CHAPTER II
LITERATURE REVIEW

2.1 Theoretical Framework

2.1.1 Audiovisual Media

Audiovisual media is a learning medium that combines visual and audio elements which engages the senses of sight and hearing simultaneously, thereby enhancing students' understanding of the material being presented effectively (Affiati et al., 2021; Winarto et al., as cited in Manulang, Haifaturrahmah, Utami, Syaharuddin, and Ala, 2024). Audiovisual media encompasses several elements such as sound and moving images including videos, films, documentaries, and audio slides that convey meaning simultaneously through hearing and sight. Utami, Saifullah, Utama, and Wibowo (2020) note that audiovisual media can be used to represent educational content and moral values in a more engaging and effective manner. This is a highly strategic approach to providing varied learning experiences, particularly for students whose learning styles are more responsive to visual and auditory information. Through audiovisual media like documentary video, it can shape the public's perception and understanding of social, cultural, and issues in a more interactive and emotional way. According to Liao (2023), the impact of this media can shape the attitudes and opinions of the audience, not just in the delivery of information. Therefore, the ethical aspects and accuracy of audiovisual messages need to be considered in their use for public communication.

2.1.2 Audiovisual Translation

When it first emerged in the mid-1990s, Guillot (2020) noted that Audiovisual Translation (AVT) focused on establishing its place within the discipline of Translation Studies and identifying technical challenges and strategies in various forms of AVT, such as subtitling, dubbing, voice-over, and audio description. However, as it has

evolved AVT is no longer limited to purely technical and linguistic issues but has expanded into the realms of pragmatics and intercultural communication. AVT shapes representations of interpersonal relationships, politeness, and even the construction of cultural identity in the target text. Textural choices in AVT are influenced by technical features as well as the simultaneous presence of various linguistic and cultural frameworks. In this context, the translation process is determined by lexical meaning, audience expectations, and the resulting pragmatic effects. For example, research on the translation of profanity indicates a tendency toward mitigation in subtitling, yet its pragmatic effects on the audience are not always negative. This opens the door to understanding that the impact of AVT on the audience is complex and contextual. Long (2023) emphasizes the importance of considering the type of AVT used as each has distinct characteristics and challenges.

2.1.2.1 Subtitling

In Subtitling, Díaz Cintas and Remael (2021) defined it as a translation activity that delivers a written rendition of the original spoken interaction, typically displayed at the bottom of the screen. It conveys the dialogue between speakers and verbal information presented visually on screen (e.g. written letters or signs), and audibly (e.g. songs or voice over narration). In linguistic parameters, they classified it as intralingual and interlingual subtitles. Intralingual subtitles, or same-language subtitling (SLS), are subtitles that use the same language as the original dialogue in the content which is commonly used for Subtitle for the Deaf and Hard-of-Hearing (SDH). Intralingual subtitles are also used as a tool for foreign language learning, such as helping viewers understand specific accents or pronunciations that are difficult to grasp, improve their vocabulary and listening comprehension, and also to expand their cultural understanding through authentic audiovisual content. Meanwhile, interlingual subtitles translate dialogue from the Source Language (SL) into a different Target Language (TL), and can be divided into monolingual, bilingual, or multilingual subtitles.

2.1.2.2 Dubbing

Cary (1960, as cited in Díaz Cintas and Remael, 2021) described Dubbing as a *traduction totale* due to its complex linguistic demands, which involves replacing an audiovisual work's original dialogue track with a new track in the TL. Dubbing aims to convince viewers that on-screen characters speak the viewer's language. The dubbing process, according to Chaume (2026), involves several stages carried out by different professionals in the audiovisual industry. The process begins when a client (e.g. television station or streaming platforms), commissions a dubbing company to translate audiovisual content into the TL. A translator first produces a rough translation of the original dialogue, which is then adapted into a synchronized dubbing script that matches the lip movements, timing, and natural speech of the on-screen characters. Afterward, voice actors perform the translated dialogue under the supervision of a dubbing director and sound engineer in the recording studio. The recorded voices are then mixed with the original soundtrack, including music and sound effects, before undergoing quality control to ensure linguistic accuracy, synchronization, and audio quality prior to distribution.

2.1.2.3 Voice Over

Voice Over (VO) according to Díaz Cintas and Remael (2021) is a technique where the original speaker's words remain faintly audible while a translation is spoken aloud. Listeners will hear a few seconds of the original language, then the soundtrack's volume is lowered so the translation in the TL can be spoken over it. This technique, where all the dialogue is read in monotone voice by a man, is widely used in documentaries, news, and nonfiction programs to respect the original text faithfully. VO is common for television in Eastern European Countries like Russia and Poland.

