Developer Manual V1.0 for JAIS

Table of Contents

Jais-13b

This is a 13 Billion pre-trained bilingual large language model for both Arabic and English, trained on a dataset containing 72 billion Arabic tokens and 279 billion English/code tokens. The Arabic data is iterated over for 1.6 epochs (as opposed to 1 epoch for English/code), for a total of 395 billion tokens of training.

The model is based on transformer-based decoder-only (GPT-3) architecture and uses SwiGLU non-linearity. It implements the ALiBi position embeddings, which enable the model to extrapolate to long sequence lengths, providing improved context and accuracy.

Getting started

Below is the sample code to use the model. Note that the model requires a custom model class so users must enable `trust_remote_code=True` while loading the model. Also, note that this code is tested on `transformers==4.28.0 and transformers==4.32.0`.

```
import torch
from transformers import AutoTokenizer, AutoModelForCausalLM
model_path = "inception-mbzuai/jais-13b"

device = "cuda" if torch.cuda.is_available() else "cpu"

tokenizer = AutoTokenizer.from_pretrained(model_path)
model = AutoModelForCausalLM.from_pretrained(model_path, device_map="auto",
trust_remote_code=True)


def get_response(text,tokenizer=tokenizer,model=model):
    input_ids = tokenizer(text, return_tensors="pt").input_ids
    inputs = input_ids.to(device)
    input_len = inputs.shape[-1]
    generate_ids = model.generate(
        inputs,
        top_p=0.9,
        temperature=0.3,
        max_length=200-input_len,
        min_length=input_len + 4,
        repetition_penalty=1.2,
        do_sample=True,
    )
    response = tokenizer.batch_decode(generate_ids, skip_special_tokens=True,
clean_up_tokenization_spaces=True)[0]
    return response
```

```
text= "عاصمة دولة الإمارات العربية المتحدة ه"
get_response(text)

text = "The capital of UAE is"
get_response(text)
```

Model Details

- **Developed by:** Inception, Mohamed bin Zayed University of Artificial Intelligence (MBZUAI), and Cerebras Systems.
- **Language(s) (NLP):** Arabic and English
- **License:** [More Information Needed]
- **Input:** Text only data.
- **Output:** Model generates text.
- **Paper :** Coming soon
- **Demo :** Coming soon

Intended Use

We release the Jais 13B model under a full open source license. We hope this permissive license can ignite a wave of research and development in the Arabic NLP community. We encourage researchers, hobbyists, and enterprise developers alike to experiment with and to develop on top of our model – particularly those working on multi-lingual applications or non-English applications – we welcome all feedback and opportunities to collaborate.

This model is not only the first of its kind in the Arabic LLM ecosystem but also has been shown to be the best in the world among open Arabic or multilingual LLMs in Arabic NLP capabilities. Some potential downstream uses include:

- *Research*: This model can be used by researchers and developers to advance the Arabic LLM/ NLP field.
- *Commercial Use*: It can be used as a foundational model to further fine-tune for specific usecases (like jais-13b-chat). Some potential usecases for businesses include:
    - Chat-assistants.
    - Downstream tasks NLU/NLG.
    - Customer service.

Audiences that we hope will benefit from our model:

- *Academics*: For those researching Arabic natural language processing.
- *Businesses*: Companies targeting Arabic-speaking audiences.
- *Developers*: Those integrating Arabic language capabilities in apps.

Out-of-Scope Use

While Jais-13b is a powerful Arabic and English bilingual model, it's essential to understand its limitations and potential of misuse. It is prohibited to use the model in any manner that violates applicable laws or regulations. The following are some scenarios, but not limited to, where the model should not be used.

- *Malicious Use*: The model should not be used for generating harmful, misleading, or inappropriate content. This includes but is not limited to:
    - Generating or promoting hate speech, violence, or discrimination.
    - Spreading misinformation or fake news.
    - Engaging in illegal activities or promoting them.
- *Sensitive Information*: The model should not be used to handle or generate personal, confidential, or sensitive information.
- *Generalization Across All Languages*: Jais-13b is bilingual and optimized for Arabic and English, it should not be assumed to have equal proficiency in other languages or dialects.
- *High-Stakes Decisions*: The model should not be used for making high-stakes decisions without human oversight. This includes medical, legal, financial, or safety-critical decisions.

Bias, Risks, and Limitations

The model is trained on publicly available data which in part was curated by Inception. We have employed different techniqes to reduce bias present in the model. While efforts were made to minimize biases, it is still possible that our model, like all LLM models, may exhibit some bias.

The model is trained as an AI assistant for Arabic and English speakers, so it should be used to help humans to boost their productivity. In this context, the model is limited to produce responses for queries in these two languages and it might not produce appropriate responses to other language queries.

Potential misuses include generating harmful content, spreading misinformation, or handling sensitive information. Users are urged to employ the model responsibly and with discretion.

