

**DANA PENYELIDIKAN STRATEGIK:
REQUEST FOR PROPOSAL (SRF-RFP)**

TAJUK	PEMBANGUNAN <i>LARGE LANGUAGE MODEL</i> (LLM) DI SEKTOR KESIHATAN BAGI KEGUNAAN KEMENTERIAN KESIHATAN MALAYSIA (KKM): Transformasi Penjagaan Kesihatan Berpandukan Kecerdasan Buatan (AI)
PENYATAAN MASALAH	Kerajaan sentiasa berusaha untuk menambahbaik perkhidmatan kesihatan yang diberikan kepada orang awam agar lebih efisien dan efektif. Antara cabaran yang dikenalpasti adalah; 1) Kompleksiti dalam pengurusan kesihatan sendiri disebabkan oleh kepelbagaian maklumat di alam maya, 2) Kompleksiti pengoperasian harian dalam memberi perkhidmatan kesihatan disebabkan oleh keperluan perkhidmatan yang semakin bertambah, dan 3) Kompleksiti membuat keputusan strategik selari disebabkan oleh kepelbagaian data dan sistem maklumat dari pelbagai sumber. Teknologi AI, seperti <i>Large Language Model</i> (LLM), merupakan salah satu cara untuk mentransformasi sektor kesihatan Malaysia dengan memperkasa kesihatan orang awam melalui pengurusan sistem kesihatan dan penyampaian perkhidmatan kesihatan yang lebih cekap dan efisien.

OBJEKTIF OBJEKTIF UMUM Membangunkan LLM dalam konteks kesihatan di Malaysia dan aplikasi berasaskan LLM yang memberi manfaat kepada rakyat, penjagaan pesakit, pengurusan sistem kesihatan, dan penyelidikan	Objektif Umum: Membangunkan LLM dalam konteks kesihatan di Malaysia dan aplikasi berasaskan LLM yang memberi manfaat kepada orang awam, penjagaan pesakit, pengurusan sistem kesihatan dan penyelidikan. <pre> graph LR A[OBJEKTIF UMUM: Membangunkan LLM dalam konteks kesihatan di Malaysia dan aplikasi berasaskan LLM yang memberi manfaat kepada rakyat, penjagaan pesakit, pengurusan sistem kesihatan, dan penyelidikan] --> B[OBJEKTIF KHUSUS 1: Membangunkan LLM menggunakan pelbagai dataset yang berkaitan dengan kesihatan di Malaysia] A --> C[OBJEKTIF KHUSUS 2: Membina aplikasi berasaskan LLM] B --> D[KEGUNAAN RAKYAT: memperkasa pesakit dalam menjaga kesihatan sendiri] B --> E[PENJAGAAN PESAKIT: pengurusan dan rawatan pesakit yang lebih efisien] B --> F[PENGURUSAN SISTEM KESIHATAN: memudah cara pemprosesan maklumat dan penyelidikan] C --> D C --> E C --> F </pre> Objektif Spesifik: 1) Membangunkan LLM menggunakan pelbagai dataset KKM yang berkaitan dengan kesihatan di Malaysia. Selain daripada LLM, penggunaan berpotensi <i>Small Language Model</i> (SLM) boleh dipertimbangkan sebagai satu pilihan untuk pembangunan model klinikal untuk KKM. 2) Membina aplikasi berasaskan LLM atau SLM dalam beberapa bidang penggunaan seperti berikut:
---	--

	a) Kegunaan Orang Awam: Memperkasa pesakit dalam menjaga kesihatan sendiri. Contohnya, membina <i>Question-Answering Chatbot</i> di laman web KKM atau di sistem <i>Personal Health Record</i> (PHR) untuk mencari maklumat perubatan. b) Penjagaan pesakit: Pengurusan dan rawatan pesakit yang lebih efisien. Contohnya, membina <i>automated text generation</i> untuk menghasilkan <i>discharge summary</i> dan <i>referral letter</i>. c) Pengurusan sistem kesihatan dan penyelidikan: Memudahkan proses pemprosesan maklumat dan penyelidikan dengan mengenal pasti pesakit berdasarkan kriteria penyelidikan. Contohnya, membina <i>medical auto-coding</i> semasa proses pengkodan maklumat seperti diagnosis dan prosedur oleh pegawai rekod perubatan.
--	--

PELAKSANAAN DASAR AI	1. Data and Information Usage	
	<i>Localized Datasets</i>	• To incorporate Bahasa Melayu corpora, considering its status as the national language and its cultural and demographic significance within Malaysia's diverse multicultural society. • The use of contextually relevant local datasets in model training is emphasized, especially for applications in the public sector and healthcare domains.
	<i>Encryption and Tokenization</i>	• The protection of personally identifiable information (PII) through tokenization or masking is a critical step before data is used for analytics or AI model training. • To ensure privacy and compliance, public sector data must undergo anonymization or pseudonymization before being shared for the purpose of AI model training.
	<i>Personal Data Protection Act (PDPA)</i>	• To ensure compliance with Malaysia's data privacy laws, personal data must be properly managed, safeguarded and protected throughout its use in AI or analytics processes.

