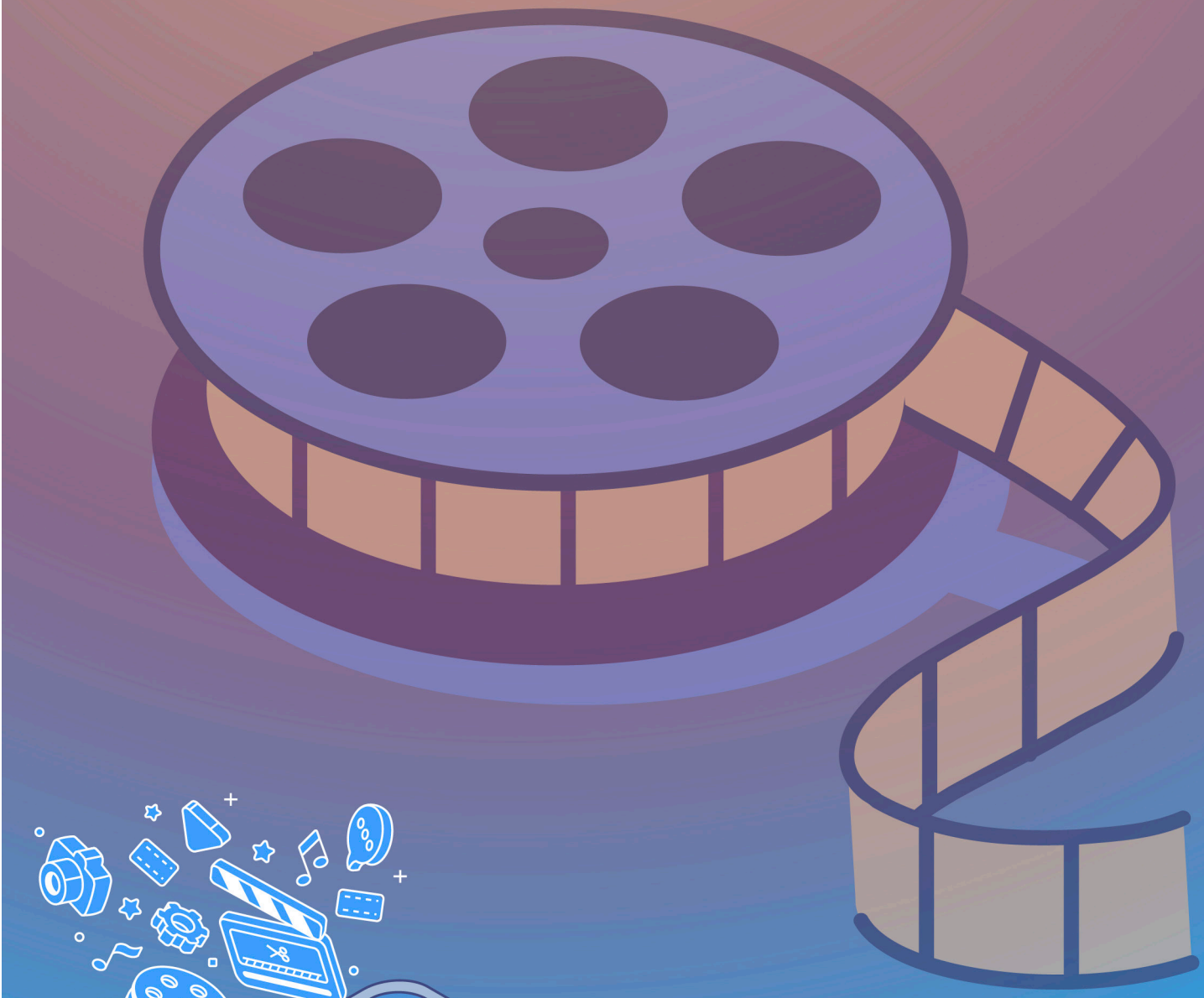
American Journal of Media, Film, Performing Arts and Music (AJMFPAM)

VOLUME 1 ISSUE 1 (2026)





PUBLISHED BY
E-PALLI PUBLISHERS, DELAWARE, USA

American Journal of Media, Film, Performing Arts and Music (AJMFPAM)
Volume 1 Issue 1, Year 2026
https://journals.e-palli.com/home/index.php/ajmfpam

When a Voice Crosses Borders: AI Dubbing, Accent Erasure and Linguistic Justice in Multilingual Screen Cultures

Munish Kumar Thakur[1*]

Article Information

Received: May 10, 2026

Accepted: November 02, 2026

Published: November 23, 2026

Keywords

Accent Discrimination, Audiovisual Translation, Digital Performance, Linguistic Justice, Synthetic Dubbing

ABSTRACT

New opportunities are opened by artificial intelligence tools like automated speech recognition, machine translation and voice cloning and synchronized dubbing. That enables the potential of reasonably low-cost access to a range of options in multiple languages for film and streaming, but it is unlikely that the constraints of linguistic intelligibility, fluency and lip synchronization are able to describe the cultural implications. This is an examination of a particular issue: the unintended change of accent, characterization and authorship of a person's speech due to discreet translation of their speech. It adopts an integrative analysis of a variety of selected and peer-reviewed research papers from the domains of audiovisual translation, sociolinguistics, speech technology and media ethics in order to create a framework of vocal justice, which includes intelligibility, indexical fidelity, performative integrity, consent, attribution and redress. The paper analyzes five related tensions that exist across linguistic access and accent standardization, natural-sounding speech output and expressive changes, relations with production and differences in language resources, technical voice and audience consent and perspectives on ease of viewing, disclosure and audience selection. It attempts to remind viewers that dubbing is an ethical endeavor, isn't solely defined by the correct dialogue, and provides a set of tools and indicators that can help producers, teams and researchers evaluate a dub operationally. The content of the findings is not in fact gathered from the viewers or productions from new sources, but it is a conceptual content. The article looks at the history of the hierarchy of accents and localization as it has been known so far and proposes a linguistically inclined approach to responsible intercultural screen translation when using synthetic dubbing.

