a i media

SUBMISSION IN COMMENT

Consumer and Governmental Affairs Bureau Seeks Comment on Petition for Declaratory Ruling and Petition for Rulemaking Filed by Telecommunications For The Deaf and Hard of Hearing et al. On Live Closed Captioning Quality Metrics and The Use of Automatic Speech Recognition Technologies

(DA 19-776)

CG Docket No. 05-231
RM-11848

October 15th, 2019

Ai-Media Inc.
241 W Federal St
Youngstown, OH 44503
(213) 337-8552

Marlene H. Dortch
Secretary
Federal Communications Commission
445 12th Street SW
Washington, DC 20554

CG Docket No. 05-231
RM-11848

October 15th, 2019

Dear Ms Dortch,

Thank you for the opportunity to participate in this call for comments.

Ai-Media Inc (Ai-Media) is a values-led business that makes over 70,000 hours of content each year accessible through live and offline captioning, transcription, subtitling and translation services, and descriptive video / audio description.

We are a technology vendor and service provider to the captioning industry in the United States, Canada, and around the world. We deploy a wide range of technologies from fully-automated to human-curated captioning for live and recorded captioning – in English, Spanish, French, Portuguese, and many other languages.

Our Facebook community is over 1 million strong, celebrating stories of access and inclusion.[1]

We also provide captioning for a range of non-broadcast environments, including schools, colleges and universities, workplaces, conferences, webinars and events, as well as social media platforms such as Facebook, YouTube, Vimeo and Twitch.

Ai-Media has 15 years' experience delivering live and recorded captioning for broadcast, five years' experience with NER, and eighteen successive quarters of historical NER audit data of our live captioning quality, averaging 99.5%.[2]

In 2013 Ai-Media became the first captioning provider in the world to appoint an independent auditor to report on captioning quality across all of our services – to better provide an objective, verifiable measure of the overall quality of captioning provided to our clients, and viewers.

We believe this to be far more effective at providing an overall picture of captioning quality than a complaints-based system.

---

[1] https://www.facebook.com/aimediaAUS/

[2] NER stands for "Number, Edition, Recognition", and is a method of assessing captioning quality that has been adopted in several US states and numerous countries around the world (https://en.wikipedia.org/wiki/NER_model). For Ai-Media audit results, see https://www.ai-media.tv/external-captioning-quality-audit/

Recommendation No. 1: Support objective technology-neutral metrics for caption quality

Ai-Media strongly supports the key element of the Petition for Declaratory Ruling and/or Rulemaking on Live Closed Captioning Quality Metrics and the Use of Automatic Speech Recognition Technologies (*Consumer Groups Petition*)[3], that "The Commission should develop objective technology-neutral metrics for caption quality."

In a rapidly-evolving technology landscape, captioning quality *measurement* must be decoupled from the *method* of captioning preparation. The former should not change. The latter will be constantly evolving.

We submit that the Commission should set the objective "North Star" on how captioning quality should be assessed. It then rests with industry participants to determine when, and in what circumstances, emerging technologies (such as electronic newsroom technique (ENT) and automatic speech recognition (ASR)) are ready for prime-time. The question for industry will be a clear one: "Can this automated technique deliver the minimum acceptable captioning quality benchmark?".

With this framework in place, viewers (and the regulator) should not be concerned about how captions are produced – simply whether the objective quality threshold has been achieved.

What are the elements of an objective technology-neutral metric for caption quality?

The chosen metric of captioning quality should have the following elements in order to be effective:

1. **Objective**: clear rules and standards that are easy to follow, regardless of the method of captioning preparation.

2. **Consistent and replicable**: different assessors should provide the same score for the same sample.

3. **Aligned with viewer perceptions of quality**: higher scores should align with viewer perceptions of better quality.

4. **Clear decision rules**: The objective metric should provide clear decision-rules to live captioners on which trade-offs inherent in live captioning will maximize quality.

5. **Incorporates the qualitative attributes** of captioning quality mandated by the Commission: The metric should score samples more highly that provide captions are "accurate, synchronic, complete, and correctly placed".

---

[3] Petition for Declaratory Ruling and/or Rulemaking on Live Closed Captioning Quality Metrics and the Use of Automatic Speech Recognition Technologies, CG Docket No. 05-231; MB Docket No. RM-11065.

6. **Effective in both English and Spanish**: Spanish language users who are deaf or hard of hearing should have the same level of access as those for whom English is a first language.

7. **Empirical evidence**: Prior to adoption by the Commission, the metric should be backed by convincing empirical evidence in theory and in practice.

Recommendation No. 2: Consider the adoption of the NER method as the Commission's objective captioning quality metric

The NER method scores well on the above criteria.

