

Understanding the Quran:

a new [Grand Challenge for Computer Science and Artificial Intelligence](#) proposed by:

Eric Atwell, Kais Dukes, Abdul-Baquee Sharaf, School of Computing, Leeds University;
Nizar Habash, Dept of Computer Science, Columbia University, New York, USA;
Bill Louw, Dept of English, University of Zimbabwe;
Bayan Abu Shawar, Faculty of Computer Studies, Arab Open University, Jordan;
Tony McEnery, Faculty of Arts and Social Sciences, Lancaster University;
Wajdi Zaghouni, Linguistic Data Consortium, University of Pennsylvania, USA;
Mahmoud El-Haj, School of Computer Science and Elec Engineering, University of Essex

The Quran is held by Muslims to be a single-authored text, the direct words of God (Allah), conveyed by the angel Gabriel to Mohammed 1355-1378 years ago, and later transcribed verbatim to be used as the sole authoritative source of knowledge, wisdom and law. The challenge for computer scientists is to represent this knowledge, wisdom and law in computer systems: to build intelligent systems which can answer any question with knowledge from the Quran, and can help society, both Muslim and non-Muslim, to understand and appreciate the Quran.

The Quran is a concise data-set, a text of less than 80,000 words, sequenced in chapters and verses. Muslims hold that the original data format was spoken Classical Arabic, captured faithfully in a sophisticated transcription system: it was vital to transcribe accurately the exact value and relative location of every consonant, vowel, and pause, every morpheme, affix and clitic, every word, verse, and chapter. Access to the Quran has traditionally been through the text: many Muslims learn to memorise and recite the verbatim data-set. Access to the underlying knowledge, wisdom and law requires interpretation and inference; much knowledge is encoded via subtle use of words, grammar, allusions, links and cross-references. For over a thousand years, scholars have sought to extract knowledge and laws from the text, and have built up a much larger Tafsir or corpus of analyses, interpretations and inference chains. Computer Science and Artificial Intelligence presents the opportunity to re-analyse the text data, extract and capture the underlying knowledge in a Knowledge Representation and Reasoning formalism, and enable automated, objective inference and querying.

Other religions also have defining books, for example the Christian Bible, and the Jewish Tanakh, which could also be amenable to Artificial Intelligence analysis and modelling. The Quran makes a good first case study as it is more concise and more homogeneous. The Bible and Tanakh are larger collections of texts by a range of authors over a longer period with a variety of literary genres including allegories, historical narrative, poetry, genealogy, and explicit exposition of various types of law; whereas the Quran is a single text in a consistent genre and style. Consequently, the Tafsir or canon of Quranic interpretations is narrowly focussed on this smaller core text, whereas the body of commentaries on the Bible and Tanakh is broader and less homogenous. Of course, lessons learnt in computing research on the Quran as a first case study could later be extended to the Bible, Tanakh, and other religious texts.

Information Retrieval and Information Extraction systems already exist, so why is Understanding the Quran a grand challenge?

1) Understanding Islam is a major societal issue:

- In Western media and WWW, one of the commonest collocations of “Islamic” is “terrorist”, fuelled by and fuelling conceptions that Islam is a threat; Western schools, universities and the

general public need an objective, impartial online Quran Expert to learn about Islam and understand its implications for society.

- Some non-Arabic-speaking Muslims may also be ignorant of the deeper meanings in the Quran, despite memorising the sounds of the verses. An impartial online Quran Expert could help them question and understand the teachings of the Quran for themselves.

2) Current systems can, in principle, answer “factoid” questions from the source text, like “Are all angels male?”; but many potential questions are more difficult and contentious to answer via text-match, requiring a new Knowledge Representation and Reasoning formalism capable of capturing complex, subtle knowledge encoded in the Classical Arabic, and inferencing in new ways which mirror the thousand-year-old traditions of scholarly analysis and interpretation.

3) In principle, we could use any book as training data for Knowledge Extraction research. The Quran stands out as the source of a large collection of analysis and interpretation texts, known as Tafsir, which could provide a Gold Standard “ground truth” for AI knowledge extraction and knowledge representation experiments. An additional sub-challenge is to encode the Tafsir interpretations in our Knowledge Representation formalism, so we can cross-check for compatibility and consistency with knowledge extraction results from the Quran corpus. In principle, we should aim to be able to reproduce computationally every sound inference and interpretation in the Tafsir. We may find some computational results are incompatible with specific Tafsir inferences and/or conclusions, which will shed new light on traditional interpretations. We may also find some new computational results which are not in any Tafsir interpretation, thus adding to the canon of Islamic wisdom.

4) The system will be used and relied on by billions of Muslims, and also billions of non-Muslims who want to understand the influence of Islam around the world. This is not just an issue of systems scalability and robustness, since the likes of Google and Yahoo etc can already handle huge volumes of queries. It is also vital that answers are always “correct” in that they are consistent with and supported by evidence from the source text, have a demonstrable chain of inference. We expect Google to sometimes get it wrong, and we can live with “acceptable error rates”; for Quran users, any false inference may be unacceptable.

5) Existing Quran websites offering limited search and analysis are already popular with scholars and the general public; for example, the [Quranic Arabic Corpus](#) research project has been snowed under by [volunteer contributors](#). This makes Understanding the Quran an ideal vehicle for research in computer-supported collaborative working. The potential demand / market for an online Quran Expert is huge, a flagship achievement for Computer Science which will capture the public imagination.

In summary, Understanding the Quran is a grand challenge for society, for western public education, for Muslim-world education, for knowledge representation and reasoning, for knowledge extraction from text, for systems robustness and correctness, for online collaboration. **Understanding the Quran** is a major new Grand Challenge for Computer Science and Artificial Intelligence.