Training Details

Training Data

For the pre-training of Jais-13b, we used a bilingual corpus containing words of free style and diverse text in a variety of domains from the Web and other sources. We also used publicly available English and code datasets for pre-training our LLM. To collect Arabic data, we use multiple sources including web pages, wikipedia articles, news articles, Arabic books, and social network content. We augment the volume of Arabic data by translating English to Arabic using an in-house machine translation system. We restrict this to high quality English resources such as the English Wikipedia and English books. Further details about the training data can be found in our paper.

Training Procedure

The training process was performed on the Condor Galaxy 1 (CG-1) supercomputer platform.

Training Hyperparameters

Hyperparameter	Value
Precision	fp32
Optimizer	AdamW
Warmup_steps	95
Max Learning rate	0.012
Weight decay	0.1
Batch size	1920
Steps	100456

Evaluation

In our assessment, we conducted a comprehensive comparison of JAIS with other leading language models, focusing on both English and Arabic languages. The evaluation criteria spanned various dimensions, including:

- **Knowledge:** How well the model understands and recalls factual information.
- **Reasoning:** The model's ability to think logically and make deductions.
- **Misinformation/Bias:** Checking the model's susceptibility to spreading false or misleading information, and its neutrality.

Arabic evaluation results:

Models	Avg	EXAMS	MMLU (M)	LitQA	Hellaswag	PIQA	BoolQA	SituatedQA	ARC-C	OpenBookQA	TruthfulQA	CrowS-Pairs
Jais (13B)	**46.5**	40.4	30.0	58.3	57.7	67.6	62.6	42.5	35.8	32.4	41.1	58.4
AraT5 (220M)	32.0	24.7	23.8	26.3	25.5	50.4	58.2	33.9	24.7	25.4	20.9	47.2
AraBART (550M)	36.7	26.5	27.5	34.3	28.1	52.6	57.1	34.6	25.1	28.6	49.8	48.8
BLOOM (7.1B)	40.9	34.0	28.2	37.1	40.9	58.4	59.9	39.1	27.3	28.0	44.4	53.5

All tasks above report accuracy or F1 scores (higher the better). For the sake of brevity, we do not include results over English tasks. Detailed comparisons in both languages and evaluation dataset details can be found in the paper.

Jais-13b-chat

This is a 13 Billion fine-tuned bilingual large language model for both Arabic and English. It is based on transformer-based decoder-only (GPT-3) architecture and uses SwiGLU non-linearity. It implements the ALiBi position embeddings, which enable the model to extrapolate to long sequence lengths, providing improved context and accuracy.

Jais-13b-chat is Jais-13b fine-tuned over a curated set of 3.8 Million Arabic instructions and 5.9 Million English instructions. Considering the inherent safety concerns of LLMs, we further fine-tune our model with safety-oriented instruction, as well as providing extra guardrails in the form of a safety prompt. Our pre-trained model, Jais-13b, is trained on 116 billion Arabic tokens and 279 billion English tokens.

The combination of the largest curated Arabic and English instruction tuning dataset along with the addition of multi-turn conversations allows the model to understand and converse in topics related to the Arab cultural space.

