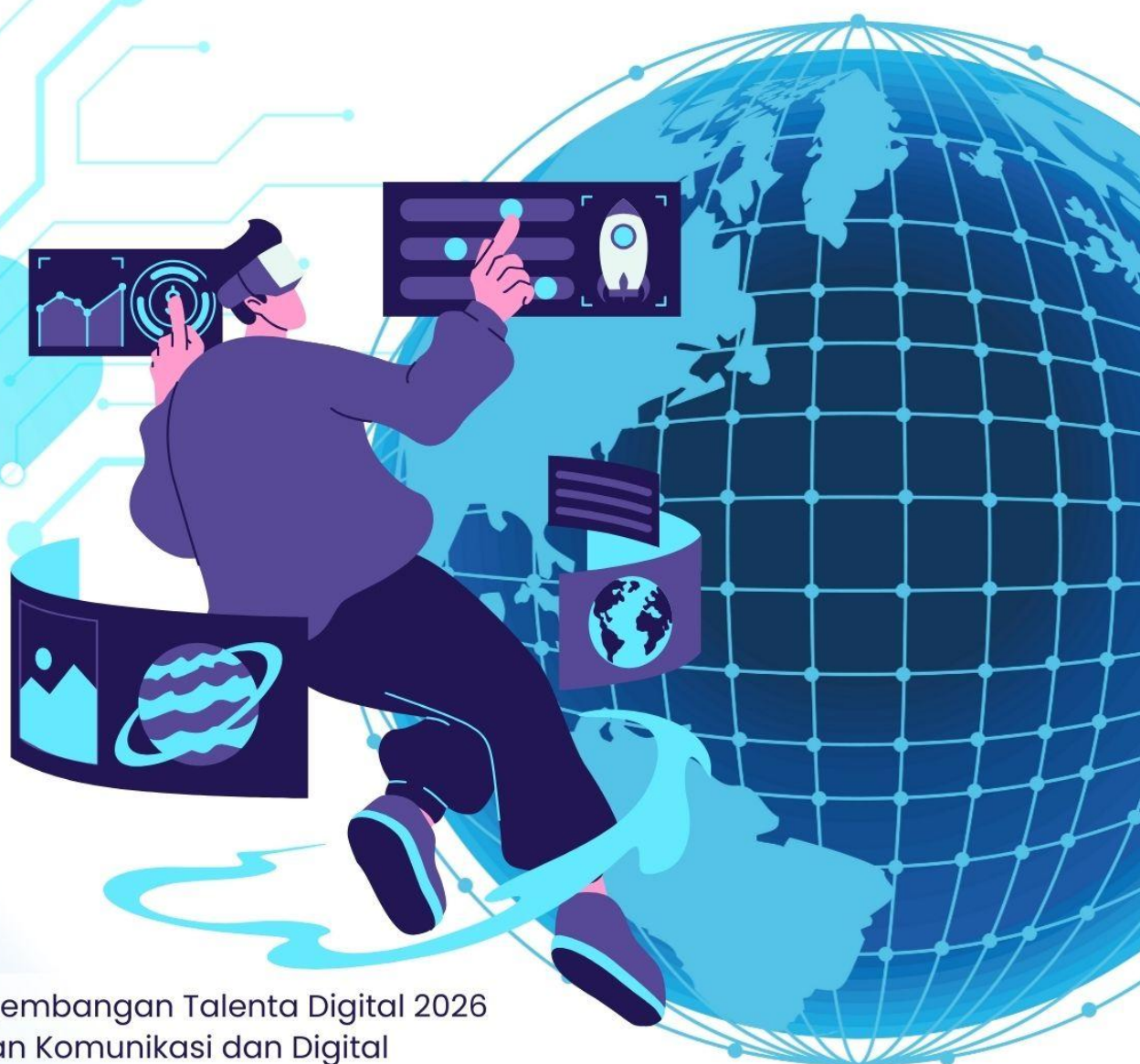


Silabus Pelatihan

LLM: Dari Teori hingga Kode



Silabus Pelatihan Program 1 - LLM: Dari Teori hingga Kode

Kementerian Komunikasi dan Digital Republik Indonesia

Digital Talent Scholarship

Tahun 2026

Disclaimer: Dokumen ini digunakan hanya untuk kebutuhan Digital Talent Scholarship Kementerian Komunikasi dan Digital Republik Indonesia. Konten ini mengandung Kekayaan Intelektual, pengguna tunduk kepada undang-undang hak cipta, merek dagang atau hak kekayaan intelektual lainnya. Dilarang untuk memproduksi, memodifikasi, menyebarluaskan, atau mengeksploitasi konten ini dengan cara atau bentuk apapun tanpa persetujuan tertulis dari Digital Talent Scholarship Kementerian Komunikasi dan Digital Republik Indonesia.

Informasi Pelatihan dan Sertifikasi	
Akademi	Digital Talent
Mitra Pelatihan	-
Tema Pelatihan	Program 1 - LLM: Dari Teori hingga Kode
Sertifikasi	Certificate of Completion DTS Komdigi
Durasi Pelatihan	21 JP
Deskripsi Pelatihan	LLM Fundamental: Dari Teori hingga Kode adalah pelatihan dasar yang dirancang untuk individu yang ingin memahami konsep, peran, dan keterampilan penting dalam pengembangan Large Language Model (LLM) sebagai bagian dari bidang Artificial Intelligence (AI) modern. Selain itu, pelatihan ini juga dirancang untuk membekali peserta dengan pemahaman dan keterampilan sistematis dalam membangun, melatih, dan mengimplementasikan model bahasa besar dari perspektif praktis dan profesional menggunakan Python. Program ini mengacu pada SKKNI No. 299 Tahun 2020, No. 103 Tahun 2023, dan No. 123 Tahun 2021, serta diselaraskan dengan okupasi TIK.DSC0704 – Insinyur Kecerdasan Artifisial (AI Engineer). Pelatihan ini cocok diikuti oleh: 1. calon profesional di bidang AI dan Data Science yang ingin memahami dasar LLM, 2. <i>developer</i> atau <i>Data Engineer</i> yang ingin memperdalam aspek bahasa alami (NLP), 3. peneliti, dosen, atau mahasiswa yang tertarik dengan bidang NLP, <i>generative AI</i>, dan LLM, 4. profesional TI yang ingin beralih karier menuju bidang <i>Generative AI Engineering</i>. Peluang Kerja dan Gaji: Berdasarkan riset pasar tenaga kerja digital, posisi <i>AI Engineer</i> atau <i>Machine Learning Engineer</i> pemula memiliki kisaran gaji awal antara Rp7.000.000 – Rp12.000.000 per bulan tergantung lokasi dan perusahaan. Untuk profesional dengan pengalaman lebih