INTRODUCTION

Language technology is increasingly influencing the relationship between the film itself and the people who can enjoy it. A serial that is produced in a language other than the viewers' native language could be broadcast in a medium which is distinctly dissimilar to the viewers' language. In historical times, this distance can be overcome by various conventions: subtitles, voiceover or dubbing. Each can assume a shape of its own to affect viewers' imaginations regarding the background of a character. As often seen on streaming platforms, what is still very much a mode of expression, not simply a way of neutralizing and conveying the text, is accents and speaking styles, as identified by the rise of English-language dubbing from abroad (Hayes, 2021; Spiteri Miggiani, 2021). In this context, the speed and scale of generative speech technologies are envisioned to be new. They also delegate who gets heard to automated systems to a certain extent. Generally, four major methods of dubbing synthetic speech are possible: recognizing the original speech, a translation of it or the rewriting of it, generation of synthetically spoken text, and application of some type of synchronization technique (temporal synchronization, lip-synchronization). Based on research, it is known that automatic systems can adapt the length of the translated sentence, adapt other characteristics of the voice similarity of the source speaker and reduce the similarity of the voice acoustic (Li *et al.*, 2024). But synchrony ratings do not provide information as to whether spoken language codes (region, class and caste) are preserved in the target language. Fiction creation/building techniques are used in the selection of register, pause, intensity, rhythm and accent in a professionally dubbed performance. While many people believe that only matching duration and lip movements is important to the localization process, Brannon *et al.* (2023) demonstrate otherwise when it comes to naturalness and translation quality requirements for professional localization. Those findings bring one researcher's job to a new and different level: to perform the sounds beyond the limits of the technology.

What becomes apparent in this article is that this linguistic representation inequality would be hidden behind the all-encompassing appearance of the ubiquity of AI dubbing. Starting from an English dub could be helpful in engrossing audience participation but a few points have to run smoothly so that they will not notice the multi-language quality that will give a political or aesthetic sense to a story. On the other hand, as equally readily happens in a brightly colored local accent on foreign characters, deliberately prototypical regional marking can reinforce stereotypes. Competence, status and belonging are issues with language ideologies for standard language, which predate AI, and are already evident in the ways competence is tested, assessed and described in the context of status and belonging (Lippi-Green, 1994). But how can programmatic systems do the same; how can they reproduce or indeed, if managed properly, can they foil them?

[1] Associate Professor & Head, Department of English, Singhania University, Rajasthan, India
[*] Corresponding author's e-mail: munish10thakur@gmail.com

This may not be a figment of the imagination for the wider field of language technology. Tatman (2017) reported that automatic captions from a specific set of videos from YouTube varied in accuracy, depending on the representation of gender within the video and on the dialect. Koenecke *et al.* (2020) reported significant differences among the word errors obtained by American English speech-recognition systems tested here (both spoken by Black and White speakers). These studies should be noted in that they are not definitive of the existing dubbing service today having such error rates. They do, however, highlight the need to monitor not only the mean quality score in a processing pipeline that appears to be relatively language-agnostic, but also the effects of the chain in relation to a linguistic community and/or linguistic variation. Every mistake made in transcribing, generating the voice, and readers' final conclusions is incorporated into the final text, voice generation, and the viewer's subsequent interpretation.

In the meantime, fake copies of voices—it becomes more than ever tangible—the personhood of the voice. No face in a text = speech + can identify a performer, can communicate emotion. McGettigan *et al.* (2025) delve into the psychological and ethical considerations of actively generating recognizable voice representations, while Chubb *et al.* (2026) examine concerns around voice and digital voice protection. Consent also has relationships with privacy & reputation, credit & accountability, compensation with voice actors (Sharma *et al.*, 2025). Screen culture is characterized by such concerns as the following: a performer's voice is an individual, another kind of artistic instrument and a commercial part of the production chain.

There are three related questions posed in the study. What about what can be lost or altered about linguistic and dramatic identity when a voice is algorithmically recreated in other language(s)? other dramatic forms? other media? multiple iterations of the same set of people and/or locations, in these multiple languages—dramas—media—iterations—? Secondly, how will the process of synthetic dubbing affect the power dynamics between the various actors involved (performers, translators, producers, platforms or audiences)? Finally, would there be any "realistic" standards which lie somewhere between "function" in multi-lingual access and "cultural" "autonomy" and "performer" "cultural" autonomy? However, the goal is to not eliminate the use of AI for localization, even if it is perceived as exploitation. Access will also be facilitated and new modes of expression created using small-language versions, possible with speech synthesis. Instead, the paper calls into doubt the idea of "more similarity implies a better translation" or "more natural implies a better translation."

Its efforts have helped to develop the concept of 'vocal justice', which complements the notion of listening as an intelligibility-oriented service. The proposed framework combines linguistically indexical ideas, screen performance cinema with the politics of the cover of datasets and the rights of identifiable speakers. The aim of the paper is to be able to digest a lot of research done individually and bring it together in one analytical model to support the meaningful and responsible decision-making process without establishing one accent, one translation strategy and one technology in each and every film. The key role in favor of this is that a convincing dubbing should not be determined by whether the audience can tell that the film has been dubbed over, but rather whether there are any sounds that have been conveyed and respect the voices that the film embodies.

LITERATURE REVIEW

In the interpretation of audiovisual translation scholarship, dubbed dialogue can be viewed as a social construction as a means of representation. Chaume (2018) uses the term methodological changes to identify the new approaches and methods, which are: It expands the sphere of research: From analysis of only textual equivalence to culture, society and reception analysis. Research into dubbing language (Pavesi, 2018) focuses on the following language features: conversational naturalness, register, creativity and source-language influence. Firstly, each of these approaches challenges the erroneous perception that a dubbed script or recording is an automatic reproduction of the original script or recording. They also provide a basis on which to evaluate machine-generated speech: The translated text should be not just a source of a narrative world; it should also be a source of a target-language culture.

Especially for the English language which is marketed as the world language but in its reality is a language with various socially distinguished accents. Hayes (2021) discusses accent variation in English dubbed films, and accent variation as a tool for developing an accent profile rather than a standardized soundscape to create a persona. Regarding intonation and rhythm, Spiteri Miggiani (2021) points out several aspects of the practical work needed to create feasible strategies for dubbing English into a reality on the streaming platform. Their work has resulted in two seemingly conflicting results. However, accents, prestige and regional, among other social differences, are not easily surfiable from one language to the other. No single accent is culturally innocent; no one's accent can completely be replaced by an accent. Khadka (2024) also depicts South Asian English as a collection of locally configured linguistic identities, and he emphasizes the conflict between international access and linguistic diversity.

