CLIPSENSE: ENHANCING BROADCAST EFFICIENCY THROUGH AI-DRIVEN VIDEO CONTENT ANALYSIS AND TAGGING

S. M. Devigne

STUDIO 47, Germany

ABSTRACT

The innovation project ClipSense focuses on the development of AI-based software to enhance efficiency in journalistic production processes, particularly for local and regional media. By utilizing technologies such as machine learning and natural language processing, ClipSense automates video analysis, tagging, and archiving. This automation allows editorial teams to reduce time-consuming routine tasks and concentrate on more creative and investigative aspects of their work.

This technical paper documents the development, challenges, and implementation of the software, incorporating feedback from end users to demonstrate its effectiveness and acceptance. Finally, the paper provides an outlook on future developments and potential enhancements of ClipSense technology, aiming to further improve adaptability and efficiency in an ever-changing media landscape.

INTRODUCTION

In today's fast-paced media landscape, the ability to efficiently and accurately produce up-to-date news is crucial for the success of TV and online editorial teams, especially at the local and regional levels. The challenges these editorial teams face are diverse and complex, ranging from limited resources to the need to quickly process and make large amounts of video material accessible. Here, the innovation project ClipSense comes into play, aiming primarily at modernizing and enhancing the efficiency of journalistic production processes through the use of Artificial Intelligence (AI).

ClipSense is an AI-powered software solution, developed to address the specific needs of local and regional media companies. By automating video analysis, tagging, and archiving, the software provides support to manage the daily challenges in editorial teams. This technology enables not only the recognition and classification of video content but also its contextual and stylistic tagging, resulting in significant time savings and quality improvements in newscast production.

The development and implementation were carried out under a funding program by the Media Innovation Center Babelsberg (MIZ) and in close cooperation with regional TV and online media. The software was presented on several industry platforms where it generated considerable interest and received positive feedback.

.

PROBLEM STATEMENT

In the world of news journalism, the efficiency of news production is crucial for the success and relevance of content. A key element of this process is a well-organized and maintained video archive, which enables quick and precise selection and use of available video material. Unfortunately, many editorial teams, especially in regional TV and online media, lack the necessary technical and personnel resources to maintain such an archive [Heesen et al. (1)]. This often leads to poorly maintained archives that significantly hinder daily work and diminish journalistic quality.

The challenge begins with the basic task of tagging and cataloging video content, which is currently mostly done manually. This manual handling is not only time-consuming but also prone to high error potential. In practice, this means that editors and journalists often have to spend a disproportionate amount of time searching for appropriate material for current reports. This is particularly problematic in an environment characterized by timeliness and speed.

In addition to the challenges of archive management, the problem exists that many archives simply cannot be effectively utilized. They often contain undiscovered, valuable content ("hidden treasures") that are difficult to access due to inadequate tagging and organization. This results in valuable footage remaining unused and unnecessarily complicating the production of new content.

The consequence of this problem is a significant impairment of journalistic quality. If journalists have to spend most of their time finding suitable material, there is less time left for the actual journalistic processing and contextualization of the news. This undermines the media's ability to react quickly and effectively to changing news situations.

In this context, the idea for ClipSense was born, an innovative tool that enables the automation of video analysis and tagging using artificial intelligence. ClipSense aims to overcome the existing challenges in managing video archives by providing efficient, systematic, and error-free tagging and archiving of video content. Implementing it in the daily routine of editorial teams could not only improve the accessibility and usability of archival material but also significantly enhance journalistic quality and efficiency in news production [Schützeneder et al. (2)].

AUTOMATION IN THE NEWSROOM AND NEEDS ANALYSIS USING AI

The integration of AI-driven tools into newsrooms marks a significant transformation in the media industry, especially for local and regional television and online editorial teams. These technologies address existing challenges such as limited personnel and financial resources while simultaneously offering solutions to optimize or substitute work processes [Grienberger et al. (3)]. By automating content management and production with advanced AI tools, not only is efficiency enhanced, but journalistic quality is also significantly improved. The introduction of ClipSense, therefore, represents a crucial and proactive measure to future-proof the media landscape and enable editorial teams to focus more effectively on content creation.

