Consumer Perception of AI-narrated Voice-Overs vs. Journalist-narrated Voice-Overs in News Videos

Anne E.G. Smink

Snr 2084563

Master's Thesis

Communication and Information Sciences

Specialization in New Media Design

School of Humanities and Digital Sciences, Tilburg University

Supervisor: Chris van der Lee

Second Reader: Tom Lentz

Jan. 15, 2024

Abstract

Artificial intelligence (AI) is increasingly being used in journalism. However, previous research has not fully addressed the perception of the use of AI-generated voice-overs in news videos. Therefore, this study aims to answer the question: 'What is the effect of a voice-over narrated by a voice-cloning AI on the perceived pleasantness and credibility of consumers compared to a journalist-narrated voice-over and does the type of news video (factual vs. human interest) influence this effect?' To investigate this, a between-subject experiment is conducted. 198 people participated in the study and filled in a survey. The participants were randomly exposed to one out of three factual news videos or one out of three human interest videos. These videos featured either an AI-narrated voice-over or a journalist-narrated voice-over and participants had to rate the voice-overs' credibility and pleasantness. A MANOVA showed a higher credibility for a human voice-over in both a factual- and human interest news video. The type of video did not influence the perceived credibility and pleasantness, but it is recommended for future research to further explore in which news contexts AI narration might be appropriate at this moment. Additionally, future studies should explore whether a greater usage of AI in the news and therefore increased exposure, could play a role in its acceptability and whether newly generated voices would be just as effective as voice cloning. This research provides insights into current perceptions of AI, whether people can distinguish an AI voice from a real human voice and gives valuable insight for journalists about the use of AI-narrated voice-overs.

Keywords: artificial intelligence (AI), text-to-speech (TTS), voice cloning, AI-narrated voice-over, journalist-narrated voice-over, factual news, human interest, pleasantness, credibility.

Consumer Perception of AI-narrated Voice-Overs vs. Journalist-narrated Voice Overs in News Videos

The use of Artificial Intelligence (AI) in journalism is becoming more prevalent. The news industry already makes extensive use of writing algorithms to generate content (Jung et al, 2017). Additionally, some major news organizations use AI-powered tools to enhance the video production process (Marconi, 2020). For instance, the BBC has created a tool called ALTO that makes use of text-to-speech (TTS) technology to produce voice-overs for video content in various languages, allowing those who do not speak English to watch BBC news reports in their language (Marconi, 2020). These developments in AI also affect or will affect other creative tasks in video storytelling at news organizations, including writing and recording a voice-over to give the video a greater sense of authenticity (Najar, 2018).

As AI keeps on playing an increasingly important role, it is essential to gain an in-depth understanding of the current use of AI and identify the challenges that exist in the field. Studies have been done to investigate the quality of AI-generated content in both written and spoken forms. Research by Miroshnichenko (2018), for instance, found that news consumers cannot tell if news has been written by people or by an AI algorithm. Research has also been conducted on spoken texts. Kit et al. (2023) compared how news is perceived with an AI-narrated voice-over vs a journalist-narrated voice-over. No negative effects of an AI-generated voice-over were found on the perception of the news, which indicates that using AI to narrate voiceovers has potential. However, the study by Kit et al (2023) was about explainer videos rather than news videos. Trust and authenticity of news videos might be of higher value for consumers than explainer videos since consumers rely on news broadcasts as the primary source of current information to keep up with important events and developments throughout the world (Cushion, 2015). Additionally, journalistic standards and ethical codes emphasize credibility in news reporting (Tumengkol, 2021). Therefore, the expectation of authenticity and credibility is likely higher for news videos than for explainer videos. This means there remains a knowledge gap about the perception of AI-narrated voice-overs and journalist-narrated voice-overs for news videos.

The type of news might also play a role in the perception of a voice-over. Research by Scott and Gobetz (1992) distinguishes news in two types: hard- and soft news. Soft news is described as articles that focus on a topic of human interest. Hard news is news where facts are the focus. Since those types of news differ in focus, they might also differ in the way the voice-over is implemented in the news. Human interest videos include captivating stories about people's lives, defined by more narrativity and evoking emotions (León, 2008; Piazza & Haarman, 2011). A human touch might therefore be expected in the voice-over. Tone and pace are aspects of a voice that reflect emotion (Pounds, 2012). Given that a voice-over for a factual news video may strive for a different feeling than one for a human interest video, those voice elements in those videos might be different.

This study aims to provide insights for news organizations that are willing to learn how to use AI for video production. According to Opdahl et al. (2023), AI might allow journalists to focus their efforts on critical thinking and creativity, enabling the effective delivery of important news while AI takes over other tasks, potentially taking over creative aspects of the video production process such as narrating voice-overs. However, little research exists on the possible risks of AI helping with or replacing creative work within journalism. This raises the question of what the repercussions are of letting AI take over or assist in creative tasks for video production, in particular, how AI-narrated voice-overs affect how people perceive the news. Voice-overs play a significant role in evoking emotions, as factors like tone or pace can influence how a video is perceived (Mogaji, 2018). Considering that tone and pace are important aspects of finding a voice-over pleasant, it becomes critical to investigate perceived pleasantness for a voice-over (Eadie & Doyle, 2005). When a voice-over is perceived as pleasant, people are likely more engaged, which is beneficial for retention (Rodero & Lucas. 2021). Additionally, measuring perceived credibility important, given its significance in journalism (Meier, 2009). Thus, to determine the possible influence of AI on how people perceive a certain news type, further research is necessary. Therefore, the following research question was formulated:

What is the effect of a voice-over narrated by a voice-cloning AI on the perceived pleasantness and credibility of consumers compared to a journalist-narrated voice-over and does the type of news video (factual vs. human interest) influence this effect?

Theoretical framework

Artificial Intelligence

Artificial can be defined as something that is fabricated (Sokolowski, 1988). To be intelligent is to be able to comprehend, learn, and adapt to accomplish goals in a variety of settings (Goertzel & Wang, 2007). When taken together, artificial intelligence is a fabricated ability to comprehend, learn, and adapt to accomplish a goal in a specific setting. Artificial Intelligence (AI) can also be defined as a technology where algorithms are designed to perform actions that would typically require human intellect (Kok et al., 2009). An algorithm is an ordered set of instructions or codes for software to perform a certain action (Coleman, 2020). According to Alkhatib et al. (2023), These algorithms can adapt and evolve based on data in the context of machine learning (ML), which is a subfield of AI.

Through ML, computers may imitate and adjust to human-like behavior. Each interaction and action taken through machine learning becomes something the system may learn from and apply the next time (Alzubi et al., 2018). Using algorithms, the system develops the ability to recognize patterns, make predictions, and carry out tasks without needing explicit programming (Janiesch et al., 2021). Training machines using provided data and algorithms is the goal of machine learning (Jakhar & Kaur, 2019). Deep learning, a form of machine learning, has recently accelerated the development and capabilities of various machine learning tools (Janiesch et al., 2021).

Deep learning is a kind of machine learning that uses computer algorithms that mimic the structure of the brain's biological neural networks (Jakhar & Kaur, 2019). Compared to simple artificial neural networks, deep neural networks frequently include advanced neurons and have multiple hidden layers organized in complicated structures. These characteristics enable deep neural networks to be provided with raw input data and automatically find features that are required for the related learning task (Janiesch et al., 2021). Neural networks can mimic human information processing and learning. The device "reads" a large amount of data and searches for patterns (Marconi, 2020). Deep neural networks perform better than shallow ML methods in most applications that need the processing of text, picture, video, and speech because DL is especially helpful in domains with big and high-dimensional data (LeCun et al. 2015). A form of processing that uses deep learning, Text-to-speech (TTS) technology, will be discussed in the following paragraph.

