

# Zaid Alyafeai

---

Zaid Alyafeai

Dhahran, Saudi Arabia

Scholar: [Zaid Alyafeai](#)

GitHub: <https://github.com/zaidalyafeai>

(+966) 506069897

[alyafey22@gmail.com](mailto:alyafey22@gmail.com)

---

## Education

- 
- **Ph.D.** at KFUPM - (Saudi Arabia) 2019 - now (3.88/4.00)
  - **M.Sc.** at KFUPM - (Saudi Arabia) 2016 - 2019 (3.81/4.00)
  - **B.Sc.** at KFUPM - (Saudi Arabia) 2011-2016 (3.84/4.00)

---

## Selected Publications

---

These are selected publications, for full list check my [google scholar page](#).

- Sanh, V., Webson, A., Raffel, C., Bach, S.H., Sutawika, L., **Alyafeai, Z.**, et al. 2022. Multitask Prompted Training Enables Zero-Shot Task Generalization. *International Conference on Learning Representations (ICLR)*.
- **Alyafeai Zaid, et al.** 2022. Masader: Metadata Sourcing for Arabic Text and Speech Data Resources. *In Proceedings of the Thirteenth Language Resources and Evaluation Conference (LREC)*.
- **Alyafeai, Zaid**, et al. 2022. "Calliar: An Online Handwritten Dataset for Arabic Calligraphy." *Neural Comput & Applic.*
- SJ Mielke, **Z Alyafeai**, et al. 2021. Between words and characters: A brief history of open-vocabulary modeling and tokenization in NLP. *arXiv preprint arXiv:2112.10508*.
- **Alyafeai, Zaid**, et al. 2021. "Arabic Compact Language Modelling for Resource Limited Devices". *Proceedings of the Sixth Arabic Natural Language Processing Workshop (WANLP)* . Located with *EACL*.
- **Alyafeai, Zaid**, et al.. 2020. "ARBML: Democratizing Arabic Natural Language Processing Tools." *Proceedings of Second Workshop for NLP Open Source Software (NLP-OSS)*. Located with *EMNLP*.
- **Alyafeai Zaid**, et al. 2020. "A fully-automated deep learning pipeline for cervical cancer classification." *Expert Systems with Applications*.

---

## Open Source

- 
- Cofounder of [ARBML](#) for democratizing Arabic NLP by revising the NLP pipeline for Arabic. We have many tools for preprocessing, tokenization, training and deployment. Our community on discord has +400 members.
  - [AttentionNN](#) Implementations of attention in many models like transformers and GPT.
  - Enriching Arabic content with two datasets: [MetRec](#) and [Calliar](#).
  - Creator of [klaam](#) and [whisperar](#), Arabic speech recognition and text-to-speech libraries.
  - Designed many easy to use [creative applications](#) for deep learning.
  - Founding member of [fihm.ai](#), where we publish Arabic articles on AI.

---

## Awards

- First at the high school competition Yemen, 2007.
- Fifth in the high school final exams, 2008.
- First at the integration star contest at KFUPM, 2012.
- Fourth at the [Barmajan hackathon](#) for our project [qawafi](#) on Arabic Poetry analysis 2022.
- Second at the [AI in Sports competition](#) for our project [Ain Sports](#) on action detection in soccer 2022.

---

## Articles & Books

- [A Gentle Introduction to TensorFlow.js](#).
- [Train a model in tf.keras with Colab, and run it in the browser with TensorFlow.js](#)
- [A dive into the Latent Space of BigGan](#).
- A book in Mathematics about [Advanced Integration Techniques](#).
- Book Chapter [Practical Deep Learning for Cloud Mobile & Edge](#) .

---

## Collaborations

- [BigScience](#) member with focus on tokenization, data sourcing and prompt engineering 2021-2022.
- [XLSR-Wav2Vec2 Fine-Tuning Week](#) for Low-Resource Languages 22-29 Mar 2021.
- [Hugging Face datasets](#) sprint 1-13 Dec 2020.
- Collaborated with researchers from Aramco for ESP failure prediction using different techniques of machine learning 2019-2021.
- I worked on distributed training of GPT-2 on 4 GPUs for Arabic poetry generation in collaboration with Kyle McDonald 2019-2020.
- Collaborated with researchers from CPG at KFUPM for modelling sand dunes 2019-2020.

---

## Invited Talks

- ARBML: PyData Riyadh ([video](#)).
- Fostering Research in Arabic NLP via Open Source Collaboration video: NLP with Friends ([video](#)).
- Enabling Research via Open Collaboration in Arabic NLP: ML Collective ([video](#)).

---

## Teaching

- Teaching assistant at the Data Science Research Summer School in collaboration with Aramco 2021, 2022, and 2023.
- ICS 104: Introduction to Programming in Python and C lab instructor at KFUPM 2022.