

# Afro-Arabic AI: Bridging Technology and Culture

Muhammad Abdul-Mageed  
The University of British Columbia & MBZUAI

X/Twitter: [@mageed](#)

Alexandria — Oct. 14-2024



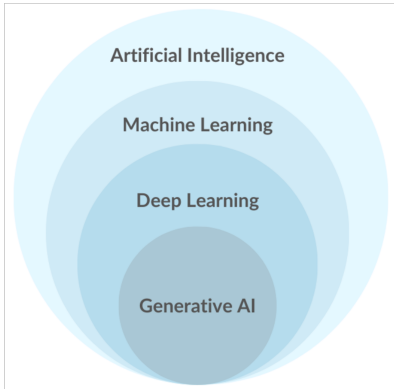
The New York Times

# An A.I.-Generated Picture Won an Art Prize. Artists Aren't Happy.



Jason Allen's A.I.-generated work, "Théâtre D'opéra Spatial," took first place in the digital category at the Colorado State Fair. via Jason Allen

# Generative AI in AI



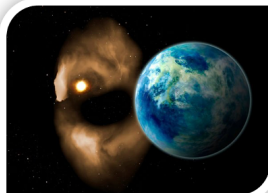
# Breakthroughs

## Biology



Spring and the shape of a protein is so basic that  
it's hard to imagine. (Credit: iStockphoto)