This antagonism is felt more strongly in multicultural or polylingual movies. Still, the movie where the Osage language features in it was studied, Persian dubbing being the context, to allow exploring several levels of preservation and neutralization of a multilingual movie, such as Rezvani Sichani (2026) has done. The analysis shows that decision-making in the secondary languages domain can affect characterization or the visual context and music tracks. Furthermore, Sivak (2026) shows that

the use of partially intelligible source speech can be as culturally significant as a 'bad' or 'mediocre' layer—in fact, it can also be a cultural practice, as in the case of the informal tradition of Ukrainian voice over seen in America. These studies hold together to trouble the taken-for-granted notion that full replacement of source sound is a universal notion of accessibility.

Few key points can be said about audience experience, if any, on the basis of technical production criteria alone. This renewed interest in reception research on dubbing and subtitling is illustrated at the same time by the lack of methodological rigor of the small studies reviewed by Antonini and Bucaria (2025) in their study of the empirical reception scholarship. Skills in assessing the audience include linguistic competence, understanding of the culture of dubbing, sensitivity to the accent and story's own rule set. An analysis of voice use in English would then give differing findings with a different audience where English is dominant, with the multi-lingual Indian and an English-language diaspora group. The variability allows us to consider comparative listening to be a more suitable approach than a 'common basis' for determining counts of the 'most natural voice'.

The more general sociolinguistic literature highlights such an acoustically nondescriptive type of vocal difference. A basic language ideology can also lead to a discriminatory conscious impact of accent as a criterion for judging an individual, as Lippi-Green (1994) explains. Recently, St. Pierre and Johnson (2026) found relationships between non-standard and foreign accents and portrayals of villainy and accent-based social anticipations within the media analyzed for the examined children. Both of these studies are focused on various institutions and genres, but both raise the issue of the lack of marking of standard speech and its ubiquitous desirability. Screen translation eradicates character history through an 'objective' model's voice, and can cause hierarchies to be made more prominent by a stereotyped (non-standard) model which a film does not champion. In the context of education, Nabung (2024) discussed cultural stereotypes and bias in the context of tertiary English language classrooms in Indonesia and called for a culturally sensitive education. Though not evidence relating to dubbing effects, this highlights the general need to look at how language practices reinforce or challenge problems of stereotypes.

The second kind of asymmetry that is imported by speech technologies is that some speech communities are over-represented in the training data compared with others, and this kind of over-representation is richer and more regular than the others. Bender and Friedman (2018) propose the data statements, which would allow one to see their design and claims being made by a language processing system that is operated by a person. Under-resourced varieties require language planning and the knowledge of their sociolinguistic nature to articulate quality issues, as reported by Lau *et al.* (2025). This work brings the attention to focus on the gap between marketing promotions of multilingual ability and the more complex, and socially oriented resources for accurately modeling dialect, prosody or mixed-language speech in a responsible manner. The number of language labels on offer is not a limiting factor on how much language is covered.

The stakes are illustrated in this one of unequal recognition. To the best of our knowledge, differences between gender and dialect (Tatman, 2017) and structured interviews as a transcribing corpus (Koenecke *et al.*, 2020) have been observed in prior research conducted on the previously mentioned platform. The effects, however, do not directly transfer to generative voice synthesis, and in many designs of synthetic-dubbing the transcription sits upstream of the generative voice synthesis. The following stage may be a correct rendition of the incorrect content of the source utterance. An erroneous intermediate transcript may then evolve into a successful final soundtrack, and result in an error being more difficult to detect by the listener.

It is helpful to know how to auto-dub. Citing concerns about speech naturalness and translation quality, and the merits of different temporality measures, Brannon *et al.* (2023) analyzed a large sample of professional localization and gave a cautionary tale: too much of what matters for linguistic and intonation aspects of speech may come at the cost of translation quality. Based on this idea, Li *et al.* (2024) propose a pause-aware method with the inclusion of translation, voice cloning and voice synchronization. It is important to appreciate that these studies reveal that engineering objectives should be negotiated and weighed in, not predetermined. Some examples illustrate how optimizing the target language for prosodicity of the source speaker's timbre may turn off the dependent natural prosodicity of the target language, and/or call the target speaker's attention to the need to agree with the performer on imprinting the source speaker's prosodics.

Where now the need for discussion on matters of improved performance. When actors turn into cultural agents and data points in the creative industries, tension starts to happen as adjusted by Thomas (2024). Sharma *et al.* (2025) create a model of privacy, reputation, accountability, consent, credit and compensation for voice actors. Familiarity with voices can also be psychologically and relationally meaningful, as demonstrated by McGettigan *et al.* (2025). These studies together cast doubt on the notion that the consent to record the original film scene translates into recording an unlimited license for the synthesis of the performer's voice in different contexts, editing or languages.

Authenticity and potential deception also come into play. Lees (2024) examines the documentary production techniques that take this to their extreme by examining examples and placing boundaries around where digital manipulation is permissible as a means to protect participants or convey difficult messages. Vaccari and Chadwick (2020) formally investigate this form of synthetic political video, and describe the way in which distorted information can instigate uncertainty about

what is happening. Forgery, however, in the entertainment world is not necessarily committing fraud; it could be a literal translation. There is a time and place for audiences to know that certain words were spoken by a familiar voice and at some point, they may expect to hear if a voice dubber was used, or if the track was recorded under a license agreement with a replica of the voice dubber.

Technical countermeasures are no cure-all for such questions. In the survey by Zhang *et al.* (2025), they discuss the issues of robustness, fairness and deployment regarding audio deepfake detection. In certain cases the nature of the sounds in an audio file can be detected to determine whether the sounds are synthetic or not; not whether it is a re-use or a performer's permission to a re-use or whether the re-use involved a culturally altered performance or composer's permission. Not all things that are worth good are always good. A "foreign" dub may be linguistically incorrect and a "foreign" dub may lead to the miscommunication of characterizations. The correct labeling of a dub doesn't mean that it isn't linguistically correct now. This separation will be necessary if a separation between authenticity, authorization and representational adequacy is necessary in a governance framework.

