

Article

Toward Annotation, Visualization, and Reproducible Archiving of Human–Human Dialog Video Recording Applications

Verena Schreyer , Marco Xavier Bornschlegl  and Matthias Hemmje 

Faculty of Mathematics and Computer Science, University of Hagen, 58097 Hagen, Germany; marco-xaver.bornschlegl@fernuni-hagen.de (M.X.B.); matthias.hemmje@fernuni-hagen.de (M.H.)

* Correspondence: verena.schreyer@fernuni-hagen.de

Abstract: The COVID-19 pandemic increased the number of video conferences, for example, through online teaching and home office meetings. Even in the medical environment, consultation sessions are now increasingly conducted in the form of video conferencing. This includes sessions between psychotherapists and one or more call participants (individual/group calls). To subsequently document and analyze patient conversations, as well as any other human–human dialog, it is possible to record these video conferences. This allows experts to concentrate better on the conversation during the dialog and to perform analysis afterward. Artificial intelligence (AI) and its machine learning approach, which has already been used extensively for innovations, can provide support for subsequent analyses. Among other things, emotion recognition algorithms can be used to determine dialog participants' emotions and record them automatically. This can alert experts to any noticeable sections of the conversation during subsequent analysis, thus simplifying the analysis process. As a result, experts can identify the cause of such sections based on emotion sequence data and exchange ideas with other experts within the context of an analysis tool.

Keywords: human–computer interaction; information visualization; emotion analysis; long-term preservation



Academic Editors: Phivos Mylonas, Cleo Sgouropoulou, Christos Troussas, Akrivi Krouska, Katerina Kabassi and Jaime Caro

Received: 27 March 2025

Revised: 21 April 2025

Accepted: 22 April 2025

Published: 26 April 2025

Citation: Schreyer, V.; Bornschlegl, M.X.; Hemmje, M. Toward Annotation, Visualization, and Reproducible Archiving of Human–Human Dialog Video Recording Applications. *Information* **2025**, *16*, 349. <https://doi.org/10.3390/info16050349>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction and Motivation

The COVID-19 pandemic significantly impacted daily life worldwide, drawing immense media attention. Beyond infections and deaths, it has caused a substantial mental health burden. Despite mild symptoms and low mortality among children and adolescents (10 to 19 years), according to [1], the pandemic led to anxiety, depression, and suicidal behavior in this group. Quarantine and social distancing increased mental health risks, with daily reports of cases and deaths heightening stress and uncertainty [2].

The pandemic accelerated the rise of mental disorders, especially among young people. The WHO [3] predicts anxiety disorders will be a major healthcare burden by 2030. Young people, already struggling with mental health, face additional challenges from digitalization and social media, which exacerbate stress and social fragmentation [4]. The isolation from the pandemic worsened these issues [5,6].

Remote work surged during the pandemic [7], introducing challenges like video fatigue and technical difficulties [8]. Psychotherapy shifted to remote sessions, offering flexibility but also presenting new challenges. Despite these challenges, remote therapy has proven effective, particularly for anxiety disorders [9,10].

Video conferencing allows for session recordings for analysis, enhancing our understanding of patient interactions. Non-verbal communication remains crucial, as highlighted

by Paul Watzlawick's communication theory [11], which emphasizes the importance of both verbal and non-verbal cues in therapy.

1.1. Motivation

This research is supervised by the chair of Multimedia and Internet Applications (MMIA) in the Faculty of Mathematics and Computer Science at the University of Hagen in collaboration with the associated Research Institute of Telecommunication (FTK) at the University of Hagen, where application-oriented research takes place. One notable project has been Smart Vortex, funded by the EU under the Seventh Framework Program, aimed at supporting corporate collaboration through interoperable tools for data stream management and analysis [12].

As part of the Smart Vortex project, the IVIS4BigData model was developed. IVIS4BigData is a reference model used to “support advanced visual user interfaces for distributed Big Data Analysis in virtual labs”. It addresses the development of information systems, data analysis, and knowledge management systems by covering the process of big data analysis from multiple data sources to enable users to participate in the information visualization process, which will help to visualize emotion sequence data [13].

The RAGE project at FTK, in cooperation with MMIA, developed the Knowledge Management Ecosystem Portal (KM-EP) to support skill development using computer game principles, where the software prototype of this project will be implemented. The KM-EP serves as a media archive and digital library, useful for managing big data analysis results [14]. The SenseCare project extended the KM-EP's uses to medical data, aiming to improve healthcare processes through sensory and machine learning technologies [15].

Based on the SenseCare project [15], the SMILE project focuses on improving mental health in young people by using sensory and machine learning technologies to gain emotional and cognitive insights [16]. The project seeks to build a sustainable resilience system based on real-world data, involving experts, citizens, and policymakers [16]. Analyses of data and progress will be supported by providing access to and using the project results to create dashboards in the KM-EP with information, statistics, and access to knowledge.

For example, emotion sequence data can be calculated from video recordings of human–human dialogs representing the emotions of the different dialog participants using emotion-recognition algorithms [17]. These data can be uploaded, visualized, analyzed, and documented using dashboards in the KM-EP [14]. For visualization, the IVIS4BigData reference model [13] covers the process of big data analysis from multiple data sources to enable users to participate in the information visualization process. By making these visualizations available, in addition to human–human dialog video recordings, experts are enabled to analyze conversations and emotion sequence data and document the analysis results in dashboards.

Currently there is no opportunity to analyze and document emotions in recorded human–human dialog videos in KM-EP by visualizing emotion sequence-data, viewing dialog recordings and providing possibilities for protocol, exchange, and discussion (Motivation Statement 1).

To support the analysis of emotion sequence data, as listening to long video recordings and conversations over again is tedious [10], experts should be able to obtain faster insight into dialog participants' emotions and interesting sections or reactions in longer emotion sequence data. Therefore, metrics calculated based on recognized emotions are useful.