## Astronomy



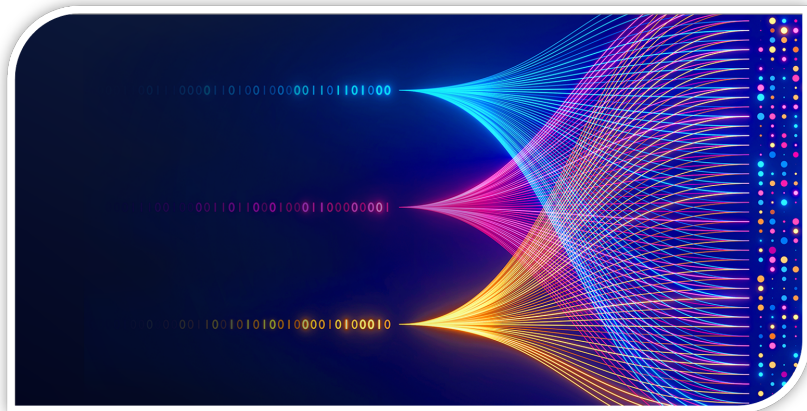
## Agriculture



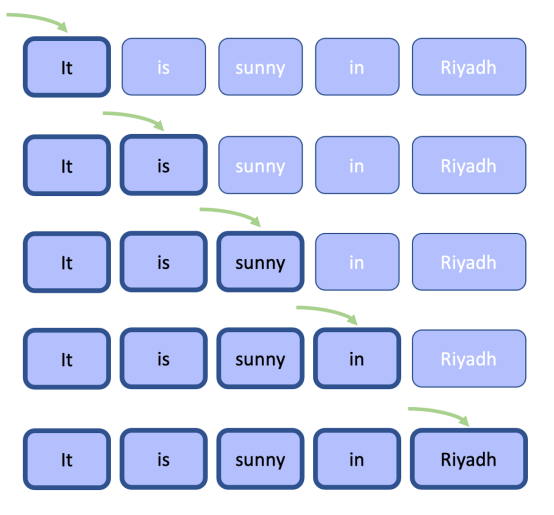
## Chemistry



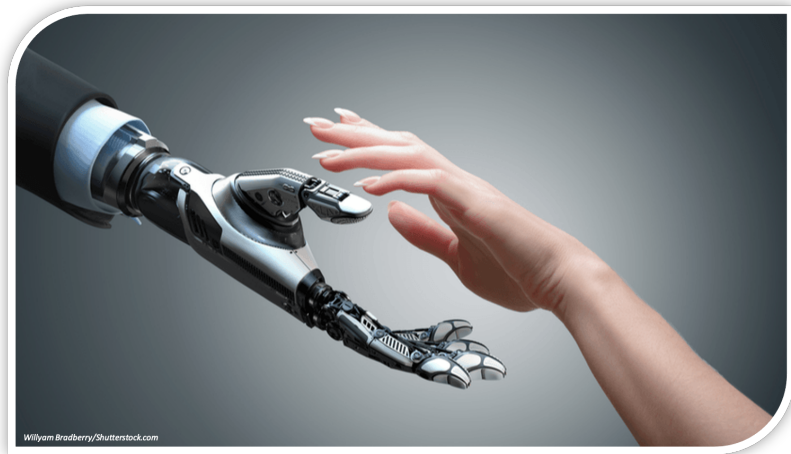
# Large Language Models (LLMs)



# Pretraining



# The Alignment Problem



# What to Align to?

## Helpful

- Attempt to perform a task concisely and efficiently
- Ask for more information when needed
- Respond with sensitivity, insight, and discretion
- Redirect ill-informed requests

## Honest

- Give accurate information
- Express appropriate levels of uncertainty
- Shouldn't deceive humans
- Be honest about itself and its knowledge
- Shouldn't withhold information

## Harmless

- Not be offensive or discriminatory
- Politely refuse to aid in dangerous acts
- Be careful when providing consequential advice

## Moral Foundations Theory

[Graham et al., 2009; Haidt, 2012]

authority - subversion

care - harm

fairness - cheating

loyalty - betrayal

sanctity - degradation

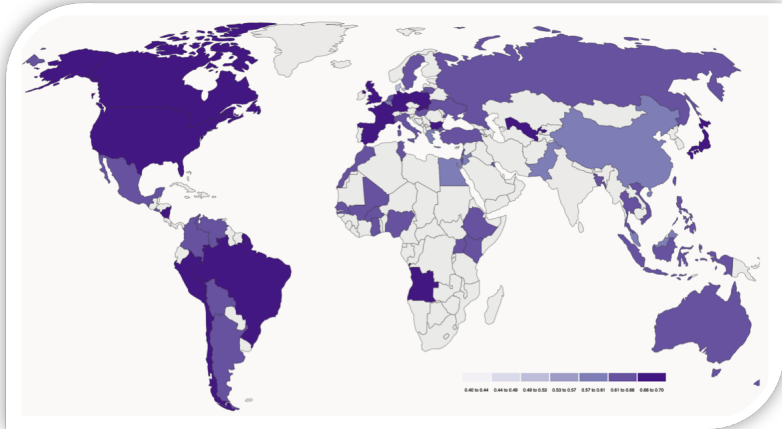
## Social Norms Theory

[Kitts and Chiang, 2008; Perkins and Berkowitz, 1986]

cultural pressure

legality

# Cultural Bias

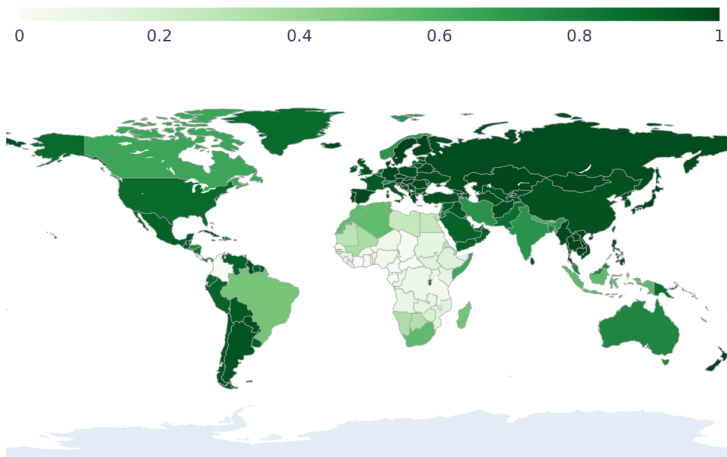


**Figure:** Responses from LMs are more similar to the opinions from certain populations, such as USA, Canada, Australia, and some European countries (Durmus et al., 2024).

# Human Language



# ChatGPT Language Bias



[Wei-Rui, C., Adebara, I., Doan, K. D., Liao, Q., & Abdul-Mageed, M. (Under Review, 2024).]