Informasi Pelatihan dan Sertifikasi	
	dari 5 tahun, rata-rata kisaran gaji yang didapatkan antara Rp12.000.000 - Rp20.000.000, bahkan untuk posisi Lead AI Engineer dapat mencapai di atas Rp20.000.000 - Rp35.000.000 per bulan. Posisi di bidang AI dan LLM memiliki permintaan yang sangat tinggi, terutama di sektor teknologi, finansial, pendidikan, dan startup digital. Berdasarkan data dari berbagai platform rekrutmen digital, ratusan lowongan <i>AI Engineer</i> dan <i>Machine Learning Engineer</i> tersedia setiap bulannya di Indonesia, menunjukkan potensi karier yang sangat menjanjikan. Pelatihan ini dirancang sebagai fondasi bagi peserta untuk siap bersaing di dunia kerja digital, baik di perusahaan nasional maupun global yang membutuhkan talenta AI yang kompeten dan adaptif terhadap perkembangan teknologi <i>Generative AI</i>.
Output Pelatihan	1. Menjelaskan konsep dasar dan ruang lingkup AI Engineering serta mengidentifikasi prinsip kerja Generative AI yang menjadi dasar berbagai aplikasi cerdas. 2. Menganalisis peran dan menilai tanggung jawab seorang AI Engineer dalam merancang dan membangun sistem AI yang etis, aman, dan berkelanjutan. 3. Memformulasikan masalah pemodelan bahasa secara statistik serta menjelaskan evolusi arsitektur dari <i>n</i>-gram hingga Transformer. 4. Menganalisis mekanisme inti Transformer termasuk self-attention, multi-head attention, dan menguraikan fungsi komponen utamanya. 5. Membandingkan dan membedakan berbagai arsitektur Transformer (<i>Encoder-Only</i>, <i>Decoder-Only</i>, <i>Encoder-Decoder</i>) serta mengasosiasikannya dengan tipe tugas NLP yang relevan. 6. Menjelaskan skala data dan kebutuhan komputasi dalam pelatihan <i>Large Language Model</i> (LLM) serta menganalisis prinsip scaling law untuk menentukan trade-off antara data, ukuran model, dan compute. 7. Menguraikan tahapan pelatihan LLM dari pretraining hingga preference alignment dan menilai teknik kompresi model seperti knowledge distillation. 8. Membandingkan strategi Fine-Tuning, Continued Pretraining, dan pelatihan dari awal untuk menentukan pendekatan adaptasi model yang paling efektif sesuai konteks dan sumber daya.
Aktivitas Pelatihan	Pelatihan akan dilakukan secara daring (<i>online</i>), peserta belajar secara mandiri (<i>self-paced learning</i>) menggunakan Platform LMS Komdigi. Untuk lulus di pelatihan ini peserta diharuskan mengikuti dan menyelesaikan: • Tes Substansi • 4 Modul Pembelajaran • Kuis Per Modul • Penilaian Akhir (Final Assessment)
Persyaratan Peserta	Peserta disarankan memiliki: 1. Pemahaman dasar pemrograman dan Python

Informasi Pelatihan dan Sertifikasi	
	2. Dasar matematika dan statistik 3. (Opsional) Familiar dengan cloud coding environment (Google Colab, Kaggle, dll.) 4. (Opsional) Familiar dengan open repository platform seperti GitHub atau Hugging Face
Persyaratan Sarana Peserta	1. Perangkat Komputasi Minimum • Laptop atau PC dengan sistem operasi: - Windows 10/11, macOS (Catalina ke atas), atau Linux (Ubuntu 20.04 ke atas) • Spesifikasi minimum: - CPU: Quad-core (Intel i5 / AMD Ryzen 5 atau lebih tinggi) - RAM: Minimum 16 GB (disarankan 32 GB untuk eksperimen model besar) - Storage: Minimal 50 GB ruang kosong 2. Perangkat Lunak dan Lingkungan Pemrograman Bahasa Pemrograman: Python 3.10 atau lebih baru • Package utama: - PyTorch - transformers, datasets, dan tokenizers dari Hugging Face - numpy, pandas, matplotlib, dan scikit-learn untuk analisis data dan visualisasi 3. Akses Internet dan Platform Pendukung Koneksi internet stabil (minimal 10 Mbps) untuk mengunduh dataset, model, dan dependensi. • Akses akun pada platform berikut (gratis atau akademik): - GitHub (untuk repository dan version control) - Hugging Face (untuk eksplorasi model open source dan eksperimen LLM) - Google Colab / Kaggle (untuk pelatihan berbasis GPU tanpa konfigurasi lokal)

