Does Modern Standard Arabic (MSA) Dominate Arabic Dialects in LLMs? A Representation-Level Analysis

Abdu Sallouh, Nicholas Popovič, Michael Färber

Contact abdu.sallouh@mailbox.tu-dresden.de

EMNLP 2026 Main Conference · Paper 6881

1 Overview

- **Large language models (LLMs)** often answer dialectal Arabic in **MSA**, even when prompted in dialect (Robinson et al., 2025).
- **The usual explanation is representational:** MSA dominates the internal space, a hidden pivot dialects are routed through.
- **We test that on** 26 Arabic varieties (25 city dialects and MSA) across five models, multilingual and Arabic-native.

Research question: Does MSA-biased generation imply MSA-dominated hidden states?

No. MSA is one variety among many in hidden space; the bias is in generation, not geometry.

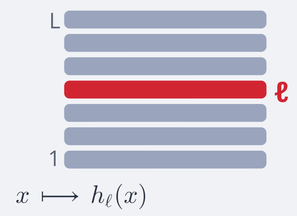
2 Approach: language-dominance probing for Arabic varieties

1 Parallel sentences

MSA	مساعدتك	أستطيع	كيف
Cairo	اساعدك	أقدر	ازاي
Beirut	ساعدك	فيني	كيف
Doha	اساعدك	أقدر	شلون
Tunis	نعاونك	انجم	كيفاش
Rabat	نعاونك	نقدر	كيفاش

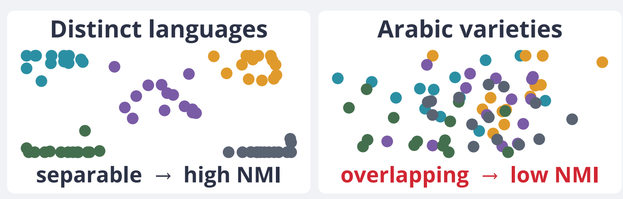
MADAR corpus (Bouamor et al., 2018): *"How can I help you?"*

2 Hidden states



Every layer of 5 LLMs, from BLOOM to Llama-3.1

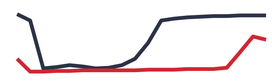
3 Principal components + Gaussian mixture



One Gaussian per variety; PUD (Parallel Universal Dependencies) as control (Shani and Basirat, 2025).

4 Two probes

Separability
normalized mutual information (NMI)
• languages • varieties



Dominance
share of pairs below 1
baseline 1/26 ≈ 0.038

$$\Lambda(x, c) = \frac{p(d_{\text{true}} \mid x)}{p(c \mid x)}$$

Λ is a likelihood ratio: x is a token, c a competing variety, d_{true} its own; **$\Lambda < 1$** means x fits c better.

3 Results: dense overlap, no dominance

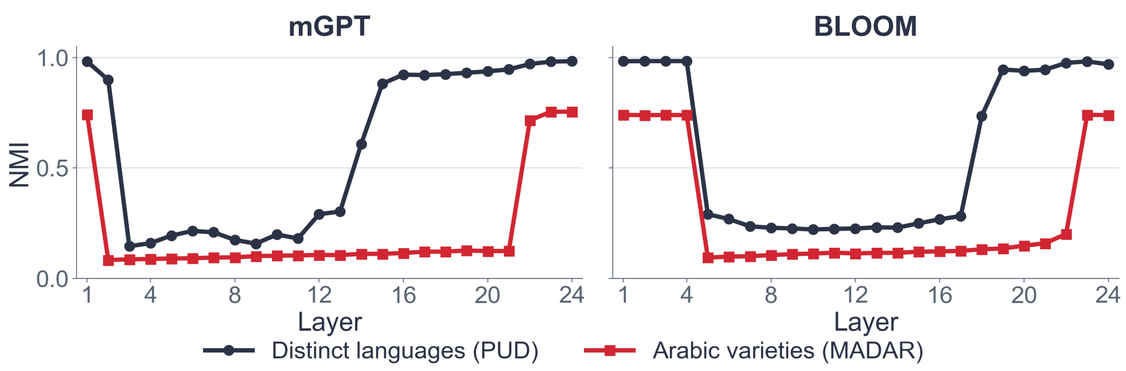


Fig. 1 NMI is low and flat for Arabic varieties (red) vs. distinct languages (navy); the low range is 1–3 layers wider.