Text to speech

Text-to-Speech (TTS) technology, which produces speech, is made possible by deep learning (Ning et al., 2019). TTS are already used in journalism for example to create multilingual voice-over tracks, as mentioned before in the introduction (Marconi, 2020). TTS is an AI-driven voice technology. TTS systems can produce sounds with a human touch like grief, enthusiasm, and empathy by adding emotional nuances to synthetic voices (Hillaire et al., 2019). An example of a TTS tool driven by Deep learning algorithms is Eleven Labs (Barrington et al., 2023). ElevenLabs can capture the emotion of human speech. The tool can clone voices or generate new ones (Elevenlabs, n.d.).

Voice-cloning is a process of duplicating someone's voice by using a computer. A person's voice can be created by getting speech examples of the person's voice. The algorithm learns from a lot of speech examples and breaks them down into phonemes, tiny speech sound components that are important to the algorithm's working since they serve as the fundamental building blocks that define the unique characteristics of the desired speech. Those phonemes are used to replicate speech (ElevenLabs Team, 2023).

Humans have already reached a point where it can be hard to distinguish AI from a human voice, since human voices with emotional nuances can be cloned, with text-to-speech technologies (TTS). This might blur the lines between an AI voice and a human voice according to Milewshi et al. (2023), potentially also in the field of journalism.

AI's Potential to Change Journalists' Roles

As already stated in the introduction, AI is becoming more prevalent in journalism. Complex generative tasks such as the generation of text, pictures, video, and speech are processes made possible by deep learning and can be useful for various journalism processes (LeCun et al. 2015; Marconi, 2020). A concern is that these new news processes might start an irreversible trend toward growing news automation, influenced by commercial interests. This might result in journalists acting more like maintainers of the news flow and deviating further from essential human skills like dealing with sources, ensuring their diversity and trustworthiness, and providing well-rounded views on unfolding events (Opdahl et al., 2023). AI taking over tasks would mean that the abilities required of journalists may change. The ability to critically evaluate news produced by or assisted by AI, or to have a clear

knowledge of what activities AI can be used for and where its limitations lie, may become more crucial.

Perception of AI use in journalism

Leading news companies, such as the Washington Post and the New York Times, use AI in some cases for certain subjects to generate whole articles from the start and publish them automatically (Longoni, 2020; Marconi, 2020). Even when AI is used to create content, the byline is rarely credited to an AI system (Dörr, 2015). Readers cannot detect whether an AI was utilized from the text alone without this notification (Köbis & Mossink, 2021; Kreps et al., 2020). In the past, publishing anonymously was seen as a way to protect journalists. However, this contradicts values like transparency and accountability. The consumer needs to know who the sources and authors are. If there is no information about a text's algorithmic nature and resources, it is hard for the audience to decide if they can trust the information (Dörr & Hollnbuchner, 2016). Longoni et al.'s (2020) research reveals that when news created by AI is labeled in that manner, consumers perceive news headlines written by AI as less accurate than those written by humans. Their research implies that people are more skeptical of AI-generated news than human-authored news. This skepticism about AI-generated news may go beyond the written word.

Perception of TTS

When looking at AI-generated speech, people can also be skeptical. Especially when people suspect they are exposed to an AI-generated voice but are not sure. Knowing whether news consumers are listening to a human, or an AI voice is crucial because that is a requirement for perceived credibility (Heiselberg et al., 2022). Besides transparency about the narrator, skepticism towards AI can also depend on the context. According to the findings of Noah et al. (2021), there is a noteworthy distinction between a human voice and a human-like computer-generated voice in the interchangeability across different usage situations. The research states that participants prefer a human voice during personal interactions instead of an AI voice.

When looking at a news-related context, A study conducted by Afonso and Almeida (2023) examined the potential benefits of the use of TTS in News Podcasts. According to their findings, a human voice was perceived as more favorable than a synthesized voice because it scored higher on

suitability in the context of the news podcast. It is therefore to be expected that a human voice would be preferred over a synthetic voice in a news-related context. However, the more human-like the synthesized voice sounded, the more people became accepting of it. Still, acceptance of a synthetic voice, even if it is extremely human-like, does not have to imply that there is a preference for an AI-generated voice. Research of Kim et al. (2022) states that AI voices are not identical to human voices. The human voice has unique acoustic characteristics, according to research conducted over the previous century (Nass and Steuer, 1993). As a result, in tiny acoustic characteristics, individuals may still detect something off. For example, in the delivery of emotion. Integrating human-like emotion into an AI voice remains a difficult job since the unique feature combinations used in voice synthesis are presently undefined according to (Yamamoto et al., 2018). Furthermore, it has been demonstrated that humans are more sensitive to human voices than to AI-voices. In an EEG study conducted by Chen (2023), people's brains showed more activity when a human voice was reading the news, compared to an AI voice reading the news. This impacts how people feel about the news when it comes to emotional or regular news. They might trust it less when it is read by AI voices. On the other hand, research by Heiselberg et al. (2022) investigated how Danish radio listeners judge the credibility of the news content after hearing a news broadcast read by an AI voice. Some participants find it an advantage that an AI-voice is not able to affect the radio news consumer intentionally or unintentionally by expressing emotions, which is beneficial for the credibility of news reading. However, other participants found that credibility was questioned when the AI voice read the news. They believe that a human voice can have a positive impact on a credible relationship between the news consumer and the news media organization. Therefore, the following hypotheses are stated:

H1 A voice-over narrated by a journalist in a news video is expected to be perceived as more pleasant compared to a voice-over narrated by a voice-cloning AI.

H2 A voice-over narrated by a journalist in a news video is expected to be perceived as more credible compared to a voice-over narrated by a voice-cloning AI.

Heiselberg et al. (2022) state that whether the message and the source behind the news broadcast are considered credible is the most important factor in determining if an automated news reading is credible. Also, according to Noah et al. (2021), it is necessary to inform users that the voice

they are listening to is generated by AI can reduce the chance of deception. This suggests that when people are not sure whether they are listening to AI or a human, they might doubt the source and find it less credible. That is the reason why the following hypothesis is stated:

H3: Individuals who suspect they are exposed to a voice-over narrated by a voice-cloning AI are likely to perceive the voice-over as less credible.

The difference in perception of TTS use in human interest and factual news

According to Reinemann et al. (2011) Soft news such as human interest, has a more emotional style, suggesting people expect a human touch in the voice-over of human interest videos. Affect values are widely used by news journalists; nevertheless, in hard news stories, journalists' emotions must be avoided to improve factuality and objectivity, which are beneficial to credibility (Appelman and Sundar, 2015; Stenvall, 2014). According to Stenvall (2014), emotions can simply be present in the events being reported. For example, the emotional impact of a tragic event. However, journalists seek to tell the events with a focus on facts and information rather than personal feelings. A human touch in the voice-over might therefore be less expected. This suggests that an AI-narrated factual news video might be perceived as more pleasant and credible compared to an AI-narrated voice-over, which leads to the following hypotheses:

H4 Voice-overs narrated by a voice-cloning AI in factual news videos are perceived as more pleasant compared to voice-overs narrated by a voice-cloning AI in human interest videos.