Currently, there are no automatically calculated metrics to support the analysis and documentation of emotion sequence data in human–human dialogs (Motivation Statement 2).

Session analysis data like these metrics support the analysis and documentation of human–human dialogs, like patient interviews, and can be stored together with session analysis documents, for example, in patient medical records. AI-supported documentation may help therapists by improving transparency and accuracy, which also applies to emotion recognition [18]. Therefore, it is necessary to be able to export analysis and documentation results, as well as video recordings and emotion sequence data, from the KM-EP to store them together with patient medical records. As patient medical records are also archived, the exported data must meet the conditions to also be digitally preserved in the long term. This is also applicable to any other human–human dialog that needs to be preserved in the long term.

There are currently no methods to make packages of emotion sequence data, video recordings, or session analysis data and documents out of human–human dialogs from analysis and documentation in the KM-EP that are available in the long term (Motivation Statement 3).

In the next section, the motivation statements introduced in this section will be addressed by identifying research questions.

1.2. Research Questions

In the previous section, the motivation for providing a dashboard with automatically calculated metrics and long-term digital preservation was introduced. The aim of creating a dashboard in the KM-EP is to enable emotion analysis and the documentation of human–human dialogs, such as with patient interviews, and facilitate information sharing between experts. This research project will be guided by the following problem descriptions and research questions.

Addressing *Motivation Statement 1*—the missing dashboard in the KM-EP—to analyze and document human–human dialogs in the KM-EP to visualize emotion sequence data; view video recordings; and support protocols, exchanges, and discussions, an expert tool needs to be developed. Analysis and documentation can, for example, be accomplished by annotating or segmenting data, as well as by exchanging information with other experts by using comments. Experts in this context could be medical diagnostic experts, computer science emotion analysis experts, or emotion experts, where each of them has a different point of interest in the data shown in such a dashboard. The medical diagnostic expert will be more interested in the human–human dialog, and the additional emotion sequence data will help in diagnosis, while the computer science emotion analysis expert is more likely to make an analysis if the recognized emotion sequence data in the video recordings are appropriate to the reactions shown by the conversation participants in the video recordings in order to improve emotion detection for instance. The emotion expert will be able to use the emotion sequence data to identify interesting parts of the conversation and analyze the emotions shown by the conversation participants in the video recordings. Dashboard experts for the same or even different groups of interest can exchange information and discuss. The dashboard, therefore, needs to support viewing video recordings, visualizing emotion sequence data, analysis and documentation, and exchanges between experts. Therefore, experts should be able to highlight interesting time sequences in the conversation by creating annotations or segmenting the conversation into different time sections. To enable an exchange, a dashboard should support creating comments addressing the emotion analysis, annotations, segments, or other comments. The results of the analysis and documentation are session analysis documents. The dashboard can then be used for patient interviews, as well as any other human–human dialogs that are recorded on video for which emotion sequence data have been analyzed and imported.

How can we provide a dashboard in the KM-EP to visualize emotion sequence data with video recordings and support experts in analyzing and documenting emotions in human–human dialogs in these videos? (Research Question 1)

Experts can thus use this dashboard to perform a subsequent analysis of the conversations and to analyze and document them accordingly. Since video recordings of human–human dialogs like patient interviews or discussion rounds in the SMILE project can, for example, last for an hour, analysis and documentation should be facilitated by an automated emotion analysis in advance. Therefore, it is helpful to calculate metrics based on the recognized emotions in the video recordings to simplify the analysis. This includes the calculation of metrics that indicate something about the emotions recognized in the conversation itself. For example, the human–human relationship between dialog participants in a recorded conversation can be evaluated, which, according to [19], is important for the success of treatment. To analyze human–human relationships, calculating metrics representing the mathematical relationship between the different emotion sequence datasets of dialog participants can be helpful. Session analysis data like mathematical relationships can then also be compared across different human–human dialog sessions, for example, to analyze the differences in the mathematical relationships of emotion sequence data for the same dialog participants across multiple sessions. In this research, only the mathematical relationships between emotion sequence data will be considered. Evaluating human–human relationships may be accomplished by future work or manually by experts using the calculated metrics.

How can we provide a dashboard in the KM-EP to calculate and visualize metrics representing the mathematical relationships between the emotion sequence data of human–human dialog participants recorded in videos? (Research Question 2)

Since the metrics on the mathematical relationships between the emotion sequence data of human–human dialog participants recorded in videos can, for example, provide feedback for treatment success in patient interviews, it can be useful to store these session analysis data in the patient’s medical records along with the session analysis documents and data of the patient’s interviews. Digital preservation also makes it possible to store video recordings and, along with them, emotion sequence data in the form of structured text files or visualizations. Therefore, packages of session analysis data and documents, video recordings, and the emotion sequence data of human–human dialogs need to be archived. Session analysis data are automatically generated data, like the metrics of *Research Question 2*, while session analysis documents are the results of analysis and documentation by experts using the dashboard in *Research Question 1*. The information archived should be constantly available and able to be loaded for sharing with third parties. For example, the long-term preservation of the patient’s medical records is relevant as well since there are various regulations regarding the retention period of patient medical records, depending on where the patient is located. For example, according to §9 para. 3 of the professional regulations of the Bavarian Chamber of Psychotherapists, the retention period is 10 years. Therefore, methods and standards are to be developed to maintain the information and to guarantee its accessibility in the long term.

How can we provide a dashboard in the KM-EP to archive packages of video recordings, emotion sequence data, and session analysis data and documents for human–human dialogs in the long term? (Research Question 3)

1.3. Methodology and Approach

To address our research questions, a methodical approach to research is necessary. Therefore, the multimethodological approach for information system research by Nuna-

maker et al. [20] will be used. This approach includes four research strategies (theory building, observation, experimentation, and system development) that are required to stay on the cutting edge of technology and to gain acceptance in organizations for research results. One of these strategies is the phase of theory building where new ideas and concepts are developed and conceptual frameworks are constructed. The phase of observation uses research methods, studies, and surveys to help formulate hypotheses and obtain a universal feeling for the challenges of the research domain by evaluating the current state of the art in science and technology. This phase of experimentation applies experiments and simulations to validate methods and theories. The last strategy is the system development phase, which consists of system design methods using prototyping, the development of products, and technology transfer [20].

