VICTORIA UNIVERSITY OF WELLINGTON TE HERENGA WAKA | Wellington Faculty of Engineering Te Wāhanga Ahunui Pūkaha | Ethics and Sustainability Symposium

Code of Ethics for Automatic Speech Recognition

John Eleigio Cecilio

Abstract
Automatic Speech Recognition (ASR), which turns human speech into text, has had a long history of development and continues to improve and expand. Both academic and industry discussions on the technology present ethical issues regarding the privacy, security, and use of user data, as well the accuracy of ASR for different demographics. This paper reviews the literature on such ethical issues and proposes a code of ethics that addresses them. A case study on Google ASR products is conducted, evaluating that the industry still has improvements to make, especially in user data collection and use, and in accuracy.

Keywords: Automatic speech recognition; Voice recording; Privacy; Ethics; Engineering.

1. Introduction
As the world evolves and advances towards implementing more '*smart*' technologies, devices that employ automatic speech recognition are made abundant. Automatic speech recognition (ASR) technologies have advanced as significantly as they have become available, raising many ethical questions and issues to be considered.

ASR has had an extensive history since its initial inception in the 1950s [1]. Modern ASR commonly adopts deep learning methods using neural networks and large learning datasets [1]. At the high level, deep learning is used to produce, after '*learning*' with a dataset, many layers of acoustic pattern recognition [1], [2]. These layers analyse various aspects of the input audio signal and appropriately produce an acoustic model. With this acoustic model, and alongside statistical language modelling [2], [3], computers can turn human speech into a processable string of words.

Recently, ASR is commonly used for interacting with intelligent virtual assistants such as Amazon's *Alexa*, Apple's *Siri*, and Google's *Google Assistant*. Devices listen for 'wake-words' which then prompt the virtual assistant to further listen for commands and assist in daily life [4]. Another application for ASR is live speech captioning as an aide for individuals that are deaf or hard-of-hearing [5].

1.1. Objective
This paper reviews existing codes of ethics and the discussions regarding the ethics of ASR technologies to present a code of ethics for such technologies. Based on the review a code of ethics is introduced and applied to a case study, discussing how each principle of the code relates to its application.

2. Ethical Issues and Expectations in ASR
This section reviews the literature on the ethical issues presented by ASR, as well as the ethical expectations placed upon engineers, to identify the important considerations that the code of ethics should address.

2.1. Engineering New Zealand Code of Ethics

Engineering New Zealand (ENZ) is an organisation that serves as a professional body for New Zealand's engineers. The organisation fosters the growth of all its members and ensures that engineers maintain a professional standard in their work. To ensure professionalism and integrity from its members, ENZ has defined a code of conduct that its members are subject to uphold [6]. Its principles are as follows:

1. Take reasonable steps to safeguard health and safety.
2. Have regard to effects on environment.
3. Report adverse consequences.
4. Act competently.
5. Behave appropriately.
6. Inform others of consequences of not following advice.
7. Maintain confidentiality.
8. Report breach of Code.

While not directly addressing the ethical considerations of ASR, the principles of the code of ethics of ENZ are broad and should apply to all kinds of engineers, including those who work in the development of ASR. Its principles may serve as a guide when proposing a code of ethics that is focused on ASR specifically.

2.2. Privacy and Collection of User Conversations

One of the biggest ethical concerns in ASR is whether the audio input of users is collected and stored beyond its analysis. There are also concerns regarding how much of the users' speech and conversations are recorded and collected.

ASR technologies often record the audio input from its users to be sent to company servers, primarily for the sake of improving their ASR systems. With the prevalent use of machine learning methods in ASR, access to many large datasets is beneficial in improving accuracy. Huang et al. [7] exclaim, recounting the past and future of ASR technologies, the opportunity of procuring data made available by the abundance of ASR in everyday devices. It is emphasized that an increase in data used to train ASR systems significantly reduces the word error rate (WER) percentage [7]. In technologies since 2010, it is estimated there has been a decrease of around 21 percent in the WER of modern ASR systems due to an increase in training data [7].