H5 Voice-overs narrated by a voice-cloning AI in factual news videos are perceived as more credible compared to voice-overs narrated by a voice-cloning AI in human interest videos.

Method

Design

To assess the hypotheses, an experiment with a between-subject design was conducted. The type of voice-over (journalist-narrated or AI-narrated), the type of news video (human interest or factual), and whether participants thought they were exposed to AI-narrated voice-over or journalist-narrated voice-over (perception of the narrator) served as the independent variables. The dependent variables were perceived pleasantness and perceived credibility.

Participants

A sample of $N = 256$ participated in the survey. However, 58 responses were removed as they were incompletely filled out. $N = 198$ people were included in the study and randomly exposed to one of the twelve conditions as can be seen in Table 1.

Table 1

Conditions

Condition	News type	Narrator
1	Factual news video 1	Journalist-narrated
2	Factual news video 1	AI-narrated
3	Factual news video 2	Journalist-narrated
4	Factual news video 2	AI-narrated
5	Factual news video 3	Journalist-narrated
6	Factual news video 3	AI-narrated
7	Human interest news video 1	Journalist-narrated
8	Human interest news video 1	AI-narrated
9	Human interest news video 2	Journalist-narrated
10	Human interest news video 2	AI-narrated
11	Human interest news video 3	Journalist-narrated
12	Human interest news video 3	AI-narrated

The participants were approached through SurveySwap and social media channels including Instagram, LinkedIn, and Facebook. Also, leaflets were handed out with a QR code for the survey at Tilburg University. Participants in the study ranged in age from 17 to 84 years old, ensuring a broad representation across several age groups. Most participants (63.1%) were young adults aged 17 to 35.

Middle-aged adults aged 36 to 55 represented 17.7%, whereas 11.1% were older adults (55 and over). 36,4% of the participants live or lived in Brabant.

Stimuli

The factors that were manipulated in this research were the type of narrator and type of news video. Six news videos were obtained from Omroep Brabant and were selected based on sharing the same narrator and the same form of video: YouTube Shorts. This ensured that the videos could not have a significant difference in terms of video length and word count of the voice-over. The videos were divided evenly into two groups: three of them were categorized as factual news videos and the other three as human interest videos. This was done to enhance ecological validity. Three videos were classified as human interest because of their emphasis on human experiences, narrativity, and emotion (León, 2008; Piazza & Haarman, 2011). The categorization of the other three videos is attributed to their factual focus. The voice-over focused mostly on presenting facts (Gobetz, 1992; Sahu & Majumdar, 2017).

The subjects of the human interest news videos can be seen in Table 2, while the subjects of the factual news videos are shown in Table 3. There were two different voice-over options for every video: the original journalist voice-over or an AI voice-over, resulting in a total of 12 videos.

Table 2

Information about Human Interest News Videos

Human interest news videos	About	Wordcount	Length of video
Brabantse blaaskapel naar Oktoberfest in Dubai	A brass band from Brabant is asked to perform at an Oktoberfest in Dubai.	48	0:41
Knalprijs voor winnaars miljoenenjacht	Winston from the Postcode Lottery surprises Maxime at her house with a check of	74	0:37

Human interest news videos	About	Wordcount	Length of video
	almost 200.000 euros and a confetti cannon, but the confetti explodes in her face.		
Trampolinespringer Milco valt in de prijzen	Milco became two timed Dutch champion trampoline jumping. He talks about how much effort it takes to achieve that success.	58	1:00

Table 3

Information about The Factual News Videos

Factual news videos	About	Wordcount	Length of video
Weer brand in de Efteling	A second fire incident during the Raveleijn show in the Efteling amusement park. The theatre was temporarily evacuated.	59	0:18
Auto valt stil op spoorwegovergang	A car broke down on a railway crossing in Bergen op Zoom. The driver managed to escape but the vehicles were damaged.	75	0:25
Ongeluk tussen motoren: motorrijders zwaargewond	An incident during a motorcycle tour where two motorcyclists got serious injuries in a crash involving four motorcycles in Andel.	51	0:20

Tool

The tool that was used to duplicate the voice-over of the journalist was an online Generative AI Text to Speech & Voice Cloning by ElevenLabs.

Measurement

To measure the voice-over's credibility of the news video, a voice credibility measurement scale was used. Research by Appelman and Sundar (2015) investigated how message credibility can be measured within and outside journalism. They proposed ten adjectives that describe message credibility. This research has selected the terms that can be used as subscales to measure voice credibility. Those adjectives are shown in Table 4.

Table 4

Voice Credibility Measurement Scale

Indicator	Item	Statement	5-point Likers scale
Voice credibility	Authorative	The news video's narrator sounded authoritative.	Strongly disagree –strongly agree
Voice credibility	Reliable	The news video's narrator sounded reliable.	Strongly disagree –strongly agree
Voice credibility	Trustworthy	The news video's narrator sounded trustworthy.	Strongly disagree –strongly agree
Voice credibility	Authentic	The news video's narrator sounded authentic.	Strongly disagree –strongly agree
Voice credibility	Believable	The news video's narrator sounded trustworthy.	Strongly disagree –strongly agree

Internal consistency of the 5-point Likert scale measuring perceived credibility was analyzed by conducting a reliability analysis, indicating it was statistically good with Cronbach's α coefficient .82.

A voice pleasantness measurement scale was created to measure the pleasantness of the voice-over of the news video, as shown in Table 5. Kit et al (2023) used the terms comprehensibility, pleasantness, naturalness, and human likeness for their Likert scale to investigate the perception of the voice-over in explainer videos. Schreibelmayr and Mara (2022) state that the more realistic a voice is, the more pleasant a voice-over is. That is why naturalness and human likeness were chosen as subscales of pleasantness in this research because those terms are strongly related to realism (Schreibelmayr & Mara, 2022). Eadie and Doyle (2005) state that voice pleasantness is strongly influenced by suprasegmental information like intonation and pace. Those terms are also chosen as subscales. Clarity is also an indicator that focuses on pleasantness according to the study of Burgoon et al. (1990). The terms pace, clarity, and intonation that were used as subscales are also subscales of the auditory perceptual analysis. This is a scale to measure voice quality. According to Burgoon et al. 1990, voice quality influences voice pleasantness, which indicates that the terms are suitable subscales to measure pleasantness with.

Table 5

Voice Pleasantness Measurement Scale

Indicator	Item	Statement	5-point Likert scale
Pleasantness	clarity	I understood what the voice-over was telling very well	Strongly disagree –strongly agree
Pleasantness	Naturalness	The news video's narrator sounded natural.	Strongly disagree –strongly agree

Indicator	Item	Statement	5-point Likert scale
Pleasantness	Human-likeness	The news video's narrator sounded human-like.	Strongly disagree –strongly agree
Pleasantness	pace	The narration's tempo and rhythm were well-maintained.	Strongly disagree –strongly agree
Pleasantness	intonation	The narrator's intonation aligned with the content	Strongly disagree –strongly agree

Internal consistency was evaluated by conducting a reliability analysis, indicating that the 5-point Likert scale measuring perceived pleasantness was statistically good with Cronbach's α coefficient of .83.

Procedure

The survey began with an informed consent form, which the participant had to read and agree with to participate in the study.

Before exposure to the news videos, the participants were asked about certain demographic information such as age and whether the participant lives or lived in Brabant.