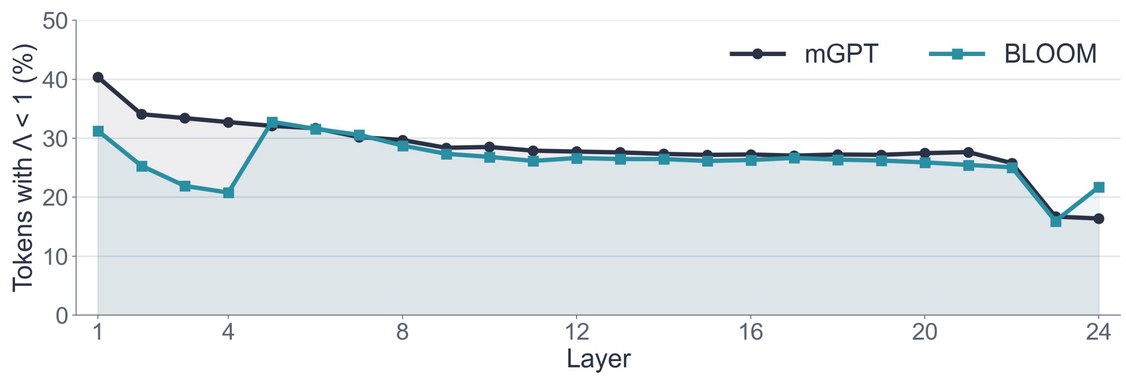


Fig. 3 Tokens with Λ < 1 stay near 30 % with no middle-layer spike: ambiguity, not a directional pull.

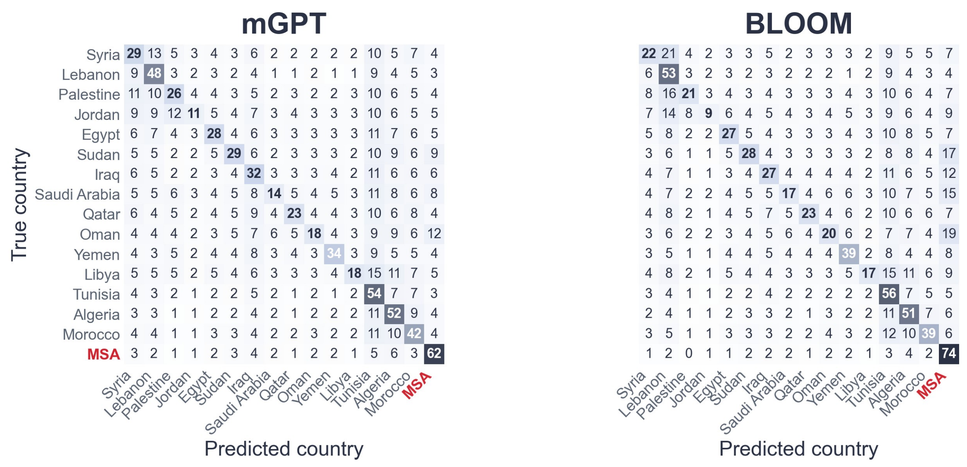


Fig. 2 Clear diagonals: dialect identity is preserved. Cities pooled by country (size-normalised); off-diagonal mass sits with neighbours.