2.1.2.4 Audio Description

Díaz-Cintas and Remael (2015, as cited in Díaz Cintas and Remael, 2021) stated that Audio Description (AD) is an accessibility service that provides verbal accounts of the pertinent visual aspects of an artwork or media product so that blind and visually impaired audiences can fully understand its structure and content. This would help people with sight loss to follow the plot, perceive characters' body language and facial expressions, appearance, their psychology, and locations where the action takes place. According to Prime Video Audio Description Style Guide (2024), the creation of AD involves visual content analysis, script writing, timing (spotting), recording, and both mixing and mastering.

2.1.3 Audiovisual Translator

Díaz Cintas and Remael (2020) argues that the demand for audiovisual translation will continue to grow as companies and organizations increasingly recognize the value of adapting their content into multiple languages to extend their global reach. According to Chaume (2020, as cited in Paramita and Sampurna, 2025), audiovisual translators work within multimodal environments where dialogue, visual cues, sound effects, and on-screen texts interact simultaneously to construct meaning. Gottlieb (2004, as cited in Paramita and Sampurna, 2025) says translators must work within limitations such as subtitle duration, synchronization, reading speed, and limited subtitle space in the screen. These conditions require them to use specific translation strategies, make the text shorter (reduction), use different words (paraphrasing), make it have the same meaning as the original (equivalence), change it to closer to TL (adaptation), or even make it longer (expansion) to ensure readability without losing the original meaning.

The professional profile of audiovisual translators has evolved substantially in recent years. According to Díaz Cintas and Remael (2020), linguistic competence, sociocultural awareness, and subject knowledge alone are no longer sufficient for operating effectively in this profession. Audiovisual translators are expected to possess

technical expertise, be familiar with information and communication technologies, and demonstrate proficiency in using increasingly sophisticated subtitling software and digital tools. They are also required to adapt quickly to new technologies and workflows as the audiovisual industry continues to evolve.

Audiovisual translators may work as freelancers or in-house professionals and provide their services to various sectors, including film and television production companies, streaming platforms, educational institutions, and multimedia industries. Technological advancements have further enabled audiovisual translators to collaborate remotely with clients and companies across different countries, creating a highly competitive yet increasingly accessible professional environment (Díaz-Cintas and Remael, 2020). Despite the growing importance of audiovisual translation in the global media industry, formal education and training in this field have traditionally received limited attention in many universities. Díaz Cintas and Remael (2020) points out that audiovisual translation was long learned mainly through hands-on professional experience rather than formal academic programs, creating a gap between industry demands and educational preparation. Consequently, audiovisual translators are expected to continuously develop both their linguistic and technological competencies to remain competitive in an ever-changing professional ecosystem.

2.1.4 Video as Educational Tool

Educational media are tools used to support the process of constructing students' knowledge and skills. Alwahibee (2019) said that one relevant approach in this context is the concept of scaffolding, which is rooted in Lev Vygotsky's theory of the Zone of Proximal Development (ZPD) and was later applied by cognitive psychologist Jerome Bruner. This concept explains that students can achieve a higher level of understanding through structured support that is gradually reduced as their learning independence increases, as said by Nurhayati, Mulyana, and Martadiputra (2016, in Purwasih and Rahmadhani, 2021). In reality, this support can be realized

through various media, including educational videos that allow for the presentation of information in a gradual and contextual manner. Teachers or the education system should find new ways to help students facing specific difficulties in regular learning by using different methods to support learning, according to Vygotsky (in Taber, 2018). Scaffolding can serve as a support for students' development in reaching the next stage as said by Raymond (2000, as cited in Alwahibee, 2019).

Various tools have been designed for this purpose, including books, audio, images, and videos, which are used to improve learning outcomes. According to Purnaningsih (2017), educational tools like these facilitate students' understanding and mastery of the material. Additionally, Molder, Anderton, Howell, Goodwin, Behrman, and Ellefson (2025) employed the Cognitive Theory of Multimedia Learning (CTML) framework to evaluate the effectiveness of three types of educational videos: animations, formal lecture videos, and documentary short films. The study's findings indicate that documentary short films have a positive impact on the formation of science identity, perception of the speaker's knowledgeability and relatability, attitudes toward science, and students' emotional engagement. These findings indicate that the documentary format is effective in increasing student engagement while deepening their understanding of the learning material. This forms the basis for research on the development of documentary media aimed at creating an effective educational tool that improves student engagement and understanding of AVT.