# What is Culture?



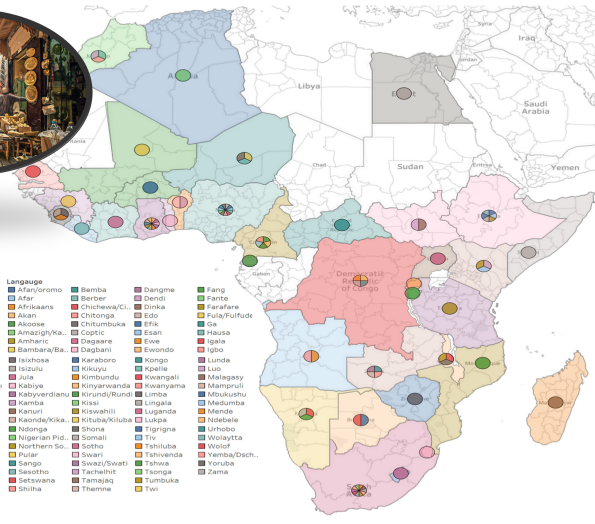
Encompasses the **ways people think, act, and communicate**, including **language, religion, social habits, arts, and traditions**



# Arab World



# African Languages



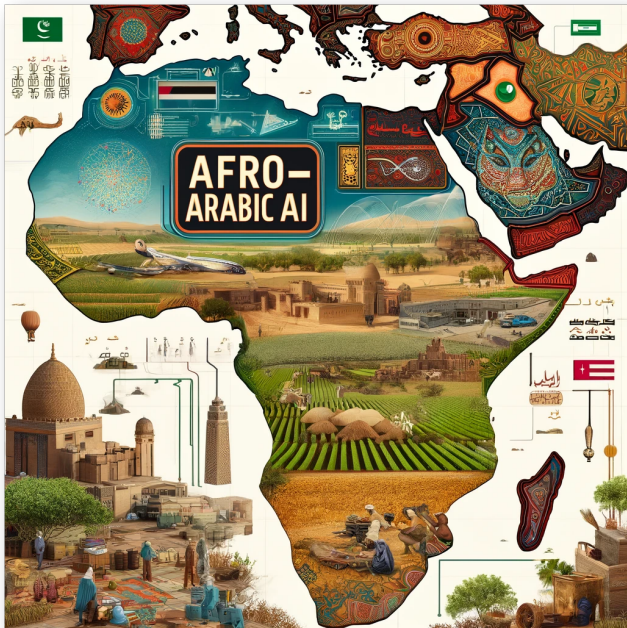
# Africa



# African People









# Partnerships

## Pressing Problems

Misinformation Conflict

Climate Change

### Ensuring Full Literacy



Social Sciences and  
Humanities Research  
Council of Canada

Canada

### InterPARES TrustRI

# Ensuring Full Literacy



[40 Academics (Co-Leads, Co-Applicants, Collaborators) across 13 institutions.  
70 trainees; 24 partner organizations (industry, community, etc.)]

# Massively Multilingual LLMs for Africa



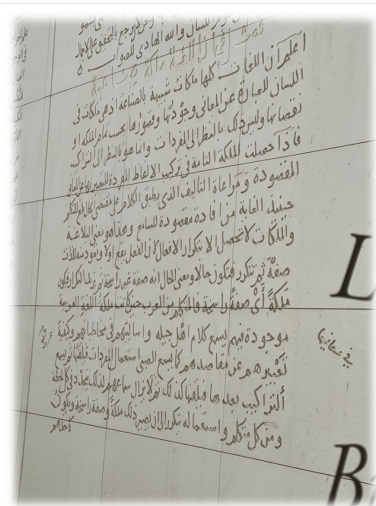
# Language Generation for 517 African Languages





[98 organizational partners in 4 continents: universities, archives, businesses, corporations  
129 co-applicants (academics); 128 collaborators (professionals); 2 postdocs; ~ 60 GRAs/year.]

