompt yang lebih detail, seperti *"Art Nouveau style, a girl holding flower"*, mampu menghasilkan elemen visual yang spesifik, seperti pola bunga dan garis melengkung. Negative prompt seperti *"bad anatomy, ugly face, low quality, NSFW Content"* sangat efektif untuk menghilangkan elemen yang tidak diinginkan, seperti distorsi anatomi atau detail wajah yang kurang estetik. Kombinasi prompt utama dengan negative prompt meningkatkan CLIP Score rata-rata dari 0.65 menjadi 0.76, menunjukkan peningkatan signifikan dalam kesesuaian antara deskripsi dan gambar yang dihasilkan.

Tabel 1.
Penggunaan Prompt Terhadap Clip Score

Tahap	Prompt yang digunakan	Negative Prompt	Rata-Rata CLIP Score
Generasi awal	<i>"Art Nouveau style, a girl holding flower"</i>	-	0.36
Negative Prompt	<i>"Art Nouveau style, a girl holding flower"</i>	<i>"bad anatomy, ugly face, low quality, NSFW Content"</i>	0.69
Spesifik Prompt	<i>"Art Nouveau style, a girl holding flower, surrounded by an ornate and decorative circular frame, muted pastel tones, soft lighting, and detailed textures"</i>	<i>"bad anatomy, ugly face, low quality, NSFW Content"</i>	0.77
Setelah Fine-Tune	<i>"a girl holding flower"</i>	<i>"bad anatomy, ugly face, low quality, NSFW Content"</i>	0.826

Setelah proses fine-tuning menggunakan metode DreamBooth, model diuji dengan berbagai variasi prompt untuk mengevaluasi fidelitas dan personalisasi hasil. Misalnya, prompt *"Art Nouveau style, a hijab girl holding flower"* menghasilkan gambar yang menonjolkan ciri khas gaya ini, seperti motif dekoratif dan nuansa warna pastel. Variasi lain, seperti *"Art Nouveau style, a yunani man holding flower"*, menunjukkan bahwa model dapat menghasilkan gambar dengan atribut gender yang berbeda, tetapi tetap mempertahankan elemen gaya yang diinginkan. Fine-tuning terbukti meningkatkan akurasi visual, dan tambahan negative prompt semakin memperbaiki detail, seperti ekspresi wajah dan tekstur pakaian. Seperti yang ditunjukkan pada table 2.

Tabel 2.
Penggunaan Variasi Prompt

Prompt yang digunakan	Negative prompt	Hasil generasi
<i>"Art Nouveau style, a hijab girl holding flower "</i>	-	
<i>"Art Nouveau style, a hijab girl holding flower "</i>	<i>"bad anatomy, ugly face, low quality, NSFW Content"</i>	
<i>"Art Nouveau style, a yunani man holding flower "</i>	<i>"bad anatomy, ugly face, low quality, NSFW Content"</i>	

Implementasi sistem berbasis web menggunakan framework Streamlit menjadi bagian penting dari penelitian ini. Aplikasi ini dirancang untuk memudahkan pengguna memasukkan deskripsi teks melalui antarmuka sederhana dan langsung menghasilkan gambar. Proses generasi ditampilkan secara real-time, memberikan umpan balik seperti *"Loading model..."* dan *"Generating image..."* sebelum hasil akhir ditampilkan. Fitur ini memberikan kemudahan bagi pengguna non-teknis untuk mengeksplorasi visualisasi karakter dengan berbagai variasi prompt.

Regularisasi dropout seperti pada table 3, diterapkan selama proses fine-tuning untuk mencegah overfitting. Dropout digunakan pada lapisan CLIP encoder, menonaktifkan secara acak sejumlah neuron selama pelatihan untuk meningkatkan generalisasi model. Pengujian menunjukkan bahwa dropout dengan tingkat 0.1 menghasilkan gambar yang lebih bervariasi tanpa mengurangi kualitas visual. Contoh hasil menunjukkan peningkatan pada elemen estetika, seperti variasi warna dan pola yang lebih beragam, dibandingkan model tanpa dropout.