While it is evident that collecting user speech would be significant in improving ASR, the collection of such data remains an ethical grey area. The companies which produce ASR make their collection policies available online, outlining how audio data may be collected. For example:

- Apple's policy on their ASR systems [8] states that by default, speech recordings may be sent to their servers while processed transcripts are always sent. Alongside this, other data such as contact names and names of devices and people in photos are also sent. Such data is not directly associated with the user's AppleID, but instead with a random identifier. There is no way to disable these behaviours without terminating the use of Apple's ASR systems.
- Amazon's policy on their ASR systems [9] states that by default, speech recordings are sent to Amazon servers and are associated with the user's Amazon account. Users may choose to configure their account so that older recordings are automatically deleted, or to opt for none of their recordings to be saved. User speech and audio recordings may be manually reviewed by a human for improving Alexa's ASR.
- Google's policy on their ASR systems [10], [11] states that by default, audio recordings are not saved, however other information including personal details

and searched terms from processed speech are collected. Users may opt-in for the collection of their audio recordings and are given options for its management.

Despite giving a sense of choice and control for speech data collection, these options are not made clear to the users of ASR products. A thorough study by Javed et al. [12] found that out of 150 participants who use Alexa, only 36.28% of them acknowledged that Alexa could store their audio recordings to Amazon's servers. Additionally, only 16.8% of the 150 participants acknowledged that they could delete audio recordings collected by their Alexa [12]. Regardless of if user control and choice exist, such control is meaningless if users are not made aware of their options.

Another ethical concern that may be compromising the privacy of ASR technology users is how much of their audio is recorded. Apple, Amazon, and Google all state that recording and analysis of speech only occur after their respective wake-word is heard [8]-[10], however, this may not always be the case. In their research, Javed et al. observed that Alexa had recorded audio without the wake-word being spoken and had stored the audio on their servers [12]. Additionally, cases of virtual assistant users find that their device is activated without the wake-word [13], [14], potentially due to misdetections or hardware faults. Users' privacy may be significantly compromised if user speech and audio are recorded beyond their control.

2.3. Security of Collected User Speech and Information

The discussion in the previous subsection makes it clear that the collection of user audio input is beneficial in the improvement of ASR technologies. Therefore, if the collection of such data occurs for such purposes, the security of collected data must be considered.

The voice recordings analysed with ASR and collected by companies are often personal and identifiable in nature. Dictating a text message or asking a virtual assistant about location-related information – such as the weather - are common scenarios and potentially provide a lot of personal information. Given the personal nature of voice and audio recordings from ASR devices, the security of these recordings must be ensured if they are persisted. The collection of user speech data provides another attack surface.

There are cases of the security of collected data being compromised. In 2019, Dutch audio data collected by Google through their Google Assistant was leaked to the public by one of their language reviewers [14], [15]. Information security calls for adequate security not only in aspects such as encryption of stored data but also for appropriate policies for the people who manage the data.

Privacy policies relating to a company's internal data processor staff are equally important as privacy policies for their customers. The General Data Protection Regulation (GDPR) [16], is the European Union's data privacy and security law and outlines what is required of entities that control any data. A privacy policy that appropriately considers and fulfils the articles stated in the GDPR better ensures information security.

2.4. Use of Collected User Speech Data for Profiling

The collection of user speech data also brings up ethical issues regarding its use. Audio recordings or transcripts for tasks such as adding items to shopping lists or searching for restaurants in their area may be collected by companies. If companies collect user speech data, such information would be effective in forming a profile of the user's preferences and serving them targeted advertising.

Takano et al. [17] propose a method that is more accurate in using ASR to obtain a user's browsing activity and build a profile of the user's preferences. Their method can profile users and make recommendations using their ASR activities without the user's knowledge [17]. While the paper does not explicitly mention targeted advertising, such methodology can be used for user profiling for the sake of serving targeted advertisements. Targeted advertising is an ethical consideration that potentially exploits the privacy of users for the sake of marketing.

2.5. Bias and Inaccessibility in ASR

Technology should serve all of humanity, and no user demographic should be restricted access to technology. While not perfect, ASR systems have continually improved, and their average word error rate (WER) has decreased. There are, however, disparities in the accuracy of ASR systems between demographics.