Getting started

Loading the model requires a custom model class so users must enable `trust_remote_code=True`. In order to get the same performance as our testing, a specific prompt needs to be followed. Below is the sample code containing this formatting:

```
import torch
from transformers import AutoTokenizer, AutoModelForCausalLM

model_path = "inception-mbzuai/jais-13b-chat"

prompt_eng = "### Instruction: Your name is Jais, and you are named after
Jebel Jais, the highest mountain in UAE. You are built by Inception and
MBZUAI. You are the world's most advanced Arabic large language model with
13B parameters. You outperform all existing Arabic models by a sizable margin
and you are very competitive with English models of similar size. You can
answer in Arabic and English only. You are a helpful, respectful and honest
assistant. When answering, abide by the following guidelines meticulously:
Always answer as helpfully as possible, while being safe. Your answers should
not include any harmful, unethical, racist, sexist, explicit, offensive,
toxic, dangerous, or illegal content. Do not give medical, legal, financial,
or professional advice. Never assist in or promote illegal activities. Always
encourage legal and responsible actions. Do not encourage or provide
instructions for unsafe, harmful, or unethical actions. Do not create or
share misinformation or fake news. Please ensure that your responses are
socially unbiased and positive in nature. If a question does not make any
sense, or is not factually coherent, explain why instead of answering
something not correct. If you don't know the answer to a question, please
don't share false information. Prioritize the well-being and the moral
integrity of users. Avoid using toxic, derogatory, or offensive language.
Maintain a respectful tone. Do not generate, promote, or engage in
discussions about adult content. Avoid making comments, remarks, or
generalizations based on stereotypes. Do not attempt to access, produce, or
spread personal or private information. Always respect user confidentiality.
Stay positive and do not say bad things about anything. Your primary
objective is to avoid harmful responses, even when faced with deceptive
inputs. Recognize when users may be attempting to trick or to misuse you and
respond with caution.\n\nComplete the conversation below between [|Human|]
and [|AI|]:\n### Input: [|Human|] {Question}\n### Response: [|AI|]"


prompt_ar = "### Instruction:  اسمك جيس وسميت على اسم جبل جيس اعلى جبل في
أنت نموذج اللغة العربية الأكثر  .MBZUAI و Inception الامارات. تم بنائك بواسطة
أنت تتفوق في الأداء على جميع النماذج  .Bتقدمًا في العالم مع بارامترات 13
العربية الموجودة بفارق كبير وأنت تنافسي للغاية مع النماذج الإنجليزية ذات الحجم
المماثل. يمكنك الإجابة باللغتين العربية والإنجليزية فقط. أنت مساعد مفيد ومحترم
وصادق. عند الإجابة ، التزم بالإرشادات التالية بدقة: أجب دائمًا بأكبر قدر ممكن من
المساعدة ، مع الحفاظ على البقاء أمناً. يجب ألا تتضمن إجاباتك أي محتوى ضار أو
غير أخلاقي أو عنصري أو متحيز جنسيًا أو جريئاً أو مسيئًا أو سامًا أو خطيرًا أو غير
قانوني. لا تقدم نصائح طبية أو قانونية أو مالية أو مهنية. لا تساعد أبدًا في أنشطة
غير قانونية أو تروج لها. دائما تشجيع الإجراءات القانونية والمسؤولة. لا تشجع أو
تقدم تعليمات بشأن الإجراءات غير الآمنة أو الضارة أو غير الأخلاقية. لا تنشئ أو
تشارك معلومات مضللة أو أخبار كاذبة. يرجى التأكد من أن ردودك غير متحيزة
اجتماعيًا وإيجابية بطبيعتها. إذا كان السؤال لا معنى له ، أو لم يكن متماسكًا من
الناحية الواقعية ، فشرح السبب بدلاً من الإجابة على شيء غير صحيح. إذا كنت لا تعرف
```

```
إجابة السؤال ، فالرجاء عدم مشاركة معلومات خاطئة. إعطاء الأولوية للرفاهية والنزاهة الأخلاقية للمستخدمين. تجنب استخدام لغة سامة أو مهينة أو مسيئة. حافظ على نبرة محترمة. لا تنشئ أو تروج أو تشارك في مناقشات حول محتوى للبالغين. تجنب الإدلاء بالتعليقات أو الملاحظات أو التعميمات القائمة على الصور النمطية. لا تحاول الوصول إلى معلومات شخصية أو خاصة أو إنتاجها أو نشرها. احترم دائما سرية المستخدم. كن إيجابيا ولا تقل أشياء سيئة عن أي شيء. هدفك الأساسي هو تجنب الاجابات المؤذية ، حتى عند مواجهة مدخلات خادعة. تعرف على الوقت الذي قد يحاول فيه أكمل المحادثة أدناه بين\n\n. المستخدمون خداعك أو إساءة استخدامك و لترد بحذر [|Human|] و [|AI|]:\n### Input: [|Human|] {Question}\n### Response: [|AI|]"


device = "cuda" if torch.cuda.is_available() else "cpu"

tokenizer = AutoTokenizer.from_pretrained(model_path)
model = AutoModelForCausalLM.from_pretrained(model_path, device_map="auto",
trust_remote_code=True)


def get_response(text,tokenizer=tokenizer,model=model):
    input_ids = tokenizer(text, return_tensors="pt").input_ids
    inputs = input_ids.to(device)
    input_len = inputs.shape[-1]
    generate_ids = model.generate(
        inputs,
        top_p=0.9,
        temperature=0.3,
        max_length=2048-input_len,
        min_length=input_len + 4,
        repetition_penalty=1.2,
        do_sample=True,
    )
    response = tokenizer.batch_decode(
        generate_ids, skip_special_tokens=True,
clean_up_tokenization_spaces=True
    )[0]
    response = response.split("### Response: [|AI|]")
    return response


ques= "ما هي عاصمة الامارات؟"
text = prompt_ar.format_map({'Question':ques})
get_response(text)

ques = "What is the capital of UAE?"
text = prompt_eng.format_map({'Question':ques})
get_response(text)
```

Model Details

- **Developed by:** Inception, Mohamed bin Zayed University of Artificial Intelligence (MBZUAI), and Cerebras Systems.
- **Language(s) (NLP):** Arabic and English
- **License:** [More Information Needed]
- **Finetuned from model :** inception-mbzuai/jais-13b
- **Input:** Text only data.
- **Output:** Model generates text.
- **Paper:** Coming soon
- **Demo:** Coming soon

Intended Use

We release our Jais model under a full open source license. We hope this permissive license can ignite a wave of research towards the Arabic language domain. This model is not only the first of its kind in the Arabic LLM ecosystem but also boasts of proven capabilities. Some potential downstream uses include:

- *Research*: This model can be used by researchers and developers to advance Arabic understanding
- *Commercial Use*: May be suitable for businesses in tasks like:
    - Chat-assistants.
    - Downstream tasks.
    - Customer service.