After watching the news video, participants were asked to fill in 5-point Likert scales in which news perception in terms of credibility and pleasantness was measured. Participants were also asked if they were familiar with the narrator since familiarity can influence credibility. People are inclined to make mental shortcuts when judging news headlines, according to a study by Pennycook and Rand (2021). Familiarity is one of the main ways to intuitively believe in the news. When they are familiar with that narrator, they might trust that narrator more and therefore influence news perception. Control questions were also asked about perception of the news itself by asking how credible the news itself

was. Participants might already be skeptical about certain news subjects which could influence the credibility and pleasantness of the voice-over as well.

Participants were told at the end of the survey whether they had been listening to an AI-narrated voice or an actual journalist's voice, depending on their condition. Also, the purpose of the experiment was explained, and a disclaimer was provided.

Data analysis

To study the research question and hypotheses, statistical tests were performed in the statistical program SPSS version 29.

A two-way MANOVA was performed to measure H1, H2, H3, H4 and H5.

H1 *(A voice-over narrated by a journalist in a news video is expected to be perceived as more pleasant compared to a voice-over narrated by a voice-cloning AI)* and H2 (*A voice-over narrated by a journalist in a news video is expected to be perceived as more credible compared to a voice-over narrated by a voice-cloning AI)* were related to the main effect.

H3 (*Individuals who suspect they are exposed to a voice-over narrated by a voice-cloning AI are likely to perceive the voice-over as less credible),* H4 (*Voice-overs narrated by a voice-cloning AI in factual news videos are perceived as more pleasant compared to voice-overs narrated by a voice-cloning AI in human interest videos.),* and H5 *(Voice-overs narrated by a voice-cloning AI in factual news videos are perceived as more credible compared to voice-overs narrated by a voice-cloning AI in human interest videos.)* were related to the interaction effect.

Results

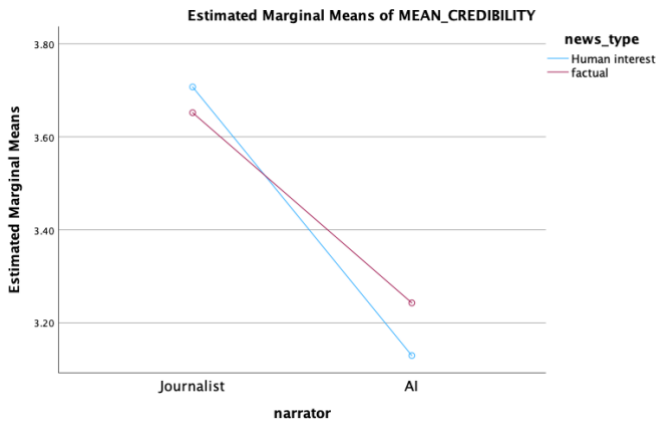
A Multivariate Analysis of Variance (MANOVA) was performed to examine the effects of the independent variables (news type, type of narrator, and perception of narrator).

H1 proposed that a voice-over narrated by a journalist in a news video was expected to be perceived as more credible compared to a voice-over narrated by a voice-cloning AI. The analysis revealed a significant main effect $F(2,189) = 9.778$, $p < .001$; Wilks $\mathbf{\Lambda}$ .906. A voice-over narrated by a journalist scored significantly higher on perceived credibility ($M = 3.64$) than a voice-over narrated by a voice-cloning AI ($M = 3.28$), suggesting that a voice-over narrated by a journalist in a news video is perceived as more credible.

A main effect was also found for H2 which stated that a voice-over narrated by a journalist in a news video was expected to be perceived as more pleasant compared to a voice-over narrated by a voice-cloning AI $F(2,189) = 9.778$, $p < .001$; Wilks **Λ** .906. A voice-over narrated by a journalist scores significantly better on perceived pleasantness ($M = 3.92$) than a voice-over narrated by a voice-cloning AI ($M = 3.56$). This indicates that a voice-over narrated by a journalist in a news video is perceived as more pleasant ($p < .001$).

Figure 1

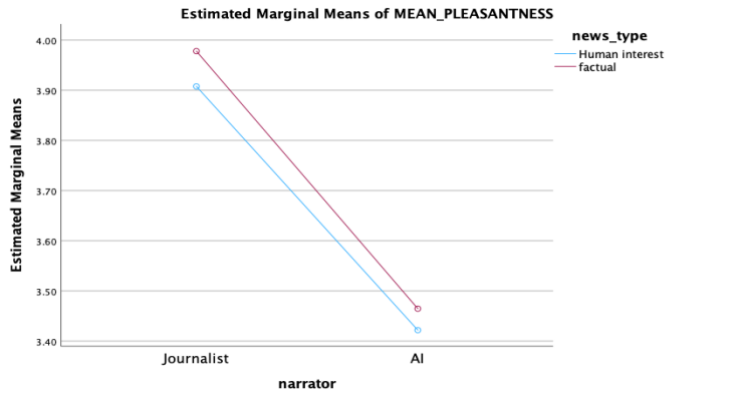
The Difference in Perceived Credibility Between Journalist- and AI-Narrated Voice-overs



Note: Line chart displaying the difference between journalist- and AI-narrated voice-overs in perceived credibility with no significant effect of news type.

Figure 2

The Difference in Perceived Pleasantness Between Journalist- and AI-narrated Voice-overs



Note: Line chart displaying the difference between journalist- and AI-narrated voice-overs in perceived pleasantness with no significant effect of news type.

An interaction effect was found for H3 which states that individuals who suspect they are exposed to a voice-over narrated by a voice-cloning AI are likely to perceive the voice-over as less credible. Participants who thought they were exposed to a voice-over narrated by a voice-cloning AI indeed perceived the voice-over as less credible $F(2,189) = 8.671$, $p < .001$; Wilks **Λ** .916. This indicates that a journalist-narrated voice is perceived as more credible than an AI-narrated voice.

H4 hypothesized that a voice-over narrated by a voice-cloning AI in factual news videos would be perceived as more pleasant compared to a voice-over narrated by a voice-cloning AI in human interest videos. The analysis revealed no significant interaction effect: $F(2,189) = .166$ $p = .847$; Wilks **Λ** .995. This implies that no difference was found between the pleasantness score of the AI-narrated factual news video ($M = 3.61$) and the pleasantness score of the AI-narrated human interest video ($M = 3.49$).

H5 proposed that a voice-over narrated by a voice-cloning AI in factual news videos would be perceived as more credible than a voice-over narrated by a voice-cloning AI in human interest videos. The credibility score of AI-narrated factual news videos ($M = 3.340$) and the credibility score of AI-narrated human interest videos ($M = 3.2$) showed no significant interaction effect: $F(2,189) = 0.166$, $p = .857$; Wilks **Λ** .995. This result suggests that no difference was found between the credibility score of AI-narrated factual news videos and the credibility score of AI-narrated human interest videos.

