Tarteel Initiative White Paper

In the name of Allāh, the Lord of Mercy, the Giver of Mercy

Executive Summary 1

Introduction 2

Phase I: Tarteel 10,000 Challenge 2

Phase II: Tarteel ML 4

Phase III: Tarteel Labs 6

Privacy and Release Plan 7

Timeline 7

Contributors 8

References 8

Executive Summary

Tarteel is an initiative to collect, analyze, and build machine learning (ML) products for the recitation of the Qurʾān. The initiative consists of 3 phases: in the first phase, Tarteel is building the world's **first open-source dataset of crowd-sourced Qur'ānic recitations**. The goal is capture the way that men and women from across the world recite the Qurʾān. In the second phase, Tarteel will perform quality checks on the data, and release the data externally as well as to internal developers to train **machine learning models that convert recitation to text**. Tarteel-built machine learning models will also be released. In the third phase, developers at Tarteel will release build **applications and interactive demos** based on these trained models. As we discuss in the Timeline below, these phases will be partially concurrent. Throughout the process, the team at Tarteel work with sincerity, transparency, and proficiency as we, God-willing, make it easier for Muslims around the world to memorize, recite, and interact with the Qurʾān.

Introduction

The memorization (*ḥifẓ*) and verbal recitation (*tilāwah*) of the Qurʾān are integral parts of a Muslim's life. The Qurʾān is recited by Muslims during the five daily prayers and children raised in Muslim households often begin the process of memorizing select portions of the Qurʾān in Arabic from an early age.

The ubiquity of the recitation of the Qurʾān presents interesting opportunities for (1) data collection, (2) machine learning-based analysis, and (3) the development of software designed to aid Muslims in their endeavors to memorize the Qurʾān correctly. The Tarteel Initiative aims to solve these three problems. In particular,

- **Data collection**: Whereas there are plenty of audio recordings from trained reciters (*qaris*) reciting the Qurʾān, we do not really have data on how ordinary Muslims recite the Qurʾān. We aim to build the world's first public dataset of crowd-sourced Qurʾānic recitation coupled with textual annotations and demographic background information.
- **Machine learning**: The data that we collect can be used to train machine learning models that help us understand the way that the Qurʾān is recited and its relation to the text. The principal ML problem that Tarteel envisions is building a speech2text model for the recitation of the Qurʾān that is based on data collected from ordinary Muslims.
- **Apps and Software**: Having a machine-learning model on hand allows us to create all sorts of applications for our end users, to help them memorize and recite the Qurʾān with greater precision and feedback.

The Tarteel Initiative consists of three phases, corresponding to each of the above challenges, which we describe in greater detail ahead. These phases do not necessarily one after the other (see Timeline section), though each stage informs and guides the next.

Phase I: Tarteel 10,000 Challenge

Gathering high-quality data of Qurʾānic recitations is necessary to appreciate the way ordinary Muslims recite the Qurʾān around the world, and is a prerequisite to performing machine learning-based analysis. In general, we know that the more data machine learning algorithms are provided -- and the more representative this data is of the final deployment -- the better the algorithm can learn and thus accurately infer certain outcomes and results. Many datasets have been collected for image recognition, natural language processing, and even medical data -- however, nothing substantial has been developed for applications associated with the Qurʾān.

Tarteel has launched an initiative (Tarteel 10,000 Challenge) to develop the first public, comprehensive, crowd-sourced audio dataset for Quranic recitation, annotated with the chapter

and verse being recited. In addition, we collect background information about the reader so that we understand the eventual demographic range to which the machine learning algorithms we develop can eventually be deployed

Specifically, in addition to the audio recording, we ask the user to provide the following information:

- Age
- Gender
- Background/Ethnicity
- Reading Type
- Platform (web/mobile)

The first phase of the Tarteel 10,000 challenge has focused on collecting 10,000 verses from people around the world[1]. This phase has focused on incentivizing individuals to read and recite at least 5 ayahs through an intuitive and approachable user interface. It also serves as a means of determining what features should be added via user feedback. The Tarteel 10,000 challenge is accessible via a website (www.tarteel.io) as well as Android and iOS mobile apps.

During the course of the Tarteel Challenge, we are publicly sharing the demographic background of our users (as an example, see Fig. 1 below). The latest demographic information can be accessed at https://www.tarteel.io/about/.