2.1.5 Documentary Video Production

Video is an audiovisual media presenting information through a combination of moving images, sound, and text, systematically organized to convey a specific message. In an educational context, video has long been used as a learning support media to represent the complexity of the teaching-learning process more realistically. Video can capture visual details and the learning context, thereby fostering deep reflection and evidence-based reasoning (Alsawie and Alghazo, 2010; Star and

Strickland, 2008 in Lin, 2025). One form of video widely used in education is the documentary film, which specifically emphasizes the presentation of reality and factual events as a source of learning. Nelson (2018, in Warden and Stirling, 2025) described 'documentary' as truth rather than the truth, as it documented these traces of truthfulness from everyday life that are captured, arranged, and presented. Reid and Sanders (2021) said documentary video as a medium provides flexibility for filmmakers to explore topics openly, personally, and constructively without being limited to a single fixed framework or structure. In his book, "Introduction to Documentary," Nichols (2024) classifies documentaries into seven modes, such as poetic, expository, observational, participatory, reflexive, performative, and interactive. In this research, the documentary adopts the expository and participatory modes. The expository mode is employed through narration and direct explanations regarding audiovisual interpretation, while the participatory mode is evident through the interaction between the filmmaker and the interviewee during the interview. This combination supports the educational objective while maintaining the authenticity of the representation.

2.1.5.1 Pre-production

The pre-production stage is a crucial initial stage of making the documentary video that focuses on planning and preparation before the filming process begins. According to Rabiger and Hurbis-Cherrier (2020), pre-production includes developing the concept and objectives of the video, preparing the script, selecting suitable locations, arranging the production schedule, assembling the production team, and determining the technical and visual approach of the video production. Thorough planning at this stage plays an important role in minimizing both technical and conceptual errors during the production process. In documentary videos, ideas must be supported by adequate research to ensure the information presented is accurate and factual. Reid and Sanders (2021) added that to create an observational documentary filmmakers could use the time spent in pre-production to find potential candidates for

interviews and to write questions for them to answer. Subsequently, these questions are developed into a storyline that acts as the framework for concepts visualization and vision of the documentary. Reid and Sanders (2021) also emphasize transforming ideas into actionable goals. After the idea is gathered through storyline, filmmakers may create storyboards, shot lists, or camera placement plans. A well prepared pre-production process enables filmmakers to conduct production activities more effectively while maintaining the creative and informational quality of the documentary video.

2.1.5.2 Production

The production stage is the implementation of all the planning carried out during the pre-production stage. During this stage, all visual and audio elements are recorded based on the previous designed concept which are in storyline, script, and storyboard that have been developed. According to Reid and Sanders (2021) in "Documentary Making for Digital Humanists," the production stage focuses on the practical process of filmmaking, including camera operation, shot composition, recording contextual footage, camera settings, interviews, lighting, camera movement, and audio recording. They also emphasize that both technical and creative considerations are needed to ensure the footage effectively communicates meaning and supports the intended message of the documentary.

2.1.5.2.1 Shooting

Shooting in the production stage focused on the process of capturing footage to record footage, activities, and interviews as the primary content of a documentary. According to Reid and Sanders (2021) effective shooting requires filmmakers to understand camera stabilization, shot composition, contextual footage, and camera settings. These aspects help communicate ideas visually within the documentary narrative. Effective shooting strengthens the documentary's realism and engagement on how audiences interpret information.

1. **Camera Stabilization**

Reid and Sanders (2021) emphasize the importance of stabilizing the camera to create and watchable footage. Stable shots can be achieved through the use of tripods or controlled handheld techniques. Handheld shooting can create immediacy and realism, while tripods provide steadier and more formal visuals, particularly for interviews. They also discuss camera movements such as panning, tilting, handheld tracking, and dolly shots that mostly used to follow action, reveal information, and create spatial continuity within a scene. Pan is the motion of a camera from left to right (horizontal), while tilt swinging up and down (vertical).