Tabel 3.
Penggunaan Drop Out dengan Variasi Prompt

Prompt yang digunakan	Kondisi	Hasil generasi
<i>"Art Nouveau style, a girl holding flower"</i>	Sebelum Regularisasi Drop Out	
<i>"Art Nouveau style, a girl holding flower"</i>	Sesudah Regularisasi Drop Out	
<i>"Art Nouveau style, a man with pigeon "</i>	Sebelum Regularisasi Drop Out	
<i>"Art Nouveau style, a man with pigeon "</i>	Sesudah Regularisasi Drop Out	

Penurunan skor akibat dropout menunjukkan model menjadi lebih bervariasi, meningkatkan generalisasi meski akurasi sedikit berkurang karena elemen acak. Pada kasus *"a man with pigeon,"* model tanpa dropout cenderung terlalu mengandalkan fitur tertentu, seperti pengenalan wajah pria, tanpa menghasilkan gambar merpati. Dropout memperbaiki eksplorasi gaya, pewarnaan, dan garis, membuat model lebih kreatif dan ekspresif dalam elemen artistik sesuai prompt.

Tabel 4.
Perbandingan Hasil Generasi pada tiap Tahapan

Prompt Inggris		Training 200	Iterasi	
Prompt Indonesia		Training 2000	Iterasi	
Negative Prompt		Regularisasi Drop Out		
Spesifik Prompt		Local Streamlit	dengan	

Dalam pengujian sistem generasi gambar, waktu yang dibutuhkan untuk menghasilkan gambar menunjukkan variasi yang signifikan bergantung pada spesifikasi perangkat keras yang digunakan. Pada konfigurasi CPU i5 1135G7 dengan RAM 16 GB, rata-rata waktu yang dibutuhkan adalah sekitar 9 menit, menunjukkan kinerja yang optimal tanpa kendala signifikan selama proses generasi gambar. Sebaliknya, pada konfigurasi yang lebih rendah, yaitu menggunakan CPU i3 6006U dengan RAM 12 GB, waktu yang dibutuhkan meningkat menjadi rata-rata 25 menit. Hal ini mencerminkan pengaruh langsung dari kapasitas pemrosesan yang lebih rendah terhadap efisiensi sistem. Selain itu, pengujian menggunakan GPU NX450 dengan VRAM 2 GB menunjukkan waktu rata-rata sekitar 12 menit. Meskipun lebih cepat dibandingkan dengan CPU i3, hasil pengujian ini mengungkap adanya peningkatan konten NSFW yang lebih tinggi. Hal ini kemungkinan besar disebabkan oleh perbedaan dalam cara GPU menangani proses generasi gambar, yang memengaruhi sensitivitas terhadap konten visual tertentu.

KESIMPULAN

Penelitian ini menunjukkan bahwa deskripsi teks dapat diterjemahkan menjadi gambar bergaya Art Nouveau menggunakan diffusion models melalui fine-tuning

DreamBooth pada Stable Diffusion v1.5. Model ini menangkap detail khas Art Nouveau, seperti garis melengkung dan pola floral, serta menghasilkan gambar yang akurat dengan bantuan prompt spesifik dan negative prompt. Penggunaan regularisasi dropout mencegah overfitting, menjaga kualitas hasil yang konsisten. Penelitian selanjutnya disarankan untuk mengeksplorasi lebih banyak gaya seni, menggunakan dataset yang lebih luas, mendukung bahasa lokal, serta mengoptimalkan arsitektur model dan perangkat keras untuk meningkatkan efisiensi dan relevansi sistem.

