



From Articles to Abstracts: Arabic Summarization with TinyLlama

Batoul Hamdan

Objective

01

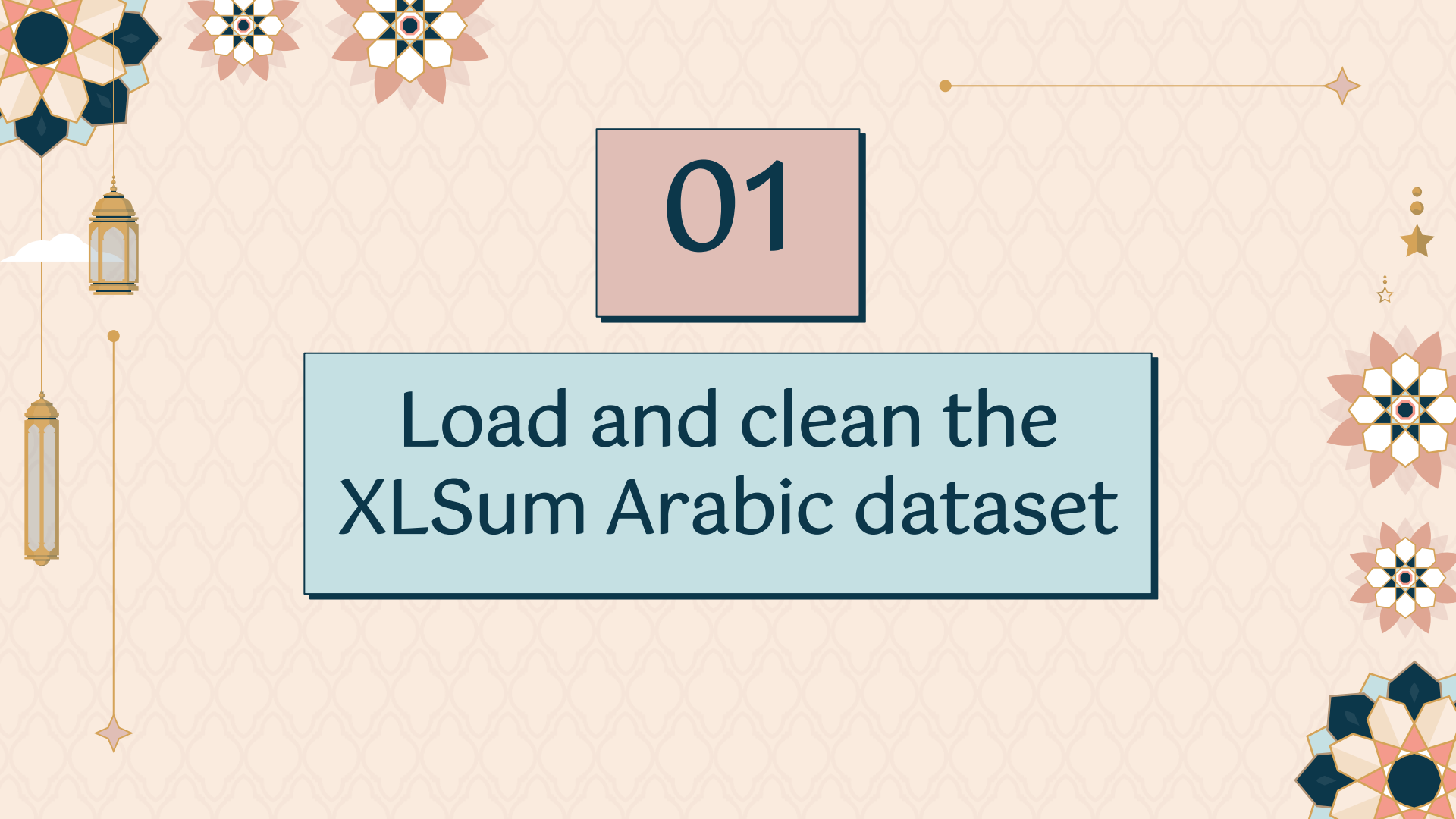
Load and clean
the XLSum
Arabic dataset

02

Load the
TinyLlama-1.1B-
Chat model

03

Fine-tune
TinyLlama
using LoRA for
summarization



01

Load and clean the
XLSum Arabic dataset

XLSum – Arabic Subset



- Created by BBC News
- Covers a wide range of topics: politics, science, culture, etc.
- High-quality, human-written summaries
- ~60,000 Arabic article-summary pairs