# Digitization of Cultural Artifacts



# Cultural Heritage



فبينما لم اذ ابتعدت عنه على اعراب الاله الا اني لا ابرك واجتهد  
القران الذي يعجز عن الشاهد المشهور الا اني لم ابرك ولا ابرك ولا  
والجهد والهدى والهدى وما اشتهه ما اشتهه ما اشتهه ما اشتهه  
الهدى ما اشتهه ما اشتهه ما اشتهه ما اشتهه ما اشتهه ما اشتهه  
الهدى ما اشتهه ما اشتهه ما اشتهه ما اشتهه ما اشتهه ما اشتهه  
والهدى ما اشتهه ما اشتهه ما اشتهه ما اشتهه ما اشتهه ما اشتهه  
تتلمذوا بعد ذلك المبعوثين لئلا ياتهم ما اشتهه ما اشتهه ما اشتهه

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ  
الْحَمْدُ لِلَّهِ الَّذِي هَدانا  
لِإِسْمَائِيلَ الَّذِي يُؤْمِنُونَ بِالْعَجَبِ  
وَيُؤْمِنُونَ بِالْآيَاتِ  
وَيُؤْمِنُونَ بِالْآيَاتِ  
وَيُؤْمِنُونَ بِالْآيَاتِ  
وَيُؤْمِنُونَ بِالْآيَاتِ  
وَيُؤْمِنُونَ بِالْآيَاتِ  
وَيُؤْمِنُونَ بِالْآيَاتِ

وَالْحَمْدُ لِلَّهِ الَّذِي هَدانا  
لِإِسْمَائِيلَ الَّذِي يُؤْمِنُونَ بِالْعَجَبِ  
وَيُؤْمِنُونَ بِالْآيَاتِ  
وَيُؤْمِنُونَ بِالْآيَاتِ  
وَيُؤْمِنُونَ بِالْآيَاتِ  
وَيُؤْمِنُونَ بِالْآيَاتِ  
وَيُؤْمِنُونَ بِالْآيَاتِ  
وَيُؤْمِنُونَ بِالْآيَاتِ



United Nations  
Educational, Scientific and  
Cultural Organization



## UNESCO Archives

# Language Translation



حنين بن إسحاق



Μαθηματική Σύνταξις

Κλαύδιος Πτολεμαῖος

τὸν μὲν ὁμαίωρον τῆ διὰ μέσον τῶν ἰσθμίων κύκλον τὸν  $ΑΒΓΔ$  περὶ κέντρον τὸ  $Ε$  καὶ διάμετρον τὴν  $ΑΕΓ$ , τὸν  $δ'$  ἐκ' αὐτοῦ φερόμενον ἐπικύκλιον, ἐφ' οὗ κινεῖται ὁ ἀστέρ, τὸν  $ΖΗΘΚ$  περὶ κέντρον τὸ  $Α$ , φερόμενον καὶ οὕτως αὐτόθεν ἴσται, διδοῖ τοῦ ἐπικύκλιου ὁμαλῶς διαφερόμενον τὸν  $ΑΒΓΔ$  κύκλον ὅσῳ ἀπὸ τοῦ  $Α$  λόγον ἔνεκα ἐπὶ τὸ  $Β$  καὶ τοῦ ἀστέρος τὸν ἐπικύκλιον, δεῦν μὲν κατὰ τὸν  $Ζ$  καὶ  $Θ$  γίνεται ὁ ἀστέρ, ἀδιαφύροσ φανέσται τῆ  $Α$  κέντροσ τοῦ ἐπικύκλιου, δεῦν δὲ κατὰ Ἑλλαν, οὐκίται, ἀλλὰ κατὰ μὲν τοῦ  $Η$  φέρει εἰσὶν γινόμενος πλείονα δὲξῃ πεποθησῖαι κίοναν τῆς ὁμαλῆς τῆς  $ΑΗ$  περιφερείας, κατὰ δὲ τοῦ  $Κ$  εἰσάουσι ὁμοίως τῆς  $ΑΚ$  περιφερείας.



