

INTERNATIONAL PRESS RELEASE

## InstaDeep and iCompass Announce TunBERT - The First AI-based Tunisian Dialect System



TUNIS, TUNISIA, MARCH 15.03.2021: InstaDeep and iCompass today proudly announced a collaboration on a Natural Language Processing (NLP) model for underrepresented languages aimed at applying the latest advances in AI and Machine Learning (ML) to explore and strengthen research in the fast emerging Tunisian AI tech ecosystem.

The NLP project consists of developing a language model for Tunisian dialect, TunBERT, and evaluating it on several tasks such as sentiment analysis, dialect classification, reading comprehension, and question-answering. *“We’re excited to reveal TunBERT, a joint research project between iCompass and InstaDeep that redefines state-of-the-art for the Tunisian dialect. This work also highlights the positive results that are achieved when leading AI startups collaborate, benefiting the Tunisian tech ecosystem as a whole”*, said Karim Beguir, CEO and Co-Founder of InstaDeep.

### Empower underrepresented languages

Bidirectional Encoder Representations from Transformers (BERT) has become a state-of-the-art model for language understanding. With its success, available models have been trained on Indo-European languages such as English, French, German etc., but similar research for underrepresented languages remains sparse and in its early stage. Along with jointly writing and



debugging the code, iCompass and InstaDeep's research engineers have launched multiple successful experiments. *“This fruitful collaboration aims to push forward and advance the development of AI research in the emerging and prominent field of NLP and language models. Our ultimate goal is to empower Tunisian talent and foster an environment where AI innovation can grow, and together our teams are pushing boundaries”* said Dr. Hatem Haddad, CTO and Co-Founder of iCompass.

## NeMo toolkit

TunBERT is developed based on NVIDIA's NeMo toolkit, which the research team used to adapt and fine-tuned the neural network on relevant data to pre-train the language model on a large-scale Tunisian corpus, taking advantage of the BERT model that was optimised by NVIDIA. TunBERT's pretraining and fine-tuning steps converged faster and in a distributed and optimised way thanks to the use of multiple NVIDIA V100 GPUs. This implementation provided more efficient training using Tensor Core mixed precision capabilities and the NeMo Toolkit. Through this approach, the contextualized text representation models learned an effective embedding of the natural language, making it machine-understandable and achieving tremendous performance results. Comparing the NVIDIA-optimised BERT model results to the original BERT implementation shows that the NVIDIA-optimised BERT-model performs better on the different downstream tasks, while using the same compute power.

## NVIDIA GTC

A member of NVIDIA Inception, - an acceleration program designed to nurture AI startups, InstaDeep has been accepted to present this research at the upcoming NVIDIA GPU Technology Conference (GTC) in April, in a talk titled “Building a Pre-Trained Contextualized Text Representation Model for Underrepresented Languages: Tunisian Dialect Use Case”. The session will be jointly presented by Nourchene Ferchichi of InstaDeep and Dr. Hatem Haddad of iCompass. GTC is a free event and will take place online between 12-16 April, 2021. [Register here to attend](#).

- END -

For more information contact:

InstaDeep: [communications@instadeep.com](mailto:communications@instadeep.com) or call +44(0)20 3890 7528.

iCompass: [team@icompass.digital](mailto:team@icompass.digital) or call +216 54 911 797



#### ABOUT INSTADEEP

Founded in 2014, InstaDeep is today an EMEA leader in decision-making AI products for the Enterprise, with headquarters in London, and offices in Paris, Tunis, Lagos, Dubai and Cape Town. With expertise in both machine intelligence research and concrete business deployments, the Company provides a competitive advantage to its partners in an AI-first world. Leveraging its extensive know-how in GPU-accelerated computing, deep learning and reinforcement learning, InstaDeep has built products, such as its novel DeepChain™ platform, that tackle the most complex challenges across a range of industries. InstaDeep has also developed collaborations with global leaders in the Artificial intelligence ecosystem, such as Google DeepMind, Nvidia and Intel. The Company is part of Intel's AI Builders program and was named an Elite Deep Learning Partner by NVIDIA.

#### ABOUT iCOMPASS

iCompass is a Tunisian AI startup founded in 2019, specialized in NLP products. iCompass uses the latest Deep Learning and Reinforcement Learning technologies to develop Digital Reputation Analysis, Chatbot and Voicebot services. iCompass is also a leader in the MEA region thanks to its innovative NLP R&D work. Linguistic barriers are no longer an obstacle and iCompass newly developed programs make it possible to process even Arabic and African dialects.