	2. Local Talent and Technology Adoption <table border="1"> <tr> <td data-bbox="501 255 772 1339"><i>AI Sandbox</i></td><td data-bbox="772 255 1476 1339"> • <i>Develop a common and reusable AI sandbox with regulatory compliance.</i> • <i>A cost-effective AI sandbox designed by local talent to support a wide range of industries, including healthcare, education, agriculture, and more.</i> • <i>Leverage local companies with homegrown technologies to implement AI solutions that are skilled and expert in the respective industry.</i> • <i>Offloading expensive High Bandwidth Memory (HBM) and Graphics Double Data Rate (GDDR) memory to cost-effective flash memory. Eliminates the need for large numbers of high-cost and power-hungry GPU cards during AI model training.</i> </td></tr> <tr> <td data-bbox="501 1339 772 1906"><i>Private and Public Cloud</i></td><td data-bbox="772 1339 1476 1906"> • <i>A private cloud should be utilized for AI model development and training, given that it's an iterative process requiring secure handling of protected and sensitive datasets throughout its lifecycle.</i> • <i>Support the adoption of cloud-based AI, federated learning and edge AI compute for AI applications roll out</i> </td></tr> </table>	<i>AI Sandbox</i>	• <i>Develop a common and reusable AI sandbox with regulatory compliance.</i> • <i>A cost-effective AI sandbox designed by local talent to support a wide range of industries, including healthcare, education, agriculture, and more.</i> • <i>Leverage local companies with homegrown technologies to implement AI solutions that are skilled and expert in the respective industry.</i> • <i>Offloading expensive High Bandwidth Memory (HBM) and Graphics Double Data Rate (GDDR) memory to cost-effective flash memory. Eliminates the need for large numbers of high-cost and power-hungry GPU cards during AI model training.</i>	<i>Private and Public Cloud</i>	• <i>A private cloud should be utilized for AI model development and training, given that it's an iterative process requiring secure handling of protected and sensitive datasets throughout its lifecycle.</i> • <i>Support the adoption of cloud-based AI, federated learning and edge AI compute for AI applications roll out</i>
<i>AI Sandbox</i>	• <i>Develop a common and reusable AI sandbox with regulatory compliance.</i> • <i>A cost-effective AI sandbox designed by local talent to support a wide range of industries, including healthcare, education, agriculture, and more.</i> • <i>Leverage local companies with homegrown technologies to implement AI solutions that are skilled and expert in the respective industry.</i> • <i>Offloading expensive High Bandwidth Memory (HBM) and Graphics Double Data Rate (GDDR) memory to cost-effective flash memory. Eliminates the need for large numbers of high-cost and power-hungry GPU cards during AI model training.</i>				
<i>Private and Public Cloud</i>	• <i>A private cloud should be utilized for AI model development and training, given that it's an iterative process requiring secure handling of protected and sensitive datasets throughout its lifecycle.</i> • <i>Support the adoption of cloud-based AI, federated learning and edge AI compute for AI applications roll out</i>				

		<i>based on the baseline model releases.</i> • <i>To support Malaysia's Cloud First Policy (MyGovCloud), enabling AI-as-a-Service.</i>
	<i>Explainable and Responsible AI</i>	• <i>All AI solutions shall comply and support Explainability (XAI), Auditability and Traceability.</i> • <i>Proposed register for certified AI solutions (MySTI) for safe deployment.</i> • <i>Model could be used and shared with public and private sectors, eventually with open-source communities in the future.</i>
	3. Data and Information Security	
	<i>Data-in-Use</i>	• <i>PII data protection through data pseudonymization, tokenization, encryption, etc.</i> • <i>Role-based access control (RBAC) for data query and access.</i> • <i>Automated backup with version control.</i>
	<i>Data-in-Transit</i>	• <i>All communication among systems must be secured using TLS, HTTPS, VPN, etc.</i>

		• <i>API calls must be secured and authenticated such as using OAuth2, signed tokens (end-points security), etc.</i>
	<i>Data-at-Rest</i>	• <i>Must use strong encryption (AES-256+) for data storage.</i> • <i>Automated backup with version control.</i>
	<i>Classification and Audit Log</i>	• <i>All public sector datasets must be classified (such as confidential, restricted, public, etc.) and access must be logged and audited regularly.</i>
JANGKAAN HASIL PROJEK (KES GUNAAN)	1) Klinikal LLM (CLLM) untuk Sektor Kesihatan yang khusus untuk bidang kesihatan dalam konteks Malaysia yang dihoskan secara berpusat, selamat, boleh dikemaskini berdasarkan perkembangan dan keperluan semasa, serta menjadi repositori language model untuk pelbagai projek AI di masa hadapan. 2) Membina aplikasi untuk kegunaan orang awam, seperti <i>mobile app</i> yang mampu memaparkan status kesihatan diri (<i>health status</i>), penilaian risiko kesihatan diri (<i>health risk assessment</i>), pengesyoran cara hidup sihat (<i>healthy lifestyle recommendations</i>); dan mempunyai <i>Question-Answering Chatbot</i> untuk mencari maklumat kesihatan dan perkhidmatan kesihatan.	