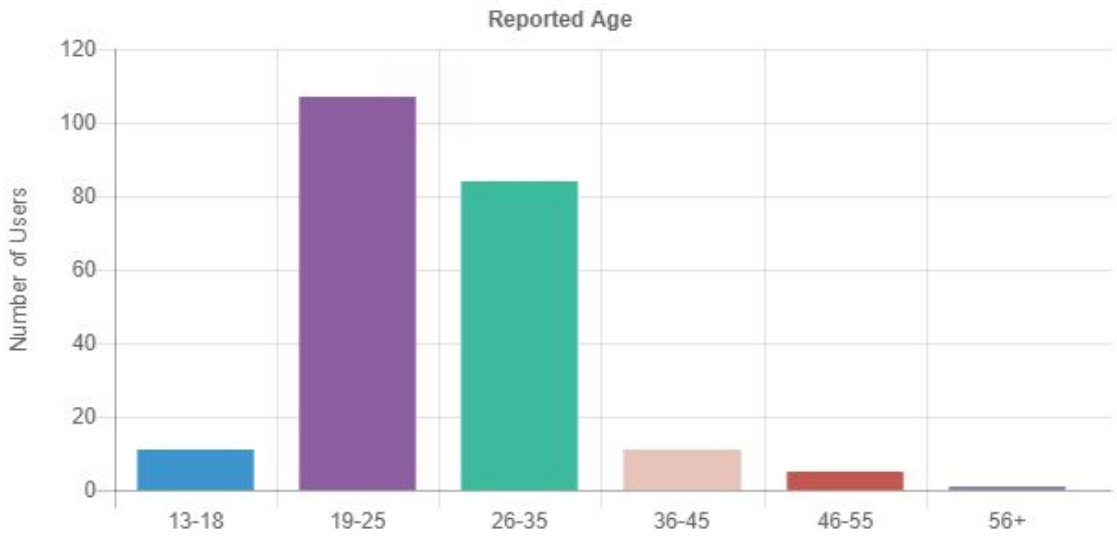


Fig 1: The reported age of Tarteel users, as of Oct. 2018

[1] The number of verses (10,000) was chosen to roughly correspond to 50 hours of recordings, the size of the smallest datasets for training speech-to-text models in the literature, e.g. see [1].

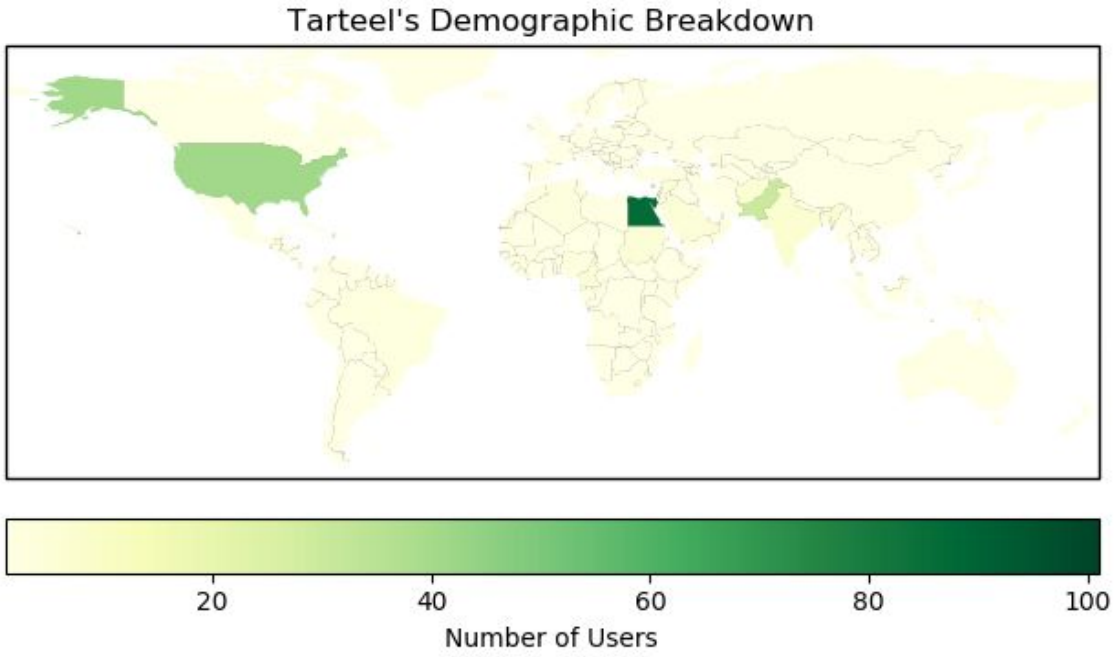


Fig 2: Global distribution of Tarteel users, as of Oct. 2018

Our goal is to diversify our dataset as much as possible, which is why we have taken steps to increase accessibility. For example, eweb platform includes an option to show the transliteration for an ayah, for users who cannot read Arabic natively. We hope that a large and diverse user base will provide a solid dataset to train a machine learning model for downstream applications.