Koenecke et al. [18] perform a detailed study examining the accuracy of five of the top ASR systems for both white and black speakers. They observed that with 42 white speakers and 73 black speakers, there was a greater WER for black speakers across all ASR systems at an average of 0.35 WER versus 0.19 WER for white speakers [18]. The study also investigated whether a difference in commonly used vocabulary was a major cause of the WER disparity between races. It was found that Google's vocabulary database contained 98.7% of words used by black speakers compared to 98.6% for those of white speakers, inferring that the disparity is more factored by ASR systems' capabilities in creating acoustic models from black speakers [18]. The use of audio samples from various demographics must be considered when improving ASR systems so that the technology is effective for all.

3. Code of Ethics

This section introduces a code of ethics, having considered the issues and expectations as identified from the reviewed literature. This code of ethics is intended for any entity that partakes in the development/ improvement, production, and sale of technology with ASR.

3.1. Collection of User Speech Data must only occur under the consent and terms as determined by Users

Entities partaking in any activities related to ASR should ensure that any collection of user speech data in any format and for any purpose is made unmistakably clear for users. To ensure user privacy, ASR devices and systems must first prompt users to configure data privacy settings before any ASR use, with all options denying data collection by default. Users should have control over any saved data and whether they wish for their data to be identifiable to them.

3.2. User Speech Data must only be used for purposes as determined by the User

Entities that have access to any user data obtained through ASR systems must ensure that the user data is never used for anything that the user has not explicitly determined. For example, a user who has agreed to share their speech data only for the improvement of ASR should never have their speech data be used to build a marketing profile. Use of ASR systems or devices is not sufficient consent for the free use of user data; users must explicitly state their consent.

3.3. Control of User Speech Data must be secured by appropriate cyber security practices

Entities that collect user data obtained through ASR systems must secure that user data using appropriate tools and methods, as well as define appropriate security and privacy policies, to best minimize the security of user data being compromised. Proper

information security principles should be followed. Security should be regularly reviewed, and appropriate incident response strategies should be defined.

3.4. Automatic Speech Recognition systems should not be any less accessible between any demographics or categories

Entities partaking in any activities related to ASR should strive for improving ASR systems to be equally accessible for all. ASR algorithms and systems should be trained for improvement with datasets that considers all demographic speech differences such as accents and vocabulary. If speech data from a certain demographic in amounts required for training is unobtainable from consenting users of ASR, such data should still be ethically sourced. This could be achieved through means such as the employment of individuals from the demographic to provide audio samples.

4. Applying Code of Ethics to a Case Study – Google ASR and Assistant

Google is currently amongst the world leaders in developing and commercialising ASR systems and devices. ASR is used for interacting with the *Google Assistant*: Google's virtual assistant is present on many devices. This section analyses whether Google, in its widespread activities with ASR, successfully upholds the proposed code of ethics for ASR.

4.1. Collection of User Speech Data must only occur under the consent and terms as determined by Users

As previously presented, Google does not collect voice and audio activity from its services without being activated by the user, however, transcripts from such audio are automatically collected [10], [11]. This goes against the principle requiring the user to explicitly consent to any data collection from ASR. Google does, however, allow for intuitive management of user data in their user settings and allows for users to delete and disable the collection of their data.

There have been cases acknowledged by Google in which their devices are activated and listen for user speech or audio when not intended [10], [13]. This goes against the principle as users did not wish to be listened to, and the device may potentially be sending user data without their knowledge. While not malicious or intentional, their devices must be constantly improved to prevent unintentional recording from occurring, to ensure user privacy and control.

4.2. User Speech Data must only be used for purposes as determined by the User

Google only collects user audio data from their ASR services if the user has allowed for it and they are transparent in how that data is used to improve their ASR systems [10], [19]. Beyond the improvement of ASR, Google may use this data, as well as transcripts from audio, to create a profile of the user's preferences to serve recommendations and targeted ads [11]. User profiling is on as long as the user is providing their activity, which goes against the code, although the serving of targeted ads and the collection of any user activity may be disabled in user settings. Unfortunately, disabling collection of user activity data to prevent profiling will also prevent collection of user data to be used for improving ASR.

4.3. Control of User Speech Data must be secured by appropriate cyber security practices

Google stores collected user interactions with the assistant on their cloud servers [10]. Despite a case such as the data leak of Dutch user audio data [14], [15], Google strive for improving their security to ensure that user data is safeguarded.