The methodological framework introduced by Nunamaker et al. [20] will be used to ensure a structured approach while addressing the defined research questions—*Research Question 1*, *Research Question 2*, and *Research Question 3*—by planning and implementing specific R&D projects addressing our research goals, which will also use this methodological framework. This research will provide an initial overview of the planned R&D projects for this research program to partially address these research questions and to ensure interoperability between these R&D projects. These contributions will be summed up within the research project. Therefore, the R&D projects need to be planned and assigned to the research questions.

Having introduced the relevance of analyzing the emotions of conversation participants in video conferences with human–human dialogs, relevant challenges and motivations for emotion analyses in human–human dialogs were identified in this section. Subsequently, the research questions were presented as the basis of the research project. Next, the initially selected state of the art in science and technology will be presented. Although the summed-up research work will be structured according to the research questions and goals, this paper will only provide an initial overview of topics like analyzing emotions in video recordings, the trustworthiness and reproducibility of emotion analyses, and the Open Archival Information System for long-term preservation. We will further research these topics in future R&D projects, just as will topics like similarity searches. Based on the initially introduced state of the art and research questions, a draft for the management of emotion analyses and a sketch of the planned emotion analysis viewer user interface will be presented in Section 3. This paper will conclude by discussing the topics and providing an outlook on future R&D projects.

2. Initially Selected State of the Art in Science and Technology

An overview of the fundamentally relevant terms and existing technologies and models will be provided by presenting the current state of science and technology regarding emotion recognition in video recordings of human–human dialogs, information visualization, emotion analysis, trustworthiness and reproducibility, and long-term archiving. This is part of the observation goals according to the methodological framework by Nunamaker et al. [20].

2.1. Analyzing Emotions in Video Recordings of Human–Human Dialogs

Maier et al. [17] developed emotion-recognition algorithms inspired by the SenseCare project [15] at the University of Hagen. These algorithms can recognize emotions from various media, including video recordings, and can calculate them in real time during video conferences. Emotions are categorized into eight types (e.g., “happy”, “sad”, and “anger”) or measured in terms of arousal and valence. The output formats, which include softmax and independent values, produce numbers between 0 and 1 for each emotion

category or arousal and valence. In the softmax format, the sum of the emotion category values for each analyzed frame is 1. Different visualizations are possible based on these formats.

The emotion recognition engine [17] uses a Convolutional Neural Network (CNN) trained to classify images into eight emotion classes (neutral, happy, sad, surprise, fear, disgust, anger, and contempt) based on pixel data [17]. CNNs, modeled after the biological processes of the human brain [21], are particularly effective for processing audio and image data [21,22]. The SenseCare engine uses a MobileNetV3-based CNN, which is even more efficient for mobile devices [23]. The model was created and trained using Tensorflow and Keras [17].

Emotion sequence data from the [17] algorithm is available as time series data, which consists of a sequence of observations ordered temporally [24]. An emotion sequence is a finite series of emotion data, $e_1, e_2, \dots, e_m$, where m is the length of the emotion sequence of frames in the video recording analyzed. Each emotion datum corresponds to a frame and is classified into one of eight emotion classes. The order of the data reflects the temporal sequence of the video frames, making emotion sequence data a type of time series data.

To use the emotion sequence data, the modular architecture of Alphonsus Keary [25] is applicable, which was developed to enhance emotion recognition through the integration of visual and physiological data. This architecture consists of modules data acquisition, data preprocessing, feature extraction, data fusion, classification, and output generation. The first module, called data acquisition, involves collecting visual information like facial expressions—just like SenseCare—but also physiological signals like heart rates. Data preprocessing is used to clean and normalize data to reduce noise and ensure consistency for subsequent analyses. This step may be useful for improving the quality of the data for further analyses like calculating mathematical relations. The module feature extraction is used to identify relevant features out of the preprocessed data like, for example, the mathematical relationships between the emotion data sets of two dialog participants. The next module is data fusion, which combines features and data into a unified representation of the emotional state. This can be achieved using an emotion viewer where the video recording can be viewed and emotion data, as well as metrics like the mathematical relationships, can be visualized. This also enables a contextual understanding of emotions during the video conference. The next module, called classification, is used to put the recognized emotions into categories. This step is not required here, as the emotion classification is already completed by the emotion recognition engine [17]. The last module is output generation, which delivers the emotional states to user interfaces. This may include the emotion viewer, as well as the export of the analysis results, for example, using the long-term preservation of the results in a patient's medical records [25].

2.2. Trustworthiness and Reproducibility of Emotion Analyses

Additional key components of emotion analyses are their trustworthiness and reproducibility. AI systems must prove their trustworthiness to be accepted and used effectively. To achieve this, the IVIS4BigData model is extended in the research project presented in [26] to integrate analysis processes and data exploration supported by AI.

To extend this model, the dimensions of trust between individuals have been mapped to a systems context. While trust in human–human interactions builds upon reliability, honesty, and competence, trustworthiness in human–computer interactions builds upon reproducibility, validity, and capability. Reliability is defined as “the quality of individuals of performing consistently” [26], while reproducibility involves “obtaining consistent computational results using the same input data, computational steps, methods, code, and conditions of analysis” [26]. Honesty means an individual stays true to the facts, while

validity is the verification of the correctness of data and results. The third dimension is competence, which is defined as the ability and skill to solve problems and the willingness to do so, while in a systems context, it is defined as capability, which is the potential of a system to achieve a certain goal [26].