The Necessity of a Well-Maintained Video Archive

A well-organized and precisely tagged video archive is essential for the production of up-to-date news. However, many editorial teams face the problem of lacking resources to maintain such archives. ClipSense addresses this by automating the tagging and archiving of video content, thereby relieving editorial teams so that they can concentrate on journalistic work.

AI as an Opportunity for Local Journalism

AI tools offer significant opportunities in local journalism, which many consider to be the "base camp of democracy." Without the support of such technologies, local journalism could fall behind, leading to the loss of an important element of public discourse and societal information. Through the targeted use of AI tools, local media can not only survive but also significantly enhance their journalistic quality and efficiency.

Needs Analysis and Requirement Specification

Before implementing the AI-driven tool, a detailed needs analysis was conducted to understand the specific requirements of local and regional editorial teams. This analysis identified inefficient management of video archives and the time-consuming search for relevant material as the main challenges. Based on these findings, targeted functions were developed to optimize workflows in editorial teams and allow editors more capacity for the creative design of content. The analysis emphasized the importance of local journalism as an essential foundation of democratic society and public discourse, underlining the need to maintain its functionality and quality through technological support.

Problem / Challenge	Solution Considerations	Facts and Steps Towards the Solution
Insufficient resources for archive management	Automation of archiving through AI-supported software	ClipSense would enable automated analysis, tagging, and archiving of videos.
Labor-intensive manual tagging and sorting	Use of machine learning and natural language processing	It would be developed to automate content and design recognition as well as journalistic tagging.
Slow and error-prone newscast production	Integration of AI tools to optimize production workflows	ClipSense would enhance efficiency and quality in production through faster and more precise tagging.
Accessibility issues in existing archives	Use of advanced algorithms for content analysis	ClipSense would analyze and categorize content, thereby improving the retrieval of archival material.
Preservation of journalistic quality	Provision of tools to support editorial work	By freeing up space for journalistic activities, the quality of reporting would be improved.
Adapting to changing media landscape	Regular updates and enhancements of software functionality	There would be future-oriented development with continuous adjustments to new requirements.
Integration into existing editorial systems	Development of a flexible API interface	ClipSense would be designed to be easily integrated into various production environments.

Table 1 – Results of the Needs Analysis and Derived Considerations

DEVELOPMENT PROCESS

The innovation project ClipSense was supported by the Media Innovation Center Babelsberg (MIZ) and successfully implemented by STUDIO 47, a regional TV broadcaster [Becker (4)]. Under the leadership of Sascha Devigne, the editor-in-chief of STUDIO 47, and with the technical expertise of Peter Kläs, managing director of feine apps GmbH, the project aimed to automate key processes in newscast production through the use of Artificial Intelligence. This automation focused particularly on the areas of video analysis, tagging, and archiving to enhance journalistic efficiency and quality.

Project Implementation and Role Distribution

Figure 1 - Scrum session of the development team during the project at STUDIO 47.

The project management was handled by Sascha Devigne, who brought extensive experience in media production into the strategic planning and coordination of the interdisciplinary team. Jörg Zeiler, as managing director of STUDIO 47, was responsible for financial controlling, ensuring budget compliance, and providing commercial transparency towards the MIZ. Peter Kläs was in charge of the technical direction, including the development of the client-based software solution that featured specific algorithms for recognizing and classifying video content.

Technical Implementation and Functionalities

The development of ClipSense involved several key components: automated video analysis, content tagging, and the recognition of shot sizes to optimize the editing process. A crucial aspect was the integration of an API interface, which enabled seamless integration of the analyzed content into existing MAM and editorial systems. This integration was crucial for improving operational efficiency and ensuring compatibility with various production environments.

