Experts

Bahasa Arab di Sebalik Kod Rahsia Revolusi AI

Fri, 12/05/2025 - 09:39

Pengalaman ini membawa saya kepada sebuah perjalanan peribadi untuk memahami peranan kritikal bahasa Arab dalam gelombang Revolusi Kecerdasan Buatan (AI) hari ini, suatu peranan yang pada mulanya saya pandang remeh.

Pada tahun 2012, ketika pertama kali melangkah ke makmal perisian di Pusat Bahasa Moden (PBM), Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA), fokus utama saya ialah kod Python dan JavaScript. Bahasa ibunda saya, bahasa Arab, bahasa puisi kuno, kitab suci, dan jutaan penutur terasa seperti tinggalan budaya yang indah, tetapi tidak relevan dalam dunia teknologi moden. Kami bekerja keras untuk membangunkan AI untuk bahasa-bahasa Eropah, namun apabila saya cuba memasukkan bahasa Arab ke dalam projek Pemprosesan Bahasa Asli (NLP) sedia ada, sistem kami gagal. Ia gagal secara menyeluruh, meninggalkan saya dengan rasa gusar dan persoalan: Adakah bahasa Arab hanya akan menjadi penumpang di pinggir jalan raya AI, atau adakah ia memegang kunci kepada reka bentuk sistem yang lebih pintar? Saya mula sedar, saya telah melihat bahasa Arab sebagai hanya satu set data yang perlu diproses, sedangkan hakikatnya, ia adalah ‘Kod Sumber Kuno’ yang bukan sahaja memerlukan pemecahan, tetapi juga penghormatan mendalam terhadap struktur intipatinya yang unik.

Revolusi Model Bahasa Besar (LLM) seperti GPT dan LLaMA, bergantung sepenuhnya pada kemampuan mereka untuk memahami dan menjana bahasa manusia. Sementara bahasa seperti bahasa Inggeris, dengan susunan kata yang agak tegar, mudah diproses. Bahasa Arab menawarkan satu cabaran dan peluang linguistik yang berbeza iaitu morfologi yang kaya. Bahasa Arab dibina atas sistem akar konsonan tiga huruf (*trilateral root system*) yang membawa makna semantik teras. Akar ini kemudiannya diubah suai melalui imbuhan dan vokal untuk menghasilkan puluhan perkataan dengan makna yang berkaitan. Sebagai contoh, akar {? ? ?} membawa makna asas ‘menulis’, daripadanya lahir Kitab (buku), K?tib (penulis) dan Maktab (pejabat). Bagi AI, satu kata dasar dalam bahasa Inggeris mungkin diterjemahkan kepada sedozen bentuk yang berbeza dalam bahasa Arab. Cabarannya, ini mewujudkan masalah ‘kelangkaan data’ di mana AI melihat pelbagai bentuk perkataan yang berkaitan sebagai entiti yang berbeza. Walau bagaimanapun, keunikan ini juga adalah peluang emas.

Apabila algoritma AI berjaya memecahkan kod akar ini, ia mampu memahami rangkaian hubungan semantik yang jauh lebih dalam dan tersusun. AI yang dilatih untuk mengenali akar {? ? ?} tidak hanya melihat tiga perkataan berbeza namun ia melihat satu konsep yang kompleks seakan ‘komunikasi bertulis’ yang diekspresikan dalam pelbagai cara, sekali gus membolehkan AI membina pemahaman konteks yang lebih kaya, berpotensi memintas masalah yang dihadapi oleh model yang bergantung semata-mata pada susunan kata.

Usaha untuk melatih LLM berbahasa Arab, seperti projek JAIS, bukan hanya bertujuan untuk menyediakan perkhidmatan terjemahan atau ringkasan, tetapi ia adalah tentang membina AI yang benar-benar boleh berfikir dan berinteraksi dalam bahasa Arab pada tahap yang bernuansa. Untuk mencapai tahap ini, jurutera terpaksa membangunkan teknik tokenisasi dan pemodelan baharu yang

direka khas untuk menangani morfologi Arab yang kaya, sambil mengambil kira varian dialek, sintaksis yang fleksibel, dan kerumitan seperti kekurangan penanda vokal yang boleh mengubah makna perkataan secara drastik (contoh: ‘ilm (ilmu) vs ‘alam (bendera)).

Menyelesaikan teka-teki bahasa Arab ini bukan hanya meningkatkan kualiti pemprosesan bahasa tersebut, tetapi ia adalah latihan untuk membina AI yang lebih universal. Jika sebuah model boleh menguasai sistem linguistik yang kompleks dan tersusun seperti bahasa Arab, ia akan menjadi lebih bersedia untuk menangani kerumitan mana-mana bahasa lain di dunia. Akhirnya, peranan bahasa Arab dalam AI melangkaui aspek teknikal dan menyentuh tema universal ‘Identiti Budaya dalam Era Digital dan Keterangkuman (*Inclusivity*) Teknologi’.

Apabila teknologi AI berkembang pesat, risiko 'Kematian Budaya Digital' semakin membesar. Bagi dunia Arab dan Islam, bahasa bukan sekadar alat komunikasi. Ia adalah saluran kepada warisan sains, falsafah, dan sastera yang luas. Dengan membina AI yang kuat dalam bahasa Arab, kita bukan sahaja mencipta aplikasi; kita mendigitalkan semula perpustakaan sejarah dan memastikan suara jutaan penutur relevan dan didengari dalam ekonomi pengetahuan global. Metafora ‘Kod Sumber Kuno’ kini menjadi lebih jelas.

Warisan linguistik dan struktur mendalam bahasa Arab bukanlah beban, tetapi harta karun. Ia menyediakan model baharu untuk kecerdasan buatan, mengajar kita bahawa kefahaman sebenar datang daripada menghormati kerumitan dan keunikan bukan hanya pemprosesan data secara pukal. Revolusi AI masa depan memerlukan pelbagai jenis 'kecerdasan' linguistik, dan tanpa ragu, bahasa Arab adalah sebahagian penting daripada cetak biru untuk membina AI yang benar-benar pintar, inklusif, dan mencerminkan kekayaan intelektual seluruh peradaban manusia.



Oleh: Mohammad Baihaqi Hasni
E-mel: baihaqi@umpsa.edu.my

Penulis adalah Guru Bahasa Kanan, Pusat Bahasa Moden (PBM), Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA).

Rencana ini merupakan pandangan peribadi penulis dan tidak menggambarkan pendirian rasmi Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA).

- 82 views

View PDF