Before the
Federal Communications Commission
Washington, D.C. 20554

In the Matter of:		
Internet Protocol Captioned Telephone Service	)	CG Docket No. 13-24
Telecommunications Relay Services and Speech-to-Speech Services for Individuals with Hearing and Speech Disabilities	)	CG Docket No. 03-123

Petition for Rulemaking to Require Option for Communications Assistants by Stand-Alone Automatic Captioning Providers

via electronic filing

May 31, 2024

Samuelson-Glushko Technology Law & Policy Clinic (TLPC) at Colorado Law

Vivek Krishnamurthy, Director
vivek.krishnamurthy@colorado.edu

Counsel to TDIforAccess, Inc. (TDI)

On behalf of:

TDIforAccess, Inc. (TDI)
AnnMarie Killian—amkillian@tdiforaccess.org
Wilmington, DE
https://TDIforAccess.org

National Association of the Deaf (NAD)
Law and Advocacy Center
Zainab Alkebsi, Policy Counsel—zainab.alkebsi@nad.org
Silver Spring, MD
https://www.nad.org

Hearing Loss Association of America (HLAA)
Barbara Kelley, Executive Director – bkelley@hearingloss.org
Neil Snyder, Director of Public Policy—nsnyder@hearingloss.org
Rockville, MD
https://www.hearingloss.org

Summary

The Communications Act requires the Commission to implement telecommunications relay services (TRS) that enable people with hearing or speech disabilities or who are Deafblind to communicate in a manner that is functionally equivalent to voice telephone users. However, the Federal Communications Commission's (FCC or Commission) 2018 Declaratory Rulemaking allowing providers to use automatic speech recognition (ASR) to generate captions for Internet Protocol Captioned Telephone Services (IP CTS) without providing a communications assistant (CA) as backup did not and could not establish that ASR meets this standard of functional equivalence. Further, research conducted *since* 2018 shows that ASR continues to frequently misinterpret speech with accents, dialects, or patterns that deviate from standard American English, or when used to recognize speech in environments with background noise. In such circumstances, ASR error rates can be as high as one in six words for the worst-affected users, making such conversations nearly impossible to follow.

Given this data and the experiences of users since the adoption of the 2018 ruling, we seek reversal—through a Commission rulemaking—of the Commission's original decision to allow IP CTS providers to rely solely on ASR for the delivery of IP CTS. Specifically, considering ASR's limitations, we seek a rule that would require all certified IP CTS providers to give users the option to select (at the start of a call) or switch (mid-call) to a live CA whenever the user finds ASR performance to be unsatisfactory.

We further urge the Commission to act expeditiously to complete its open proceeding on the establishment of clear, technology-neutral performance goals and metrics for IP CTS, which will assist the Commission in making assessments of each provider's ability to meet the functional equivalence standard, regardless of whether they use CAs or automated systems to generate captions.

Table of Contents

Discussion

The above-signed organizations respectfully petition the Federal Communications Commission (FCC or Commission) to revise its rules to require providers of Internet Protocol Captioned Telephone Services (IP CTS) using automated speech recognition (ASR) technologies to also make available to their users a communications assistant (CA) as a backup.[1] We do so in view of the Communications Act's functional equivalence requirement for telecommunications relay services (TRS),[2] and the evidence furnished below demonstrating that ASR-only IP CTS services fail to meet this required standard of service in many circumstances.

Specifically, we ask the Commission to initiate a new notice-and-comment rulemaking to:

1. Require IP CTS providers that rely on ASR—whether they provide ASR-only captions or provide both ASR-only and human assisted captions—to give users the option to select a CA to generate captions at the start of a call or at any point thereafter, as needed to achieve functionally equivalent communication; and
2. Refrain from certifying additional ASR-only providers of IP CTS until a new rule requiring providers to give users the option to select CAs is adopted.

Additionally, we renew our request for the Commission to fulfill its commitment to establish technology-neutral performance goals and metrics for the provision of IP CTS,[3]

---

[1] We file this petition pursuant to Section 1.401 of the Commission's rules. 47 C.F.R. § 1.401.

[2] 47 U.S.C. § 225(a)(3); *see also* 47 C.F.R. § 64.601(a)(43).

[3] *See, e.g.,* Letter from Blake Reid, Counsel to Telecommunications for the Deaf and Hard of Hearing, Inc. (TDI), to Marlene H. Dortch, Secretary, FCC, CG Docket No. 03-123 et al., at 1 & nn.1-2 (filed Mar. 26, 2021), https://www.fcc.gov/ecfs/filing/10326769717167 (Accessibility Advocacy and Research Organizations March 2021 *Ex Parte*) (citing numerous filings by the organizations); Comments of Accessibility Advocacy and Research Organizations on the Application of InnoCaption for Certification as a Provider of Internet Protocol Relay Service, CG Docket Nos. 13-24 & 03-123, at 1 & n.2 (filed Dec. 2, 2022), https://www.fcc.gov/ecfs/search/search-filings/filing/1202331113216 (Comments of Accessibility Advocacy and Research Organizations on InnoCaption Application) (citing other filings by the organizations related to IP CTS applications).

as set forth in the Commission's 2018 Notice of Inquiry and proposed in its 2020 Further Notice of Proposed Rulemaking. These goals and metrics should apply to all forms of IP CTS without regard to whether CAs, ASR systems, or some combination thereof are used to generate captions.

I. Background

Section 225 of the Communications Act requires the Commission to implement telecommunications relay services (TRS) that are functionally equivalent to voice telephone services.[4] In a 2018 Notice of Inquiry, the Commission sought comment on whether the definition of functional equivalence proposed by consumer groups in 2011 serves as an appropriate definition of the term.[5] Consumer Groups urged a definition of "functional equivalence" for TRS that enables users to:

> participate equally in the entire conversation with the other party or parties and […] experience the same activity, emotional context, purpose, operation, work, service, or role (function) within the call as if the call is between individuals who are not using relay services.[6]

[4] 47 U.S.C. § 225(a)(3).

[5] *Misuse of Internet Protocol (IP) Captioned Telephone Service; Telecommunications Relay Services and Speech-to-Speech Services for Individuals with Hearing and Speech Disabilities*, CG Docket Nos. 13-24 and 03-123, Report and Order, Declaratory Ruling, Further Notice of Proposed Rulemaking, and Notice of Inquiry, 33 FCC Rcd 5800, 5869, ¶ 158 (2018) (cited hereinafter as *2018 IP CTS Declaratory Ruling* when referencing the Declaratory Ruling, *2018 IP CTS Further Notice of Proposed Rulemaking* when referencing the Further Notice of Proposed Rulemaking, and *2018 IP CTS Notice of Inquiry* when referencing the Notice of Inquiry).

[6] *Consumer Groups' TRS Policy Statement – Functional Equivalency of Telecommunications Relay Services: Meeting the Mandate of the Americans with Disabilities Act,* contained in Letter from Tamar Finn, Counsel to Telecommunications for the Deaf and Hard of Hearing, Inc. (TDI), to Marlene H. Dortch, Secretary, FCC, CG Docket Nos. 03-123 and 10-51, Attach. at 1 (filed Apr. 12, 2011).