Exploratory analysis

An exploratory analysis was conducted to investigate if participants were aware of their exposure to AI-generated content. The frequency of mistaken recognition could be an indicator of the quality of AI-generated voices since repeated occurrences of misjudgment may signal an improvement in accuracy. A chi-square test ($p < .001$) shows a significant difference in the distribution of people thinking they were listening to an AI-narrated voice-over or a journalist-narrated voice-over between the AI narrator and the human narrator. Table 6 shows this distribution. 59 out of the 95 exposed to an AI-narrated voice-over thought it was a real human voice. An online binomial test calculator showed that the probability of exactly, or more than, 59 (K) out of 95 (n) is $p < .01$. This implies that

significantly more people who were exposed to an AI-narrated voice-over thought it was a real journalist. In addition, among participants exposed to the journalist-narrated voice-over, 11 out of 103 thought they were listening to AI. The binomial test showed that the probability of exactly, or less than, 11 (K) out of 103 (n) is $p < .001$. This suggests that significantly more people who were exposed to a journalist-narrated voice-over thought the journalist was indeed a real journalist. In essence, the AI voice was recognized as an AI voice more frequently. However, a significantly larger portion of the individuals still did not realize they were dealing with an AI voice.

Table 6

Distribution Of Perceived Vs. Actual Narrator

Count	Narrator			
		Journalist	AI	Total
MEAN_AI_OR_NOT	AI	11	36	47
	JOURNALIST	92	59	151
Total		103	95	198

Note: shows the participant's suspicion of exposure to or journalist-narrated voice-over AI (MEAN_AI_OR_NOT) in either a journalist-narrated condition or AI-narrated condition (narrator).

Discussion

The research question addressed in this study was: 'What is the effect of an AI-narrated voice-over on the news perception of consumers compared to a journalist-narrated voice-over, and does the type of news (factual- vs. human interest news) influence this effect?' The goal of this study was to learn more about the influence of AI on people's perceptions of news videos, with a particular focus on voice-overs because there was a lack of information in the existing literature on this topic. The aim was to get insights into whether people can tell the difference between AI and human narration, the consequences of AI use for news perception, and provide guidance on how journalists can cope with AI.

One of the main findings of the study was that people perceived a journalist-narrated voice-over as significantly more pleasant and credible than an AI-narrated voice-over (H1 and H2). This is in line with the expectation that people would perceive a journalist-narrated voice-over as more pleasant and credible than an AI-generated voice-over. However, participants were often wrong about the type of narrator they were listening to. Most of the participants who were exposed to the AI-narrated voice-over thought they were listening to a real human, and still, the journalist condition scored higher on pleasantness and credibility. This seems contradictory since participants who thought they were exposed to an AI-narrated voice perceived the voice-over as less credible (H3). One possible explanation for these contradictory results could be that people indeed did not hear that the AI-narrated voice was not human, but they were skeptical of the possibility of AI use and hence perceived it as less pleasant and trustworthy. Being transparent about using AI might have caused a more credible and pleasant perception of the AI narrator. This is in line with the research of Dörr and Hollnbuchner (2016). They state that in the lack of information about a content's algorithmic origin, it becomes challenging for the audience to judge the credibility of the information. Another explanation could be that participants unconsciously heard something was off while listening to the AI voice-over but did not realize it was not a human. This might depend on small cues in intonation, pace, or emotion. According to Hirai et al. (2015), variation in emotional tone and intonation of an AI voice affects the quality of the voice. Listeners may unconsciously notice a difference when these small cues differ from human-like patterns, adding to a feeling that "something is off" even if they cannot explicitly identify it as AI-generated.

Another insight was that whether a news video is human interest or factual does not influence the perception of pleasantness and credibility. This is not in line with the expectation that factual news videos would be perceived as more credible as well as pleasant compared to an AI-narrated human interest video. Accidents were shown in the factual news videos, where the narrator might have added a human touch or passed on a little emotion in the voice-over. The human interest videos, on the other hand, might have needed a little more emotion. As a result, the line between human interest and factual videos might not have been clear enough. The videos that were chosen and how they were assigned to either human interest or factual might have been the factors why no significant result was

found. One of the reasons that the human interest videos were labeled as such, was because of the emphasis on emotion. According to Stenvall (2014), emotions can simply be present in the events being reported, as well as in factual news. This implies an overlap between the two concepts. Also, other studies like Reinemann et al. (2011) and Glogger (2018) state that it is difficult to distinguish the concepts of hard news and soft news and that they have a significant overlap. This might have played a role in why no significant result was found in the news type.

The insights that are gained from this research could be valuable for journalists since it sheds light on how people perceive an AI-narrated voice-over and emphasizes the challenge of distinguishing AI-narration and human narration. It also suggests that journalists should be transparent about using AI.

Limitations and future research

The use of the ElevenLabs tool to generate AI-narrated voice-overs presents an important limitation to the study since the generated voice-overs might have had the potential to be even more human-like. The variety in outcomes after each generation, due to the amount of input data, highlight the challenge of reaching the highest potential human-like quality (Elevenlabs, n.d.). This limitation must be known since results might have been different when the quality of the AI-narrated voice-over was more human-like which may have affected the participants' perception.

Another limitation has to do with the question: 'Did you think you were exposed to an AI-narrated voice-over?' This question may have unintentionally influenced the participants' perceptions by directing their attention to the possibility of having encountered an AI-narrated voice. Participants who did not initially consider listening to an AI-narrated voice-over while they were listening to one, might have started to notion when reading the question and filled in they thought it was AI. This limitation is noteworthy since it might have influenced the results' validity when it comes to testing H3.

The third limitation is that every participant was only exposed to one video. It is recommended that future research use a within-subjects approach, with each participant seeing multiple videos, either only with AI-narrated voice-overs or only journalist-narrated voice-overs, to enhance internal validity.

Further research is recommended to explore the specific contexts in which AI-generated might be preferred or perceived as just as credible and pleasant compared to a human voice. For instance, examining this in the context of factual news videos where information always follows a consistent formula and may not require a human touch, would be valuable. Investigating various contexts can contribute to a better understanding of the situations in which AI voices are accepted or even preferred, providing information on their possible uses and limitations.

Another suggestion for future studies involves delving deeper into attitudes toward AI use in journalism and exploring how familiarity with AI impacts these attitudes. According to Komiak and Bensbasat (2006), familiarity improves the need to adopt via cognitive and emotional trust. Exploring if people who are familiar with AI voices are more accepting of AI might give insights into AI use in journalism.

Lastly, future studies should explore whether a completely AI-generated voice would be just as acceptable in terms of credibility and pleasantness as the voices of journalists, considering the ethical issues associated with voice cloning. If broadcasters prefer to clone voices, strict ethical guidelines and agreements are recommended.

Conclusion

In conclusion, this study shows a preference for a human journalist in terms of perceived credibility and pleasantness in the voice-over of news videos compared to a voice-over narrated by a voice-cloning AI. The findings do not show a significant influence of the type of news on pleasantness and credibility. Based on the outcomes of this study, it is not yet recommended to use voice-over automation for news videos. Further research is necessary to explore which context AI narration could suit, whether familiarity with AI makes it more acceptable over time, and if generating a new voice instead of cloning would be just as effective. This is important because of ethical concerns associated with cloning voices and it requires careful consideration and strict guidelines. Regardless of the use of AI voice-overs in news videos, human engagement continues to be necessary for transparency and ethical considerations.