A model of trustworthiness in AI-based big data analysis is shown in Figure 1. It illustrates the knowledge process between big data analysis systems and their environment. The technical aspects of AI in the big data analysis process are on the right, while the “Data Intelligence” layer addresses trustworthiness. The “AI-Model Deployment” layer manages data and models for training and evaluation. The four layers “Data Integration”, “Data Analytics”, “Data Interaction”, and “Data Effectuation” represent the technical analysis process “to improve the effective benefit of big data technologies” [26]. AI can support the overall knowledge management and knowledge engineering process by being integrated into tasks like mapping data schemas, correcting data, explaining results, improving visualizations, and supporting interactions [26].

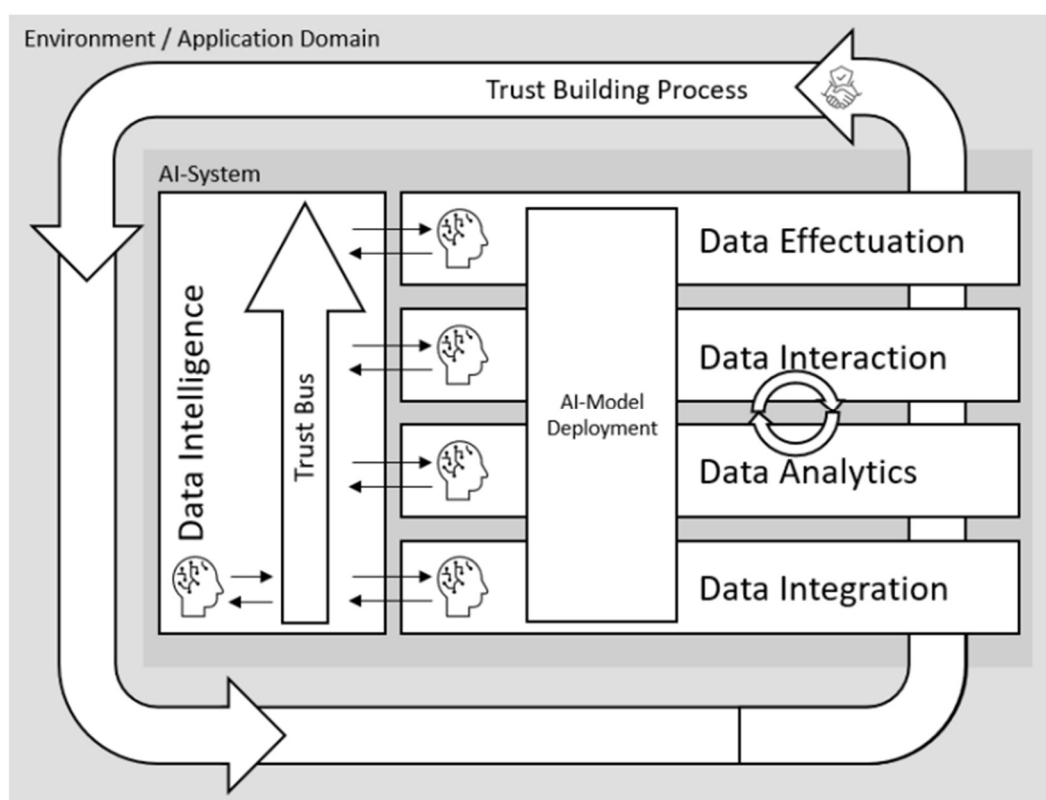


Figure 1. Trustworthy AI-based Big Data Management (TAI-BDM) reference model [26].

AI is, therefore, integrated in all layers, which requires all layers to have individual trust components based on their output within the “Data Intelligence” layer. To generate an overall trustworthy process, the layers’ trustworthiness needs to be integrated into an overall “Trust Bus”, which stores aspects of trustworthiness to be able to evaluate the trustworthiness of the overall big data analyses based on AI. The “Trust Building Process” represents past interactions between individuals and the systems that influenced trust in the AI-based system [26].

This research project focuses on the “Data Analytics” and “Data Interaction” layers, where emotion analyses will be visualized for users. Reproducibility is crucial for understanding how the data were calculated by the emotion-recognition algorithm [17]. This involves comparing emotion analyses to determine the effects of different processing methods or configurations. Influencing factors need to be stored, visualized, and compared

to achieve transparency. Reproducibility means “obtaining consistent computational results using the same input data, computational steps, methods, code, and conditions of analysis” [26]. Freire et al. [27] introduced the PRIMAD model to define influencing factors more precisely. PRIMAD stands for platform, research goal, implementation, method, actor, and data. This model captures variable components during data-oriented experiments to determine if the same results were achieved under different conditions. Sharing PRIMAD components is necessary for verifying reproducibility. Adjusting data can demonstrate the robustness and generalizability of the method [27].

To analyze emotion sequence data, it is necessary to store them with metadata and video recordings. This allows us to assign the emotion sequence data of different participants to one human–human dialog in the emotion analysis series. The data need to be visualized, and interactions like annotating should be supported (*Research Question 1*). Metrics for the mathematical relationships between emotion sequence data can help analyze human–human dialogs, such as when evaluating emotional connectedness (*Research Question 2*). The emotional connectedness of two people in a dialog is a concrete form or interpretation of a mathematical relationship between the emotion sequence data of the conversation participants. For this purpose, models and algorithms of time series analysis will be used regarding similarity searches and comparability (through noise suppression methods), and visualization and interaction possibilities for the exchange between experts need to be developed.

2.3. Open Archival Information System

Emotion analyses should be available in the long term in order to be, for example, stored in a patient’s medical records (*Research Question 3*). For this purpose, the Open Archival Information System (OAIS) reference model for long-term preservation and access to digital information will be considered. It includes best practices for saving data, managing metadata, replicating data, and more to secure the integrity and durability of digital resources. OAIS aims to support the availability of information over an extended period, although this is made more difficult by technological change (new data formats or media, a changing user community, etc.). It does not set methods to conduct activities involved in preserving a digital archive but instead sets standards for these activities. Using this framework ensures that organizations implement the necessary structures and processes to ensure long-term access to digital information [28].