In 2018, the Commission issued a declaratory ruling authorizing the provision of TRS using ASR only, without any human involvement.[7] The Commission's ruling was based on its finding that ASR provided the ability "to engage in communication by wire or radio 'in a manner that is functionally equivalent' to voice communication services."[8] This finding was flawed in 2018 and it remains flawed now.

The obligation to ensure functionally equivalent telephone communication services extends to *all* individuals eligible to use TRS. Accordingly, as overseer of the IP CTS program, the Commission must consider the effectiveness of ASR not only for those scenarios in which it will work best, but also for those in which it is likely to fail. For the latter group of calls, the Commission must take action to ensure access to live assistants at any point during a call to achieve effective communication.

As detailed below, the evidence before the Commission in 2018 failed to support its conclusion that ASR technology could satisfy the statutory requirement of functional equivalence when providing IP CTS independently, without human assistance as a backup option. Moreover, research conducted since 2018 indicates that ASR continues to exhibit significant difficulties in achieving accuracy for a significant percentage of calls made by IP CTS users—particularly when these individuals make calls to or receive calls from individuals whose dialect, accent, and speech patterns diverge from the norms of standard American English. It is for this reason that we now petition the Commission to reverse its decision to allow use of ASR without requiring certified IP CTS providers to offer CAs as a backup option.

---

[7] *2018 IP CTS Declaratory Ruling*, 33 FCC Rcd at 5827, ¶ 48. As in the 2018 IP CTS Declaratory Ruling, this Petition uses ASR to refer to fully automated speech recognition without human involvement while acknowledging that automated speech recognition engines can also play a role in IP CTS services using CAs.

[8] *Id.*

II. The 2016 MITRE study does not support the Commission's 2018 decision to permit ASR-only IP CTS.

The Commission's 2018 Declaratory Ruling authorizing the use of stand-alone ASR flowed from its heavy reliance on a 2016 study conducted by the MITRE Corporation.[9] This study, conducted in two phases,[10] found that at least one extant ASR system was equivalent or better than three of the four existing IP CTS services at providing an accurate, expedient, and usable relay service.[11] Drawing upon this study, and in particular MITRE's report summarizing Phase 2 of this study,[12] the Commission outlined several benefits of ASR—including lower transcription delays, equivalent accuracy, and better privacy—to support its conclusion that ASR was a viable alternative to human-assisted TRS in achieving functional equivalence.[13] The Commission also noted ASR's lower operating costs, aligning with Congress's directive to provide TRS efficiently.[14] As discussed in detail below, however, the Commission's conclusions regarding ASR's universal efficacy were not supported by the MITRE study or any other data.

A. The Commission ignored key findings of the MITRE study and misconstrued others.

The Commission selectively pulled content from MITRE's report to reach its result, choosing to rely on certain findings, while ignoring or misconstruing others that were key to MITRE's conclusions. For example, the Commission's justification for approving

---

[9] *Id.* at 5827-28, ¶ 49.

[10] The first phase "captured results from controlled user assessments and established a baseline of usability metrics," while the second described the results of a "usability assessment of alternative speech recognition technologies" and provided "qualitative and quantitative measures for device and caption performance." MITRE Corporation, *Internet Protocol Caption Telephone Service (IP CTS) – Summary of Phase 2 Usability Testing Results* at ii (2016), CG Docket Nos. 03-123 and 13-24 (posted Apr. 11, 2018) (cited hereinafter as MITRE Phase 2 Summary), https://docs.fcc.gov/public/attachments/FCC-18-79A1.pdf.

[11] *Id.*

[12] *Id.*

[13] *2018 IP CTS Declaratory Ruling*, 33 FCC Rcd at 5827-31, ¶¶ 49-51, 57.

[14] *Id.* at 5832, ¶ 59.

ASR placed heavy emphasis on the technology's ability to deliver transcriptions quickly, noting the shorter delay between utterances and ASR-generated transcriptions as compared to CA-generated captions.[15] The Commission also pointed to the expressed preference of some consumers for ASR-generated captions.[16]

In so doing, the Commission selectively ignored that 100 percent of the participants in MITRE's Phase 2 study stated that "if they could only improve one characteristic, they would choose accuracy over speed,"[17] and that "[a]ll participants would prefer to have a Communication [Assistant] on IP CTS calls because they believe this will improve accuracy."[18] Indeed, the 2016 MITRE report noted that study participants "were willing to accept some delay *if it assured more accurate transcripts*," and "expressed that it is more important to receive the correct information the first time and reduce the need to repeat information during the call."[19] To this end, MITRE even recommended that IP CTS providers "[d]evelop and design a feature that allows for real-time, on-demand switching from an automated capability to CA-based services as needed by users."[20] The Commission did not provide for such a mandate either in its 2018 rulemaking or at any subsequent time.

Further, according to MITRE, the intent of its report was to:

> provide the FCC with recommendations on the viability of alternative speech recognition technologies for use in IP CTS environments from the user's perspective via usability feedback and comprehension scoring.[21]

---

[15] *Id.* at 5827-28, ¶ 49.

[16] *Id.* (noting that "[c]onsumer surveys conducted by MITRE also indicated that some consumers prefer captions generated using ASR over captions facilitated by CAs.").

[17] MITRE Phase 2 Summary at 12.

[18] *Id.*

[19] *Id.* (emphasis added).

[20] *Id.* at 15.

[21] *Id.* at 3.

Notably, MITRE did not set out to evaluate whether ASR was sufficiently accurate or reliable to be used as an exclusive means to provide IP CTS without the involvement of a CA. To the contrary, the 2016 report expressed MITRE's hope that its study would "assist in developing minimum specifications and requirements for a fully functional automated system."[22] Indeed, MITRE explicitly cautioned the Commission to "research the feasibility of using fully automated S[peech-]T[o-]T[ext] services in place of existing IP CTS services" before permitting the deployment of such services.[23]

In other words, the self-described purpose of MITRE's Phase 2 study was to "provide the FCC with recommendations on the *viability* of alternative speech recognition technologies for use in IP CTS environments" rather than to *affirm* or *certify* that such technologies were ready for prime time.[24] The Commission did not recognize these limitations in its 2018 rulemaking, however, as it relied almost exclusively on the 2016 MITRE report to justify its decision to authorize ASR-only IP CTS services.

B. MITRE's study has numerous shortcomings.

Reliance on the MITRE report to reach the conclusion that ASR-only captions would be sufficient to meet the functional equivalency standard was misplaced for other reasons, as outlined below.

MITRE's sample size was too small. MITRE's phase 2 study had a mere eleven participants, of whom four people identified as hard of hearing, four identified as having a hearing loss, and three identified as deaf.[25] While this insignificant sample size might have been sufficient to "provide the FCC with recommendations on the viability" of using

[22] *Id.* at 15.

[23] *Id.*

[24] *Id.* at 3 (emphasis added).

[25] *Id.* at 23.

ASR for IP CTS,[26] it was far too small to draw generalizable insights about the needs and preferences of the TRS user community as a whole.