Literature

Afonso, M., & Almeida, P. (2023). Combining Text-to-Speech services with conventional voiceover for news oralization. In *Communications in computer and information science* (pp. 68–79). https://doi.org/10.1007/978-3-031-45611-4_5

Alkhatib, R., Sahwan, W., Alkhatieb, A., & Schütt, B. (2023). A brief review of machine learning algorithms in forest fires Science. *Applied sciences*, *13*(14), 8275. https://doi.org/10.3390/app13148275

Alzubi, J. A., Nayyar, A., & Kumar, A. (2018). Machine Learning From Theory to Algorithms: An Overview. *Journal of physics*, *1142*, 012012. https://doi.org/10.1088/1742-6596/1142/1/012012

Appelman, A., & Sundar, S. S. (2015). Measuring message credibility. *Journalism & Mass Communication Quarterly*, *93*(1), 59–79. https://doi.org/10.1177/1077699015606057

Barrington, S., Barua, R., Koorma, G., & Farid, H. (2023). Single and Multi-Speaker Cloned Voice Detection: From perceptual to learned features. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2307.07683

Burgoon, J. K., Birk, T., & Pfau, M. (1990). Nonverbal behaviors, persuasion, and credibility. *Human Communication Research*, *17*(1), 140–169. https://doi.org/10.1111/j.1468-2958.1990.tb00229.x

Chen, G. (2023). AI Voices Reduce Cognitive Activity? A Psychophysiological study of the media effect of AI and human newscasts in Chinese journalism. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1243078

Coleman, F. (2020). *A human algorithm: How Artificial Intelligence is Redefining Who We Are*. Melville House UK

Cushion, S. (2015). *News and Politics: The rise of live and interpretive journalism*. http://ci.nii.ac.jp/ncid/BB1855490X

Deuze, M., & Beckett, C. (2022). Imagination, Algorithms and News: Developing AI literacy for Journalism. *Digital journalism*, *10*(10), 1913–1918. https://doi.org/10.1080/21670811.2022.2119152

Dörr, K. (2015). Mapping the field of algorithmic journalism. *Digital Journalism*, *4*(6), 700–722. https://doi.org/10.1080/21670\811.2015.1096748

Dörr, K., & Hollnbuchner, K. (2016). Ethical challenges of Algorithmic journalism. *Digital Journalism*, *5*(4), 404–419. https://doi.org/10.1080/21670811.2016.1167612

Eadie, T. L., & Doyle, P. C. (2005). Classification of Dysphonic Voice: Acoustic and Auditory-Perceptual measures. *Journal of Voice*, *19*(1), 1–14. https://doi.org/10.1016/j.jvoice.2004.02.002

ElevenLabs. (n.d.). *`Overview*. https://elevenlabs.io/docs/speech-synthesis/overview

ElevenLabs (z.d.). Python Guide - ElevenLabs. https://elevenlabs.io/docs/api-reference/integration-guides/python-text-to-speech-guide

ElevenLabs Team. (2023, 7 juli). *Voice conversion*. ElevenLabs Blog. https://elevenlabs.io/blog/voice-conversion/

Glogger, I. (2018). Hard and Soft News – A Twofold Approximation to a Key Concept in Journalism Research. https://kola.opus.hbz-nrw.de/files/1731/Dissertation.pdf

Goertzel, B., & Wang, P. (2007). *Advances in artificial general intelligence: Concepts, Architectures and Algorithms : Proceedings of the AGI Workshop 2006*. IOS Press.

Heiselberg, L., Blom, J. N., & Van Dalen, A. (2022). Automated news reading in the neural age: audience reception and perceived credibility of a news broadcast read by a neural voice. *Journalism Studies*, *23*(8), 896–914. https://doi.org/10.1080/1461670x.2022.2052346

Hillaire, G., Iniesto, F., & Rienties, B. (2019). Humanising Text-to-Speech through emotional expression in online courses. *Journal of interactive media in education*, *2019*(1). https://doi.org/10.5334/jime.519

Hodari, Z., Watts, O., & King, S. (2019). Using generative modelling to produce varied intonation for speech synthesis. *arXiv preprint arXiv:1906.04233*.

Jakhar, D., & Kaur, I. (2019). Artificial intelligence, machine learning and deep learning: Definitions and differences. *Clinical and Experimental Dermatology*, *45*(1), 131–132. https://doi.org/10.1111/ced.14029

Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, *31*(3), 685–695. https://doi.org/10.1007/s12525-021-00475-2

Jung, J., Song, H., Kim, Y., Im, H., & Oh, S. (2017). Intrusion of software robots into journalism: The public's and journalists' perceptions of news written by algorithms and human journalists. *Computers in Human Behavior*, *71*, 291–298. https://doi.org/10.1016/j.chb.2017.02.022

Kim, J., Xu, K., & Merrill, K. (2022). Man vs. machine: human responses to an AI newscaster and the role of social presence. *Social Science Journal*, 1–13. https://doi.org/10.1080/03623319.2022.2027163

Kit, L. W., Yuin-Y, C., Zulkifli, B., & Nie, K. S. (2023). Perception of university students towards the use of artificial intelligence-generated voice in explainer videos. https://fslmjournals.taylors.edu.my/wp-content/uploads/SEARCH/SEARCH-2023-Special-Issue-SEARCH-Conf2022/SEARCH-2023-P6-15-SEARCHConf2022.pdf

Köbis, N., & Mossink, L. (2021). Artificial intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry. *Computers in Human Behavior*, *114*, 106553. https://doi.org/10.1016/j.chb.2020.106553

Kok, J. N., Boers, E. J., Kosters, W. A., Van der Putten, P., & Poel, M. (2009). Artificial intelligence: definition, trends, techniques, and cases. *Artificial intelligence*, *1*, 270-299.

Komiak, & Benbasat. (2006). The effects of personalization and familiarity on trust and adoption of recommendation agents. *Management Information Systems Quarterly*, *30*(4), 941. https://doi.org/10.2307/25148760

Kreps, S. E., McCain, R. M., & Brundage, M. (2020). All the news that's fit to fabricate: AI-Generated text as a tool of media misinformation. *Journal of experimental political science (Print)*, *9*(1), 104–117. https://doi.org/10.1017/xps.2020.37

Kühne, K., Fischer, M. H., & Zhou, Y. (2020). The human takes it all: humanlike synthesized voices are perceived as less eerie and more likable. evidence from a subjective ratings study. *Frontiers in Neurorobotics*, *14*. https://doi.org/10.3389/fnbot.2020.593732

LeCun, Y., Bengio, Y., & Hinton, G. E. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

León, B. (2008). Science related information in European Television: A study of prime-time news. *Public Understanding of Science*, *17*(4), 443–460. https://doi.org/10.1177/09636625056073089

Longoni, C., Fradkin, A., Cian, L., & Pennycook, G. (2022). News from generative artificial intelligence is believed less. *2022 ACM Conference on Fairness, Accountability, and Transparency*. https://doi.org/10.1145/3531146.3533077

Marconi, F. (2020). 2. ENABLERS: THE AI TECHNOLOGIES DRIVING JOURNALISTIC CHANGE. In *Columbia University Press eBooks* (pp. 55–128). https://doi.org/10.7312/marc19136-005

Meier, K. (2009). TRANSPARENCY IN JOURNALISM. Credibility and trustworthiness in the digital future. *Actas II Congreso the Future of Journalism*. https://edoc.ku.de/13483/

Milewski, K., Zaporowski, S., & Czyżewski, A. (2023). Comparison of the ability of neural network model and humans to detect a cloned voice. *Electronics*, *12*(21), 4458. https://doi.org/10.3390/electronics12214458

Miroshnichenko, A. (2018). AI to bypass creativity. Will robots replace journalists? (The answer is "Yes"). *Information*, *9*(7), 183. https://doi.org/10.3390/info9070183

Mogaji, E. (2018). Emotional appeals in advertising. In *Emerald Publishing Limited eBooks* (pp. 25–46). https://doi.org/10.1108/978-1-78756-299-820181003