An archive is defined in [28] as an institution or organization that preserves and makes records accessible. Traditional archives manage records like movies, papers, books, or maps and ensure their comprehension and authenticity for users. Traditionally, these media have been stored in long-term stable materials to keep the information secure and to control access. There are three entities that interact with an archive: producers, consumers, and management. Producers are individuals or systems providing information. Consumers are individuals or systems that access information. A special subgroup of consumers is the “designated community”, a defined user group that should be able to understand the stored information. The management sets overall policies and guidelines for the content stored in the archive. These individuals determine how the archive should be managed but are not involved in the day-to-day operation of the archive. It is possible for an individual or a system to act simultaneously as a producer and a consumer, depending on whether it supplies or retrieves data in the form of information objects [28].

Data objects, either physical or digital, contain at least one bit. Representation information provides data objects with additional meaning. For long-term preservation, archives should capture all available representation information to avoid costly re-identification later. Information transmission to an OAIS as shown in Figure 2 involves information

packages, which include content information (data object and representation information) and the Preservation Description Information (PDI) needed to preserve the content. Packaging information combines these elements, and descriptive information aids in search and retrieval functions. There are three types of information packages: Submission Information Packages (SIPs), Archival Information Packages (AIPs), and Dissemination Information Packages (DIPs). SIPs are sent by producers to the archive, AIPs are stored in the archive, and DIPs are sent to consumers [28].

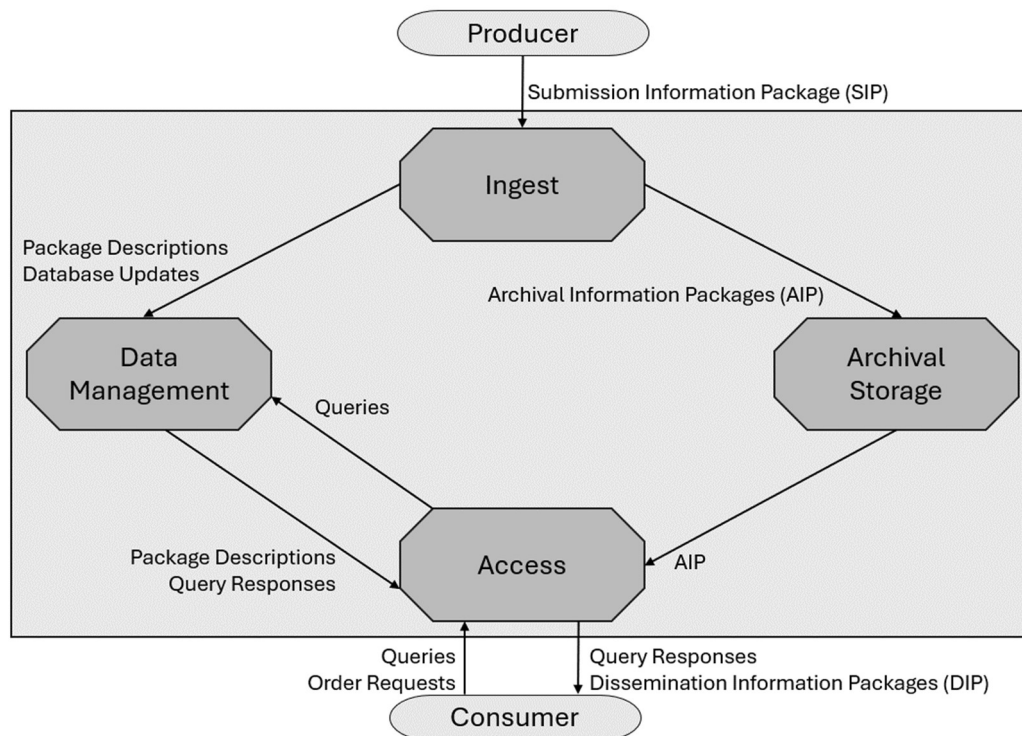


Figure 2. High-level data flow in an OAIS according to [28].

To make emotion analyses available in the long term, the necessary activities of the OAIS model should be modeled, implemented, and evaluated. To do so, emotion sequence data, along with the video recordings, visualizations, user annotations, and other metadata—such as information according to the PRIMAD model [27]—must be converted into an emotion data package. The contents of the emotion data package and how these need to be converted will be investigated as part of this research project.

This section provided an initially selected overview of the state of the art in science and technology, including emotion recognition from video recordings, the trustworthiness and reproducibility of emotion analyses, and long-term archiving using the OAIS reference model. The next section will initially sketch the modeling and design of a system to support emotion analysis with annotations, visualizations, metrics for mathematical relations, and long-term archiving, which is part of the theory-building goals according to the methodological framework by Nunamaker et al. [20].

3. Initially Sketched Modeling and Design

To illustrate the software prototype to be developed as part of this research project, a model based on the Trustworthy AI-based Big Data Management (TAI-BDM) reference model [26] for the management of emotion analyses is used as a starting point. Based on the model, the envisioned use cases are explained, and an interface sketch is presented.

The model for managing emotion analyses shown in Figure 3 supports the trust-building process for the emotion recognition algorithm executed by the emotion engine

as the four layers of the TAI-BDM Reference Model [26], as it enables us to execute the four technical layers, “Data Integration”, “Data Analytics”, “Data Interaction”, and “Data Effectuation”. By managing emotion analysis series containing multiple emotion analysis data, like video recordings of human–human dialogs, the recognized emotions can be stored. Therefore, one analysis series represents one dialog, and the video recording is assigned to the analysis series, while the emotion analyses contain the emotion sequence data and assigned metadata. The emotion analyses can contain the emotion sequence data of different conversation participants but also different results from calculating emotion sequence data with different parameters, on different platforms, or with other changed factors according to [27] in order to analyze the different results.