Project Phases and Milestones

The development process of ClipSense was divided into five work packages, starting with detailed analysis and planning, through design and implementation, to evaluation, integration, and the final rollout. Each phase was carefully planned to meet the needs and requirements of the end-users, with particular attention paid to the usability and functionality of the software. The final phase included training for editorial staff and the implementation of the software in the production environments of STUDIO 47, ensuring comprehensive use and acceptance in daily operations.

Work Package	Period	Milestones
WP 1: Analysis and Planning	June 2023	Requirements analysis and specification of the software functions of ClipSense; Identification of required resources and technologies; Definition of project scope; Creation of a detailed project plan
WP 2: Design	July – August 2023	Design of architecture and data models; Design of user interface and UX; Definition of database structure and/or schemas; Research of algorithms and models for video content recognition and tagging; Design of production workflows and tools
WP 3: Implementation	August – November 2023	Development of backend and frontend functions of ClipSense (UI); Integration of video content recognition and tagging modules (APIs); Implementation of newscast production workflows and tools; Creation of tests and quality assurance processes
WP 4: Evaluation and Integration	December 2023	Conducting system tests and troubleshooting; Integration of ClipSense into existing systems and infrastructures; Conducting user acceptance tests and feedback
WP 5: Training and Rollout	January 2024	Training of users; Deployment of ClipSense in production environments; Support in system integration and customization; Documentation of final results

Table 2 – Work packages and milestones during the individual project phases

USAGE AND APPLICATION OF CLIPSENSE

ClipSense is an AI-powered editorial software specifically designed to optimize processes in local and regional TV and online editorial teams through automation. This tool supports the automated analysis, tagging, and archiving of video content to simplify and accelerate journalistic work.

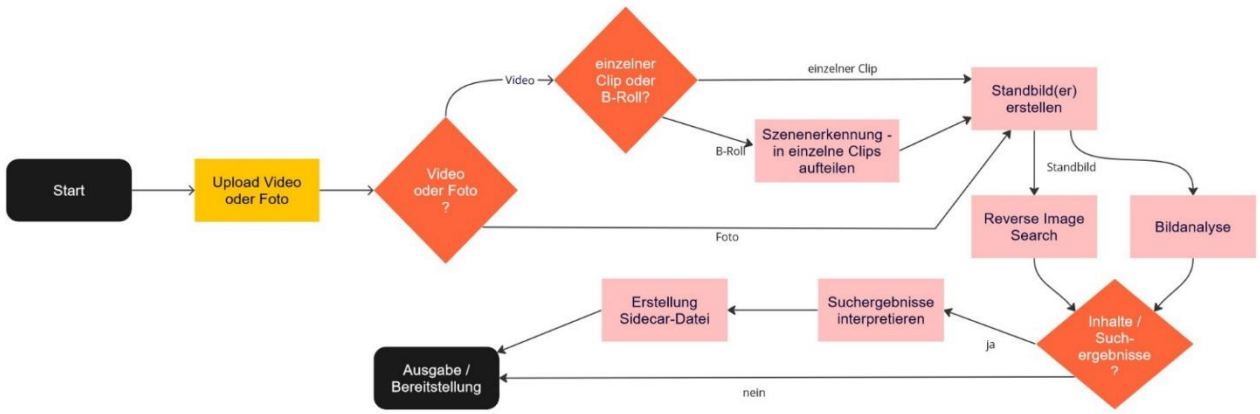

Figure 2 – Flowchart of the Functionality, Interfaces, and Work Processes of ClipSense