MITRE's testing failed to reflect a range of real-life calling conditions. The eleven participants in MITRE's study evaluated the performance of seven transcription providers based on just *two test transcripts*, without any indication of the extent to which the contents of the transcripts were representative of the diversity of communications that are facilitated by IP CTS.[27]

One test transcript was of a conversation nominating a teacher to speak at a graduation ceremony (the "Ms. Jackson" transcript),[28] while the other charted the path of someone "calling a local credit union and transcribing the [Interactive Voice Response] paths for a specific task" (the "bank" transcript).[29] The two transcripts were rated at grades 6.4 and 3.2 respectively on the widely-used Flesch-Kincaid reading ease formula,[30] were written in Standard American English, and featured stilted conversations in formal settings.[31]

In studying such limited samples and simple conversational topics, MITRE did not even try to capture the range of real-life scenarios that commonly take place over IP CTS calls. Accordingly, its study did not and could not represent the diversity of English speakers in these United States. Nor could the two conversations that took place effectively capture the varied expressions that people typically convey on a phone call—for example, expressions of care, joy, anger, worry or surprise. Other real-world

---

[26] *Id.* at 3.

[27] *Id.* at 23.

[28] *Id.* at 20, 33-34.

[29] *Id.* at 21, 34.

[30] *Id.* at 17, 20-21 (noting that the Flesch-Kincaid reading ease formula refers to "[t]he grade level (based on the U.S. education system) at which a user can understand text").

[31] *Id.* at 33-34.

conditions, such as background noises and poor line connections, were also missing from these sample calls.

The ASR systems tested by MITRE were not designed for IP CTS calls. None of the ASR systems that were tested by MITRE had been designed for IP CTS purposes. During Phase 2, MITRE used the two transcripts to compare the performance of three ASR engines with four existing certified IP CTS providers that relied on CAs to generate captions.[32] But because the ASR systems MITRE decided to test were "not developed for use during telephone calls,"[33] MITRE had to re-configure these systems specifically for the purpose of conducting its analysis.[34] This improvisational approach to testing ASR raises further questions about the reliability and generalizability of the results MITRE obtained.

MITRE's testing did not reflect diversity in users' speech. There is no indication in the MITRE study that the two transcripts that were tested were voiced by speakers with non-standard English accents, dialects, or speech patterns. MITRE's 2016 report states that in Phase 2 of its study, it *only* tested transcripts that were voiced by native English speakers.[35] MITRE states that the "Ms. Jackson" transcript was read by a MITRE employee, but we have no additional information about this user's speech characteristics. Nor do we have complete information about the speech characteristics of the individuals using the "bank" transcript.

---

[32] *Id.* at ii.

[33] *Id.* at 4, 23 (stating that the STT-1 and STT-3 dictation engines that were tested "are dictation services and not intended for use during telephone calls").

[34] *Id.* at 23 ("CAMH engineered the MITRE Usability Lab, a technology testing lab maintained by MITRE, the CAMH FFRDC operator, to simulate how captions *may* appear *if* they were developed for such use.").

[35] MITRE chose two of the seven transcripts that it used in Phase 1 of its study to be tested in Phase 2 of its study. *Id.* at 6. The Phase 2 Report notes that, of the original seven transcripts, only one transcript "was executed with both a native English speaker and a non-native English speaker." *Id.* at 19. This was transcript #4, which wasn't chosen to be tested in Phase 2 of the study, despite MITRE's acknowledgement that, in Phase 1, "[f]or all providers and SSTs (sic.), the average accuracy for the non-native speaker sample was lower than for the native speaker." *Id.* at 19-20.

* * *

All in all, the findings of MITRE's phase 2 study as presented in its 2016 report fail to provide sufficient data on the performance of ASR across the full range of dialects, accents, speech patterns, and subject matters likely on IP CTS calls across the United States. Relying on the speech patterns of just two individuals and the responses of eleven IP CTS users to gauge the accuracy of ASR systems is problematic from a statistical perspective, as it can hardly be said to be representative of the full population of IP CTS users. Even MITRE "observed that faster speech, background noise, more complex speech, computer-generated voices, and non-native English speakers all have a negative impact on accuracy,"[36] a conclusion that the Commission seemingly failed to take into consideration. Ignoring these stated shortcomings, the Commission regrettably pressed forward in its 2018 ruling to approve ASR-only IP CTS providers.

III. Despite improvements, studies show that stand-alone ASR continues to present challenges for a substantial segment of the IP CTS user community.

Research conducted in the years since the 2018 Declaratory Rulemaking further demonstrates why the Commission erred in finding that ASR was a sufficiently mature technology to enable its use as a stand-alone alternative to human-assisted IP CTS. There is little dispute that over the past several years, ASR technology has continued to improve, and that average error rates associated with ASR use have been decreasing. However, studies conducted over the six years since the Commission approved stand-alone ASR for IP CTS continue to show that ASR's performance varies significantly based on who is speaking, where they are speaking from, and the overall conditions of their real-life calling scenarios. New data confirms that these errors are not distributed uniformly over the range of calls that are facilitated by IP CTS. Rather, some calls achieve

[36] *Id.* at 19.

excellent performance and accuracy from ASR, while others using ASR face such serious challenges as to render their conversations unintelligible.[37]

In April 2021, Hamilton Relay—a certified IP CTS provider—presented the Commission its findings on the accuracy of ASR based on data gathered from 5000 randomly-selected, real-world calls.[38] Subsequently, in November 2023, Hamilton Relay presented additional data for 62,000 real-life calls.[39] Hamilton Relay's results show that the technology achieved an overall average accuracy rate of 90% in 2021, increasing to just under 94% as of last fall.[40] However, these average accuracy figures obscure the highly uneven distribution of errors across different types of calling scenarios.[41] The charts below, which contain data from each of Hamilton Relay's studies, illustrate the problem in graphic detail.[42] Each shows the word error rate (WER) of ASR divided by quartile.[43]

---

[37] Korbinian Kuhn et al., *Measuring the Accuracy of Automatic Speech Recognition Solutions*, 16 ACM Transactions on Accessible Computing 1 (2024), https://dl.acm.org/doi/10.1145/3636513.

[38] *Ex Parte* Filing of Hamilton Relay, Inc., CG Docket Nos. 03-123, 13-24, at 1 (filed Apr. 9, 2021) ("Hamilton 2021 *Ex Parte*"), https://www.fcc.gov/ecfs/document/104091240915701/1.