Najar, G. A. (2018). The changing nature of news reporting, story development and editing. *Journal of Media and Communication Studies*. https://doi.org/10.5897/jmcs2018.0631

Nass, C., & Steuer, J. (1993). Voices, boxes, and sources of messages. *Human Communication Research*, *19*(4), 504–527. https://doi.org/10.1111/j.1468-2958.1993.tb00311.x

Noah, B., Sethumadhavan, A., Lovejoy, J., & Mondello, D. (2021). Public perceptions towards synthetic voice technology. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *65*(1), 1448–1452. https://doi.org/10.1177/1071181321651128\

Opdahl, A. L., Tessem, B., Dang-Nguyen, D., Motta, E., Setty, V., Throndsen, E., Tverberg, A., & Trattner, C. (2023). Trustworthy journalism through AI. *Data and Knowledge Engineering*, *146*, 102182. https://doi.org/10.1016/j.datak.2023.102182

Pennycook, G., & Rand, D. G. (2021). The psychology of fake news. *Trends in Cognitive Sciences*, *25*(5), 388–402. https://doi.org/10.1016/j.tics.2021.02.007

Piazza, R., & Haarman, L. (2011). Toward a definition and classification of human interest narratives in television war reporting. *Journal of Pragmatics*, *43*(6), 1540–1549. https://doi.org/10.1016/j.pragma.2010.12.005

Pounds, G. (2012). Multimodal expression of authorial affect in a British television news programme. *Discourse, Context and Media*, *1*(2–3), 68–81. https://doi.org/10.1016/j.dcm.2012.03.001

Reinemann, C., Stanyer, J., Scherr, S., & Legnante, G. (2011). Hard and soft news: a review of concepts, operationalizations and key findings. *Journalism: Theory, Practice & Criticism*, *13*(2), 221–239. https://doi.org/10.1177/1464884911427803

Rodero, E., & Lucas, I. (2021). Synthetic versus human voices in audiobooks: The human emotional intimacy effect. *New Media & Society*, *25*(7), 1746–1764. https://doi.org/10.1177/14614448211024142

Sahu, I., & Majumdar, D. (2017). Detecting Factual and Non-Factual Content in News Articles. *CODS '17: Proceedings of the 4th ACM IKDD Conferences on Data Sciences*. https://doi.org/10.1145/3041823.3041837

Schreibelmayr, S., & Mara, M. (2022). Robot voices in daily life: Vocal Human-Likeness and application context as determinants of user acceptance. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.787499

Scott, D., & Gobetz, R. H. (1992). Hard News/Soft News Content of the National Broadcast Networks, 1972–1987. *Journalism Quarterly*, *69*(2), 406–412. https://doi.org/10.1177/107769909206900214

Sokolowski, R. (1988). Natural and Artificial Intelligence. *Daedalus*, *117*(1), 45–64. http://www.jstor.org/stable/20025138

Stenvall, M. S. I. (2014). PRESENTING AND REPRESENTING EMOTIONS IN NEWS AGENCY REPORTS. *Critical Discourse Studies*, *11*(4), 461–481. Https://doi.org/10.1080/17405904.2013.866588

Tumengkol, A. E. A. (2021). THE JOURNALISTIC CODE OF ETHICS AND PROFESSIONALISM OF THE MEDIA. *Jurnal Ilmiah Ilmu Komunikasi Communique*, *3*(2), 37–49. https://doi.org/10.31289/jikq.v3i2.47

Yamamoto, K., Takahashi, K., Kishiro, K., Sasaki, S., & Hayashi, H. (2018). Analysis of emotional expression by visualization of the human and synthesized speech signal sets — A consideration of audio-visual advantage-. *2018 International Workshop on Advanced Image Technology (IWAIT)*. https://doi.org/10.1109/iwait.2018.8369809

Appendix A

Survey components, exemplified by 'condition ,' including the Informed consent and the survey conclusion.

Perceptie nieuwsvideo

Start of Block: Toestemmingsverklaring

Q143 **Titel:** Inzicht verkrijgen in de perceptie van voice-overs in nieuwsvideo's

Onderzoeker: Anne Smink

Contactinformatie:
a.e.g.smink@tilburguniversity.edu

Omschrijving van het onderzoek: Deze enquête, uitgevoerd door Anne Smink, is onderdeel van de masterscriptie voor Tilburg University onder begeleiding van Chris van der Lee. Het doel van dit onderzoek is om de perceptie van voice-overs in nieuwsvideo's te begrijpen.

Onderzoeksprocedure: Als u ermee akkoord gaat om deel te nemen aan dit onderzoek, zullen er eerst een aantal persoonlijke vragen aan u worden gesteld. Vervolgens wordt er aan u gevraagd om een video te bekijken. Hiervoor is het van belang dat uw geluid aan staat. Deze video duurt maximaal een minuut lang. Na het bekijken van de video zult u de voice-over van deze video beoordelen. Daarna zult u ook de inhoud van het nieuws zelf beoordelen. De gehele procedure zal ongeveer 5 tot 8 minuten duren.

Meerwaarde: Uw deelname zou bij kunnen dragen aan het verkrijgen van inzichten die van waarde kunnen zijn binnen de journalistiek en mogelijk ook daarbuiten.

Risico's: Bij deelname aan dit onderzoek kan het zijn dat u wordt blootgesteld aan een van de video's met een gevoelig onderwerp: een treinongeluk of motorongeluk. Dit kan ongemak of emotionele gevoeligheid veroorzaken. De video toont echter geen daadwerkelijk ongeluk of verwondingen. Als u zich ongemakkelijk voelt bij het bekijken van deze beelden, wordt het afgeraden om verder deel te nemen aan dit onderzoek. Indien u besluit toch deel te nemen, dient u zich bewust te zijn van dit mogelijke risico.

Vertrouwelijkheid: Deelname aan dit onderzoek is anoniem. De gegevens die u verstrekt zullen alleen worden gebruikt voor dit onderzoek. Uw gegevens worden veilig bewaard en blijven beschikbaar voor enkel de onderzoeker en de begeleider tot aan de diploma-uitreiking. Vrijwillige deelname en recht op terugtrekken: Meewerken aan dit onderzoek is vrijwillig. U heeft het recht om u op ieder moment terug te trekken zonder gevolgen. U kunt op ieder moment contact opnemen (zie contact gegevens hieronder) met de onderzoeker om de informatie die u gedeeld heeft in te zien, aan te laten passen, of te laten verwijderen.

Contactgegevens: Indien u vragen of opmerkingen heeft, of als u extra informatie wilt verkrijgen over dit onderzoek, kunt u contact opnemen met Anne Smink via de mail: a.e.g.smink@tilburguniversity.edu. Lees dit formulier alstublieft zorgvuldig door.

Q142 Gaat u akkoord met het bovenstaande?

- ○ Ja (1)
- ○ Nee (2)

Skip To: End of Survey If Gaat u akkoord met het bovenstaande? = Nee

End of Block: Toestemmingsverklaring

Start of Block: Demografische vragen

Q1 Hoe oud bent u?

_____

Q2 Woont u/heeft u ooit in Brabant gewoond?

- ○ Ja (1)
- ○ Nee (2)

Skip To: Q4 If Woont u/heeft u ooit in Brabant gewoond? = Nee

---

Q3 Hoelang woont u/of heeft u in Brabant gewoond?