1. **Automated Video Analysis:** ClipSense uses AI technology to identify and classify image content within videos. It recognizes people, objects, places, and temporal contexts such as time of day or season. This technology enables the categorization of video material by editorial topics like sports, politics, or culture.
2. **Tagging of Video Content:** The system automatically generates tags based on the recognized image contents. These tags are assigned to the videos and stored in an indexed data structure, facilitating the archiving and reuse of the material.
3. **Recognition of Shot Sizes:** ClipSense identifies the shot sizes of video clips, such as close-ups or wide shots, and categorizes them according to a standardized nomenclature. This is particularly useful for automated video editing as it helps avoid editing errors like jump cuts or transition errors.
4. **Provision for Newscast Production:** Through an API interface, the edited video content can be directly transferred to MAM or editorial systems. This allows the automatic creation of video newscasts using the generated tags and tags.
5. **Archiving of Video Content:** After analysis and tagging, the video content is automatically stored. This ensures that the material is immediately available for future journalistic productions and can be used efficiently.

Step	Description	Output
Video Input	ClipSense receives video content, identifies whether they are individual clips or assembled B-Rolls.	Raw video input
Scene Detection	Separation of the video into individual clips, based on the analysis of cuts and scene changes.	Split video clips
Automated Video Analysis	Analysis of each clip using AI algorithms to detect people, objects, and other relevant elements.	Detected objects, people, places, times
Recognition of Shot Sizes	Determination of the camera settings for each clip (e.g., wide shot, medium shot, close-up).	Classified shot sizes
Tagging	Generation of tags based on the analysis results, associating these tags with the corresponding clips.	Tags and metadata assigned to each clip
Archiving	Storage of the clips with their tags and metadata in a footage archive.	Systematically archived and tagged video content
Provision for Newscast Production	Handover of the edited and tagged content to editorial systems via an API.	Integrated into production environments, ready for the creation of video newscasts

Table 3 – Derived Work Steps and Results of the Functional Processes

The development of ClipSense involved overcoming various technological challenges:

- **Precise Video Analysis:** By integrating advanced machine learning and AI algorithms, the team was able to achieve high accuracy in the recognition and classification of video content. These algorithms were trained on a wide range of categories and properties.
- **Efficient Tagging:** An automatic tagging system was implemented to translate recognized objects and people into meaningful tags. These tags are crucial for the quick findability and accessibility of the archived material.
- **Recognition and Standardization of Shot Sizes:** Correct identification and naming of shot sizes were enabled by specialized algorithms, allowing this information to be used for precise and error-free video editing.

Integration and Use in Editorial Systems

ClipSense was designed to seamlessly integrate into existing Media Asset Management Systems (MAM systems). The generated data and metadata, such as EXIF data or sidecar files, are automatically provided with the clips and can be easily incorporated into the editorial teams' workflow systems. This enables efficient and structured use of video content for newscast production and promotes dynamic and modern news reporting.

FUNCTIONALITY OF CLIPSENSE

ClipSense is a software solution that leverages artificial intelligence (AI) to optimize the processing and analysis of video content for local and regional TV and online editorial teams. This chapter describes the detailed technical steps, programming methods, and tools used to develop and implement ClipSense, with the goal of simplifying and enhancing the editing and management of media content.

Technical Execution and Scene Separation

The process in ClipSense begins with scene separation, where incoming video material is divided into individual scenes using the tool ffmpeg, an open-source video editing software. Ffmpeg is specifically configured to capture precise timestamps of each scene and to extract screenshots at defined points, preferably in the last third of the scene (about 75%). These screenshots form the basis for further in-depth analysis.

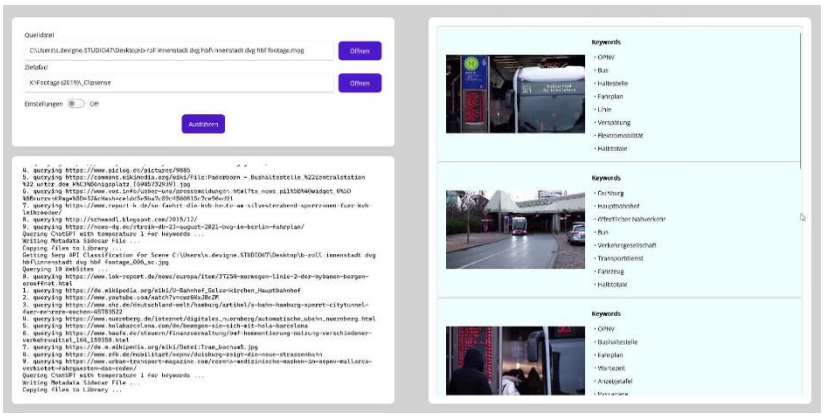

Figure 3 - Screenshot of the User Interface: Video input field (top left), console window (bottom left), output preview of the detected scene including tagging (right)

