Final Project: Ethical Impact

Our code itself isn't very useful, since it's still pretty flawed and doesn't have great accuracy. However, with a more detailed training set, the algorithm could potentially get near-perfect results, as we saw in the paper by Picone et al. And, there's been further development with this problem since the paper we studied,[1, 2] the problem has essentially been solved.

Potential benefits:

- Kanji are very difficult to learn, and there's over 2,000 of them used commonly in things like news, with even more in literature. Especially for learners of Japanese as a second language, learning the word itself on pronunciation alone isn't hard, but learning the associated kanji can be very difficult. This program would help remove some barriers to literature for people who don't yet know a large number of kanji.
- Having a kanji-to-hiragana (pronunciation) program would be very useful in text-to-speech applications.

Potential abuses/issues/harm to society:

- People may become dependent on the technology and not bother learning how to read or write kanji. This could cause a large part of Japanese language and culture to be forgotten.
- Automatically translated kanji for things like news sites could be of lower quality than translations provided by the authors themselves. (There is still a small percentage of error.)

References

[1] Wei-Bin Chang. Kanji-to-hiragana conversion based on a language model. *Acoustics, Speech, and Signal Processing*, 2001.

[2] Wei-Bin Chang and Sachiko Morishita. A statistical approach to automatic phonetic transcription of japanese orthographic words. 10(4):55–63, 2003.