DAFTAR PUSTAKA

- Baldi, J. R. T., Y. Yohannes, and ... 2023. "Penggunaan Convolutional Neural Network Sebagai Pengenalan Huruf Bahasa Ibrani." *Jutisi: Jurnal Ilmiah ...*
- Farwati, Maryani, Irenda Talitha Salsabila, Kholifah Raihanun Navira, Tata Sutabri, and Universitas Bina Darma Palembang. 2023. "Analisa Pengaruh Teknologi Artificial Intelligence (AI) Dalam Kehidupan Sehari-Hari [Analyze the Influence of Artificial Intelligence (AI) Technology in Daily Life]." *Jurnal Sistem Informatika Dan Manajemen* 11(1):41-42.
- Harahap, Fahri Aulia Alfarisi, Ronaldo Mardianson Sinaga, Khusnul Arifin, and Kana Saputra S. 2022. "Implementasi Algoritma Convolutional Neural Network Untuk Mendeteksi Penyakit Ginjal Implementation of Convolutional Neural Network for Detecting Kidney Disease." *Teknologi Informasi, Komputer Dan Aplikasinya (JTika)* 4(2):212-19.
- Hendrawati, Theresia, Dewa Ayu Putri Wulandari, and Ni Luh Wiwik Sri Rahayu Ginantra. 2024. "Penerapan Metode Stable Diffusion Dengan Fine Tuning Untuk Pola Endek Bali." *Jurnal Teknologi Informasi Komunikasi (e-Journal)* 10:8-10.
- Pongtambing, Yulita Sirinti, Felycitae Ekalaya Appa, A. Muh. Amil Siddik, Eliyah A. M. Sampetoding, Harsi Admawati, Arini Anestesia Purba, Arnold Sau, and Esther S. Manapa. 2023. "Peluang Dan Tantangan Kecerdasan Buatan Bagi Generasi Muda." *Bakti Sekawan: Jurnal Pengabdian Masyarakat* 3(1):23-28. doi: 10.35746/bakwan.v3i1.362.
- Puspita, Kharisma Galuh. 2023. "Analisis Kestabilanmodel Matematika Penyebaran Penyakit Demam Berdarah Dengue Dengan Pengaruh Treatmentdan Fogging." *MATHunesa Jurnal Ilmiah Matematika* 11(03):318-27.
- Baldi, J. R. T., Y. Yohannes, and ... 2023. "Penggunaan Convolutional Neural Network Sebagai Pengenalan Huruf Bahasa Ibrani." *Jutisi: Jurnal Ilmiah ...*
- Farwati, Maryani, Irenda Talitha Salsabila, Kholifah Raihanun Navira, Tata Sutabri, and Universitas Bina Darma Palembang. 2023. "Analisa Pengaruh Teknologi Artificial Intelligence (AI) Dalam Kehidupan Sehari-Hari [Analyze the Influence of Artificial Intelligence (AI) Technology in Daily Life]." *Jurnal Sistem Informatika Dan Manajemen* 11(1):41-42.
- Harahap, Fahri Aulia Alfarisi, Ronaldo Mardianson Sinaga, Khusnul Arifin, and Kana Saputra S. 2022. "Implementasi Algoritma Convolutional Neural Network Untuk Mendeteksi Penyakit Ginjal Implementation of Convolutional Neural Network for Detecting Kidney Disease." *Teknologi Informasi, Komputer Dan Aplikasinya (JTika)* 4(2):212-19.
- Hendrawati, Theresia, Dewa Ayu Putri Wulandari, and Ni Luh Wiwik Sri Rahayu Ginantra. 2024. "Penerapan Metode Stable Diffusion Dengan Fine Tuning Untuk Pola Endek Bali." *Jurnal Teknologi Informasi Komunikasi (e-Journal)* 10:8-10.
- Pongtambing, Yulita Sirinti, Felycitae Ekalaya Appa, A. Muh. Amil Siddik, Eliyah A. M. Sampetoding, Harsi Admawati, Arini Anestesia Purba, Arnold Sau, and Esther S.

- Manapa. 2023. "Peluang Dan Tantangan Kecerdasan Buatan Bagi Generasi Muda." *Bakti Sekawan: Jurnal Pengabdian Masyarakat* 3(1):23–28. doi: 10.35746/bakwan.v3i1.362.
- Puspita, Kharisma Galuh. 2023. "Analisis Kestabilanmodel Matematika Penyebaran Penyakit Demam Berdarah Dengue Dengan Pengaruh Treatmentdan Fogging." *MATHunesa Jurnal IlmiahMatematika* 11(03):318–27.