Integration of Web Searches and Machine Learning

A significant initial problem was the lack of specific training material for image analysis. To compensate for this deficit, a solution was developed that combines web searches and machine learning:

1. **Selection and Use of Web Searches:** Initially, a variety of image-based web search services that offer API support were evaluated. The choice fell on an API with an extensive database and robust search functions that meet the specific requirements.
2. **Extraction and Processing of Web Content:** After selection, content from relevant websites is extracted. This content is partly processed by a Large Language Model (LLM) to create a comprehensive database for further training models.
3. **Training of the LLM:** The data obtained from the web searches are used as input for training the LLMs, utilizing ChatGPT to derive relevant keywords and tags for the classification and tagging of the video material from the text data.

Integration and Use of AI-Supported Features

ClipSense integrates several AI-supported functions: from scene recognition, which divides B-rolls or footage into individual clips, to video analysis, which is performed using two combined APIs.

Programming and Technical Implementation

The programming of ClipSense was primarily done in Python, a language known for its versatility and extensive support by data analysis and AI libraries. Python scripts controlled the interaction between the web search API and the machine learning modules.

Optimization and Customization

The algorithms of ClipSense were continuously tested and adjusted, which includes fine-tuning the web search and the LLM prompts. Algorithms for recognizing shot sizes were developed to improve accuracy for later editing of the video material. Integrating the image-based results into the queries led to significant improvements in analysis accuracy.

Integration and Results

ClipSense is integrated into existing Media Asset Management Systems (MAM) via a specially developed API that allows flexible and reliable transfer of processed content and seamless integration into a variety of production tools and platforms. This API supports various video formats, thereby ensuring high compatibility with existing media production systems. The results of scene analysis and tagging are stored in a structured manner to support fast indexing and querying.

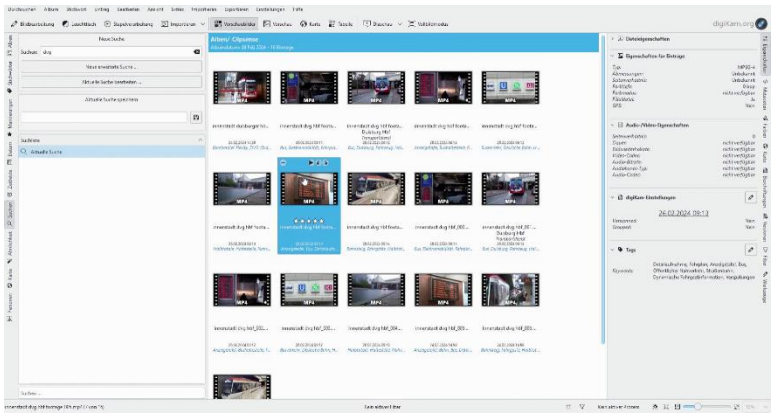

Figure 4 - Transfer and import of the analyzed scenes including tagging via API to the MAM system

Ongoing Development and Adjustment

The technical architecture is designed to be future-proof. Through the flexible API interface, regular updates and adjustments to new technological standards and requirements can be easily implemented. Future improvements will include the integration of additional data sources, such as GPS information, and further development of shot size recognition. The incorporation of contextual information from previous scenes will refine the analysis further.