## TURJUMAN: A Public Toolkit for Neural Arabic Machine Translation

El Moatez Billah Nagoudi\* AbdelRahim Elmadany\* Muhammad Abdul-Mageed\*

Deep Learning & Natural Language Processing Group

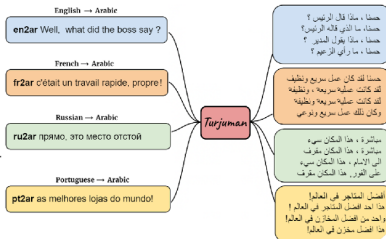
The University of British Columbia

{moatez.nagoudi,a.elmadany,muhammad.mageed}@ubc.ca



**Turjuman** is a neural machine translation toolkit. It translates from 20 languages into Modern Standard Arabic (MSA). Turjuman is described in this paper: "**TURJUMAN: A Public Toolkit for Neural Arabic Machine Translation**".

**Turjuman** exploits our **AraT5 model**. This endows Turjuman with a powerful ability to decode into Arabic. The toolkit offers the possibility of employing a number of diverse decoding methods, making it suited for acquiring paraphrases for the MSA translations as an added value.



# Toucan: Many-to-Many MT for 150 African Pairs





News Article	
	<p>بيرن - رويترز: دعا الفرنسي ميشيل بلاتيني رئيس الاتحاد الاوروبي لكرة القدم لانشاء قوة شرطة دولية متخصصة للتعامل مع أحداث الشغب المرتبطة بالرياضة. وقال بلاتيني في مؤتمر صحفي بمقر الاتحاد الاوروبي لكرة القدم في نيون بسويسرا أن هذا الجهاز قد يساعد على الحد من أعمال الشغب والفساد وفضائح المراهات والمساعدة على منع المشايخ العروفين من حضور المباريات.</p> <p>ونقل موقع الاتحاد الاوروبي على الانترنت عن بلاتيني قوله:</p> <p>« نحن نعيش في عالم عنيف اليوم وللأسف انتقل هذا العنف الى ملاعب رياضتنا المفضلة »</p> <p>واضاف لدنيا فرق شرطة للانترنت والموسيقى على سبيل المثال لكن ليس هناك قوة شرطة واحدة للرياضة التي تحظى باهمية كبرى في العالم.</p> <p>وقال بلاتيني انه اشاد بقرار السلطات الايطالية بتأجيل المباريات عقب مقتل شرطي خلال أعمال شغب في مباراة بدوري الدرجة الاولى الشهر الماضي. وأضاف « لقد كان وقتا يتعين فيه اتخاذ اجراء. لا يمكننا الاستمرار بهذا الشكل. »</p>
<b>Original:</b>	
<b>Prompt:</b>	<p>بيرن - رويترز: دعا الفرنسي ميشيل بلاتيني رئيس الاتحاد الاوروبي لكرة القدم لانشاء قوة شرطة دولية متخصصة للتعامل مع أحداث الشغب المرتبطة بالرياضة. وقال بلاتيني في مؤتمر صحفي في مدينة زوريخ السويسرية أمس الاول : « يجب ان تكون هناك شرطة في كرة القدم . »</p> <p>واضاف بلاتيني ، الذي انتخب رئيسا للاتحاد الاوروبي في العام الماضي، ان مثل هذه القوة يمكن ان تساعد الشرطة المحلية في التعامل مع