Several strands tie in culture analysis to practice and/or instruction. This volume sheds light on the language-learning means fulfilled by dubbing activities as per Buil Beltrán (2024), which opens the voice/voice(s), rhythmic production/expressive production of learners. What is relevant here is not the possibility of scaling up what is happening in the classroom to professional cinema but allowing the emphasis to lie on dubbing as an act of interpretation and performance with choices, rehearsal, and embodied attention. However, the topics of obtaining voice (for translation research) and good output (for speech engineering), of bias (for critical AI research), and of the rights of the performer (for labor scholarship) remained out of the spotlight of the reviewed scholarship. There are few integrated accounts that give the how-to for aggregating these concerns in a synthetic-dubbing decision. Such an account is expanded upon in this article.

MATERIALS AND METHODS

This research is a conceptual analysis using the approach of integrative and qualitative. The unit or object of analysis is not a particular film or series or unknown platform dataset or fresh film data or group of participants who were not consulted previously. It is an argument delivered on evidence that appeared from empirical publications concerning audiovisual translation, synthetic speech, linguistic variation, creative labor or about the authenticity of the media. If the primary research question is interdisciplinary and an objective is to make some useful conceptual combination instead of attempting to calculate a combined statistical effect, then an "integrative" approach is appropriate. To this paper, there is no survey and/or experiment or systematic meta-analysis.

In October 2026, research relevant to the content of the manuscript was identified and verified by using the pages on the publisher's website, scholarly-index pages and journal websites, bibliographic records, and pages from the Association for Computational Linguistics Anthology. Foundational works on accent ideology as well as recent works on AI dubbing and voice cloning were included, with a purposive selection. Each of the four traditions exhibited in the working collection on the stage was chosen with great care and will be discussed as a characteristic of current audiovisual translations and receptions, as well as sociolinguistic and speech data inequalities and the issues that arise regarding technical dubbing and synthetic media, and the ethical/legal concerns regarding performance. Articles from science journals (peer-reviewed) were chosen and selected full conference papers were chosen since they had made a direct technical intervention. The text(s) cited in the ultimate synthesis are indicated in the reference list.

Four steps in the analysis were used. First, there were claims of technical ability, which had no tie with claims of audience benefit or of cultural representation or legal authorization. Second, the categories of access, vocal difference, performance, 'ownership' of voice and transparency noted as 'provisional' are mentioned. Overall, and thirdly, the issue of localization – and the evidence available on that issue to both balance and detract from the simple technology-improves-accessibility argument – was discussed in terms of non-synthetic localization, as well as synthetic speech. 4th, the cross-cutting themes were then translated into a cross-cutting evaluation framework of six dimensions, and then applied as an illustration to deliberately hypothetical production situations, to show the practical application.

The review does not provide a complete scoping of reviews nor is it a PRISMA-compliant review. Since these procedures were not followed, no numerical "literature search flow" or the inter-rater reliability coefficient or the frequency of the coded themes are reported. The implications of recommendations it draws are theory-building statements, based on a documented literature, not estimates of the degree of a harm in the industry. This boundary is highly pertinent for highly dynamic speech systems where capability demonstrations and product specifications can go out of date quickly. Future research should be empirical for claims regarding specific audiences/performance in less resourceful languages.

To screen out technological determinism, the criterion for evaluation decides the fidelity of the function that the character performs as opposed to the fidelity of the speaker's acoustic dialogue. Moreover, the framework doesn't take issue that one can improve the quality of the output instead of having to get the performers' permission. A translation tool for production questions, in the form of a decision matrix, is provided. This is not a validated psychometric instrument, it is not an industry standard, but is a normative instrument suggested by

the presented article. A test of the reliability, operational definitions and feasibility of the procedure can be carried out with real multilingual productions in a subsequent study.

RESULTS AND DISCUSSION

In the synthesis, five linked tensions and a six-part framework of the challenges of vocal justice are created. This is not a statistical comparison of the projects that were successful with those that were not. This is not a conceptual comparison of successful dubbing projects vs. unsuccessful ones. The six dimensions of the framework can be interpreted as harmonious complementary decision tests: intelligibility, indexical fidelity, performative integrity, consent, attribution and redress. None of the above can replace the use of the other indicators. An easily audible voice might sound like it is from a different social class, a closely mimicked voice might appear recognizable, but can't be trusted, and an authorized voice might be technically distorted.

Access and the Danger of Accent Normalization

This facility of cross-language dubbing can get rid of the stories' significant obstacle to enjoyment completely. Use of a familiar language may foster engagement among viewers who have different literacy levels, as well as among viewers who prefer listening to reading or the audience who is less tech-savvy or with access needs. The ethical quandary begins when the issue of access is understood as "no difference in language." The study of the English dubbing by Hayes (2021) shows that accent variation can be considered as a characterizing element of a successful dubbing localization practice. Additionally, Pavesi (2018) mentions the specificity of presenting speech in a screen: express and conversational. Common events and "level speaking" may facilitate understanding and attenuate the discrepancies in which the story locates its actors socially. There are at least three different varieties of accent normalization. One of these is 'phonetic flattening': modification to a model voice used in the system for features of rhythm and prosody. The latter is sociolinguistic smoothing, namely all characters' utterances are given the same status. The third is 'interactional flattening' that can be summarized as the need to maintain a 'smooth, self-contained sentence' through the use of hesitation, overlaps, interruptions and code switching. None is inevitable, and sometimes one may (hopefully) be considered appropriate just for clarification. The danger is while normalization becomes a habit and neither the system nor commissioning team considers if the voice difference belongs in the scene.

Creating noticeable accents in target languages is hardly a good solution. The easy way for a rural character to have a sort of rural English accent would be to give them a "rural English accent," and the easy way to attract a foreign accent to a character who speaks an Indigenous language would be to let that character have a stereotypical foreign accent. The stereotypes of these solutions, however, are different in different cultures. The reasons why socially marked accents are not, and should not be, a neutral marker of personality, are presented in illustrations by Lippi-Green (1994) and St. Pierre and Johnson (2026). Preserving the difference which has a dramatic and relational meaning is therefore called: indexical fidelity and not simply a re-creation which coincides mechanically with the accent label. Who is the authority in the original text, who is the marginalized, who is intimate, and who is alienated; what evidence supports the selected translation in the target language?