_____

Q4 Ben je bekend met Omroep Brabant?

- ○ Ja (1)
- ○ Nee (2)

End of Block: Demografische vragen

Start of Block: Perceptie Omroep Brabant

Q5 Hoe vaak kijkt u naar online nieuwsvideo's van Omroep Brabant

- ○ bijna nooit/nooit (1)
- ○ Maandelijks (2)
- ○ Elke twee weken (3)
- ○ Elke week (4)
- ○ Een paar keer per week (5)
- ○ Iedere dag (6)
- ○ Een paar keer per dag (7)

Skip To: End of Block If Hoe vaak kijkt u naar online nieuwsvideo's van Omroep Brabant = bijna nooit/nooit

Q6 Ik vind de video's van Omroep Brabant betrouwbaar

- ○ Sterk mee oneens (1)
- ○ Oneens (2)
- ○ Neutraal (3)
- ○ Eens (4)
- ○ Sterk mee eens (5)

End of Block: Perceptie Omroep Brabant

Start of Block: Conditie 1

C1 U krijgt nu een korte nieuwsvideo te zien. Hierbij is het van belang om aandachtig te kijken en luisteren met uw **geluid aan**. Na de video zullen er vragen worden gesteld over de video.

C1

Q146 Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: Conditie 1

Start of Block: Betrouwbaarheid meten video 1

Q18 Geef aan in hoeverre je het eens bent met de stellingen over de **voice-over** van de nieuwsvideo die u zojuist gezien hebt.

	Sterk mee oneens (1)	Mee oneens (2)	Neutraal (3)	Mee eens (4)	Sterk mee eens (5)
De voice-over spreker klonk gezaghebbend. (1)	○	○	○	○	○
De voice-over spreker klonk betrouwbaar. (2)	○	○	○	○	○
De voice-over spreker klonk professioneel. (3)	○	○	○	○	○
De voice-over spreker klonk authentiek. (4)	○	○	○	○	○
De voice-over spreker klonk geloofwaardig. (5)	○	○	○	○	○

End of Block: Betrouwbaarheid meten video 1

Start of Block: aangenaamheid meten video 1

Q19 Geef aan in hoeverre u het eens bent met de stellingen over de voice-over van de nieuwsvideo die u zojuist gezien hebt.

	Sterk mee oneens (1)	Mee oneens (2)	Neutraal (3)	Mee eens (4)	Sterk mee eens (5)
Ik begreep goed wat de voice-over spreker vertelde. (1)	○	○	○	○	○
De voice-over spreker sprak op een natuurlijke manier. (2)	○	○	○	○	○
De voice-over spreker klonk menselijk. (3)	○	○	○	○	○
Het tempo en ritme waarin de spreker van de voice-over sprak was aangenaam. (4)	○	○	○	○	○
De intonatie van de voice-over spreker paste bij het beeld en nieuws. (5)	○	○	○	○	○

End of Block: aangenaamheid meten video 1

Start of Block: Controle vraag

Q21 Ik herkende de spreker van de voice-over van andere nieuwsberichten.

- ○ Ja (1)
- ○ Nee (2)

End of Block: Controle vraag

Start of Block: Conditie 1 2e keer

Q143 Om de inhoud van het nieuws te kunnen beoordelen, krijgt nu de mogelijkheid om dezelfde video nog eens te bekijken. Als dit niet nodig is kunt u verder klikken.

Q144

End of Block: Conditie 1 2e keer

Start of Block: Inhoud meten 1

Q42 Geef aan in hoeverre u het eens bent met de stellingen over de **inhoud** van de nieuwsvideo die u zojuist gezien hebt.

	Sterk mee oneens (1)	Mee oneens (2)	Neutraal (3)	Mee eens (4)	Sterk mee eens (5)
De nieuwsvideo was compleet, er ontbrak geen belangrijke informatie. (1)	○	○	○	○	○
De inhoud van de nieuwsvideo kwam betrouwbaar over. (2)	○	○	○	○	○
De inhoud van de nieuwsvideo kwam professioneel over. (3)	○	○	○	○	○
De inhoud van de nieuwsvideo was objectief. (4)	○	○	○	○	○
De inhoud van de nieuwsvideo kwam authentiek over. (5)	○	○	○	○	○
De inhoud van de nieuwsvideo kwam geloofwaardig over. (6)	○	○	○	○	○

End of Block: Inhoud meten 1

Start of Block: AI of mens 1

Q40 Dacht u dat de voice-over stem gegenereerd was met Artificial Intelligence?

- ○ Ja (1)
- ○ Nee (2)

Q41 Waarom dacht u dit?

_____

End of Block: AI of mens 1

Start of Block: End of Survey menselijke stem 1

End of survey Bedankt voor uw deelname aan deze enquête. Uw beoordeling van de video's van Omroep Brabant, waarin de voice-over ofwel een echte mens ofwel door AI gegenereerd was, wordt zeer op prijs gesteld.
U heeft zojuist geluisterd naar een stem van een mens: namelijk de stem van [redacted]. Hij is werkzaam bij Omroep Brabant en spreekt regelmatig voice-overs in.

Door uw mening te delen, heeft u geholpen inzicht te krijgen in de perceptie en herkenning van door AI gegenereerde stemmen. Het doel van dit onderzoek is om de huidige impact van AI op journalistiek en de mogelijke moeilijkheden te begrijpen en hoe de bijbehorende risico's en uitdagingen aangepakt kunnen worden.

Disclaimer:
De video's die in dit onderzoek worden gebruikt, zijn afkomstig van Omroep Brabant. In de video's zijn veranderingen aangebracht in voice-over en achtergrondgeluiden, om specifieke onderzoeksdoeleinden te ondersteunen. Dit onderzoek wordt niet uitgevoerd namens Omroep Brabant. De inzichten die uit dit onderzoek worden verkregen, hebben betrekking op de perceptie van voice-overs in nieuwsvideo's en kunnen waardevol zijn voor diverse partijen, waaronder nieuwsorganisaties zoals Omroep Brabant. Omroep Brabant heeft ingestemd met het gebruik van de video's voor dit onderzoek en is op de hoogte van het doel van het onderzoek.

End of Block: End of Survey menselijke stem 1

Appendix B

Table B1

Human interest video links

Human interest news videos	Link video with AI voice-over	Link with voice-over journalist
Brabantse blaaskapel naar Oktoberfest in Dubai	https://youtube.com/shorts/rQHM6zGGrlQ	https://youtube.com/shorts/3gm6EPY5THA
Knalprijs voor winnaars miljoenenjacht	https://youtube.com/shorts/GOtX_AY3PTk	https://youtube.com/shorts/pf4TT7XqNTM
Trampolinespringer Milco valt in de prijzen	https://youtube.com/shorts/Jk2kKXZ32mM	https://youtube.com/shorts/ZYD0QNdYwcs

Table B2

Factual news video links

Factual news video	Link video with AI voice-over	Link with voice-over journalist
Weer brand in de Efteling	https://youtube.com/shorts/C_9mH8Ty390	https://youtube.com/shorts/IqkjZTDWnAU
Auto valt stil op spoorwegovergang	https://youtube.com/shorts/fhGT6zLLUzc	https://youtube.com/shorts/3LBw86Rs9Ys
Ongeluk tussen motoren: motorrijders zwaargewond	https://youtube.com/shorts/QqDts2QYb0I	https://youtube.com/shorts/gU3d035KOfM