الاحداث التي تقع أثناء المباريات. وقال بلاتيني : « يمكن ان يكون هناك تعاون بين الشرطة الوطنية والشرطة الدولية (الانتربول) والشرطة الاوروبية (اليوروبول) وشرطة الامم المتحدة »</p> <p>واضاف « سيكون هناك تنسيق كامل بين هذه الاطراف ، مشيرا الى ان الشرطة الوطنية ستكون مسؤولة عن الامن في الملاعب، بينما ستتحمل الشرطة الدولية مسؤولية الامن »</p> <p>وتابع بلاتيني « ان قوة الشرطة التي يمكن انشاؤها ستكون قوية بما فيه الكفاية لمواجهة اي نوع من العنف. »</p> <p>وقال « لا نريد ان نرى احداثا مماثلة لما حدث في كأس العالم الاخيرة في جنوب افريقيا » ، في اشارة الى اعمال العنف التي رافقت المباراة النهائية بين اسبانيا وهولندا في ١١ تموز (يوليو) الماضي في جوهانسبرغ، والتي اسفرت عن مقتل مشجعين اثنين واصابة أكثر من ١٥٠ آخرين بجروح.</p>
<b>Generated:</b>	
<b>Original:</b>	<p>يمكن لقائمة النظام الغذائي للبحر الأبيض المتوسط ، التي تتضمن الكثير من الفواكه والخضروات والقليل من الدهون المشبعة ، أن تمثل بديلا مقيدا وقوي التأثير عن أنواع الحمية الأخرى. ففي دراسة أجريها باحثون تابعون لجامعة هارفارد، تم إخضاع ١٠١ رجلا وإمرأة من البدناء لحمية على الأغذية منخفضة الدسم أو حمية البحر الأبيض المتوسط. وبعد ١٨ شهرا، زاد وزن المتطوعين الذين اتبعوا الحمية قليلة الدهون بمقدار ٦ أرطال، وفقد المتزنون بحمية البحر المتوسط ٩ أرطال من أوزانهم. وعند نهاية الدراسة تبين أن ٢٠ بالمئة فقط من أصحاب الحمية عن الدهون وصلوا صحتهم مقابل نصف أصحاب حمية البحر الأبيض المتوسط الذين واطبوا عليها. كما أثبتت دراسة أخرى أن حمية البحر الأبيض المتوسط أكثر فعالية في التخلص من الوزن الزائد.</p>
<b>Prompt:</b>	<p>يمكن لقائمة النظام الغذائي للبحر الأبيض المتوسط ، التي تتضمن الكثير من الفواكه والخضروات والقليل من الدهون المشبعة ، أن تمثل بديلا مقيدا وقوي التأثير عن أنواع الحمية الأخرى. أظهرت بعض الدراسات أن حمية البحر المتوسط يمكن أن تقلل من خطر الإصابة بسرطان القولون والمستقيم. ومع ذلك ، هناك حاجة إلى مزيد من البحوث لتحديد ما إذا كانت هذه الحمية يمكن أن تساعد في الوقاية من السرطان أم لا. على أي حال ، فإن اتباع نظام غذائي غني بالفواكه والخضروات والحبوب الكاملة والبقوليات والمكسرات واللحوم الخالية من الدهن قد يقلل من مخاطر الإصابة بالسرطان. يمكن لنظام غذائي صحي ومتوازن أن يساعد في الحفاظ على وزن صحي للجسم ، مما قد يساعد أيضا في الحماية من بعض أنواع السرطان ، بما في ذلك سرطان الثدي والقولون والبروستاتا.</p>
<b>Generated:</b>	