XLSum - Example



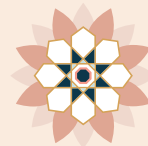
Article

الحكومة أعلنت عن خطة جديدة لتحسين
التعليم في المناطق الريفية...



Summary

الحكومة تطلق خطة لتطوير التعليم في
المناطق الريفية.



Cleaning Dataset

Removing unnecessary columns and
keeping important only

| summary | text |
|---------|------|
| | |

Splitting Dataset

XLSum Dataset

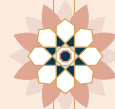
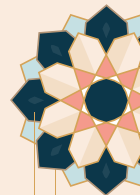
Splitting it into 2


Train

30%

Validation

70%





02

Model & Setup

Base Model



TinyLlama

TinyLlama-1.1B-Chat

Lightweight

~1.1B parameters

Pretrained on diverse
multilingual data

Fine-Tuning Adapter

- LoRA (Low-Rank Adaptation)
- Efficient adaptation with minimal memory usage
- Only updates small trainable matrices

Quantinization

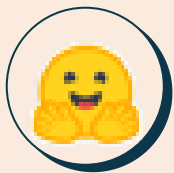


4-bit (nf4) for lower
memory usage &
faster training



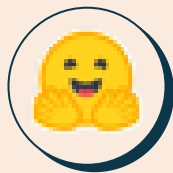
Enabled via
bitsandbytes

Libraries Used



Transformers

model & tokenizer



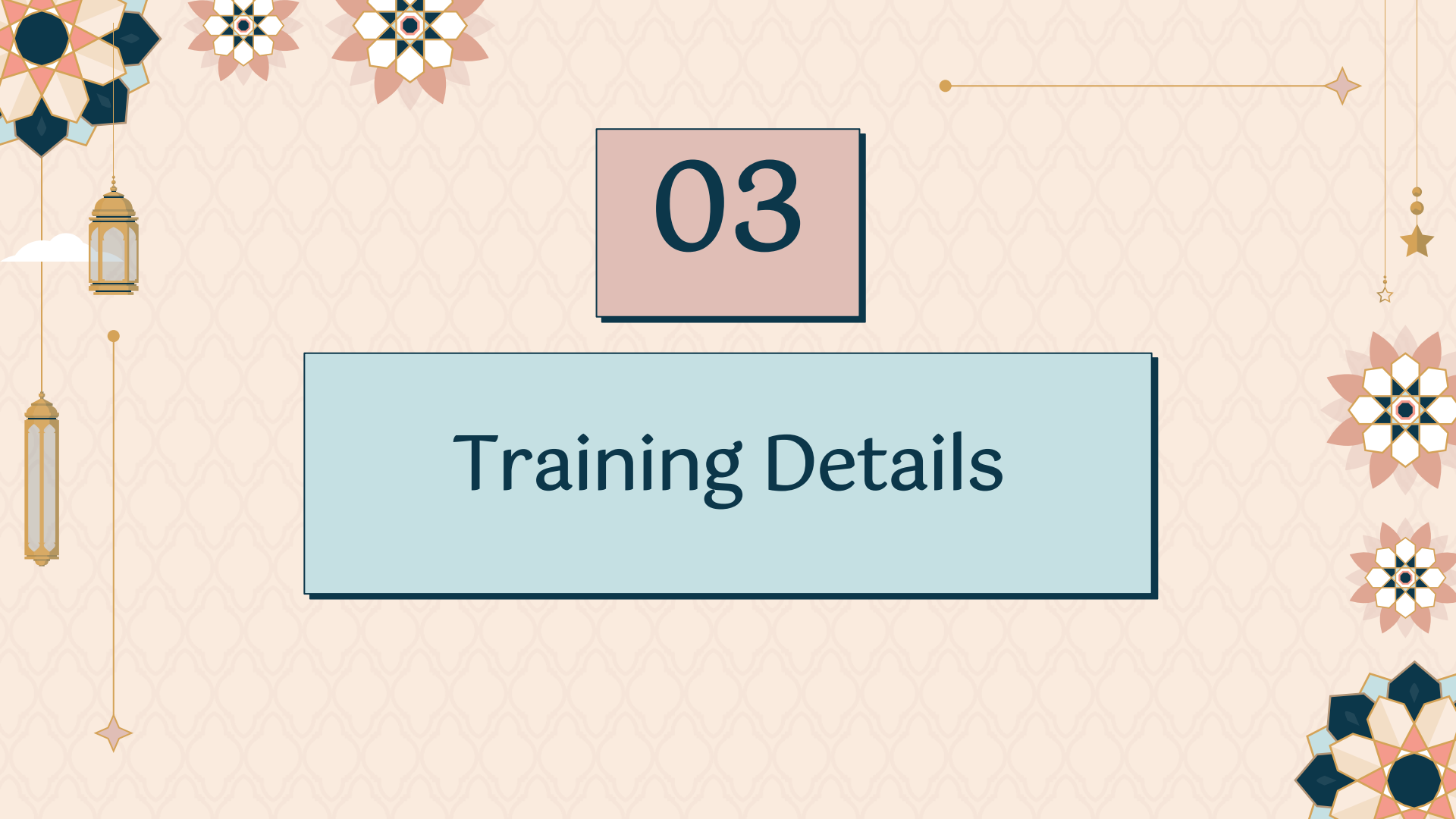
PEFT

apply LoRA adapters



TRL

Training utilities
(SFTTrainer)



03

Training Details

Training Settings

- Batch size: 8
- Epochs: 5
- Learning rate: $5e-5$
- Gradient accumulation: 2 steps
- Max sequence length: 128 tokens

Fine-Tuning Method

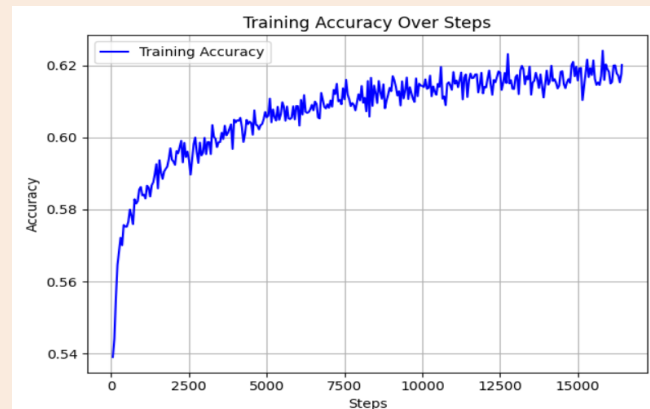
- SFTTrainer from TRL
- Supervised fine-tuning (SFT) using LoRA adapters
- Mixed precision (fp16 or bf16) for faster training