As part of our efforts to create a high-quality dataset, we will also carry out validation steps on the recorded audio. We envision two kinds of quality control (QC): **automated QC** and **reviewer-based QC**. The automated QC will use recording length and frequency spectra to determine if the submitted audio recording consists of human speech, as opposed to empty recordings or background noise. The reviewer-based QC will be carried out internally (and may in the future be extended to external reviewers) to confirm that the correct and complete verse is recited for every audio file.

Phase II: Tarteel ML

Upon collection of a training dataset, the next phase in Tarteel Initiative is to begin to train machine learning models for various downstream applications. The primary model that we are interested in training is a machine learning model that converts spoken recitation into text, described in the box below.

ML Task 1: Recitation to Text (*Primary Task*)

- The input data to this model will be the paired audio recordings that are provided by users, along with the corresponding text of the chapter and verse being recited.
- The architecture that we are considering are bi-directional long short-term memory (LSTMs) trained using the connectionist temporal classification (CTC) loss, similar to prior work presented in [2]. This loss is illustrated in Fig. 3, and notably allows the training of recitation and text data on a sentence/verse-level, without requiring explicit labeling of the corresponding points in the audio and text.

We also include some secondary models that the Tarteel team will consider training, as they may provide some illumination into how the Qurʾān is recited -- for example, by discovering in which ways users from different parts of the world recite Quran differently.

ML Task 2: Recitation Classification

- The input data to this model will be the paired audio recordings that are provided by users, along with the self-reported ethnicity or other demographic information provided.
- The architecture that we are considering are bi-directional long short-term memory (LSTMs) trained using the cross-entropy loss, which can be seen as adaptation of prior work, e.g. presented in [2].

ML Task 3: Style Transfer

- The input data to this model will be the paired audio recordings that are provided by users, along with the corresponding verses recited by professional reciters (qāris).
- The architectures that we are considering is an extension of the tacotron architecture, as proposed by [3].

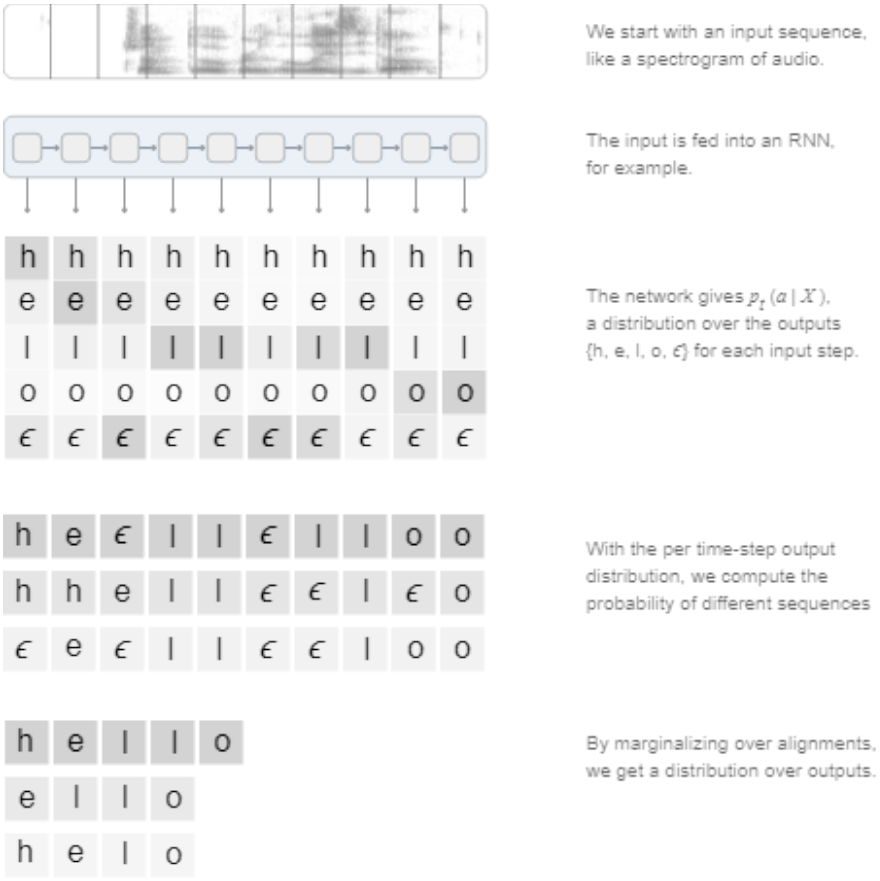


Figure 3: An explanation of the CTC loss, adapted from [4].