Naturalness, Synchronization and Performative integrity

Intelligibility, acoustic naturalness and synchronization are fundamental concerns in much speech engineering, to which they grant appropriate attention. These are the features that a dub needs to possess; without them the dub can be fatiguing or useless. However, when performer eccentricities are not considered toward optimal results, they can, and do, become errors awaiting correction. Professional dubbing is not simply a matter of timing and assumption of lip movements, caution Brannon *et al.* (2023). Technical processes can indeed be beneficial for HLT, as shown by Li *et al.* (2024). Changes which appear to be improvements on the way things are done can truly change the dramatic impact of a sentence that is not complete or a beat before a confession.

Differentiating between the lexical meaning and the performed meaning. Don't assume that when the person says, "I'm okay," they really are! This may vary from pitch to pitch, bodily tension to tension, and context to context, reflecting defiant, fearful, resigned or sarcastic language. A synthetic system would usefully take the sentence, reasonably simulate the properties of the voice in the spectral realm and yet perform the wrong interpersonal meaning. Performance integrity does not simply involve making copies of every contour (breath and pitch): some adjustments are a part of the production process of the cross-language activity. To be identified, reviewed and justified it demands consequential changes to the expressive texture of a performance. Translator/voice directors do not duplicate blindly the decorativeness of what machines produce but rather interpretatively prompt the dramaturgy.

This also works for code switching (where a shift within a language from one language to another could be linked to intimacy, resistance, humor, hierarchy, and a shift in audience). In Rezvani Sichani's (2026) study on alternative Persian dubs, it is shown that dubbing treatment of a third language can impact the multilingual characterization. The representativeness of AI dubbing is a point that can be observed and even predicted by any automated system as such, as it is not peculiar to the field of dubbing in general, but is present in the work described. The situation where the scene is switched from majority language to minority language, or vice versa, may be taken by using only one target-language soundtrack which will destroy

the effect of the language change. These include short preservation of source material, evident transposition of pitch and register, and omission/translation of important words, and others as discussed with the participants involved – that is, the performers/actors, who are aware of the languages involved.

Uneven Datasets and the Unequal Right to be Heard

If the speech resources are more strongly mobilized from the high-resource varieties of the standard, the possibilities for a widespread use of dubbing will not be better. Bender and Friedman (2018) talk about the advantages of documentation of language information for the support of scientific transparency and clarification of inconsistencies between the language community in the training data and the language community aimed at by the intended user community. Lau *et al.* (2025) demonstrate that errors in multilingual datasets may be at varying magnitudes and happen for a variety of different reasons, and that some of these reasons are related in part to the institutional context of a language. In synthetic dubbing, errors are possible such as incorrect names, lessened dialect differences, and unintelligible pronunciation, or the routine switching of culturally specific terms for the more standard ones.

Technological inequity accrues. If a dialect is not well known, a poor account is included in the script translated. If there isn't sufficient evidence of code switching or figurative language, the following processing step might be able to "simplify" instead of comprehend the dialogue. When there is not much training material for the expression in speech synthesis, the final voice could be extremely neutral, too cheerful sounding. The general danger of using the speech recognition underlying the indicators stems from two sources: one is the examples provided by Tatman (2017) and Koenecke *et al.* (2020) of how benchmarks can obscure characteristics of social groups; the other is the danger of reinforcing the stereotype associated with the child. There is no intention to translate them in other languages, but the reasoning can be transferred. Expressive reporting has to take place by dialect, gender presentation (if applicable), age range, other speaker features, such as protecting privacy and community involvement.

This is the case in particular in states that have internal linguistic diversity. A platform may excel in coverage of a nationally recognized pronunciation/dialect of a given language, and lack coverage for a minority variety or locally salient pronunciation, or have a mixed-language dialogue altogether. This results in a seeming paradox – the more choices you give about language, the fewer choices you get in each language. We would like to mention that a language button might be available, but it is essential to recognize the difference between buttons that are available and those that work sufficiently across various social environments – a difference that needs to be recognized in an ethical localization process. For a film full of jokes, cuts, emotionally disturbed speech, whispered conversations and news speech, it's not enough to test it out as a formal news speech. The test does not assess general sentence skills, but rather the components of the language of a realistic dramatic world.

Consent, Creative Labor and the Ownership of Vocal Likeness

One cannot simply use a voice in a finished production, time after time, without repercussions. It's not to be overlooked by the public and has the expressive power of any art. Thomas (2024) puts actors in the role of a "cultural agent" as well as "potential data sources" in a new synthetic-media landscape. Sharma *et al.* (2025) share their observations of how privacy, reputation, accountability, consent and compensation and credit for voice-actor experiences are interdependent. Their arguments suggest that it would be a bad idea to leave out details beyond the first recording session in your permission. Questions that are relevant are the following: what languages, characters, projects, territories and distributors are authorized, what are the model versions and what are the future versions to be taken.

Of course, it should be fine-grained, consistent and time-limited content permissions within "practicable contractual and technical limits." A licensed performance of a performance by another actor may be required in the same production, even if the source actor does not want to give a derivative performance. Instead, it's best to talk to a professional dubbing actor, which is not paying every actor to dub the films, but can afford one actor to do training for one dubbing job only. Either way, it does not work to tell the audience that they like the end result of whatever the outcome may be. McGettigan *et al.* (2025) attempted to understand the interpersonally salient aspect of familiar voices and Chubb *et al.* (2026) examined the lack of legal protections for unfixed voice. There are many different jurisdictions and different jurisdictions have different laws so it can be very hazardous to base one license on all the ethical and contractual considerations!

Also, issues with the distribution of labor. Based on the experience and knowledge of other people who have been in the field, specifically translators, dialogue writers, directors, sound editors and actors who can make the appropriate decisions when they make a finished text, the traditional localization seems to be relying on these individuals' expertise. It doesn't mean that human participation in the synthesis has to come to a stop once it gets faster, it is a transformation that it could undergo instead. One flow might be for specialist translators/cultural experts to do the work, with the human artists doing the work in the middle and assisted synthesis used for lower-risk content that was approved by the specialist translators/experts. There's another one that will be able to change the savings and pay in other languages. To ensure equity, it's not a pure technical vocation but good governance and compensation. For the final artifact, there must be written attribution of human and synthetic contribution in terms accessible to the audiences and participants.