2. **Shot Composition**

In composition, there is an important element known as the 'rule of thirds,' Reid and Sanders (2021) also included the use of head room, looking room, and framing techniques to create balanced visuals in that rule. Proper composition allows the subject to appear natural within the frame. The arrangement of visual elements also influences audience perception of a scene and supports the communication of narrative meaning. Therefore, different shot types and angles function as visual strategies to convey information and establish context within documentaries. Reid and Sanders (2021) also discuss the use of wide shots, mid-shots, and close up as important elements of cinematic visual language; the same goes for the camera angles. Wide shots provide environmental context, mid-shots emphasize interaction and expression, and close-ups highlight emotional or significant details. The variation of shot sizes creates visual rhythm and helps structure the documentary narrative. Meanwhile, filmmakers can communicate strength or vulnerability by shooting from a low or high angle.

3. **Camera Settings, Lenses, Focus, and Exposure**

Reid and Sanders (2021) explain that filmmakers must understand camera settings such as aperture, shutter speed, ISO, and frame rate to produce visually effective footage. Exposure determines how much light reaches the camera sensor, affecting image brightness and detail visibility. Proper exposure ensures that details in bright and dark areas remain visible. Frame rate also affects the representation of movement and visual smoothness, while lens selection influences perspective and spatial perception. They recommend shooting at 24 frames per second (fps) with a shutter speed of 1/50 or 1/48 to achieve natural cinematic motion. Lens focal length influences both magnification and spatial perception, where longer focal lengths create stronger zoom effects and shallower visual space. Aperture affects both focus depth and light intake, the smaller aperture allows more light to enter the camera, making it suitable for stylized footage with blurred backgrounds.

4. **Contextual Footage**

Capture contextual footage should be planned by the filmmakers in addition to primary interview or activity footage according to Reid and Sanders (2021). Contextual footage helps audiences better understand the environment, situation, and atmosphere surrounding the subject. Such footage also supports editing continuity and enriches the visual storytelling of the documentary.

2.1.5.2.2 Recording Audio

Audio is an essential component of documentary production because it conveys information, atmosphere, and emotional meaning. Reid and Sanders (2021) explain that

recording audio on site requires careful attention to environmental conditions, microphone placement, and unwanted noise. Clear audio quality is necessary because poor sound can reduce audience understanding even when visuals are effective. According to them, there are three aspects of recording audio.

1. **Recording Sound on Site**

   On site audio recording involves capturing dialogue, ambient sound, and environmental audio directly from the filming location. Filmmakers may use built-in microphones, external microphones, or lavaliere microphones depending on production needs. Environmental awareness is important because microphones capture surrounding sounds that may interfere during recording.

2. **Voice Over and Commentary**

   Voice over and commentary are commonly used in documentaries to provide explanations, connect scenes, and guide audience understanding. Voice over narration helps deliver contextual information that may not be fully represented visually. The narration should complement the visuals without overpowering the documentary's observational quality.

3. **Engineering Ambience**

   Ambience creating realistic soundscapes, contributes to the atmosphere of a scene and strengthens the audience's sense of presence within the documentary environment. Properly engineered ambience helps maintain continuity between scenes and enhances immersion.

2.1.5.3 Post-Production

Post-production is the final stage in the documentary filmmaking process, focusing on processing all recorded footage into a complete documentary video. According to Reid and Sanders (2021), post-production is the process of arranging

footage into a pre-planned sequence, a new creative phase in which a documentary film can be reconstructed, developed, and even take on a narrative form different from the initial plan. Through the editing process, filmmakers can discover new meanings, alter the story structure, and adjust the film's rhythm based on the footage obtained. They also explain that this stage demands flexibility and critical evaluation of the production results. The script can be rewritten, the sequence of scenes can be altered, and footage not originally intended for inclusion can create new intellectual or emotional meanings.

2.1.5.3.1 Video Editing

Editing is the process of arranging and combining video and audio clips to create a coherent narrative. In their discussion of post-production (Reid and Sanders, 2021), the editing workflow divided into three stages, including rough cut, fine cut, and final cut editing. The first stage is the rough cut, which is assembled to provide a general overview of the documentary's narrative structure and flow. This stage typically focuses on issues of pacing, transitions, and sections of the footage that are not yet well-structured. Reid and Sanders (2021) also explain that structural issues can be significant enough to require revisions, the addition of new footage, or changes to specific sections. This involves the courage to cut scenes or sequences that do not make a significant contribution to the visuals, meaning, or narrative of the documentary. Once the rough cut is complete, the editing process moves on to the fine cut stage. This stage focuses on refining the editing rhythm, the timing between cuts, and the audio and visual continuity within the film. During the fine cut stage, filmmakers begin ensuring that the narrative flows clearly and that the audience remains engaged throughout the film. This stage also involves developing the soundscape, selecting music, and refining other audio elements. They explain that music and sound design significantly influence the emotional atmosphere of a documentary and can improve the narrative impact of the video. The final stage of post-production is the final cut according to Reid and Sanders (2021). This stage involves putting the finishing touches on the documentary before the film is declared complete. All editing is finalized,

temporary elements are replaced with their final versions, and any remaining technical or narrative issues are resolved at this stage. The audio must also be balanced so that dialogue is clear, while the music and ambient sounds continue to support the atmosphere without distracting from the film's main message. This last stage concludes all visual and audio elements together into a complete documentary ready for publication or screening to audiences.