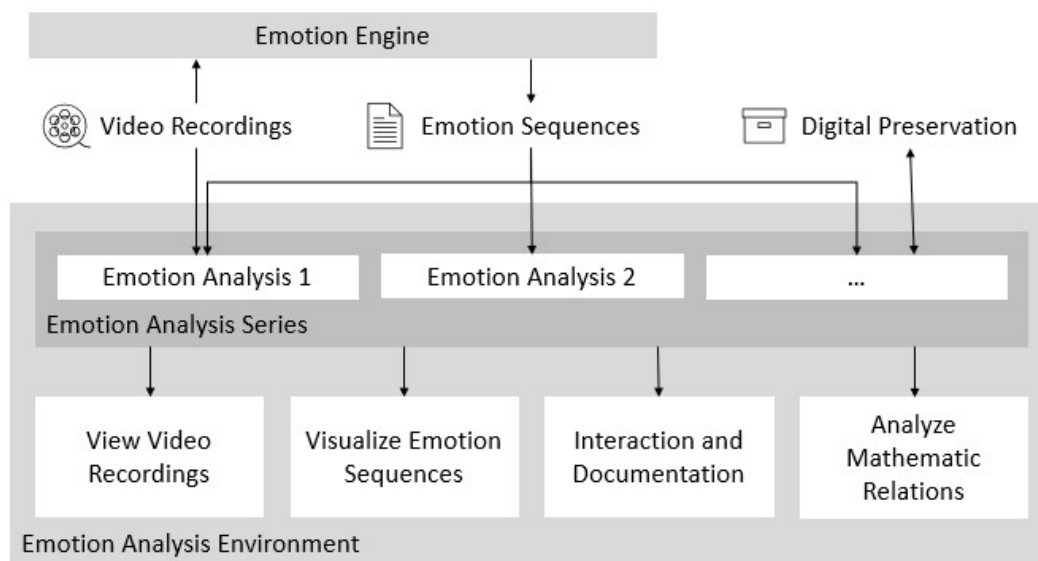


Figure 3. Initial draft of the emotion analysis management model.

The emotion analyses can be analyzed and compared using the emotion analysis environment by viewing video recordings, visualizing emotion sequences, interacting with the analysis results, and documenting them or analyzing mathematical relations by viewing calculated metrics. Experts might create annotations, segments, or comments to exchange knowledge or the results of analyzing the existing analysis with other experts. The results might then help to improve emotion recognition.

In contrast to the TAI-BDM reference model [26], there is no artificial intelligence used to support the technical layers; instead, it supports creating explainable results created by the emotion engine.

To address the research questions, the emotion analysis management model presented is to be implemented in a software prototype with the following initial use cases. The emotion analysis series containing emotion analyses needs to be managed with metadata, video recordings, and emotion sequence data. To analyze the data of the video recordings, they need to be viewed; emotion sequence data need to be visualized; and experts should be able to interact with the visualizations and document results by, for example, annotating or segmenting time periods (*Research Question 1*). To enable exchange, experts should be able to comment on the whole human–human dialog, segments, and annotations. To support the analysis, automatic metrics like mathematical relationships between emotion sequence data will be calculated (*Research Question 2*). To make the analysis results available over a long period of time, the future software prototype should not only support long-term digital preservation by implementing the OAIS reference model introduced in Section 2.3

but also support reimporting the emotion analysis series for further analyses (*Research Question 3*).

For the analysis, an emotion analysis viewer will be designed by sketching a graphical user interface, as shown in Figure 4. At the top of the viewer, the user should be able to select the analysis series and the analyses to be analyzed and compared, as well as which panels should be displayed. The video recording can be viewed in the emotion analysis viewer, where the segments and annotations should be highlighted in the timeline and listed next to the video recording. In addition, the segments and annotations should also be highlighted in the timeline of diagrams visualizing the emotion sequence data (here exemplified by a visualization using a line chart). It should be possible to select multiple different visualizations and view them simultaneously at the bottom of the viewer. The comments should be displayed by an overlay when the corresponding elements, such as annotations or segments, are selected. The comments referring to the whole analysis series will be shown equally to the annotations and segments but without time periods. The analysis series, with their included analyses, will be managed outside of the viewer to upload files, manage metadata, and import or export the analysis series.



Figure 4. Example emotion analysis viewer graphical user interface sketch.

The viewer may be used by different types of expert users. For example, emotion data experts may want to analyze the data quality of the emotion sequence data calculated by the emotion recognition algorithms. For therapists, the emotion data and their visualizations and metrics may support reviewing video recordings of therapy sessions in order to take additional notes. Therefore, different visualizations can be selected by the users to apply to their needs, while the video recording and the panel for annotations, segments, and comments will always be available to provide context to the emotion data and to document analysis results.

To sum up this section on the conceptual modeling of a tool for the annotation, visualization, and reproducible archiving of emotion sequence data from human–human dialog video recordings and to follow the methodology of this research project, first, a model for the management of emotion analyses was presented. Subsequently, a sketch of a possible graphical interface enabling experts to annotate and visualize emotion sequence data was presented, which enables the analysis of emotion sequence data from video

recordings of human–human dialogs. The emotion management model is, therefore, presented as a use case of the TAI-BDM reference model [26].

4. Discussion and Outlook

Motivated by Motivation Statement 1 (there are no existing opportunities to analyze and document emotions in recorded human–human dialog videos in the KM-EP), Motivation Statement 2 (there are no automatically calculated metrics to support the analysis), and Motivation Statement 3 (there are no methods to make packages of emotion sequence data, video recordings, or session analysis data and documents available in the long term), the planned research project on the annotation, visualization, and archiving of human–human dialog video recordings was presented in this paper. The research is based on the presented research questions.