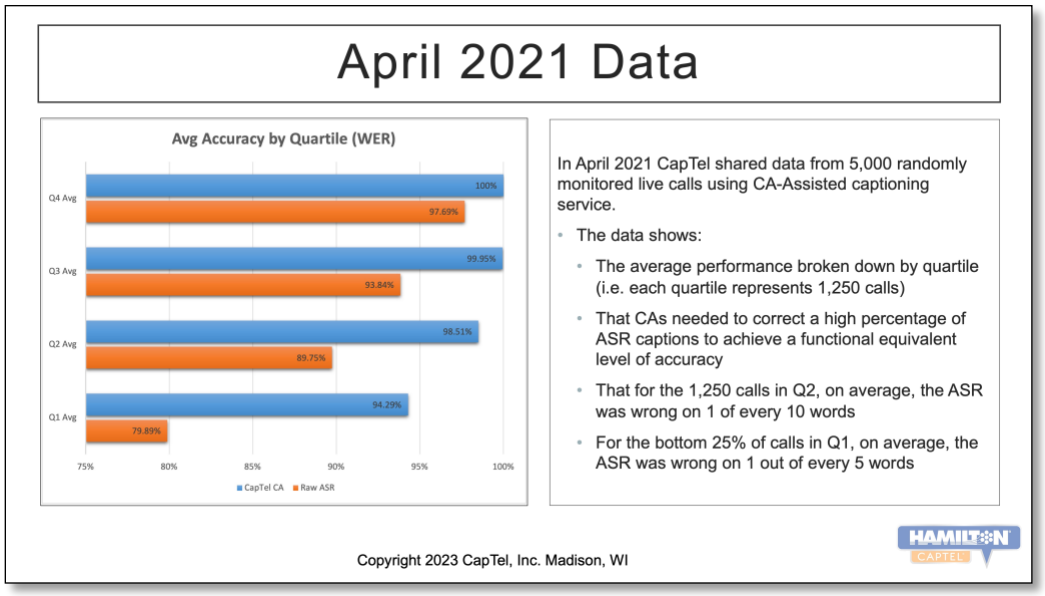
[39] *Ex Parte* Filing of Hamilton Relay, Inc., CG Docket Nos. 03-123, 13-24, at 8-10 (filed Apr. 17, 2024) ("Hamilton 2024 *Ex Parte*"), https://www.fcc.gov/ecfs/document/10417533701363/1.

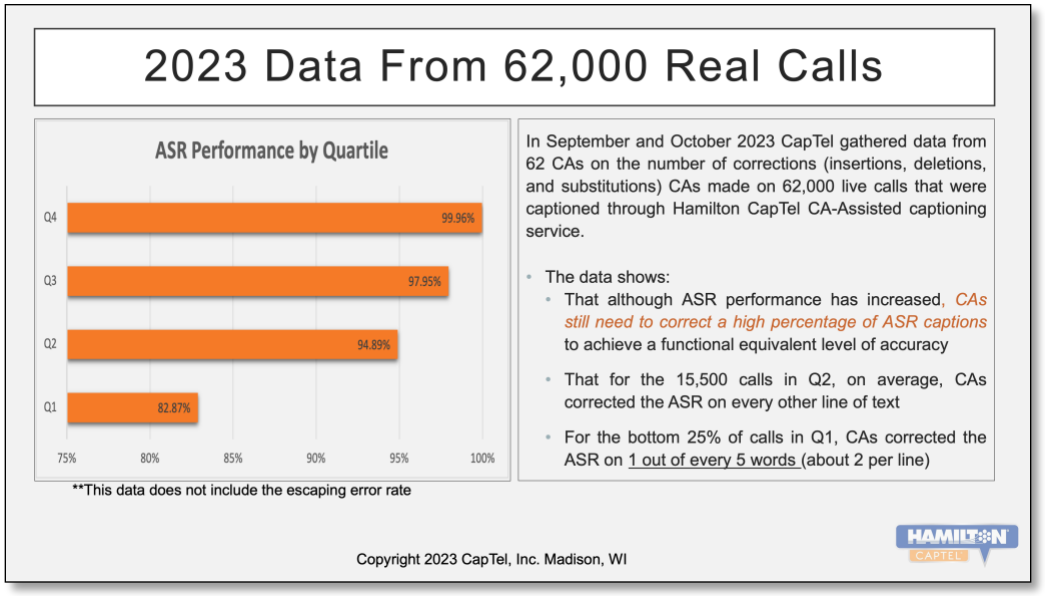
[40] Hamilton 2021 *Ex Parte*, at 13; Hamilton 2024 *Ex Parte*, at 8-9.

[41] Hamilton 2021 *Ex Parte*, at 17; Hamilton 2024 *Ex Parte*, at 8-9.

[42] These charts are reproduced from Hamilton's 2023 submission to the Commission. *See* Hamilton 2024 *Ex Parte*, at 7-9.

[43] Hamilton Relay was able to obtain this data because its CA-assisted captioning service first uses ASR to generate a transcript, after which the CA makes any necessary corrections to these machine-generated transcriptions. The error rates therefore reflect the number of corrections that Hamilton's CAs need to make to the raw output of its ASR system. It should be noted that because some errors do escape notice, the actual ASR error rates are likely slightly higher than indicated.





For calls in the top (fourth) quartile, ASR is performing very well. In 2021, the technology had an accuracy of 97.69%, which increased to an astonishing 99.96% in 2023. For calls in the bottom quartile, however, the story is very different. In 2021, ASR had an accuracy of just 79.89%, which increased to only 82.87% by the fall of 2023.

The table below summarizes the 2021 and 2023 accuracy data on a quartile-by-quartile basis. It also expresses the percentage accuracy by quartile in terms of a ratio of one in how many words can be expected to contain an error:

Quartile	**2021 Accuracy Rate (Ratio)**	**2023 Accuracy Rate (Ratio)**
Fourth (Top 25%)	97.69% (1 in 43)	99.96% (1 in 2500)
Third	93.84% (1 in 16)	97.95% (1 in 49)
Second	88.75% (1 in 9)	94.89% (1 in 20)
First (Bottom 25%)	79.89% (1 in 5)	82.87% (1 in 6)
Average (mean)	90.04% (1 in 10)	93.92% (1 in 16)

The above table reveals that ASR's word error rate as of last fall was only one error in 2500 words for calls in the top quartile, but nearly one error in every six words for calls in the bottom quartile. By way of illustration, if petitioners were in the top quartile and had used ASR to generate this petition from speech, we would expect just three errors over the body of this text. But if they were in the bottom quartile, one of every six words would be incorrectly transcribed. To understand just how much a one-in-six word error rate can garble communications, consider the following well-known passage:

> When in the coarse of human events, it becomes unnecessary four one people to resolve the political bands which have connected them with another, and assume Hmong the powers of the earth, the separate unequal station to witch the Laws of Nature and of Nature's God and title them, a decent respect to the piñons of mankind requires they should declare the casas which impel them to separation.[44]

What explains the massive discrepancy between the performance of ASR in the top and bottom quartiles? Studies conducted since 2018 show that the accuracy of ASR can be extremely high when such systems are tested under ideal conditions, such as when

[44] The passage is, of course, the first sentence of the Declaration of Independence with common speech recognition errors introduced every one in six words. *See* The Declaration of Independence, ¶ 1 (U.S. 1776), https://www.archives.gov/founding-docs/declaration-transcript.

they are used to transcribe the speech of white males speaking standard American English in a quiet environment.[45] When conditions diverge from this ideal, however, the accuracy of ASR suffers greatly; specifically, studies show that ASR's capabilities vary considerably based on the dialect, accent, speech of the speaker, and the amount of background noise present.