Evaluation



1.33

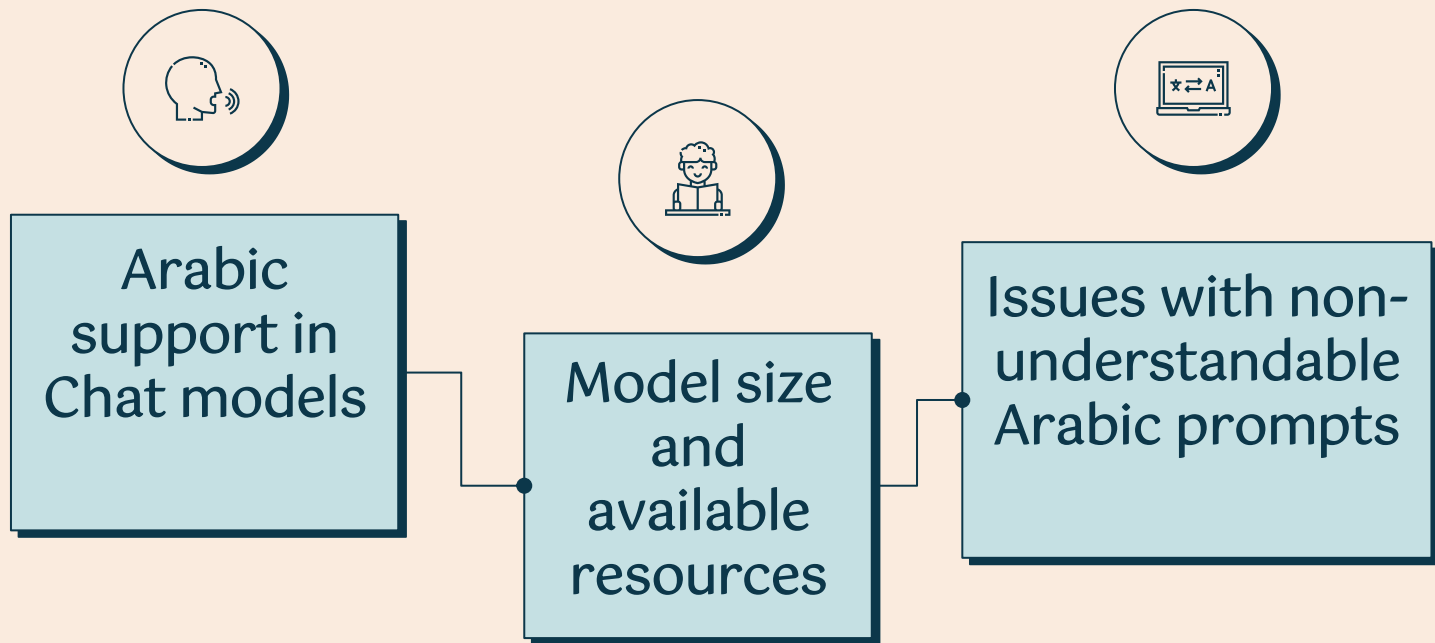
Train Loss



62%

Mean Accuracy

Challenges



Demo 1

Summarize this article:

... التكنولوجيا الحديثة أصبحت جزءاً لا يتجزأ من حياتنا اليومية

أعلنت منظمة التكنولوجيا الحديثة أنها تعتقد أنها تعمل في مجال التكنولوجيا الحديثة.
الحديثة والتكنولوجيا الحديثة أصبحت جزءاً لا يتجزأ من حياتنا اليومية.

Demo 2

Summarize this article:

شهدت السنوات الأخيرة تطورًا هائلًا في مجال التكنولوجيا، حيث أصبحت الأجهزة الذكية مثل الهواتف المحمولة والحواسيب اللوحية جزءًا لا يتجزأ من حياتنا اليومية. لم يعد استخدام الإنترنت مقتصرًا على الترفيه أو التواصل الاجتماعي فقط، بل أصبح وسيلة رئيسية للعمل والتعليم والتجارة. ساهمت تطبيقات الذكاء الاصطناعي وتعلم الآلة في تحسين الخدمات الصحية من خلال تحليل البيانات الطبية والتشخيص المبكر للأمراض.

وعلى الصعيد الاقتصادي، أدت التكنولوجيا إلى نشوء مجالات عمل جديدة مثل التسويق الرقمي، وتطوير التطبيقات، والعمل عن بُعد، مما أتاح فرصًا وظيفية واسعة للأفراد حول العالم. كما ساهمت تقنيات الواقع الافتراضي والواقع المعزز في تدوير طريقة تفاعلنا مع البيئة الرقمية.

في المقابل، أثارت هذه التغيرات السريعة تحديات عدة تتعلق بالخصوصية الرقمية وأمن المعلومات، إذ أصبحت الهجمات السيبرانية أكثر تعقيدًا وخطورة. لذلك، بات من الضروري تعزيز الوعي المجتمعي بأهمية حماية البيانات الشخصية، وتحدثت السياسات والقوانين التي تنظم استخدام التكنولوجيا.