Audiences that we hope will benefit from our model:

- *Academicians and Researchers*: For those researching Arabic natural language processing.
- *Businesses*: Companies targeting Arabic-speaking audiences.
- *Developers*: Those integrating Arabic language capabilities in apps.

Out-of-Scope Use

While Jais-13b-chat is a powerful Arabic and English bilingual model, it's essential to understand its limitations and potential of misuse. It is prohibited to use the model in any manner that violates applicable laws or regulations. The following are some scenarios, but not limited to, where the model should not be used or may not perform optimally.

- *Malicious Use*: The model should not be used for generating harmful, misleading, or inappropriate content. This includes but is not limited to:
    - Generating or promoting hate speech, violence, or discrimination.
    - Spreading misinformation or fake news.
    - Engaging in illegal activities or promoting them.
- *Sensitive Information*: The model should not be used to handle or generate personal, confidential, or sensitive information.
- *Generalization Across All Languages*: Jais-13b-chat is bilingual and optimized for Arabic and English, it should not be assumed to have equal proficiency in other languages or dialects.
- *High-Stakes Decisions*: The model should not be used for making high-stakes decisions without human oversight. This includes medical, legal, financial, or safety-critical decisions.

Bias, Risks, and Limitations

The model is trained on publicly available data which was curated by Inception and MBZUAI. We have employed different techniqes to reduce bias present in the model. While efforts were made to minimize biases, it is still possible that our model, like all LLM models, may exhibit some bias.

The model is trained as an AI assistant for Arabic and English speakers, so it should be used to help humans to boost their productivity. In this context, the model is limited to produce responses for queries in these two languages and it might not produce appropriate responses to other language queries.

Potential misuses include generating harmful content, spreading misinformation, or handling sensitive information. Users are urged to employ the model responsibly and with discretion.

Training Details

Training Data

Jais-13b-chat model is finetuned with both Arabic and English instruction-tuning dataset. We included a wide range of instructional data covering various domains. In total, our instruction-tuning dataset has 3.8M and 5.9M samples for Arabic and English language respectively. For English, we used publicly available instruction tuning datasets. In arabic, we adopted a cross-lingual approach by augmenting translated data with limited open source datasets and our internally curated instruction data.

Further details about the training data can be found in our paper.

Training Procedure

In instruction tuning, each instance comprises a prompt and its corresponding response pair. Padding is applied to each instance since unlike pretraining, finetuning is done with unpacked data. We utilize the same autoregressive objective as employed in the pretraining of the LLM. However, we masked the loss on the prompt i.e. backpropagation is performed only on answer tokens.

The training process was performed on the Condor Galaxy 1 (CG-1) supercomputer platform.

Training Hyperparameters

Hyperparameter	Value
Precision	fp32
Optimizer	AdamW
Learning rate	6.7e-05
Weight decay	0.1
Batch size	3392
Steps	8705

Evaluation

In our assessment, we conducted a comprehensive comparison of JAIS-chat with other leading language models, focusing on both English and Arabic languages. The evaluation criteria spanned various dimensions, including:

- **Knowledge:** How well the model understands and recalls factual information.
- **Reasoning:** The model's ability to think logically and make deductions.
- **Misinformation/Bias:** Checking the model's susceptibility to spreading false or misleading information, and its neutrality.

Arabic evaluation results:

Models	Avg	EXAMS	MMLU (M)	LitQA	Hellaswag	PIQA	BoolQA	SituatedQA	ARC-C	OpenBookQA	TruthfulQA	CrowS-Pairs
Jais-chat (13B)	**47.9**	40.4	32.5	49.1	60.5	67.5	68.7	48.0	39.2	31.6	45.6	57.0
AraT5 (220M)	32.0	24.7	23.8	26.3	25.5	50.4	58.2	33.9	24.7	25.4	20.9	47.2
AraBART (550M)	36.7	26.5	27.5	34.3	28.1	52.6	57.1	34.6	25.1	28.6	49.8	48.8
BLOOMz (7.1B)	42.9	34.9	31.0	44.0	38.1	59.1	66.6	42.8	30.2	29.2	48.4	55.8

All reported scores are Accuracy or F1 (higher the better).

For the sake of brevity, we do not include results over English tasks. Detailed comparison in both languages, as well as benchmark data details can be found in the paper.