SYSTEM TESTING AND IMPLEMENTATION

The ClipSense project posed significant technical challenges, particularly regarding the precise analysis of video content under various environmental conditions. Overcoming these challenges was crucial for the successful implementation and operation of the software.

Development and Implementation Phase

The development was divided into two main phases: backend and frontend development. In the second half of the project, the focus shifted to implementing additional modules for comprehensive video analysis and improving the user interface to ensure intuitive operation. The final phase included the integration of specialized algorithms for object recognition and adaptation to various shot sizes and media formats.

System Tests and Quality Assurance

Concurrent with the development phase, extensive system tests were conducted to ensure the functionality of all components. These tests aimed to identify and correct any errors to guarantee stable and reliable software. Additionally, the tests were expanded to assess both technical reliability and user acceptance. User acceptance tests [Stoll (5)] were carried out to collect direct feedback from end users, which was crucial for optimizing usability.

Integration into Existing Infrastructures

A key component of the project was the integration into the existing system landscape of the users. This required close collaboration with external partners and the development of a flexible API interface, allowing ClipSense to be seamlessly integrated into existing editorial and media management systems. This integration ensured that ClipSense could be effectively adopted without disrupting the existing workflows, thereby enhancing the overall efficiency and productivity of the media production processes.

EVALUATION, FEEDBACK, AND ACCEPTANCE OF AI DEPLOYMENT

Throughout its development and testing phases, the ClipSense project underwent extensive evaluations and received feedback that highlighted the acceptance and applicability of AI-supported software in media production. By being showcased at various industry platforms and relevant professional conferences, ClipSense generated broad attention and positive responses.

Feedback from Conferences and Media Reports

ClipSense was featured at events such as the Medientage Mitteldeutschland in Leipzig, the Local Journalism Congress in Berlin, the MIZ Media Innovation Pitch in Babelsberg, and the MediaTechHub Conference in Potsdam [Effenberg and Briegel (6)]. These presentations attracted considerable attention, with the innovation and technical advancements of the software being particularly emphasized.

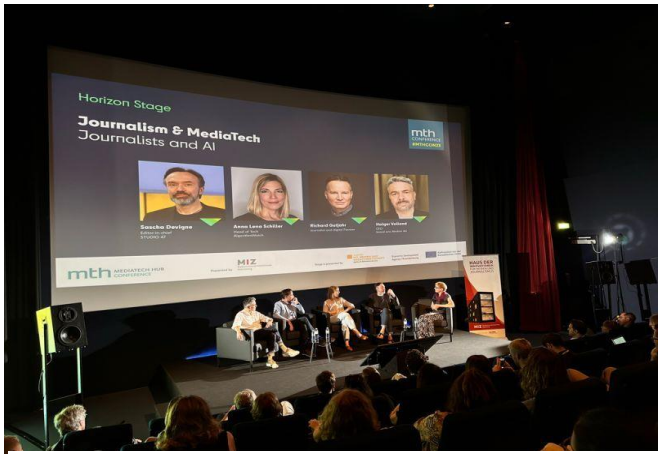

Figure 5 - Panel discussion at the MTH Conference on the use of ClipSense

Media coverage by professional journals such as Kress Pro, DJV Journal, and the

Digital Publishing Report [Meier (7)] highlighted the innovative aspects, especially the application of artificial intelligence in newscast production. These publications helped to sharpen awareness within the media industry and stimulated discussions about the use of AI in media production.

Practical Application and User Feedback

In practical applications, ClipSense was tested in cooperation with regional TV and online media in Germany (including Glomex, ProSiebenSat.1 Media; ems TV) and Austria (including Tirol TV, R9). These partnerships were crucial for fine-tuning the software by enabling it to be tested and adjusted under real production conditions. The direct integration into the workflows of the broadcasters allowed for the collection of valuable feedback from end-users, which was crucial for the further development and optimization of the software.