Like other systems based on machine learning, ASR engines utilize models that are trained on large collections ("corpora") of spoken data to analyze speech and predict a sequence of sounds, words, and phrasal structures based on the given speaker's signal.[46] Typically, these datasets utilize "standardized" speech, lacking representation of speech characterized by less standard dialects, regional or non-native and accents, and speech affected by disabilities.[47] Correspondingly, as discussed below, the performance of ASR systems in the real world varies based on the data that their models are trained upon.[48]

---

[45] *See generally* Constantin Spille et al., *Comparing Human and Automatic Speech Recognition in Simple and Complex Acoustic Scenes*, 52 Comput. Speech & Language 123 (2018), https://doi.org/10.1016/j.csl.2018.04.003 (discussing ASR performance in complex, noisy, and multi-speaker situations); Mikel K. Ngueajio & Gloria Washington, *Hey ASR System! Why Aren't You More Inclusive?*, *in* HCI International 2022 – Late Breaking Papers: Interacting with eXtended Reality and Artificial Intelligence 421, 423 (Jessie Y.C. Chen et al. eds., 2022), https://doi.org/10.1007/978-3-031-21707-4_30.

[46] Joshua L Martin & Kelly Elizabeth Wright, *Bias in Automatic Speech Recognition: The Case of African American Language*, 44 Applied Linguistics 613, 614 (Aug. 2023), http://dx.doi.org/10.1093/applin/amac066.

[47] Siyuan Feng, Bence Mark Halpern, Olya Kudinga & Odette Scharenborg, *Towards inclusive automatic speech recognition*, (last visited May 29, 2024), https://www.sciencedirect.com/science/article/pii/S0885230823000864 (noting that "ASR systems are typically trained on speech from native speakers of a 'standard' variant of that language, inadvertently discriminating not only the speech of non-native speakers but also that of speakers of regional or sociolinguistic variants of the language."(citations omitted). This article also noted that ASR finds the speech of children to be more challenging than adult speech "due to children's shorter vocal tracts, slower and more variable speaking rate and inaccurate articulation."

[48] *Id.*

A. Dialects

ASR systems are often trained on datasets that prioritize standard or mainstream dialects. The corollary is that such systems tend to perform poorly in recognizing the speech of individuals with non-standard or less common dialects.[49]

Dialect refers to a specific form of pronunciation, vocabulary, and grammar making up a speech pattern that is specific to a particular region or group of people.[50] American English, African American Vernacular English, Indian English, and Australian English are among the many widely spoken dialects of the English language.[51] Most ASR systems are trained primarily on corpora made up of speech samples of speakers of American English, however.[52] This impacts the accuracy of these systems in recognizing the speech of individuals who speak other English dialects.

Consider the performance of ASR systems in recognizing the speech of speakers of African American Vernacular English (AAVE), which is spoken by an estimated 80% of African Americans in the United States.[53] The accuracy of ASR systems in recognizing the speech of AAVE speakers is far lower than for speakers of Standard American English.[54] As a consequence, even for identical utterances, the word error rate for speakers who are black is *twice* as high as for speakers who are white.[55] And study after study has shown that even the newest, most advanced ASR systems exhibit far higher error rates when

---

[49] *See* Rachel Dorn, *Dialect-Specific Models for Automatic Speech Recognition of African American Vernacular English* 1 (2019), https://doi.org/10.26615/issn.2603-2821.2019_003.

[50] Mathew Jones, *Dialect vs. Accent: Definitions, Similarities, & Differences*, Magoosh (last visited Apr. 13, 2024), https://magoosh.com/english-speaking/dialect-vs-accent-differences-and-examples/.

[51] *See generally* Melanie Röthlisberger & Benedikt Szmrecsanyi, *Dialect Typology: Recent Advances*, *in* Handbook of the Changing World Language Map 131 (Stanley D. Brunn & Roland Kehrein eds., 2020) (describing the various modern English dialects and their geographical distributions).

[52] *See* Dorn, *supra* note 49, at 19.

[53] *Id*. at 1.

[54] Ngueajio & Washington, *supra* note 45, at 423.

[55] *Id.*

processing the speech of black and minority-community speakers as opposed to white speakers of Standard American English.[56]

In 2022, the Commission acknowledged this algorithmic bias. Citing a *New York Times* article reporting on research findings that showed ASR-only systems misidentified words spoken by black individuals at a substantially higher rate than words spoken by white people,[57] the Commission noted that such bias could affect the accuracy with which ASR transcribes voices belonging to members of minority communities. Similarly, in a separate statement to the 2022 IP CTS Order, FCC Commissioner Starks raised concerns about such bias, questioning the extent to which IP CTS captions generated on a fully automated basis could meet the functional equivalence standard.[58] Regrettably, the Commission did not act on these concerns to ensure a CA option when ASR failed. By failing to so address the discrepancies in accuracy across these different types of speakers, the Commission runs the risk of permanently institutionalizing algorithmic bias into its TRS program.

B. Accents

ASR systems also struggle with recognizing the speech of individuals who speak with non-standard accents. Accents are a style of pronunciation that may be specific to a

---

[56] *Id* at 425-26.

[57] *Internet Protocol Captioned Telephone Service Compensation; Misuse of Internet Protocol (IP) Captioned Telephone Service*; *Telecommunications Relay Services and Speech-to- Speech Services for Individuals with Hearing and Speech Disabilities*, CG Docket Nos. 22-408, 13-24 and 03-123, Notice of Proposed Rulemaking and Order on Reconsideration, 37 FCC Rcd 15243, 15248-49, ¶ 16 (2022) (hereinafter *2022 IP CTS Notice on Rates)*, https://docs.fcc.gov/public/attachments/FCC-22-97A1.pdf (citing Cade Metz, *There Is a Racial Divide in Speech-Recognition Systems, Researchers Say,* N.Y. Times (Mar. 23, 2020), https://www.nytimes.com/2020/03/23/technology/speech-recognition-bias-apple-amazon-google.html.

[58] Specifically, in discussing the possible need for separate rates for IP CTS provided via CAs versus fully automated ASR, FCC Commissioner Geoffrey Starks wrote that he "remain[s] concerned about potential algorithmic bias in ASR" as "studies have shown that speech recognition systems make far more errors when transcribing the speech of people of color than of their white counterparts." *2022 IP CTS Notice on Rates*, Statement of Commissioner Geoffrey Starks, 37 FCC Rcd at 15273.

country, region, or social class.[59] Speakers of American English (a dialect) may speak with the "General American English" accent (an umbrella term for the accent used by most speakers of English in the U.S)[60] or they may well have a German accent, an Arabic accent, or a Tagalog accent—based on the impact of the speaker's first language on their pronunciation of English.

ASR word error rates vary considerably based on a speaker's accent, even if they are speaking a mainstream dialect of English.[61] For example, the ASR systems that underlie Google and Amazon's smart speakers—the systems the Commission admired in its 2018 rulemaking for having word error rates as low as 4.9%[62]—perform far worse when attempting to recognize the speech of individuals with accents other than General American English. Testing conducted by the *Washington Post* in 2018 showed that individuals with a Midwestern or Southern accent were only 2% and 3% (respectively) less likely to receive accurate responses from their smart devices than individuals with a General American accent; however, this likelihood rose to 30% for individuals with "non-native" English accents.[63]

Even when ASR systems accurately transcribe the speech of individuals with non-standard accents, such individuals are often forced to repeat themselves before their speech is accurately transcribed.[64] In fact, the average number of attempts that individuals with non-standard accents must make is nearly double that of individuals who have a standard accent, demonstrating that low word error rates can sometimes

---

[59] Jones, *supra* note 50.