- 3) Membina aplikasi untuk **pengurusan pesakit**, seperti membina peringatan/*alerts* apabila terdapat hasil ujian yang membahayakan pesakit; membangunkan *automated text generation* untuk menghasilkan *discharge summary* dan *referral letter* daripada rekod pesakit; dan membina sistem *clinical decision support* yang memberi panduan kepada pegawai perubatan berasaskan *Clinical Practice Guideline* (CPG) yang melalui *Question-Answering Chatbot*.
- 4) Membina aplikasi untuk **pengurusan sistem kesihatan**, seperti automasi koding perubatan dalam ICD dan SNOMED CT; penghasilan *Disease Related Group* (DRG); mengenalpasti kohort pesakit untuk penyelidikan klinikal; dan penggunaan *Question-Answering Chatbot* untuk pencarian maklumat statistik kesihatan Malaysia.

KEGUNAAN RAKYAT

Status dan penilaian risiko kesihatan diri
(Health status and risk assessment)

Pengesyoran cara hidup sihat
(Healthy lifestyle recommendations)

Chatbot untuk mendapatkan maklumat tentang perkhidmatan kesihatan

PENGURUSAN PESAKIT

Peringatan dan *Alerts* apabila terdapat hasil ujian membahayakan

Automasi penyediaan *discharge summary* dan surat rujukan

Clinical decision support (CDS) untuk beri panduan kepada pegawai perubatan berasaskan *Clinical Practice Guideline* (CPG) melalui Chatbot

PENGURUSAN SISTEM KESIHATAN

Automasi koding perubatan seperti ICD dan SNOMED CT

Chatbot untuk mencari statistik kesihatan Malaysia

Penghasilan *Disease Related Group* (DRG)

Mengenalpasti kohort pesakit untuk penyelidikan klinikal

JUSTIFIKASI	1) Menyokong Pelan Hala Tuju Kecerdasan Buatan Negara 2021-2025 dengan keutamaan penjagaan kesihatan. 2) Menyokong Kertas Putih Kesihatan untuk transformasi sistem kesihatan Malaysia melalui pemerkasaan pendigitalan di bidang kesihatan. 3) Menjadi asas kepada pembangunan teknologi kesihatan AI, seperti: a) Membolehkan orang awam mendapatkan perkhidmatan kesihatan dan maklumat tentang perkhidmatan kesihatan dengan lebih mudah; b) Memperkasa orang awam dengan panduan kesihatan sendiri yang diperibadikan melalui maklumat yang mudah difahami; c) Meningkatkan kecekapan proses harian dan membantu dalam membuat keputusan klinikal untuk memenuhi keperluan penjagaan pesakit; d) Memperkasa pengurusan sistem kesihatan dan penyelidikan melalui strategi berasaskan data, justeru menyokong keputusan strategik agensi; dan e) Mempersiapkan negara dengan kapasiti pemprosesan data untuk transformasi pembiayaan kesihatan.
--------------------	--

SPESIFIKASI MATRIKS DAN SKOP KERJA	Projek ini bertujuan untuk membangunkan LLM klinikal dengan ketepatan berkonteks dan pematuhan yang tinggi. Hasilnya merangkumi sokongan diagnostik dan analisis data yang lebih baik. Penanda aras merangkumi ketepatan, kerelevanan respons, dan pematuhan privasi. Matriks pengukuran menilai prestasi berdasarkan ketepatan (<i>accuracy</i>), ingatan (<i>recall</i>), skor (<i>score</i>), kependaman (<i>latency</i>), dan kepuasan pengguna (<i>user satisfaction</i>) melalui pengesahan klinikal dan ujian perintis dunia sebenar. 1) KKM Klinikal <i>Large Language Model</i> (CLLM) a) Mengenalpasti, mengumpul, dan memproses pelbagai <i>dataset</i> yang berkaitan dengan kesihatan di KKM. Pemprosesan ini termasuk memastikan anonimisasi identiti pesakit atau pengguna dalam <i>dataset</i> yang akan digunakan; b) Menyediakan dan menyerahkan semula <i>dataset</i> yang telah diproses (<i>training dataset</i>), bagi kegunaan KKM pada masa hadapan; c) Menyediakan perkhidmatan <i>hosting</i>, perkakasan dan perisian yang diperlukan untuk pembangunan dan penggunaan LLM yang berskala; d) Melatih model LLM yang berskala, efisien, dan sesuai dengan konteks penjagaan kesihatan Malaysia; e) Menguji prestasi model untuk memastikan keberkesanan dan ketepatan keputusan yang dihasilkan;
--	---