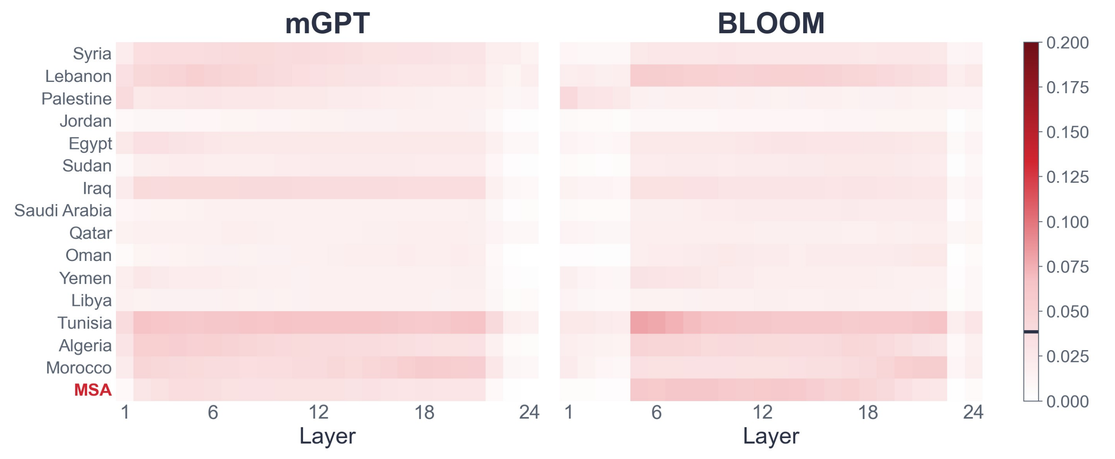


Fig. 4 No hub. Every row sits at the uniform baseline 1/26 ≈ 0.038 (dark tick), MSA included.
Mean dominance score per city. Countries with several cities. Syria: Aleppo, Damascus · Jordan: Amman, Salt · Egypt: Alexandria, Aswan, Cairo · Iraq: Baghdad, Basra, Mosul · Saudi Arabia: Jeddah, Riyadh · Libya: Benghazi, Tripoli · Tunisia: Sfax, Tunis · Morocco: Fes, Rabat

4 Other models and a supervised check

All models: BLOOM · mGPT · Llama-3.1 · Qwen2.5 · ArabianGPT

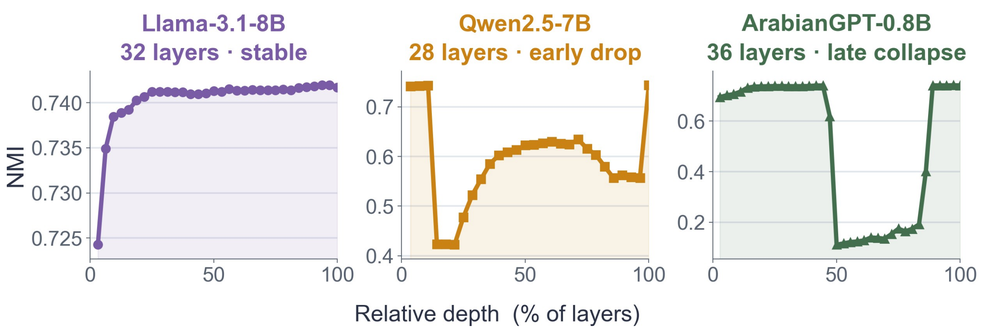


Fig. 5 Separate y-axes: each panel keeps its own scale, as labelled above.

Supervised linear probe

Dialect labels stay decodable from the same hidden states, so low NMI means overlap, not erasure.

BLOOM	**0.19 – 0.25**
mGPT	**0.14 – 0.21**
chance (1/26)	**0.038**

Robustness checks

The picture does not depend on the preprocessing choices:

- POS filtering and prefix removal keep the trends
- PCA dimensionality varied (0.5–3×)
- Cross-lingual PUD replication reproduces prior results

Takeaway Generation bias ≠ representational dominance.
Analyses of multilingual and dialectal LLMs should separate what models prefer to generate from how they organise closely related varieties.

Limitations travel-domain MADAR only; 400 (BLOOM, mGPT) or 100 sentences per dialect; rule-based POS tagging. Findings are exploratory.

What's next larger, non-travel-domain corpora; more Arabic-centric models; a trained POS tagger; the same curves for other close language families.

Scan for the paper and code
abdu.codes/ArabicLLM

Key references
Shani and Basirat. 2025. Language dominance in multilingual large language models. BlackboxNLP.
Bouamor et al. 2018. The MADAR Arabic dialect corpus and lexicon. LREC.
Robinson et al. 2025. ALQASIDA: Analyzing LLM quality and accuracy systematically in dialectal Arabic. Findings of ACL.
Keleg et al. 2025. Revisiting common assumptions about Arabic dialects in NLP. ACL.
Elshabrawy et al. 2025. When alignment hurts: Decoupling representational spaces in multilingual models. arXiv:2508.12803.

FUNDED BY Federal Ministry of Research, Technology and Space
Software Campus, grant 01IS23070
SACHSEN
This measure is funded by the Federal Government following a resolution of the German Bundestag, and co-financed with tax funds on the basis of the budget passed by the Saxon State Parliament.