## Dialect

ما نخلص شغلنا و نخلص يا هاني بيه و لا ايه؟  
هاني بيه: ايوة يا مدام.

مدام: انا عايزة اتكلم معاك شوية في موضوع مهم جدا جدا و مش عارفة اقولهولك ازاي  
هاني: خير ان شاء الله يا ماما في حاجة ولا ايه؟.

ممدوح: بصي يا ستي انا مش عايزك تزعلي مني عشان انا بحبك و عمري ما حبيت حد قدك

## Tweets

غرد: اللي بيبي يرحل ترى الباب مفتوح.. ماني ب شاري خاطر انسان هالمره!.....

غرد: عدم اللامؤاخزة .. عدم التدخل في شؤونهم!!#حقيقه \_لايمكن \_انكارها

غرد: ماغاديش #كاس \_العالم \_حيوحشنا 🍷🍷

غرد: كيفج يا قلبي!!

# JASMINE, the 'Poet'



Al-Mutanabbi Statue in [Baghdad](#)

## Poetry

**Original:**

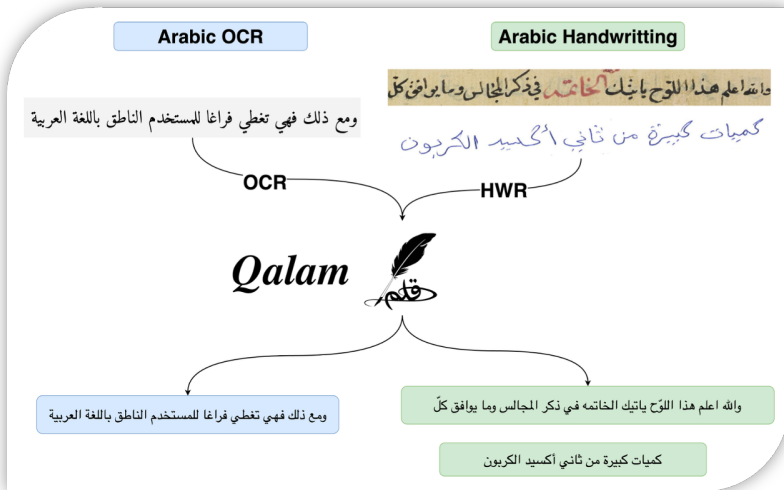
واحر قلباه ممن قلبه شيم \*\* ومن بجسمي وحالي عنده سقم  
مالي اُكتم حبا قد برى جسدي \*\* وتدعي حب سيف الدولة الأمم  
إن كان يجمعنا حب لغرته \*\* فليت أنا بقدر الحب نقسم  
قد زرتة وسيوف الهند مغمدة \*\* وقد نظرت إليه والسيوف دم  
فكان أحسن خلق الله كلهم \*\* وكان أحسن مافي الأحسن الشيم  
فوت العدو الذي يمتته ظفر \*\* في طيه أسف في طيه نعم  
قد ناب عنك شديد الخوف واضطنعت \*\* لك المهابة ما لا تصنع بهم  
ألزمت نفسك شيئا ليس يلزمها \*\* أن لا يواريهم أرض ولا علم  
أكلما رمت جيشا فائتني هربا \*\* تصرفت بك في آثاره الهم

**Prompt:**

واحر قلباه ممن قلبه شيم \*\* ومن بجسمي وحالي عنده سقم  
مالي اُكتم حبا قد برى جسدي \*\* وتدعي حب سيف الدولة الأمم  
إن كان يجمعنا حب لغرته \*\* فليت أنا بقدر الحب نقسم

**Generated:**

لا تحسبوا رقصي بينكم طربا \*\* فالطير يرقص مذبوحا من الألم  
يا لائمي في هواه والهوى قدر \*\* لو لم تذق منه حلوا لم تتبق  
قد صار قلبي قابلا كل صورة \*\* فمرعى لغزلان ودير لرهبان  
وبيت لأوثان وكعبة طائف \*\* وألواح توراة ومصحف قرآن  
أدين بدين الحب أنى توجهت \*\* ركائبه فالحب ديني وإيماني  
من كان يخضب خده بدموعه \*\* فنحورنا بدمائنا تتخضب



# Peacock: A Family of Arabic Multimodal LLMs





ما المواد المستخدمة في صناعة مقبض الجنبية الظاهر في الصورة؟



Question in MSA

يبدو أن مقبض الخنجر مصنوع من الخشب والمعدن. ومع ذلك، بدون مزيد من المعلومات أو السياق، لا يمكن تحديد المادة الدقيقة المستخدمة في صنع المقبض بدقة.



Peacock  
answer  
in MSA

مقبض الجنبية الموجود في الصورة يبدو أنه مصنوع من مادة فاخرة مثل القرن أو الفضة، وقد يكون مُطعمًا بالذهب كما يظهر من الزخارف البراقة.



GPT4

GPT-4  
answer  
in MSA

# Egyptian Dialect



هي الصورة دي فيها ايه؟



Question in  
Egyptian Dialect

الصورة فيها ثلاث أولاد صغيرين قاعدين على سرير، وبيقروا كتاب. في شويه صور تقريبًا في الخلفية، ممكن تكون بتضفي عنصر زخرفي في الصورة.