Addressing the first research question (*Research Question 1*), on how to provide a dashboard in the KM-EP to visualize emotion sequence data with video recordings and to support experts in analyzing and documenting emotions in human–human dialogs, suitable visualization techniques must be researched before being modeled, implemented, and evaluated. To enable the analysis and documentation of emotion sequence data, suitable interaction techniques must be offered that support data exploration and annotation and, therefore, documentation and exchange with other experts. For this purpose, the research project must first research the basics of information visualization, as well as the different visualization techniques and interaction possibilities, and examine their applicability to the context of emotion sequence data from human–human dialogs. Subsequently, a dashboard must be modeled, implemented, and evaluated that offers various visualization techniques and interaction options for experts to examine or compare emotion analyses. Interaction techniques for documenting emotion analyses, such as annotation or segmentation, must be considered.

The second research question (*Research Question 2*), on how to provide a dashboard in the KM-EP to calculate and visualize metrics representing the mathematical relationships between the emotion sequence data of human–human dialog participants recorded in videos, requires research on algorithms for calculating these metrics. For example, similarity search algorithms can be used for this purpose. These metrics can be used to assess the mathematical relationships based on the similarity or dissimilarity of the emotion sequence data, and the emotional connection may then be assessed by corresponding emotion experts. To this end, the algorithms should be configurable and executable via an execution workflow in a graphical user interface, but the results should also be visualized in a dashboard. To ensure that the metrics for the mathematical relationships of emotion sequence data are not falsified by measurement-related fluctuations in emotion recognition, it should also be possible to reduce noise before calculating the metrics for the mathematical relationships. Algorithms must, therefore, be researched that enable noise reduction and the calculation of mathematical relations and are suitable for the context of the emotion sequence data to create an execution workflow and a dashboard.

Focusing on the third research question (*Research Question 3*), on how to provide a dashboard in the KM-EP to archive packages of video recordings, emotion sequence data, and session analysis data and documents of human–human dialogs in the long-term, the OAIS reference model will be applied. An initial overview of the OAIS reference model has already been provided in this paper. In the further progress of this research project, it remains to be investigated, modeled, implemented, and evaluated regarding how emotion data packages consisting of video recordings, emotion sequence data, and session analysis data and documents of human–human dialogs can be converted into packages according to the standard; which metadata are relevant and should be archived with the packages; and

how these packages can be re-imported in order to continue the investigation of emotion analyses. In addition, compression techniques for reducing storage requirements are to be researched and, if applicable to the context of the research project, modeled, implemented, and evaluated.

The emotion analysis management model presented in Figure 3 addresses the three research questions. However, the model is only conceptual and is only intended to provide an overview of the planned research work. During the research project, a more detailed examination of the state of the art in science and technology on the individual research questions is required, as well as the modeling, implementation, and evaluation of the various aspects of the research project based on this. The research project should be based on the PRIMAD model according to [27] and the TAI-BDM reference model for trustworthiness [26].

Author Contributions: Conceptualization, V.S.; methodology, M.H. and M.X.B.; validation, V.S.; investigation, V.S.; writing—original draft preparation, V.S.; writing—review and editing, V.S., M.H. and M.X.B.; visualization, V.S.; supervision, M.H. and M.X.B.; project administration, V.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article; further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial intelligence
KM-EP	Knowledge Management Ecosystem Portal
CNN	Convolutional Neuronal Network
TAI-BDM	Trustworthy AI-based Big Data Management
OAIS	Open Archival Information System
PDI	Preservation Description Information
SIP	Submission Information Package
AIP	Archival Information Package
DIP	Dissemination Information Package

References

1. Pavone, P.; Ceccarelli, M.; Taibi, R.; La Rocca, G.; Nunnari, G. Outbreak of COVID-19 infection in children: Fear and serenity. *Eur. Rev. Med. Pharmacol. Sci.* **2020**, *24*, 4572–4575. [[PubMed](#)]
2. Theberath, M.; Bauer, D.; Chen, W.; Salinas, M.; Mohabbat, A.B.; Yang, J.; Chon, T.Y.; Bauer, B.A.; Wahner-Roedler, D.L. Effects of COVID-19 pandemic on mental health of children and adolescents: A systematic review of survey studies. *SAGE Open Med.* **2022**, *10*, 20503121221086712. [[CrossRef](#)] [[PubMed](#)]
3. Juruena, M.F.; Eror, F.; Cleare, A.J.; Young, A.H. The role of early life stress in HPA axis and anxiety. In *Advances in Experimental Medicine and Biology*; Springer: Amsterdam, The Netherlands, 2020; pp. 141–153. [[CrossRef](#)]
4. Winstone, L.; Mars, B.; Haworth, C.M.A.; Kidger, J. Types of Social Media Use and Digital Stress in Early Adolescence. *J. Early Adolesc.* **2023**, *43*, 294–319. [[CrossRef](#)]
5. Usher, K.; Durkin, J.; Bhullar, N. Eco-anxiety: How thinking about climate change-related environmental decline is affecting our mental health. *Int. J. Ment. Health Nurs.* **2019**, *28*, 1233–1234. [[CrossRef](#)] [[PubMed](#)]