[60] *See generally* Cynthia G. Clopper et al., *Perceptual Similarity of Regional Dialects of American English*, 119 J. Acoustical Soc. Am. 566, 556-57 (2006), https://doi.org/10.1121/1.2141171.

[61] Ngueajio & Washington, *supra* note 45, at 425-26.

[62] *2018 IP CTS Declaratory Ruling*, 33 FCC Rcd at 5828-29, ¶ 51.

[63] Drew Harwell, *The Accent Gap*, Wash. Post (July 19, 2018), https://www.washingtonpost.com/graphics/2018/business/alexa-does-not-understand-your-accent/.

[64] Ngueajio & Washington, *supra* note 45, at 426.

provide a misleading measure of the true accuracy of ASR.[65] These discrepancies can take a toll on both IP CTS users and the persons with whom they communicate who speak English with non-standard accents.[66] For example, a recent study of speakers of AAVE found that 90% of them felt angry, self-conscious, and dissatisfied because ASR systems forced them to strain or repeat their speech multiple times in order to work.[67]

C. Noise

Studies also show the significant impact that noise can have on the accuracy of ASR. Environmental noise—from the din of traffic to the cries of a baby—can disrupt speech signals in speech processed by an ASR system and cause inaccuracies in transcription output.[68] Research is underway to improve the performance of ASR in noisy conditions using a range of techniques, from improving "speech-to-noise" ratios in the inputs that are fed into ASR systems[69] to using novel computational techniques to filter noise from speech.[70] For the time being, however, noise continues to be a problem that impacts the accuracy of ASR systems.

Aggravating this situation is that—as holds true for various other environmental factors—the disruptive effects caused by excessive noise fall disproportionally on those in lower income American communities. Studies show that environmental noise in our country is positively correlated with locations that house increased proportions of people of color, and are often determined by an area's socio-economic status, as measured by

---

[65] *Id.*

[66] *Id* at 424.

[67] *Id.*

[68] Sheng-Chieh Lee et al., *Threshold-Based Noise Detection and Reduction for Automatic Speech Recognition System in Human-Robot Interactions*, Sensors, June 28, 2018, at 1, https://doi.org/10.3390/s18072068.

[69] *Id* at 2.

[70] Patrick Eickhoff et al., *Bring the Noise: Introducing Noise Robustness to Pretrained Automatic Speech Recognition*, *in* Artificial Neural Networks & Mach. Learning – ICANN 2023 376, 377 (Lazaros Iliadis et al. eds., 2023), https://doi.org/10.48550/arXiv.2309.02145.

factors that include poverty rates, the rate of high school completion, and the proportion of renters versus homeowners.[71] This is in part because such communities are more likely to be in proximity to high-traffic roads and industrial facilities, both of which produce noisy environments.[72]

D. Non-standard Speech

Non-standard speech can also impact the effectiveness of ASR systems. Dysarthria,[73] speech slurring, stuttering, and a range of dysfluencies (such as repetitions, interjections, and revisions of speech) can all affect the accuracy of ASR, whether these arise from congenital factors (such as cleft palate or cerebral palsy) or from diseases and injuries (such as brain traumas or injuries to the mouth and throat).[74]

Research has shown that the accuracy and functionality of ASR diminishes as an individual's speech diverges from standard speech patterns based on a range of characteristics associated with these and other speech disabilities.[75] In addition, the accuracy of ASR is impacted by changes in the rate of speech, with higher rates of speech

---

[71] Joan A. Casey et al., *Race/Ethnicity, Socioeconomic Status, Residential Segregation, and Spatial Variation in Noise Exposure in the Contiguous United States*, Env't. Health Persps., July 25, 2017, at 2, https://doi.org/10.1289/EHP898.

[72] *Id* at 7.

[73] Dysarthria "refers to a group of neurogenic speech disorders characterized by 'abnormalities in the strength, speed, range, steadiness, tone, or accuracy of movements required for breathing, phonatory, resonatory, articulatory, or prosodic aspects of speech production.'" *Dysarthria in Adults*, Am. Speech-Language-Hearing Ass'n, https://www.asha.org/practice-portal/clinical-topics/dysarthria-in-adults/ (last visited May 27, 2024).

[74] *Speech and Language Disorders*, Penn Med. (Feb. 24, 2022), https://www.pennmedicine.org/for-patients-and-visitors/patient-information/conditions-treated-a-to-z/speech-and-language-disorders.

[75] This study looked at increased abnormalities in dysarthria severity, nasality, vocal quality, articulatory precision, and prosody (the patterns of stress, rhythm, and rise/fall of the voice in speech).
Ming Tu et. al., *The Relationship Between Perceptual Disturbances in Dysarthric Speech and Automatic Speech Recognition Performance*, 140 J. Acoustical Soc. Am. 416, 418-19 (2016), https://doi.org/10.1121/1.4967208.

being associated with lower accuracy.[76] The same is true of the impact on ASR accuracy of casual speech slurring (defined as a "reduction of pronunciation of certain phonemes, or syllables") and disfluencies such as false starts, repetitions, hesitations, and filled pauses.[77]

Nearly 18 million American adults have experienced a speech disability or non-standard speech in the last 12 months,[78] and nearly one in 14 American children between the ages of 3 and 17 have a developmental language disorder.[79] When an ASR engine fails to accurately transcribe the speech of these individuals, IP CTS users need the option of switching to a CA who can better facilitate the communication on the call.

IV. The Commission's failure to establish performance goals and metrics for IP CTS has impeded its ability to ensure the functional equivalence of such services.

The considerable discrepancy in ASR performance between the top and bottom quartiles of IP CTS calls, as described in the prior section, highlights the importance of having effective performance goals and metrics to evaluate the accuracy of IP CTS by providers who rely solely on ASR. When the Commission first approved fully automated ASR as a permissible means of generating IP CTS captions in 2018, it promised that it would require applicants for IP CTS certification to prove that their ASR service quality matches that of CA-assisted IP CTS "with respect to captioning transcription delays, accuracy, speed, and readability."[80] But this commitment lacked substance then and continues to lack substance now, as the Commission has yet to establish metrics to assess

---

[76] M. Benzeghiba et al., *Automatic Speech Recognition and Speech Variability: A Review*, 49 Speech Commc'n 763, 766-67 (2007), https://doi.org/10.1016/j.specom.2007.02.006.

[77] *Id*. at 766.

[78] *Quick Statistics About Voice, Speech, Language*, Nat'l Inst. on Deafness and Other Commc'n Disorders (Mar. 4, 2024), https://www.nidcd.nih.gov/health/statistics/quick-statistics-voice-speech-language.

[79] *Id*.

[80] *2018 IP CTS Declaratory Ruling*, 33 FCC Rcd at 5834, ¶ 63.

the accuracy or readability of transcriptions for any form of IP CTS—whether provided by CAs or ASR.