2.1.5.3.2 Subtitle Production

Subtitle production is an important part of the post-production stage in audiovisual media. This process involves creating, editing, synchronizing, and reviewing subtitles to ensure that the translated text can be understood clearly by the audience while remaining synchronized with the visual and audio elements of the video (Díaz Cintas and Remael, 2020). In subtitle production, subtitlers usually work based on subtitle guidelines or style guides provided by broadcasting companies or subtitling agencies. For instance, over-the-top (OTT) media streaming like Netflix offer open access to all their subtitle style guides in multiple languages published through its Partner Help Center. Netflix is used as an example in this study because it is among the OTT platforms that have expanded the global circulation of audiovisual content and increased the demand for multilingual localization services, making subtitle production an increasingly important aspect of audiovisual translation (Díaz Cintas and Remael, 2020). These guidelines contain technical and linguistic standards such as subtitle formatting, reading speed, punctuation, character limitations, and translation consistency to maintain subtitle quality across audiovisual productions. In general, Netflix guidelines regulate aspects such as a maximum of two subtitle lines, approximately 42 characters per line (CPL), subtitle synchronization, speaker identification, italics usage, and reading speed limitations to ensure readability and viewer comfort. According to Netflix Timed Text Style Guide: General Requirements, the recommended maximum line length of around 42 characters is intended to maintain subtitle readability across different screen sizes and devices while preventing subtitles

from occupying excessive screen space (Netflix, 2022). Netflix also generally recommends a reading speed of around 20 characters per second (CPS) for adult content, although this may vary depending on the target audience and genre. Reading speed limitations are applied to ensure that viewers have enough time to read subtitles naturally without missing important visual scenes or dialogue information. Subtitle synchronization is also important because subtitles must appear and disappear in accordance with speech timing and shot changes to maintain viewing comfort and audiovisual coherence. Furthermore, the use of a maximum of two subtitle lines is intended to reduce visual overload and prevent subtitles from covering significant visual elements on screen, allowing viewers to simultaneously process both verbal and visual information effectively. Italics are commonly used to indicate off-screen dialogue, song lyrics, internal thoughts, or foreign-language expressions, while speaker identification helps viewers distinguish between multiple speakers in the same scene. In the Indonesian Timed Text Style Guide, Netflix also provides language-specific guidance related to the use of formal and informal Indonesian expressions, slang adaptation, cultural references, abbreviations, and the translation of honorifics in order to maintain naturalness and cultural appropriateness for Indonesian audiences (Netflix, 2025). For example, the guideline emphasizes that subtitles should sound natural and conversational in Indonesian while still preserving the meaning and tone of the original dialogue.

Díaz Cintas and Remael (2021) in their book, “Subtitling: Concepts and Practices” explaining the subtitling process with several steps. The process usually begins when a production company, distributor, television station, or streaming platform assigns a project to a language service provider (LSP) or subtitling agency. Afterward, the audiovisual material, dialogue list, and supporting documents are reviewed to ensure the completeness and accuracy of the information that will be translated. The next step is spotting or timing, which refers to determining when subtitles should appear and disappear on screen according to the dialogue, shot

changes, and audiovisual rhythm. In some productions, spotting is carried out directly by the subtitler, while in modern workflows it is often completed by technicians who prepare subtitle templates to support the translation process. Once the template and video materials are available, the subtitler begins translating the dialogue into the target language while also considering other relevant audio and visual elements such as songs, on-screen text, and background sounds.