1. **Objective**: NER has clear rules and standards that are easy to follow. Recognizing that not all errors are equal, NER scores errors (with weightings) as being minor (0.25), standard (0.5) or major (1.0). Correct edits are not penalized. There is a global system of NER evaluation, with certification available for NER assessors.

2. **Consistent and replicable**: Among trained NER assessors, the variation in scores between different assessors is very low. In Ai-Media's experience, this is often as low as 0.1%. Ai-Media certifies all its captioners internally according to the NER system and has aligned its certification system with that of the Live Reporters International Certification Standard (LiRICS) as overseen by Prof Pablo Romero Fresco.[4]

3. **Aligned with viewer perceptions of quality**: NER is a system that has been designed with viewer perceptions of quality at its core. Its credibility rests on robust data – collected over many years, across different settings and languages – that align higher scores with improved perceptions of quality. Most recently, in Canada, Canadian NER guidelines have been adopted with the following minor modifications to the classification of error types as follows[5]:

   The six error types and their deductions are:

   - **Correct Edition (CE)** (-0.0). CE is scored when captions are different from the verbatim audio but retain its full meaning, without interrupting words/phrases.
   - **Omission of Main Meaning (OMM)** (-0.5)
     The captions have lost the main idea presented in the audio's Independent Idea Unit (see page 2).
   - **Omission of Detail (OD)** (-0.25). Here, the captions have lost one or more modifying meaning, affecting a Dependent Idea Unit (see page 2) but not the main idea.
   - **Benign Error (BE)** (-0.25) A captioned word or phrase is incorrect, causing an interruption in the reading. However, the viewer can readily figure out the original meaning, from context (in video) or similarity to the real word.
   - **Nonsense Error (NE)** (-0.5) A captioned word or phrase is incorrect and the viewer can't figure out the original meaning. If the impact of the word/phrase is to alter or omit the meaning of the idea unit, OD, OMM or FIE will be scored instead.
   - **False Information Error (FIE)** (-1.0) The captions make sense, but the information they present is different from the verbatim. The caption viewer cannot tell that the meaning is false.

---

[4] http://galmaobservatory.webs.uvigo.es/services/certification/
[5] Canadian NER Evaluation Guidelines, Version: 2018-11-02.

4. **Clear decision rules**: As a provider of high-quality live captioning services at scale, watched by millions of people, a key priority of Ai-Media is to establish clear, robust decision rules for live captioners. The NER scoring system provides this clarity as follows:

   - With the current state of the technology, errors in live captioning are inevitable (either due to editing by the captioner, or misrecognition by the software). By weighting each error, and with the empirical evidence of thousands of NER assessments, clear decision rules have emerged on how errors should be treated by live captioners.

   - Correcting a live captioning error on air itself contains a trade-off. By choosing to correct an error on-screen, the captioner foregoes captioning new live content. *The decision rule is that captioners should only correct an error that misleads the viewer* (the FIE error). If captioners correct errors that score (0.5) or (0.25) they will likely omit information that will detract further from captioning quality and result in a lower NER score overall.

5. **Incorporates the qualitative attributes** of captioning quality mandated by the Commission: NER scores samples more highly that provide "accurate, synchronic, complete, and correctly placed" captions, by considering the context to the viewer. For example, some spoken content (such as "play by play" sports commentary) is designed to describe the content *immediately* on screen. In the context of live captioning with inherent delays of 3-5 seconds, captioning such content is always misleading, as the description of what is on screen will not match the captions.

6. **Effective in both English and Spanish**: Spanish language users who are deaf or hard of hearing should have the same level of access as those for whom English is a first language. The NER framework works across multiple languages including Spanish.

7. **Empirical evidence**: Since its inception almost a decade ago, the NER method has emerged as the global de facto standard for captioning quality measurement.

   - The NER system of caption quality has been deployed in several US states, and multiple international jurisdictions around the world including the UK, Spain, France, Italy, Switzerland, Germany, Belgium and Australia.[6]

   - In September 2019 the Canadian regulator, the CRTC, became the most recent to adopt NER, officially replacing the verbatim standard.[7]

   - Ai-Media is not aware of any jurisdiction that has moved away from NER having adopted it.

---

[6]Romero-Fresco & Martinez, "Accuracy Rate in Live Subtitling – the NER Model",
http://captiontest.com/roehampton%20NER-English.pdf
[7] https://crtc.gc.ca/eng/archive/2019/2019-308.htm

Conclusion

Ai-Media thanks the Commission for the opportunity to comment on the Consumer Groups Petition.

We look forward to working with all stakeholders on this important reform to drive improvements in quality in live captioning across the United States.

Sincerely,

Tony Abrahams
Co-founder & CEO