Peacock  
answers  
in  
Egyptian  
Dialect

الصورة تُظهر ثلاثة أطفال يجلسون على سرير كبير وهم منهمكون في قراءة كتاب كبير مفتوح أمامهم. الكتاب يبدو كأنه يحتوي على صور ومعلومات، ربما عن فيلم أو موضوع محدد يهتمون به. الغرفة مضاءة بضوء النهار الطبيعي.



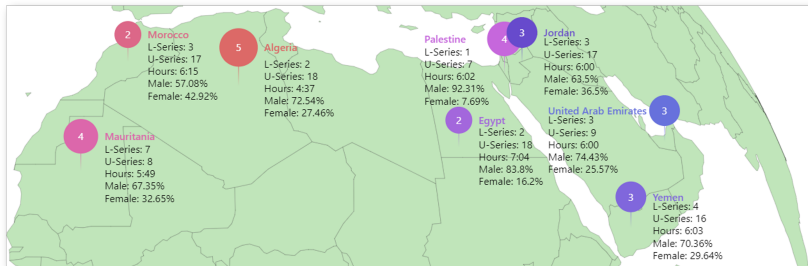
GPT4

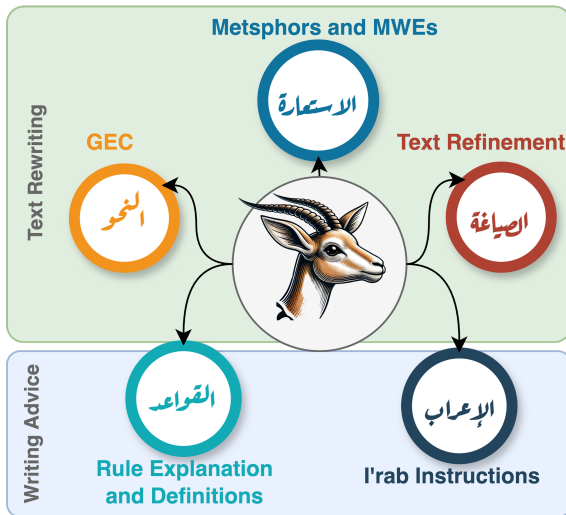
GPT-4  
answers  
in MSA

# Arabic Culture



# Casablanca





# Moroccan Cultural Heritage



# Moroccan Dialect



أهلا وسهلا بيك! الرباط هي عاصمة المغرب  
وكتطل على المحيط الأطلسي. فيها بزاف ديال  
الاماكن السياحية الي خاصك تزورها بحال باب  
الرواح، صومعة حسان، القصر الملكي، الوداية،  
شالة وغيرها. الرباط مدينة هادئة وكتعجب  
الناس الي كيحبو الهدوء والراحة. شنو بالضبط  
بغيتي تعرف على الرباط؟



لا، للأسف ماشي ممكن تزور شالة فالليل حيت  
كتسد فالساعة 6 و نص ديال العشية. كنصحك  
تمشي زوال باش تستافد وتشوف المعلمة كاملة  
بالنهار.

سلام، بغيت نسولك على  
مدينة الرباط



شكرا، واش نقدر نزور شالة فالليل؟

شكرا، واش هادا جامع الفنا؟



# Research Group



Muhammad



Moatez



AbdelRahim



Fakhraddin



Samar



Chiyu



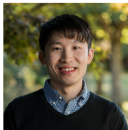
Bashar



Peter



Tawkat



Weirui



Ife



Alcides



Gagan



Sang



Karima



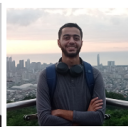
Ahmed



Abdul



Samir



Abdelrahman



Ganesh



Ahsan

# Acknowledgements



Natural Sciences and  
Engineering Research  
Council of Canada

Canada

INNOVATION.CA

CANADA FOUNDATION FOR INNOVATION | FONDATION CANADIENNE POUR L'INNOVATION



Social Sciences and Humanities  
Research Council of Canada  
Conseil de recherches en  
sciences humaines du Canada

Canada

Google



Canada  
Research  
Chairs

Chaires  
de recherche  
du Canada

Canada



Penn



Microsoft  
Research



compute  
canada

