



Why we still need advances in systems for Automatic Speech Recognition (ASR) - SF 01_01_24

If you ask the proverbial “man or women in the street” about speech recognition, they would likely say, “it’s already been done” and refer to Alexa as an example. And if you’re a business with a need for speech to text is satisfied by general language models published and as Microsoft, Google and AWS cloud services, then you might agree too.

But is has ASR really been “done”? I asked our good friend ChatGPT a couple of questions on this subject, to get a better understanding. The conclusion was easy. The conventional ASR used today can definitely benefit from alternative solutions, that can address its known limitations, constraints and dependencies. This is precisely why the novel invention for Automatic Speech Recognition, developed and patented by the University of Oxford, Faculty of Linguistics and Phonetics, is being brought to market today as *FlexSR. The FlexSR system directly addresses and radically improves 9 out of 10 of the problems highlighted against each of the questions posed and answered in the Chat GPT Q&A, as follows.

Question: What problems still need to be overcome in speech recognition AI machine learning and large models?

ChatGPT Answer

While significant progress has been made in speech recognition using AI and machine learning, there are still several challenges and areas for improvement. Some of the ongoing issues include:

1. Accents and Dialects:

- Speech recognition systems may struggle with accents, dialects, and variations in pronunciation. Training models to understand a diverse range of linguistic nuances is an ongoing challenge.

2. Background Noise:

- Recognizing speech in noisy environments remains a challenge. Real-world scenarios often involve background noise, and improving the robustness of models to accurately transcribe speech in various acoustic conditions is a priority.

3. Speaker Variability:

- Variability in individual voices, pitch, and speaking styles can impact the accuracy of speech recognition systems. Systems need to adapt to different speakers effectively.

4. Limited Data for Underrepresented Languages:

- Many speech recognition models are trained on datasets that are biased toward dominant languages and dialects. There is a need for more diverse and representative datasets, especially for underrepresented languages, to improve accuracy across a broader linguistic spectrum.



5. **Context Understanding:**

- Improving contextual understanding is crucial. Speech recognition systems often struggle to interpret the broader context of a conversation, leading to errors in understanding meaning and intent.

6. **Multilingual Support:**

- Expanding support for multiple languages and dialects is an ongoing challenge. Building models that can effectively understand and transcribe speech in a wide range of languages is an area of active research.

7. **Adaptation to New Vocabulary:**

- Adapting speech recognition models to understand and adapt to new vocabulary, slang, or industry-specific terms is a challenge. Continuous learning and adaptation are essential for real-world applications.

8. **Data Privacy and Security:**

- Privacy concerns related to storing and processing audio data must be addressed. Ensuring that speech recognition systems comply with privacy regulations and protect user data is a critical consideration.

9. **Real-Time Processing:**

- Achieving real-time processing for large-scale speech recognition tasks, especially on edge devices with limited resources, is a technical challenge. Low-latency requirements in applications like voice assistants and transcription services require efficient algorithms.

10. **Interpretable Models:**

- Increasing the interpretability of speech recognition models is important for understanding why certain errors occur and for building trust in the technology, especially in critical applications.

Addressing these challenges involves a combination of advances in model architectures, the development of more diverse and extensive datasets, improvements in training methodologies, and a focus on real-world applications to ensure that speech recognition systems perform well across various scenarios and user demographics. Ongoing research and development efforts are essential to making continuous improvements in this field.



Question What are the limitations and data collection and processing of current AI and machine learning for automatic speech recognition for a variety of languages and accents for a variety of different domains?

ChatGPT Answer

The limitations and challenges in data collection and processing for automatic speech recognition (ASR) across languages, accents, and domains are multifaceted. Here are some key considerations:

1. Data Bias:

- Training datasets may be biased toward certain languages, dialects, or accents, leading to reduced accuracy for underrepresented linguistic variations.
- Lack of diversity in training data can result in biased models that perform poorly for speakers with accents or in regions not well-represented in the training set.

2. Underrepresented Languages:

- Limited availability of training data for certain languages and dialects makes it challenging to build accurate ASR models for those linguistic variations.
- Building ASR models for languages with fewer resources requires substantial efforts in data collection and may still result in less accurate performance.

3. Domain Specificity:

- ASR models trained on general datasets may not perform optimally in specialized domains with specific jargon, terminology, or industry-specific language.
- Adapting ASR models to different domains requires domain-specific training data, which may not always be readily available.

4. Accents and Pronunciation Variability:

- Accents and pronunciation variations can pose challenges, especially if training data does not adequately represent these variations.
- ASR systems may struggle with understanding speakers with non-standard accents, leading to decreased accuracy.

5. Background Noise and Environmental Factors:

- ASR models may not generalize well to various acoustic environments and background noise conditions.
- Real-world scenarios often involve unpredictable noise levels, making it challenging to ensure accurate recognition in diverse settings.

6. Limited Multilingual Models:

- Building effective multilingual models that can handle multiple languages seamlessly is challenging. Many existing models are optimized for specific languages, and extending them to new languages can be non-trivial.



7. Privacy Concerns:

- Collecting and processing audio data for training ASR models raises privacy concerns. Ensuring compliance with data protection regulations and addressing user privacy is crucial.

8. Interference from Other Modalities:

- In certain applications, such as transcription services for videos, ASR models need to handle interference from other modalities, like background music or overlapping speech from multiple speakers.

9. Resource-Intensive Training:

- Training large-scale ASR models can be computationally expensive and resource-intensive. This may limit the accessibility of training such models for smaller organizations or researchers with limited resources.

10. Domain Adaptation Challenges:

- Adapting pre-trained models to specific domains may be challenging, particularly if the target domain has limited training data. Fine-tuning on domain-specific data is essential but may not always be sufficient.

Addressing these challenges requires ongoing efforts in data collection, model development, and research. Open collaboration, diverse dataset creation, and attention to ethical considerations, including privacy and fairness, are crucial for advancing ASR systems that are more accurate, robust, and inclusive across languages, accents, and domains.

*FlexSR is the patented system brought to market by FLEXSR LIMITED. See the additional papers on FlexSR, such as “An Improved Approach to Automatic Speech Recognition – FlexSR Whitepaper” addressing problems highlighted in this paper.