Transparency, Authenticity and Audience Agency

Synthetic dubbing does not need to fool the audience to be able to spark their questions about authenticity. It makes it difficult for the viewer to determine what voice might have been used to dub into English, if one was used, and whether it was the original actor's voice or if an actor for the film was given an imported dubbing voice. Lees (2024) provides examples of cases where digitally altered AV media has been used creatively and protectively, while Vaccari and Chadwick (2020) provide examples of protecting the question of authenticity by the use of synthetic AV media. These are different levels of operation, but constrained by proportional disclosure that if it is a production it shouldn't tell you that any performance operating a machine isn't a performance at all.

There must be choice if there are disclosures. If a new technical note is buried, or if there is any human dubbing, or if the production will be designed to keep the code-switching, then an audience will get little information about these from such a note. Viewers may be required where appropriate: Original-language audio with captioning, oral dubbing (where available) in some human voice with appropriate label, recorded synthetic dubbing. This will be a difference that matters to disabled viewers and users who use languages from which they suffer from not speaking native: subtitles or dubbing are not accessible to everyone. Antonini and Bucaria (2025) caution about the potential dangers of audience-taste presumptions, based on an audience's inadequate reception research. Ideally, there should be leeway for other states and an excellent interface should not be regarded as the final "final" state. The solution is not so simple, as acoustic watermarking and/or synthesized audio detection can detect authenticity for proof of production origin. In their research (2025), Zhang *et al.* explore technical challenges for audio deepfakes. Any voice marked socially in an ideal detection would not be able to give the editor any feedback about a flattening, whether it was there or not; nor could it supply information to the editor on whether the translator was named or how it was decided to consent. Governance needs therefore to have at least two records, a provenance record (for the production and creation of the audio) and an authorization-and-translation record (for the acceptance of the voice as well as the translation of meanings). These records would be for different users, and shouldn't all be boxed under 'AI used.'

A Six-Dimension Framework for Vocal Justice

The multilingual soundtrack in the framework presented in this document is a consequence of negotiation in the public sphere of people, places and relationships. It contains 6 dimensions that correspond to various possible failures, but not in the order of universal scores. Intelligibility refers to how well listeners, in a dialogue, can interpret the sentences in the context. The indexical fidelity is related to the relationship and the social meaning in original relation to the linguistic differences that are maintained. Performative integrity – whether or not consequential elements of acting are delivered. Consent is an approval of the use of a voice likeness and the associated data. Attribution is whether or not man and man-made contributions are recognized. Redress discusses the level of importance that a harmful or unauthorized rendition can be met with by the performer, a representative of the community, or a production partner. It is intended to infuse ethical motivations which are nebulous into the manageable realm of production decision-making. Table 1 shows the production tests and supporting evidence for each dimension.

Often intelligibility and indexical fidelity are at odds. Sometimes, a mixed-language expression can help some readers to understand the expression, and by itself might be necessary for another reader to get a sense for understanding some relationship. It does not stipulate which side has to win in the end. Instead, it must acknowledge the trade-off and choose a language production that would catch the attention of the audience and would be acceptable to the storytelling: such as an unobtrusive subtitle, motivated dialogue gloss, accent/register change, maintenance of the source language phrase. It can then be assessed by using target-audience listening tests, such as those by listeners who are familiar with the listening culture of the target audience. This process is more than a mere description of an automatic choice being "culturally neutral".

Sometimes performance integrity will be conflicting with a numerical set of consequences associated with the temporal alignment. Other times, unsymmetrical pauses and speech rhythms are engaged in an attempt to express vulnerability and/or resistance. The team might adjust the translated sentence to suit the original shot; or might allow a small time lag if the deletion of a pause would upset the performance. The findings of Brannon *et al.* (2023) may serve as benchmarking data to make such a judgment, so may the type of engineering concerns involved in bringing a machine design to fruition as described by Li *et al.* (2024). Thus, the problem of creative optimization is rather a constrained co-authorship than the problem of just searching for minimum error of mouth-synchronization.

The other 3 scales are accountability. If a voice can only be identified, then training and deployments are not conducted without the permission of the individual(s) concerned and are freshly reviewed if/when there is a significant change in an approved use. The distinction will be drawn between the author of the performance and the translator/adapter, the voice supplier and the technologies employed to make the delivery. Redress must include a documented procedure for reporting suspect accent mapping, mistranslations (culturally) and/or unauthorized copies. Technical systems can change the artist's voice vulnerability as highlighted by Chubb *et al.* (2026) and Sharma *et al.* (2025). Then a credible framework following the LQ review process must have contractual and organizational safeguards.

Applying the Framework: Three Hypothetical Production Cases

Ideally, you should have in mind a family drama that unfolds in more than one language – what we call the "city" language and a "home" language of lower status. Even with a high intelligibility score, a synthetic target-language dub can fail to get a good indexical fidelity score. The reviewer needs to identify and preserve the sense of intimacy or exclusion by writing the same language, indicating register changes and/or adding subtitles. Not a reference to a movie, but a representative example/demonstration. As it is supported by the research on multilingual dubbing and processing third language dialogue (Rezvani Sichani, 2026).

Another case includes the recording of the popular show-business figure's own voice for various regions around the world for the fabrication of a copy. There has to be scope of usage approved and documented, be it timbral similarity, as there is a need to prove a right to use them in this manner; scope of usage approved and documented, compensation and the individual's right of objection in future usage is required, and they must be used in a wide release. Source voice does not necessarily have a better alternative than a sensitive new performance by the actor of the target language. This should be based on the artistic appropriateness and informed consent of the character regarding the section of character to be retained and the successful marketing of the film, not merely because the marketing line can preserve the original star's voice in every market (McGettigan *et al.*, 2025; Sharma *et al.*, 2025; Thomas, 2024).

The case of a hypothetical performance is fast-paced comedy in a so-called "low-resource" variety. An upstream recognition system can't handle local diction, and the lapped-up version of the dub will have a nice flowing mono version of the jokes. It is possible that the reader (or listener) will not be able to tell that there are errors, because the synthetic voice sounds convincing. In this case, there is a high level of intelligibility and a low level of accuracy in translation. Examples of a suitable response include getting a checklist from the source language speaker who knows that language, getting an independent speaker of the source language to review the checklist, and, finally, pilot-listening sessions with appropriate groups. Support for the diagnostic approach includes aspects of Bender and Friedman (2018), Lau *et al.* (2025) and Koenecke *et al.* (2020).