Rencana Pelatihan					
No	Modul	Topik	Outcome	JP	Pemadanan Unit Kompetensi
1	1.1 Pengantar GenAI Engineering untuk Aplikasi LLM: Pengenalan dan konsep dasar penting	1.1 Pengantar GenAI dan AI Engineering: proses end-to-end pengembangan aplikasi AI berbasiskan LLM 1.2 Konsep matematika penting	❖ Menjelaskan seluk-beluk AI Engineering serta mengidentifikasi komponen dan prinsip kerja model Generative AI yang menjadi dasar	5	J.62AIN00.004.1 - Membuat Rencana Proyek Solusi Artificial Intelligence (AI) J.62AIN00.002.1 - Menentukan Sasaran Teknis Solusi Artificial

Rencana Pelatihan					
No	Modul	Topik	Outcome	JP	Pemadanan Unit Kompetensi
		untuk memahami LLM 1.2.1 Matriks: Operasi, Perkalian, Faktorisasi Matriks, Ekstensi ke Tensor 1.2.3 Backpropagation dan autograd 1.2.4 Vektor: Operasi vektor, jarak / kemiripan vektor 1.2.5 Peluang: Distribusi peluang, sampling 1.3 Pengenalan Pytorch 1.3.1 Implementasi berbagai operasi dalam Pytorch 1.3.2 Implementasi model sederhana dalam Pytorch 1.3.3 Implementasi training dengan Pytorch 1.3.4 Konsep lanjutan Pytorch	berbagai aplikasi berbasis AI. ❖ Menganalisis peran dan menilai tanggung jawab seorang AI Engineer dalam merancang, membangun, dan mengimplementasikan sistem berbasis AI secara etis dan berkelanjutan. ❖ Menerapkan konsep-konsep dasar AI Engineering untuk mempersiapkan diri mengikuti modul-modul lanjutan dalam program ini.		Intelligence (AI)
2	1.2 Fundamental LLM: Dari Pemodelan Bahasa Statistik ke Arsitektur Transformer	2.1 Memahami masalah pemodelan bahasa secara statistik 2.2 Pemodelan bahasa menggunakan n-gram dan Kneser-Ney smoothing 2.3 Mekanisme attention (self-attention → multi-head attention) dan perbandingannya dengan pendahulunya	❖ Memformulasikan masalah pemodelan bahasa secara statistik (P(W) dan menjelaskan prinsip dasar serta keterbatasan model n-gram sebagai fondasi historis. ❖ Menganalisis keterbatasan fundamental arsitektur rekuren	6	J.62AIN00.002.1 - Menentukan Sasaran Teknis Solusi Artificial Intelligence (AI) J.62AIN00.004.1 - Membuat Rencana Proyek Solusi Artificial Intelligence (AI) J.62AIN00.015.1 - Memasang Solusi Artificial Intelligence (AI)

Rencana Pelatihan					
No	Modul	Topik	Outcome	JP	Pemadanan Unit Kompetensi
		(RNN, LSTM, GRU, dst.), serta evolusinya (misalnya FlashAttention) 2.4 Arsitektur Transformer dan komponen penyusun sebuah LLM 2.5 Evolusi arsitektur Transformer (encoder-decoder, seperti BART, T5; encoder-only, seperti BERT, ROBERTA)	(RNN/LSTM) dalam menangani ketergantungan sekuensial jangka panjang. ❖ Menjelaskan evolusi konseptual dari attention standar ke self-attention dan multi-head attention sebagai mekanisme inti untuk menangkap konteks. ❖ Menguraikan arsitektur Transformer orisinal (Vaswani et al., 2017), mengidentifikasi fungsi spesifik dari setiap komponen penyusunnya (misalnya, Positional Encoding, Encoder Stack, Masked Decoder Stack). ❖ Membandingkan dan membedakan (compare and contrast) tiga keluarga arsitektur Transformer (Encoder-Only, Decoder-Only, Encoder-Decoder) dan mengasosiasikannya dengan tipe tugas NLP yang sesuai (misalnya, NLU vs. NLG vs. Seq2Seq).		