As we intend to release the recordings and demographic information (see privacy and release plan), we hope this inspires other machine learning analyses as well.

Phase III: Tarteel Labs

Ultimately, the purpose of collecting data and training machine learning models on Quranic recitation is to build applications that can improve the relationship of Muslims with the Quran. Here, we describe several such projects, that are eventual goals of the Tarteel Labs team. As we intend to release the machine learning models publicly, we hope this inspires other end user applications as well.

Hifz Companion: An automated system that can listen to a user's recitation of the Quran and correct mistakes in recitation by noting discrepancies between the written text and the recitation.

Tajweed Tutor: An automated system that goes beyond the Hifz Companion to provide tajweed-level feedback.

Follow the Imam: A screen placed outside a musallah that displays the verse the imam is reading during taraweeh, along with its translation.

Guess the Ayah: Inspired by games such as Google's Quick Draw, we can create interactive demos that encourage people to recite an ayah, and have an automated system try to guess which ayah the user is reciting. This can be used to increase user engagement, which will allow us to collect even more data.

Tarteel's final and overarching goal is to provide tools which aid, enhance and ease a Muslim's relationship with the Qurʾān. With ML models backed by strong, diverse, and clean data there is incredible potential for building smart Qurʾānic software.

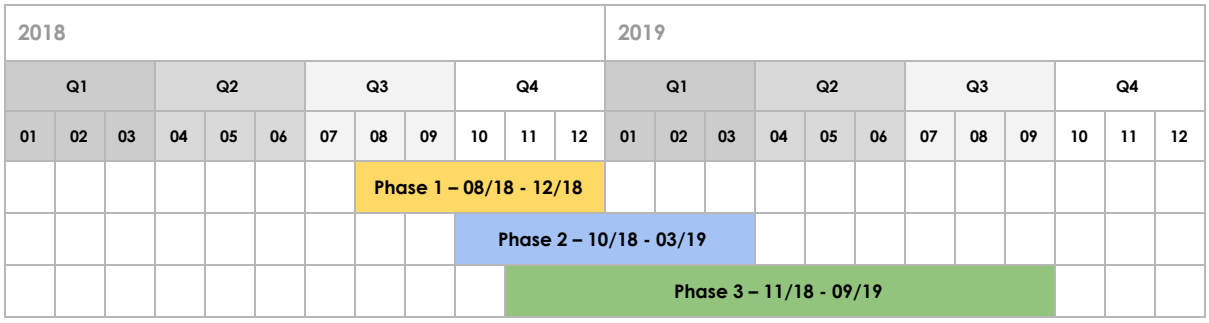
Privacy and Release Plan

The users who provide Tarteel with audio recordings of their recitations also provide a valuable trust to our team. In protect their privacy, while at the same time, creating a public dataset to be released to developers, we will take the following steps:

- Before 5,000 recordings have been collected, no recordings will be released to any developers, which will make it harder to identify individuals in recordings.
- Until 10,000 recordings have been collected, recordings will only be released to core developers, part of the Tarteel team, as part of an effort to validate quality.
- After 10,000 recordings have been recorded, Tarteel will release the recordings publicly to developers for training and analysis.

Timeline

We anticipate that the timeline for the three phases to be as follows:



Contributors

Tarteel is an open-source project hosted on GitHub, and has contributors from around the world. Principal contributors include:

- Abubakar Abid, PhD student at Stanford University
- Ali Abid, PhD student at Stanford University
- Ali Abdalla, mechanical engineer at Tesla
- Abdellatif Abdelfattah, software engineer at Twitter
- BaHaa Jr, a software engineering student at HTI
- Anas Abou Allaban, undergraduate student at Northeastern University, ECE'19
- Areeba Abid, a biomedical engineering student at Georgia Tech
- Mohamed Omran, software developer at Fixed Solutions

References

[1] Bansal, Sameer, et al. "Low-Resource Speech-to-Text Translation." arXiv preprint arXiv:1803.09164 (2018).

[2] Graves, Alex, Abdel-rahman Mohamed, and Geoffrey Hinton. "Speech recognition with deep recurrent neural networks." Acoustics, speech and signal processing (icassp), 2013 ieee international conference on. IEEE, 2013.

[3] Skerry-Ryan, R. J., et al. "Towards End-to-End Prosody Transfer for Expressive Speech Synthesis with Tacotron." arXiv preprint arXiv:1803.09047 (2018).

[4] Hannun, Awni. "Sequence Modeling with CTC." Distill 2.11 (2017): e8.