وفي الختام، لا شك أن التكنولوجيا الحديثة قدمت فوائد عظيمة للبشرية، لكن الاستفادة منها تتطلب توازنًا بين التطور التقني والحفاظ على القيم الإنسانية.

Summary: أثارت هذه التغيرات السريعة تحديات عدة تتعلق بالخصوصية الرقمية وأمن المعلومات، إذ أصبحت الهجمات السيبرانية أكثر تعقيدًا وخطورة.

Demo 3

Summarize this article:

شهد العالم خلال العقود الأخيرة تحولاً جذرياً في طريقة عيش الأفراد وعملهم وتواصلهم، بفضل التقدم السريع في مجال التكنولوجيا الرقمية. فقد أصبحت الأجهزة الذكية والتقنيات المتطورة مثل الإنترنت عالي السرعة، والذكاء الاصطناعي، والحواسيب السحابية، جزءاً لا يتجزأ من الحياة اليومية لملايين البشر حول العالم.

في مجال التعليم، ساهمت التكنولوجيا في إحداث نقلة نوعية، حيث ظهرت منصات التعلم عن بُعد التي مكّنت الطلاب من متابعة دراستهم من أي مكان في العالم. كما أصبح من الممكن تقديم محاضرات افتراضية، وتقسيم مناهج تعليمية متقدمة على الفيديوها والمحاكاة والألعاب التعليمية، مما عزز من فاعلية العملية التعليمية ونقل الفجوة بين المتعلمين.

أما في قطاع الصحة، فقد مكّن الذكاء الاصطناعي الأطباء من تشخيص الأمراض بدقة وسرعة من خلال تحليل الصور الطبية والبيانات السريرية. وأصبحت الروبوتات تُستخدم في بعض العمليات الجراحية الدقيقة، بينما ساعدت تطبيقات الهواتف في متابعة الحالة الصحية للمرضى وتنبيههم بشأن الأدوية أو مواعيد الفحوصات.

من الناحية الاقتصادية، أدت هذه الثورة الرقمية إلى خلق فرص جديدة للأعمال، مثل التجارة الإلكترونية وتطوير التطبيقات والعمل الحر عبر الإنترنت، وساعدت الحوسبة السحابية الشركات على تخزين البيانات وتشغيل الأنظمة عن بُعد، مما قلل من التكاليف التشغيلية وسهّل التوسع في الأسواق العالمية.

في المقابل، أثار هذا التطور السريع مخاوف متزايدة بشأن أمن المعلومات والخصوصية الرقمية. إذ أصبحت الهجمات السيبرانية أكثر تعقيداً وتهدد للبيانات الحساسة الحيوية، مما استدعى تعزيز وسائل الحماية وتحديث التشريعات القانونية المرتبطة بالفضاء الرقمي. كما زادت الحاجة إلى توعية الأفراد حول كيفية حماية بياناتهم الشخصية وتبني التهديدات الإلكترونية ولا يمكن إغفال التأثير الاجتماعي لهذا التحول، حيث تغيرت طبيعة العلاقات الاجتماعية وأصبح التواصل يعتمد بشكل كبير على الوسائط الرقمية، مما أثار نقاشات حول تأثير التكنولوجيا على العلاقات الإنسانية والتفاعل الاجتماعي.

وفي الختام، فإن التكنولوجيا الحديثة رغم ما تقدمه من فوائد كبيرة، تفرض تحديات معقدة تتطلب وعياً مجتمعياً وتحولاً دوائياً لضمان استخدامها بما يخدم الإنسان ويحافظ على القيم الأساسية للمجتمعات.

Summary: شهد العالم خلال العقود الأخيرة تحولاً جذرياً في طريقة عيش الأفراد:

Conclusion

- TinyLlama can be efficiently adapted for the Arabic language
- Acceptable performance with LoRA and sufficient data
- Potential for expansion into other NLP tasks



Thanks!

Do you have any questions?

Batoul.hamdan19@gmail.com