The Commission first sought comment on IP CTS performance goals and metrics in a Notice of Inquiry (NOI) that accompanied its 2018 Declaratory Ruling approving ASR.[81] These included proposals to measure IP CTS based on: (1) transcription accuracy; (2) transcription synchronicity (i.e., caption delay); (3) transcription speed; (4) speed of answer; (5) dropped or disconnected calls; (6) service outages; and (7) usage data.[82] In its 2020 IP CTS Compensation Order, the Commission seemed to be taking the next step to adopt these measures with issuance of a Further Notice of Proposed Rulemaking (FNPRM) that sought comment on integrating specific IP CTS metrics into the Commission's mandatory minimum TRS standards.[83] In 2022, the Commission again touted the importance of adopting "measures and metrics that would allow more precise assessment of IP CTS service quality, including compliance with minimum TRS standards" and for the purpose of "assessing how well each provider and captioning approach performs in meeting the objectives of section 225."[84] To date, however, the Commission still has not moved forward in meeting its commitment to evaluate provider performance using these measures.

Absent performance goals and metrics for achieving functionally equivalent communications, the Commission simply cannot determine the extent to which its regulations and the IP CTS providers that it certifies are fulfilling the purpose and goals of the TRS program. The importance of performance goals and metrics for the TRS

---

[81] *2018 IP CTS Notice of Inquiry*, 33 FCC Rcd at 5868-74, ¶¶ 155-175.

[82] *Id.* at 5870-71, ¶ 164.

[83] *Misuse of Internet Protocol (IP) Captioned Telephone Service; Telecommunications Relay Services and Speech-to-Speech Services for Individuals with Hearing and Speech Disabilities; Structure and Practices of the Video Relay Service Program*, CG Docket Nos. 13-24, 03-123, and 10-51, Report and Order, Order on Reconsideration, and Further Notice of Proposed Rulemaking, 35 FCC Rcd 10866, 10897-907, ¶¶ 62-92 (2020) (seeking to specify the minimum TRS standards with respect to caption delay and accuracy).

[84] *2022 IP CTS Notice on Rates*, 37 FCC Rcd at 15249-50, ¶ 17.

program has been well documented over the years. In 2015, following up on a Congressional mandate to assess the TRS program as a whole, the Government Accountability Office (GAO) called on the Commission to establish performance goals and tailor performance measures around them.[85] In preparing its report, GAO spoke with several Commission officials who all acknowledged the overriding goal of TRS as being to provide functionally equivalent telecommunications.[86] However, GAO found that the Commission had not established any *performance goals*, or clear definitions of the meaning of functionally equivalent telecommunications, which GAO said was needed to guide the Commission's effort in achieving this purpose.[87] GAO further explained that individual metrics were needed to assess whether the Commission was achieving these goals.

Then-Commissioner (now Chair) Rosenworcel similarly expressed concern over the lack of performance goals and metrics, when, in her separate statement on the 2018 Declaratory Ruling, she stated that the Commission had "put[] the cart before the horse by introducing automatic speech recognition into the IP CTS program" before it had addressed its "most basic regulatory responsibilities" in determining the standard of quality a hard of hearing user can expect.[88]

What is especially disturbing is that the Commission's failure to issue performance goals and metrics has not stopped it from making unsubstantiated claims about the

---

[85] U.S. Gov't Accountability Off., GAO-15-409, *Telecommunications Relay Service: FCC Should Strengthen Its Management of Program to Assist Persons with Hearing or Speech Disabilities* at 17 (2015), https://www.gao.gov/assets/gao-15-409.pdf.

[86] *Id.*

[87] *Id.* (noting a lack of performance goals tied to "clearly define[d] desired program outcomes").

[88] *2018 IP CTS Declaratory Ruling*, Concurring Statement of Commissioner Jessica Rosenworcel, 33 FCC Rcd 5800, 5900 (2018).

merits of stand-alone ASR,[89] nor from approving providers that rely solely on ASR to generate IP CTS captions. In the six years following the 2018 Declaratory Rulemaking, the Commission has certified six new IP CTS providers that exclusively rely on ASR to provide IP CTS.[90] In response to the Commission's request for comments on each of these provider's certification applications, consumer groups consistently have raised concerns that, without technology-neutral quality metrics, the Commission cannot ensure that a particular service has the requisite level of quality to provide functional equivalence to all IP CTS users.[91] These organizations have been steadfast in urging the Commission to make good on its long-standing promise to develop performance goals and metrics for IP

[89] For example, in its 2022 IP CTS Notice on Rates, the Commission reiterated its claims from 2018 that ASR-only captioning offers better speed of answer, lower caption delay, and *a level of accuracy that is generally comparable to that of CA-assisted captioning*, notwithstanding, as shown above, data showing that this holds true only for certain demographics. *2022 IP CTS Notice on Rates*, 37 FCC Rcd at 15248-49, ¶ 16 (emphasis added).

[90] *See Telecommunications Relay Services and Speech-to-Speech Services for Individuals with Hearing and Speech Disabilities*, CG Docket No. 03-123, Memorandum Opinion and Order, DA 20-485 (May 5, 2020), https://docs.fcc.gov/public/attachments/DA-20-485A1_Rcd.pdf (MachineGenius Certification Order); *Telecommunications Relay Services and Speech-to-Speech Services for Individuals with Hearing and Speech Disabilities*, CG Docket No. 03-123, Memorandum Opinion and Order, DA 20-587 (June 4, 2020), https://docs.fcc.gov/public/attachments/DA-20-587A1.pdf (Clarity Certification Order); *Telecommunications Relay Services and Speech-to-Speech Services for Individuals with Hearing and Speech Disabilities*, CG Docket No. 03-123, Order, DA 22-442 (Jan. 4, 2024), https://docs.fcc.gov/public/attachments/DA-24-11A1.pdf (Global Caption Certification Order); *Telecommunications Relay Services and Speech-to-Speech Services for Individuals with Hearing and Speech Disabilities*, CG Docket No. 03-123, Order, DA 24-12 (Jan. 4, 2024), https://docs.fcc.gov/public/attachments/DA-24-12A1.pdf (Nagish Certification Order); *Telecommunications Relay Services and Speech-to-Speech Services for Individuals with Hearing and Speech Disabilities*, CG Docket No. 03-123, Order, DA 24-48 (Jan. 17, 2024), https://docs.fcc.gov/public/attachments/DA-24-48A1.pdf (Rogervoice Certification Order); *Telecommunications Relay Services and Speech-to-Speech Services for Individuals with Hearing and Speech Disabilities*, CG Docket No. 03-123, Order, DA 24-49 (Jan. 17, 2024), https://docs.fcc.gov/public/attachments/DA-24-49A1.pdf (NexTalk Certification Order).

[91] *See, e.g.,* Accessibility Advocacy and Research Organizations March 2021 *Ex Parte*; Comments of Accessibility Advocacy and Research Organizations on InnoCaption Application at 1-3.

