جامعة
حمد بن خليفة
HAMAD BIN KHALIFA
UNIVERSITY
عضو في مؤسسة قطر
Member of Qatar Foundation

Shaheen: An AI-based Arabic-English Machine Translation System

INCLUDING ARABIC DIALECTS IN TRANSLATION PROGRAMS

Current machine translation technologies increase access to the over 7,000 languages published on the Internet by attempting to fit as many languages as possible into one system.
This one-size-fits-all approach, however, fails to account for the lexical nuances and dialectal variations of Arabic. It is also difficult to customize conventional machine translation systems to user-specific demands, such as adaptions to particular language genres or dialects, decreasing the quality and adaptability of the translation.

USING DEEP LEARNING FOR MORE ROBUST TRANSLATION

Researchers at HBKU have developed Shaheen, a new machine translation system based on artificial intelligence that understands the nuances of Arabic dialects.

Shaheen uses components such as a state-of-theart Arabic segmenter called Farasa, as well as a specialized module for out-of-vocabulary words that uses a combination of subword translation and transliteration.

In an extensive evaluation, it outperformed all major translation providers in translating Arabic dialects.

Moreover, the technology's novel adaptation technique enables rapid customization to a domain of the user's interest such as medical, media, and travel with high quality translation.

APPLICATIONS

HBKU's novel machine translation software could improve translation in a variety of areas for both written and spoken language:

> News outlets and publications

> Social media

> Travel and tourism

> Medicine and health care
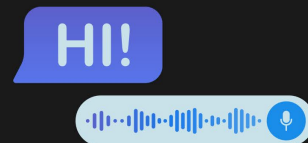
VALUE PROPOSITIONS

HBKU's system enhances Arabic-to-English translation by addressing non-standard linguistic variations, improving accessibility and usability compared to existing technologies.

Effective: Accounts for the complexities of Arabic language, including complex morphology, lexical nuances, normalization, and dialectal variations

Accurate: The system demonstrates significant improvements over existing Arabic-to-English translation technologies, especially in handling dialects and specialized genres.

Flexible: Provides flexibility to the user to integrate specific preferences based on customer data

Intelligent: Includes a means of training and benchmarking the system's capabilities

PATENT STATUS

Patent application WO2020197421A1 published, copyrights exist for this technology.

LICENSING OPPORTUNITIES

Hamad Bin Khalifa University is offering this technology for license.
For more information, please contact: innovation@hbku.edu.qa