This is a BMI reaction to these cases that shows automation is not a key ethical issue. The pivotal difference is between accountable and unaccountable transformation. It may also be possible to have a human dub into the accent in question that is being deleted, and a good synthetic dub can retain a certain distinction more. Likewise, a film which can be seen by widespread audiences can sensibly provide various translation modes instead of providing one authoritative soundtrack. The general syntactic structure that emerged, then, was the rejection of technological triumphalism and of what might be read as a "nostalgia blanket" when it comes to the past way of localizing. It examines the roles and purposes of different productions, and audiences.

Production and Evaluation Implications

In preproduction, producers should consider the linguistic risks during system selection. This is the identification of formal/informal registers, accent/dialect and names, songs, code-switching and culturally marked silence in terms of their roles in the narration. The review needs to also discern fictional characterization from claims related to actual communities. The voice team should ascribe the voice data source (person acting as the voice source, if possible, including authorization from that original voice source at the level of granularity where she is able to authorize her voice) and secure permission from that original voice source, at this level, including permission to clone her voice into the target voice language. Procedures are carried out to ensure the transparency as advocated by Bender and Friedman (2018) with emphasis on labor protective measures as developed by Sharma *et al.* (2025). The idea is to point out decisions that have the most cultural or rights issues without having to read the entire rule book per each word.

A mixed production process is probably more protected against litigation than a completely unproduced pipeline. In cases of non-standard speech, evaluations of recognition output should be conducted; in cases of narrative and social scenes, content checks for translation accuracy should be performed; in synthetic delivery, accent checks for the accent roles the commissioning ensemble may not detect should be carried out, and in accent associations in which the commissioning team may not detect accent roles, accent checks for the accuracy of those accent roles should be performed by experts in the target language. For certain situations there may be a need for re-performance such as singing, spontaneous interaction, mood or laughter, or local dialect/idioms. Some technical evidence is welcome and supplemental, not in lieu of cultural/dramatic evidence.

This needs to be broken down into two passes. Concept of a linguistic pass: Meaning, Named entities, Code switching, Register, Marked speech. There is a performance-and-rights pass that includes expressive delivery, vocal resemblance, permissions and credit, and remedy/availability. The audience pilots are then able to listen and compare listening conditions that are relevant to the listening task instead of just to the audio that is most natural. Some of the limitations of the current research on reception are pointed out by Antonini and Bucaria (2025) and future pilots should incorporate information about the recruitment criteria, language history of the participants, their listening method, their previous experience with dubbed media, or restrictions on samples used. A critical framework which comes in handy when generating further and further questions and further and further repeatable decisions is highly useful.

Proportionality is also key to governance. The

six dimensions by themselves are not a basis for implementation of requirements to independent small producers as compared to large platform producers. If it's a "low-tech" project, they may be able to get permission, have an original-language audio archive, have a log of significant translation decisions, etc., and have an accountability reviewer with a limited amount of resources. Independent audits, compensation mechanisms and formal complaint mechanisms provide better foundations at a large scale. There's a shared key requirement, which is "reason why" or "traceability" — an explanation about why a voice sounds the way it does, who authorized the voice and how errors that are apparent later, but which had a consequence, become correctable. It is easy to be critical of the culture that we are a part of, and to look inward at that culture, holding the professionals responsible without being able to trace it back to its sources or follow a chain of causes and consequences.

Last but not least, there is an empirical question, namely, whether voice options are preferred by readers and the extent of creative harm that occurs, which can only be answered by a combination of scholarly articles. The proposed standards must be field tested in other industrial and language contexts, for example, with the multilingual cinema of South Asian and African (and other) cultures. One potential design would be a comparison of 3 independent dubs of the same short scenes, one dub being synthetic, with the other 2 dubs being supervised and/or human voice and human direction. For each dub, the following measures of comprehension, perceived character identity and emotion and acceptability of the consent arrangements could be assessed. It is an idea about studies to be conducted in future and is not a study conducted here. This study might help determine if there are differences between the dimensions to be evaluated which show significant positive correlations with difference in viewer and difference in performer results, and, if there are, where do trade-offs truly exist?

Table 1: Proposed assessment matrix for vocal justice (conceptual tool, not empirically validated).

Assessment dimension	Production test	Evidence or remedy
Intelligibility	Can relevant audiences follow dialogue?	Comprehension trial; accessibility review
Indexical fidelity	Are relationships expressed through linguistic differences still legible?	Source/target sociolinguistic review
Performative integrity	Are pauses, tone and emotional intention treated deliberately?	Voice direction and scene-based review
Consent	Was the specific use of identifiable voice data authorized?	Scope-limited record of permissions
Attribution	Are translators, performers and synthetic methods credited?	Credits and track metadata
Redress	Can a stakeholder contest and correct a material problem?	Named review contact and correction protocol

CONCLUSION

Multilingual dubbing can be synthetic, enabling access in multiple languages when the meaning of the screen characters is acquired via accent, social identity and expression. In this conceptual analysis, the researcher will assert that technical ability and synchronizing are not sufficient indicators of a responsible localization. It is justice-oriented, where the two core concepts are intelligibility and indexical fidelity, which are combined with performative integrity, consent, attribution and redress to evaluate representation and responsibility. The five tensions are not about the inevitability of eliminating language difference in the process of AI dubbing, but rather potential dangers and choices.

Producers and platforms should not only record the origin of voice data, but also obtain specific consent from the performer and credit the human contributions, and offer easy means of project verification. Culturally important dialogue and accent decisions need to be checked by translators, voice directors and speakers who are familiar with the target/source-language communities. Clear information about tracks created, or not created, synthetically needs to be provided to audiences, and audiences should be given the option to have access to original-language tracks as available. The framework should be systematically examined in the future by conducting comparative listening studies, interviews with the practitioners, and cross-linguistic production evaluations. Currently, the utility of the framework is in a context-specific decision-making process that retains transparency of the choices made, not in providing a universal quality score that can be easily applied.

Declarations

Funding: The submitting author should state any funding or write "No external funding" only if accurate.

Conflicts of interest: To be completed and verified by the submitting author.

Data availability: No primary dataset was generated or analyzed for this conceptual critical-review article. Sources are listed in References.