	f) Memastikan model mematuhi garis panduan dan prinsip AI yang bertanggungjawab, termasuk privasi data, ketelusan, dan keadilan; dan g) Menambahbaik model dari semasa ke semasa mengikut keperluan dan perkembangan teknologi. 2) Aplikasi KKM Berasaskan Klinikai LLM a) Mereka bentuk dan membangunkan aplikasi kesgunaan berdasarkan LLM yang dipersetujui bersama; b) Menyediakan infrastruktur dan perkhidmatan hosting, perkakasan, perisian yang diperlukan untuk pembangunan aplikasi LLM; c) Menguji aplikasi dalam pelbagai situasi/senario untuk memastikan keandalan dan keselamatan penggunaannya; d) Melancar aplikasi kepada pengguna dan memberi latihan kepada kakitangan yang berkaitan; e) Memastikan aplikasi mematuhi standard informatik kesihatan seperti ICD, SNOMED CT, LOINC, dan MyHDW; f) Memastikan aspek keselamatan bagi aplikasi yang dibangunkan mengikut garis panduan, pekeliling, dan piawaian ISO yang berkaitan; g) Memastikan aplikasi yang dihasilkan berintegrasi dengan sistem sedia ada di KKM; dan h) Menyelenggara dan menambahbaik aplikasi mengikut keperluan dan perkembangan teknologi dan metodologi.
--	--

GARIS MASA	Pembangunan projek dibahagikan kepada tiga fasa utama; Setiap fasa pembangunan disarankan menggunakan penyelesaian pembangunan sistem yang terdapat di dalam MySTI (contoh: AIDevX yang dibangunkan oleh MIMOS). Fasa 1: Pembangunan LLM [9 bulan] • Pengumpulan dan pemprosesan data, menggabungkan dan menyediakan data termasuk memastikan anonimisasi maklumat pesakit yang sensitif; • Pemilihan model pra-latihan yang sesuai sebagai asas untuk melatih LLM yang berskala dan efisien yang disesuaikan dengan konteks penjagaan kesihatan Malaysia; dan • Menguji prestasi LLM dan penalaan halus (<i>fine-tune</i>) parameter mengikut keperluan aplikasi KKM. Fasa 2: Pembangunan Aplikasi [12 bulan] • Mereka bentuk, membangunkan, melaksana, dan menguji aplikasi berasaskan LLM; • Melancarkan aplikasi berasaskan LLM di peringkat pengguna; dan • Melaksanakan integrasi dengan sistem maklumat sedia ada di KKM. Fasa 3: Pelaksanaan Rintis dan Latihan [12 bulan] • Pelaksanaan rintis prestasi aplikasi; • Penilaian LLM dan aplikasi berdasarkan kepuasan penggunaan pengguna;
-------------------	---

	• Penalaan halus (<i>fine-tuning</i>) LLM berdasarkan maklumbalas pengguna; dan • Pemindahan teknologi dan latihan pembangunan pengetahuan kemahiran.	
NILAI DANA MAKSIMUM (RM)	Anggaran RM15 juta.	
TECHNOLOGY READINESS LEVEL (TRL)	Projek hendaklah berada minimum di peringkat TRL 3 (semasa permohonan) dan mencapai TRL 7 dan ke atas (tamat tempoh pembiayaan/projek).	
RUJUKAN	Pusat Informatik Kesihatan, Bahagian Perancangan, Kementerian Kesihatan Malaysia (KKM)	
POTENSI PEMEGANG TARUH	1) Kementerian Sains, Teknologi, dan Inovasi (MOSTI) 2) Kementerian Kesihatan Malaysia (KKM) 3) Kementerian Komunikasi (KKOM) 4) Kementerian Digital (KD) 5) Jabatan Digital Negara (JDN)	
PERMOHONAN	Saluran Permohonan	Permohonan Dana Penyelidikan Strategik – RFP (SRF-RFP) boleh dibuat melalui Sistem Dana Bersepadu MOSTI di pautan SDB: SRF-RFP
	Tempoh Permohonan	33 bulan untuk 3 fasa. Fasa 1: 9 bulan Fasa 2: 12 bulan Fasa 3: 12 bulan
	Garis Panduan Permohonan	Perincian syarat permohonan adalah seperti terkandung dalam Garis Panduan SRF-RFP di pautan GP SRF-RFP