In the past, subtitle production required multiple hardware devices, including a desktop computer, an external video player for VHS or Betacam tapes, and a television monitor to view the audiovisual material (Díaz-Cintas and Remael, 2020). Subtitlers also relied on basic subtitling programs and, in some cases, stopwatches or video toggles to determine subtitle timing manually. However, the transition from analogue to digital technology at the turn of the twenty-first century transformed subtitle production significantly. Traditional workstations that depended on several pieces of hardware were gradually replaced by a single computer equipped with specialized subtitling software (Díaz-Cintas and Remael, 2020). In modern subtitle production, subtitlers generally use specialized subtitling software such as Aegisub, OOONA, and so on, that allows translation, synchronization, subtitle positioning, and subtitle simulation to be completed within a single working system. Some subtitling programs can automatically detect shot changes, display audio waveforms for dialogue synchronization, calculate subtitle duration, monitor characters per second, and perform spelling and grammar checks. After the translation process is completed, the subtitles undergo revision and quality control (QC) to ensure there are no translation errors, typographical mistakes, timing inconsistencies, or terminology issues. At this stage, both linguistic and technical quality of the subtitles are evaluated to ensure compliance with subtitle guidelines or style guides provided by companies or distribution platforms. Once finalized, subtitle files can be exported into various formats such as .srt, .ass, and .vtt depending on the platform, and distributed through television, film, DVD, or digital streaming platforms. The development of internet

technology has also simplified file transfer processes through cloud platforms and file transfer protocol (FTP) systems.

2.2 Previous Studies

One of the primary documentary work references for this study is the documentary "Salt in Their Veins" by Hencke (2025) via YouTube. The documentary portrays the lives of the Bajo community and their deep connection to the sea through an observational approach that emphasizes authentic visual storytelling. It relies on natural sound, direct interviews, and varied camera techniques, such as wide shots for contextualization and close-ups for emotional expression, allowing audiences to interpret the story independently. This immersive storytelling style serves as an inspiration for the present study in designing a documentary that not only conveys information but also creates emotional engagement in introducing the Audiovisual Translation (AVT) profession. However, the documentary mainly focuses on socio-cultural and environmental issues rather than introducing a professional field or functioning explicitly as an educational tool. Thus, a gap remains regarding the use of documentary media to increase public understanding of emerging professions such as AVT.

The second reference is "Secrets of Christ's Tomb" by Strange (2025) via National Geographic YouTube channel, which documents the restoration process of Jesus' tomb in Jerusalem through an investigative and research-based approach. In its presentation, the documentary combines voice over narration, expert interviews, and visual reconstructions alongside on-site documentation. The structured delivery of information, supported by subtitles and visual explanations, makes complex topics accessible to wider audiences. This approach is relevant to the present study, particularly in organizing educational content in an engaging and comprehensible manner. Nevertheless, the documentary centers on historical and archaeological

themes, leaving limited exploration of how documentary media can be used to introduce professions and raise career awareness among audiences.

Another reference is “The Hidden Soul of Pekojan” by Kusuma, Sufitri, and Hermawan (2025) via Bahasa Asing Terapan YouTube channel. The documentary explores the culture and daily life of Kampung Pekojan in Semarang through observational footage and interview-based storytelling. Its use of cinematic visuals and authentic interviews enables audiences to understand the community’s identity from the perspectives of the participants themselves. This interview-centered approach is considered relevant to the development of the present documentary, which also highlights personal experiences and insights from practitioners. However, the documentary focuses primarily on cultural heritage and local identity, without addressing professional education or career-related issues, particularly in the field of AVT.

The last reference is Utami, Saifullah, Utama, and Wibowo (2020) emphasize that audiovisual media can play an important role in supporting the learning process by increasing students’ motivation and engagement. Their study, conducted in an Educational Psychology course, revealed that students who were involved in producing audiovisual materials showed stronger motivation toward learning activities. Furthermore, the integration of sound and visual elements was found to stimulate creativity and facilitate students’ participation and comprehension. While these findings confirm the educational value of audiovisual media, the research focuses on classroom learning and students’ involvement in media production. It does not explore the potential of documentary videos as educational tools, nor does it examine their use in introducing emerging professions. Therefore, the present study expands this area of research by creating a documentary video that introduces AVT as a profession, aiming to improve students’ awareness, understanding, and career interest in the field.

Although those works provide important references in visual storytelling, educational delivery, and interview-based documentary techniques, none of them specifically focus on introducing a professional field such as Audiovisual Translation (AVT) that serves purposes as an educational tool. Most existing documentaries discuss cultural, historical, social issues, or class-room related topics. Therefore, this research fills the gap by developing a documentary video as an educational tool to increase public understanding of AVT. The documentary combines informative explanations, cinematic storytelling, interviews, and subtitling elements to make the topic more accessible and engaging for audiences. This documentary video is expected to help audiences gain a clearer understanding of the real experiences, challenges, and opportunities within the AVT profession, which is still relatively unfamiliar to the wider public.