To ensure security and privacy of their cloud servers, Google regularly review and verify that they are compliant with recognized certifications and provide relevant resources to

ensure security compliance of Google services globally [20]. Amongst their list of compliances is the previously discussed GDPR, although they are also compliant with the ISO/IEC 27001, 27017, 27018, and 27701 standards [20]. These all signify that Google maintain information security - appropriately securing data and any personally identifiable information.

4.4. Automatic Speech Recognition systems should not be any less accessible between any demographics or categories

As previously reviewed, Koenecke et al. [18] study racial disparities in speech recognition, finding a WER of 0.19 for white speakers versus 0.31 for black speakers. When comparing the error rates for identical short phrases, the WER was 0.17 versus 0.11 for black and white speakers respectively [18]. WER disparity is attributed to a lack of variation and inclusion in training data for the systems that process acoustic models of speech, as in accents, cadence, and inflection as opposed to vocabulary used [18]. With the study being published very recently in 2020, Google still needs to work on training their ASR systems with inclusivity in mind so that that it is accessible for all.

5. Conclusion

ASR technology will continue to benefit humanity as advancements are made. However, reviewed literature proves that ethical issues persist with the technology. The proposed code of ethics takes these issues into consideration to mitigate them. When applying the code to a leader in ASR, it is apparent that, while the industry takes ethical considerations in ASR technology, shortcomings to the code call for steps to take in ensuring better accuracy, and user privacy and control.

The code of ethics aims to serve as a foundation in initiating further consistent discussion in the ethics of ASR, as advancing technology is constantly evolving and new ethical issues may arise.

References

[1] IBM, “What is Speech Recognition?,” IBM Cloud Education, 2 September 2020.

[2] D. Yu and L. Deng, Automatic Speech Recognition, London: Springer, 2015.

[3] R. Kuhn, D. Mori and Renato, “A Cache-Based Natural Language Model for Speech Recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 12, no. 6, pp. 570-583, 1990.

[4] M. Hoy, “Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants,” *Medical Reference Services Quarterly,* vol. 37, no. 1, pp. 88-81, 2018.

[5] J. Butler, B. Trager and B. Behm, “Exploration of Automatic Speech Recognition for Deaf and Hard of Hearing Students in Higher Education Classes,” in *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*, Pittsburgh, 2019.

[6] Engineering New Zealand, “Engineer Tools: Code of Ethical Conduct,” 1 July 2016.

[7] X. Huang, J. Baker and R. Reddy, “A historical perspective of speech recognition,” *Communications of the ACM,* vol. 57, no. 1, pp. 94-103, 2014.

[8] Apple Inc., “Ask Siri, Dictation & Privacy,” Apple Inc., 2022.

[9] Amazon.com Inc., “Alexa, Echo Devices, and Your Privacy,” Amazon, 2022.

[10] Google Nest, “Data security and privacy on devices that work with Assistant,” Google, 2022. [Online].

[11] Google, “Google Privacy Policy,” Google, 10 February 2022.

[12] Y. Javed, S. Sethi and A. Jadoun, “Alexa's Voice Recording Behavior: A Survey of User Understanding and Awareness,” in *14th International Conference on Availability, Reliability and Security*, Canterbury, 2019.

[13] A. Russakovskii, “Google is permanently nerfing all Home Minis because mine spied on everything I said 24/7,” Android Police, 10 October 2017.

[14] K. Murname, “These Are The Real Problems Revealed By The Belgian Leak Of Google Assistant Voice Recordings,” Forbes, 14 July 2019.

[15] D. Monsees, “More information about our processes to safeguard speech data,” Google, 11 July 2019.

[16] Council of the European Union, “General Data Protection Regulation (GDPR),” Proton Technologies AG., 27 April 2016.

[17] K. Takano, H. Honda and K. Fun Li, “User Preference Profiling Based on Speech Recognition for Personalized Recommendation,” in *2012 Seventh International Conference on Broadband, Wireless Computing, Communication and Applications*, Washington, 2012.

[18] A. Koenecke, A. Nam and E. Lake, “Racial disparities in automated speech recognition,” *Proceedings of the National Academy of Sciences of the United States of America,* vol. 117, no. 14, 2020.

[19] Google, “Learn how Google improves speech models,” Google, 2022.

[20] Google Cloud, “Compliance resource center,” Google, 2022.