Rencana Pelatihan					
No	Modul	Topik	Outcome	JP	Pemadanan Unit Kompetensi
3	1.3 Fundamental LLM: Komponen Inti dan Arsitektur LLM Modern	3.1 Tokenisasi dan Embedding: pra-proses fundamental LLM (BPE, trainable embeddings, positional encodings, special tokens) 3.2 Arsitektur LLM modern (autoregressive decoder-only) dan evolusinya 3.3 Proses sampling pada LLM (softmax, temperature tuning, konfigurasi sampling, Constrained Decoding) 3.4 Adaptasi untuk LLM multimodal: multi-modal projectors 3.5 Mixture of Experts (MoE): concept 3.6 Pendekatan alternatif: model state-space (MAMBA) dan Diffusion LLM 3.7 Library Transformers dan Tokenizers sebagai toolkit praktis untuk implementasi LLM	❖ Menjelaskan arsitektur Decoder-Only (Autoregressive) sebagai fondasi LLM generatif modern dan membedakannya dari arsitektur Encoder-Decoder penuh. ❖ Menguraikan proses sampling dari logits ke Softmax, dan menganalisis bagaimana parameter kontrol (seperti Temperature, Top-K, dan Top-P) memengaruhi perilaku output model. ❖ Mengidentifikasi dan membedakan evolusi serta pendekatan arsitektur lanjutan, termasuk Multimodal Projectors (untuk VLM), Mixture of Experts (MoE) (untuk efisiensi), dan Mamba (sebagai alternatif Transformer).	5	J.62AIN00.004.1 - Membuat Rencana Proyek Solusi Artificial Intelligence (AI), J.62AIN00.002.1 - Menentukan Sasaran Teknis Solusi Artificial Intelligence (AI), J.62AIN00.015.1 - Memasang Solusi Artificial Intelligence (AI)
4	1.4 Training LLM: From-Scratch, Fine-Tuning, atau Continued Pretraining?	4.1 Kebutuhan data dan compute dalam proses pelatihan LLM 4.2 Scaling laws: menyeimbangkan jumlah data dan ukuran model untuk	❖ Menjelaskan skala data dan kebutuhan komputasi yang diperlukan dalam proses pelatihan Large Language Model (LLM).	5	J.62AIN00.002.1 - Menentukan Sasaran Teknis Solusi Artificial Intelligence (AI) J.62AIN00.004.1 - Membuat Rencana Proyek Solusi

Rencana Pelatihan					
No	Modul	Topik	Outcome	JP	Pemadanan Unit Kompetensi
		memaksimalkan efisiensi compute 4.3 Tahapan melatih LLM dari awal: pretraining dan post-training 4.4 Distillation: teknik kompresi model yang ramah performa 4.5 Fine-tuning untuk adaptasi LLM ke domain atau tugas spesifik 4.6 Continued pretraining untuk menambahkan pengetahuan baru ke LLM 4.7 Pertimbangan dalam memilih antara training from scratch, fine-tuning, atau continued pretraining	❖ Menganalisis prinsip scaling law pada LLM untuk menentukan keseimbangan (trade-off) antara jumlah data, ukuran model, dan kapasitas komputasi guna mengoptimalkan penggunaan sumber daya. ❖ Menguraikan tahapan utama dalam proses pelatihan LLM, mulai dari pretraining hingga preference alignment. ❖ Menjelaskan serta membedakan berbagai teknik kompresi LLM, termasuk metode seperti knowledge distillation, dan menilai dampaknya terhadap performa model. ❖ Membandingkan pendekatan Fine Tuning, Continued Pretraining, dan pelatihan dari awal, serta menentukan strategi yang paling sesuai berdasarkan kebutuhan dan sumber daya yang tersedia.		Artificial Intelligence (AI) J.62DMI00.015.1 - Melakukan Proses Review Pemodelan



Pusat Pengembangan Talenta Digital

Kementerian Komunikasi dan Digital RI