6. Anand, N.; Sharma, M.K.; Thakur, P.C.; Mondal, I.; Sahu, M.; Singh, P.; J., A.S.; Kande, J.S.; Neeraj, M.S.; Singh, R. Doomsurfing and doomscrolling mediate psychological distress in COVID-19 lockdown: Implications for awareness of cognitive biases. *Perspect. Psychiatr. Care* **2022**, *58*, 170–172. [CrossRef] [PubMed]
7. Sherif, A. Share of Employees Working Primarily Remotely Worldwide 2015–2023. Statista. Available online: <https://www.statista.com/statistics/1450450/employees-remote-work-share/> (accessed on 18 August 2024).
8. Sherif, A. Global Challenges Experienced in Video Meetings 2021, by Category. Statista. Available online: <https://www.statista.com/statistics/1255656/challenges-experienced-online-video-meetings-global/> (accessed on 18 August 2024).
9. Bouchard, S.; Allard, M.; Robillard, G.; Dumoulin, S.; Guitard, T.; Loranger, C.; Green-Demers, I.; Marchand, A.; Renaud, P.; Cournoyer, L.-G.; et al. Videoconferencing Psychotherapy for Panic Disorder and Agoraphobia: Outcome and Treatment Processes From a Non-randomized Non-inferiority Trial. *Front. Psychol.* **2020**, *11*, 2164. [CrossRef] [PubMed]
10. Stadler, M.; Jesser, A.; Humer, E.; Haid, B.; Stippl, P.; Schimböck, W.; Maaß, E.; Schwanzar, H.; Leithner, D.; Pieh, C.; et al. Remote Psychotherapy during the COVID-19 Pandemic: A Mixed-Methods Study on the Changes Experienced by Austrian Psychotherapists. *Life* **2023**, *13*, 360. [CrossRef] [PubMed]
11. Schultz, J.-H.; Alvarez, S.; Nikendei, C. *Heidelberger Standardgespräche—Handlungsanweisungen zur Ärztlichen Gesprächsführung*; HeiCuMed: Heidelberg, Germany, 2018.
12. FTK e.V. Forschungsinstitut für Telekommunikation und Kooperation. Smart Vortex. Available online: <https://www.ftk.de/de/smart-vortex> (accessed on 21 April 2025).
13. Bornschlegl, M.X.; Berwind, K.; Kaufmann, M.; Engel, F.C.; Walsh, P.; Hemmje, M.L.; Riestra, R. *IVIS4BigData: A Reference Model for Advanced Visual Interfaces Supporting Big Data Analysis in Virtual Research Environments*. In Proceedings of the Advanced Visual Interfaces. Supporting Big Data Applications, Bari, Italy, 7–10 June 2016; pp. 1–18. [CrossRef]
14. Vu, B.; Develasco, M.; Mc Kevitt, P.; Bond, R.; Turkington, R.; Booth, F.; Mulvenna, M.; Fuchs, M.; Hemmje, M. A Content and Knowledge Management System Supporting Emotion Detection from Speech. In *Conversational Dialogue Systems for the Next Decade*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 369–378. [CrossRef]
15. FTK e.V. Forschungsinstitut für Telekommunikation und Kooperation. SenseCare: Sensor-Gestütztes Affective Computing für Verbesserte Medizinische Versorgung. Available online: <https://www.ftk.de/de/projekte/sensecare> (accessed on 26 March 2025).
16. FTK e.V. Forschungsinstitut für Telekommunikation und Kooperation. SMILE—Supporting Mental Health in Young People: Integrated Methodology for Clinical Decisions and Evidence-Based Interventions. Available online: <https://www.ftk.de/en/projects/smile> (accessed on 21 April 2025).
17. Maier, D.; Hemmje, M.; Kikic, Z.; Wefers, F. Real-Time Emotion Recognition in Online Video Conferences for Medical Consultations. In *Machine Learning, Optimization, and Data Science*; Springer: Berlin/Heidelberg, Germany, 2024; pp. 479–487. [CrossRef]
18. Falcetta, F.S.; de Almeida, F.K.; Lemos, J.C.S.; Goldim, J.R.; da Costa, C.A. Automatic documentation of professional health interactions: A systematic review. *Artif. Intell. Med.* **2023**, *137*, 102487. [CrossRef] [PubMed]
19. Universität Heidelberg. Psychotherapeutische Hochschulambulanz. Available online: <https://www.psychologie.uni-heidelberg.de/psychotherapie/hochschulambulanz/> (accessed on 21 April 2025).
20. Nunamaker, J.F.; Chen, M.; Purdin, T.D. Systems Development in Information Systems Research. *J. Manag. Inf. Syst.* **1990**, *7*, 89–106. [CrossRef]
21. LeCun, Y.; Kavukcuoglu, K.; Farabet, C. Convolutional networks and applications in vision. In Proceedings of the 2010 IEEE International Symposium on Circuits and Systems, Paris, France, 30 May–2 June 2010; IEEE: Piscataway, NJ, USA, 2010; pp. 253–256.
22. Albawi, S.; Mohammed, T.A.; Al-Zawi, S. Understanding of a convolutional neural network. In Proceedings of the 2017 International Conference on Engineering and Technology (ICET), Antalya, Turkey, 21–23 August 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1–6. [CrossRef]
23. Howard, A.; Sandler, M.; Chen, B.; Wang, W.; Chen, L.-C.; Tan, M.; Chu, G.; Vasudevan, V.; Zhu, Y.; Pang, R.; et al. Searching for MobileNetV3. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1314–1324. [CrossRef]
24. Gogolou, A.; Tsandilas, T.; Palpanas, T.; Bezerianos, A. Comparing Similarity Perception in Time Series Visualizations. *IEEE Trans. Vis. Comput. Graph.* **2019**, *25*, 523–533. [CrossRef] [PubMed]
25. Keary, A.; Delaney, K.; Hemmje, M. Affective Computing for Emotion Detection Using Vision and Wearable Sensors. Ph.D. Thesis, Cork Institute of Technology, Cork, Ireland, 2018.
26. Bornschlegl, M.X. *Towards Trustworthiness in AI-Based Big Data Analysis*; FernUniversität in Hagen: Hagen, Germany, 2024. [CrossRef]

27. Freire, J.; Fuhr, N.; Rauber, A. Reproducibility of Data-Oriented Experiments in e-Science (Dagstuhl Seminar 16041). *Dagstuhl Rep.* **2016**, *6*, 108–159. [[CrossRef](#)]
28. Management Council of the Consultative Committee for Space Data Systems. *Reference Model for an Open Archival Information System (OAIS)*; Management Council of the Consultative Committee for Space Data Systems: Laurel, MD, USA, 2020; Volume Pink Book, No. 2.1.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.