Research ethics: No human participants were recruited, and no new identifiable participant data were collected.

Use of AI-assisted tools: To be completed accurately by the submitting author in line with the journal's current disclosure policy.

REFERENCES

Antonini, R., & Bucaria, C. (2025). Beyond the screen: A systematic review of perception/reception studies in interlingual dubbing and subtitling. *Journal of Audiovisual Translation, 8*(2), 1–21. https://doi.org/10.47476/jat.v8i2.2025.371

Bender, E. M., & Friedman, B. (2018). Data statements for natural language processing: Toward mitigating system bias and enabling better science. *Transactions of the Association for Computational Linguistics, 6*, 587–604. https://doi.org/10.1162/tacl_a_00041

Brannon, W., Virkar, Y., & Thompson, B. (2023). Dubbing in practice: A large scale study of human localization with insights for automatic dubbing. *Transactions of the Association for Computational Linguistics, 11*, 419–435. https://doi.org/10.1162/tacl_a_00551

Buil Beltrán, P. (2024). Exploring the impact of didactic dubbing on speaking skills and motivation in English for specific purposes settings. *Journal of Audiovisual Translation, 7*(1), 1–26. https://doi.org/10.47476/jat.v7i1.2024.287

Chaume, F. (2018). An overview of audiovisual translation: Four methodological turns in a mature discipline. *Journal of Audiovisual Translation, 1*(1), 40–63. https://doi.org/10.47476/jat.v1i1.43

Chubb, J., Harrison, P., Tompkinson, J., & Thompson, E. (2026). How close is too close? Digital replicas and protection of the "unfixed" voice. *AI and Ethics, 6,* Article 582. https://doi.org/10.1007/s43681-026-01323-9

Hayes, L. (2021). Netflix disrupting dubbing: English dubs and British accents. *Journal of Audiovisual Translation, 4*(1), 1–26. https://doi.org/10.47476/jat.v4i1.2021.148

Khadka, B. K. (2024). The English in South Asian contexts: Exploring historical development to present role in education. *Journal of Natural Language and Linguistics, 2*(1), 179–189. https://doi.org/10.54536/jnll.v2i1.3762

Koenecke, A., Nam, A., Lake, E., Nudell, J., Quartey, M., Mengesha, Z., Toups, C., Rickford, J. R., Jurafsky, D., & Goel, S. (2020). Racial disparities in automated speech recognition. *Proceedings of the National Academy of Sciences, 117*(14), 7684–7689. https://doi.org/10.1073/pnas.1915768117

Lau, M., Chen, Q., Fang, Y., Xu, T., Chen, T., & Golik, P. (2025). Data quality issues in multilingual speech datasets: The need for sociolinguistic awareness and proactive language planning. *In Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics* (Vol. 1, pp. 7466–7492). Association for Computational Linguistics. https://doi.org/10.18653/v1/2025.acl-long.370

Lees, D. (2024). Deepfakes in documentary film production: Images of deception in the representation of the real. *Studies in Documentary Film, 18*(2), 108–129. https://doi.org/10.1080/17503280.2023.2284680

Li, Y., Guo, J., Zhang, M., Ma, M., Rao, Z., Zhang, W., He, X., Wei, D., & Yang, H. (2024). Pause-aware automatic dubbing using LLM and voice cloning. *In Proceedings of the 21st International Conference on Spoken Language Translation* (pp. 12–16). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.iwslt-1.2

Lippi-Green, R. (1994). Accent, standard language ideology, and discriminatory pretext in the courts. *Language in Society, 23*(2), 163–198. https://doi.org/10.1017/S0047404500017826

McGettigan, C., Bloch, S., Bowles, C., Dinkar, T., Lavan, N., Reus, J. C., & Rosi, V. (2025). Voice conversion and cloning: Psychological and ethical implications of intentionally synthesising familiar voice identities. *Journal of the British Academy, 13*(3), Article a31. https://doi.org/10.5871/jba/013.a31

Nabung, A. (2024). Challenges and strategies in addressing cultural stereotypes and bias in Indonesian tertiary ELT/EFL classrooms. *Journal of Natural Language and Linguistics, 2*(1), 18–29. https://doi.org/10.54536/jnll.v2i1.2775

Pavesi, M. (2018). Reappraising verbal language in audiovisual translation: From description to application. *Journal of Audiovisual Translation, 1*(1), 101–121. https://doi.org/10.47476/jat.v1i1.47

Rezvani Sichani, B. (2026). Flower moon lost in translation: Killers of Osage voices in dubbing. *Journal of Audiovisual Translation, 9*(1), 1–20. https://doi.org/10.47476/jat.v10i1.2027.393

Sharma, T., Zhou, Y., & Berisha, V. (2025). PRAC3 (Privacy, Reputation, Accountability, Consent, Credit, Compensation): Voice actors in AI data-economy. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, 8*(3), 2360–2372. https://doi.org/10.1609/aies.v8i3.36721

Sivak, K. (2026). The informal Ukrainian voiceover. *Journal of Audiovisual Translation, 9*(1), 1–20. https://doi.org/10.47476/jat.v9i1.2026.340

Spiteri Miggiani, G. (2021). Exploring applied strategies for English-language dubbing. *Journal of Audiovisual Translation, 4*(1), 137–156. https://doi.org/10.47476/jat.v4i1.2021.166

St. Pierre, T., & Johnson, E. K. (2026). Linking the depiction of accents in children's media to the development of language stereotypes. *Child Development, 97*(4), 1288–1302. https://doi.org/10.1093/chidev/aacag048

Tatman, R. (2017). Gender and dialect bias in YouTube's automatic captions. *In Proceedings of the First ACL Workshop on Ethics in Natural Language Processing* (pp. 53–59). Association for Computational Linguistics. https://doi.org/10.18653/v1/W17-1606

Thomas, S. (2024). AI and actors: Ethical challenges, cultural narratives and industry pathways in synthetic media performance. *Emerging Media, 2*(3), 523–546. https://doi.org/10.1177/27523543241289108

Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media + Society, 6*(1). https://doi.org/10.1177/2056305120903408

Zhang, B., Cui, H., Nguyen, V., & Whitty, M. (2025). Audio deepfake detection: What has been achieved and what lies ahead. *Sensors, 25*(7), Article 1989. https://doi.org/10.3390/s25071989