The feedback indicated that ClipSense significantly eased the daily work of news editorial teams. Here are some quotes:

- "I find the innovative power to be very strong, I think there are a lot of possibilities and time savings [...] I am always a fan of automation and increasing efficiency, especially in the media sector." (Herman del Campo, Managing Director, Cyber Innovations)
- "We [...] need these support tools exactly to be effective and quick also in motion picture [...] The added value is definitely there for us, and I see a very large market potential for your project because I believe everyone who works with motion picture needs it, not just the traditional regional TV stations but also those who will work online with motion picture in the future." (Tilo Winkler, Deputy Editor-in-Chief, Lausitzer Rundschau)

Figure 6 - Feedback session on ClipSense with industry experts at the MIZ

Regular Updates and Introduction of New Features

To meet the dynamic requirements of the constantly evolving media landscape, regular software updates and the introduction of new features are planned. These continuous improvements ensure that ClipSense keeps pace with technological advances and the changing needs of media producers.

Implementation of User Feedback Loops

A central component of the ongoing development strategy is the establishment of feedback loops with end users. These loops are crucial for continuously improving the software. Close collaboration with users allows not only for technical optimization but also for continually increasing the practicality and user-friendliness of the software. The interactive process enables the development team to collect direct feedback and precisely align the software with the specific needs and challenges of the users.

Future Integrations and Partnerships

Moreover, the innovation team plans to expand its partnerships with leading technology providers and media houses. These strategic alliances will be crucial for integrating

advanced technologies and embedding ClipSense into different editorial environments. Such partnerships will not only enhance the technical robustness but also broaden the application scope within the media industry, enabling ClipSense to support a wider range of journalistic activities and production scales.

CONCLUSIONS

The integration of artificial intelligence into media production processes, as demonstrated in the recent initiative, has marked a pivotal advance, particularly for local and regional news organizations. By automating key tasks such as video analysis, and archiving, AI-driven solutions can significantly enhance both the efficiency and the quality of media workflows.

Throughout its evaluation phase, ClipSense proved to be a crucial tool in reducing the manual labor involved in video content management. Feedback from these trials highlighted a substantial decrease in the time editors and journalists spent on routine tasks, enabling a greater focus on creative and investigative endeavors. This shift not only streamlined production processes but also enriched the content quality, facilitating deeper narrative development and contextual richness in journalistic outputs.

REFERENCES

1. Heesen, J., Bieber, C., Lauber-Rönsberg, A., Neuberger, C., Elmer, C., & Hühnert, T., 2023. Künstliche Intelligenz im Journalismus. Potenziale und Herausforderungen für Medienschaffende. Whitepaper aus der Plattform Lernende Systeme.

2. Schützeneder, J., Graßl, M., & Meier, K., 2024. Grenzen überwinden, Chancen gestalten: KI im journalistischen Newsroom–Bestandsaufnahme, Perspektiven und Empfehlungen für Journalismus und Politik. FES Impuls.

3. Grienberger, K., Matthes, K., & Paulus, W., 2024. Vor allem Hochqualifizierte bekommen die Digitalisierung verstärkt zu spüren. Aktuelle Analysen aus dem Institut für Arbeitsmarkt- und Berufsforschung.

4. Becker, G., 2023. KI-Jobs im Journalismus: Wettrennen gegen die Technologie. DFJV Deutscher Fachjournalisten-Verband.

5. Stoll, X., 2023. UAT: Was ist ein User Acceptance Test? HubSpot.

6. Effenberg, P. & Briegel, F., 2023. Track: Journalism & MediaTech. MTHCON23 Report.

7. Meier, S., 2023. Medien und KI: Use Cases. Digital Publishing Report.

ACKNOWLEDGEMENTS

The author would like to express his gratitude to all colleagues in the editorial and management team at STUDIO 47 for their support throughout this project. A special thank goes to Peter Kläs for developing and implementing ClipSense, as well as to the Media Innovation Center Babelsberg for their support in the development process. Additionally, the author would like to thank the International Broadcast Convention for providing the opportunity to showcase the project and its results.