CTS.[92] They also have called upon the Commission to require IP CTS providers to be more transparent (in their certification applications) regarding their ability to provide functionally equivalent services.[93]

Of particular concern to the undersigned organizations is the Commission's recent decision to certify an IP CTS provider that will rely exclusively on ASR in the closed environment of carceral facilities.[94] While acknowledging the ongoing absence of standards by which to assess the accuracy and readability of this provider's ASR,[95] the Commission concluded that its service was capable of providing "verbatim" transcriptions with "competent" typing, spelling, and grammar pursuant to its existing TRS rules.[96]

The approval of a provider under these circumstances is especially disconcerting given that logic and common sense would suggest that ASR-only IP CTS systems will perform especially poorly in a carceral context. Unfortunately, the American system of criminal justice faces well-known and extremely troubling racial and ethnic disparities in incarceration rates.[97] At the same time, as the Commission is fully aware, the research

---

[92] *See e.g.*, Letter from Blake Reid, Counsel to Telecommunications for the Deaf and Hard of Hearing, Inc. (TDI), to Marlene H. Dortch, Secretary, FCC, CG Docket Nos. 22-408 et al. (filed Dec. 14, 2022), https://www.fcc.gov/ecfs/search/search-filings/filing/1215058477374; Letter from Blake Reid, Counsel to Telecommunications for the Deaf and Hard of Hearing, Inc. (TDI), to Marlene H. Dortch, Secretary, FCC CG Docket Nos. 03-123 et al. (filed May 18, 2022), https://www.fcc.gov/ecfs/search/search-filings/filing/10518720203024.

[93] *See e.g.*, Comments of Accessibility Advocacy and Research Organizations on InnoCaption Application; Comments of Accessibility Advocacy and Research Organizations on the Application of Global Caption, Inc., for Certification as a Provider of Internet Protocol Captioned Telephone Service, CG Docket 03-123, at 1-2 (May 9, 2022), https://www.fcc.gov/ecfs/search/search-filings/filing/10509013691258.

[94] On January 4, 2024, the Commission provisionally certified Global Caption, Inc to use ASR to provide IP CTS in carceral facilities on a fully automatic basis. Global Caption Certification Order.

[95] *Id.* at ¶ 12.

[96] *Id.* Although one test report provided by Global Caption indicated that its ASR service tested better than industry average, merely being better than average does not demonstrate whether a service is providing a level of quality necessary to achieve functional equivalence. *Id.* at ¶ 13.

[97] *See* Wendy Sawyer, *Visualizing the Racial Disparities in Mass Incarceration*, Prison Pol'y Initiative (July 27, 2020), https://www.prisonpolicy.org/blog/2020/07/27/disparities (detailing the systematic racism

noted above shows that algorithmic bias compromises the ability of ASR to accurately transcribe the speech of people of color[98]— a bias that creates a racial divide for IP CTS users.[99] This means that when such individuals call friends and family who are also members of communities whose speech diverges from Standard American English, there will be an increased likelihood that ASR transcriptions will perform poorly for them. Yet these individuals will have no way of switching to a CA when this occurs. And because there is no FCC requirement for carceral facilities to offer a choice of IP CTS providers to incarcerated individuals, those residing in facilities that contract with ASR-only providers will not have the option to switch to another IP CTS provider when ASR fails them.

The gravity of this situation is heightened by the significant noise levels commonly present in carceral facilities. In noisy environments, IP CTS users typically rely on both their residual hearing and the provided captions to understand the conversation. However, when background noise interferes with their ability to hear, they must rely solely on the captions. The need for transcription accuracy under such noisy conditions—especially on telephone conversations with legal counsel, where precision is essential to fully understand and plan one's legal options and defense—underscores the inadequacy of ASR-only solutions when CAs are unavailable as backup. It is especially troubling, therefore, that notwithstanding these deficiencies, the Commission moved forward with

---

evident in the criminal justice system demonstrated by the vast disparity of people of color in prisons, jails, and mass incarceration).

[98] *2022 IP CTS Notice on Rates*, 37 FCC Rcd at 15248-49, ¶ 16; *see also* Allison Koenecke et al., *Racial Disparities in Automated Speech Recognition*, 117 Proc. Nat'l Acad. Sci. 7684, 7685 (2020), https://www.pnas.org/doi/10.1073/pnas.1915768117 (studying "state of-the-art ASR systems" developed by five major tech companies and finding an average word error rate of 35% for black speakers compared to 19% for white speakers); *supra* text accompany notes 55-56.

[99] *2022 IP CTS Notice on Rates*, 37 FCC Rcd at 15245, ¶ 7 (citing Cade Metz, *There Is a Racial Divide in Speech-Recognition Systems, Researchers Say,* N.Y. TIMES (Mar. 23, 2020), https://www.nytimes.com/2020/03/23/technology/speech-recognition-bias-apple-amazon-google.html.

approving certification for a stand-alone ASR provider in this context, one in which it should have known that the shortcomings of ASR systems would be at their greatest.

V. Conclusion

Depending on speakers and call conditions, data continues to reveal an alarmingly high error rate on IP CTS calls when stand-alone ASR technology is used to generate captions in various call scenarios. When one in every six words is incorrect, users struggle not only with nuances and emotional context, but with communications that can be entirely incomprehensible. These shortcomings disproportionately affect calls that IP CTS users make to or from people of color and those from lower socioeconomic backgrounds, often due to specific speech patterns or background noises that worsen ASR performance. This outcome is particularly acute in carceral environments, where demographic factors and loud background noises combine to severely undermine ASR transcription accuracy. The more that the Commission fails to acknowledge these limitations in ASR—largely the product of algorithms that have been trained on datasets that have disregarded people with speech disabilities and individuals with certain dialects or accents—the greater the risk of baking into the IP CTS program institutional bias that can will frustrate the principal goal of the TRS program to ensure telephone access for all.

Under the functional equivalence standard of Section 225 of the Communications Act, the Commission has an obligation to ensure that IP CTS users have telephone communications that are functionally equivalent to voice telephone users, regardless of who they call or the method used to generate captions when they place their calls. To achieve this, we call upon the Commission to:

(1) initiate a new notice-and-comment rulemaking that will require IP CTS providers using ASR for the generation of captions to give users the option to select a CA at the start of a call and to switch to a CA during when ASR fails;

(2) refrain from certifying additional ASR-only IP CTS providers until the above rulemaking is completed; and

(3) act expeditiously to complete its work on developing technology-neutral performance goals and metrics for IP CTS.

Respectfully Submitted,

/s/

Vivek Krishnamurthy, *Director*
vivek.krishnamurthy@colorado.edu

Sebastian Blitt, Madeline Finlayson, Sarah Misché, Kevin Nguyen, & Sophie Pickering, *Student Attorneys*

Samuelson-Glushko Technology Law & Policy Clinic (TLPC) at Colorado Law

Counsel to TDIforAccess, Inc.

May 31, 2024

On behalf of:

TDIforAccess, Inc. (TDI)
AnnMarie Killian—amkillian@tdiforaccess.org
Wilmington, DE
https://TDIforAccess.org

National Association of the Deaf (NAD)
Law and Advocacy Center
Zainab Alkebsi, Policy Counsel—zainab.alkebsi@nad.org
Silver Spring, MD
https://www.nad.org

Hearing Loss Association of America (HLAA)
Barbara Kelley, Executive Director—bkelley@hearingloss.org
Neil Snyder, Director of Public Policy—nsnyder@hearingloss.org
Rockville, MD
https://www.hearingloss.org