Enhancing and Re-Purposing TV Content for Trans-Vector Engagement

Deliverable 6.3 (M39)
Second Validation of Personalization Prototype
Version 1.0

DOCUMENT INFORMATION

Delivery Type	Report
Deliverable Number	6.3
Deliverable Title	Second Validation of Personalized Prototype
Due Date	M39
Submission Date	31. March 2021
Work Package	WP6
Partners	RBB, NISV, GEN, CERTH
Author(s)	Willy Lamm, Annette Wilson, Karolin Krämer (RBB), Rasa Bocyte, Dimitra Atsidis (NISV), Basil Philipp, Michael Kündig, Krzysztof Ciesielski (GEN), Kostas Apostolidis, Vasileios Mezaris (CERTH)
Reviewer(s)	Vasileios Mezaris
Keywords	Consumer Use Cases, Personalised Video Content, Video Recommendations, Video Summaries, User Engagement
Dissemination Level	PU
Project Coordinator	MODUL Technology, Am Kahlenberg 1, 1190 Vienna, Austria.
Contact Details	Coordinator: Dr Lyndon Nixon (nixon@modultech.eu) R&D Manager: Prof Arno Scharl (scharl@weblyzard.com) Innovation Manager: Bea Knecht (bea@zattoo.com)

Revisions

Version	Date	Author	Changes
0.1	02.03.2021	Willy Lamm (RBB)	Table of Contents
0.2	09.03.2021	Vasileios Mezaris, Kostas Apostolidis (CERTH)	Input for Video Summarization Web Service
0.3	10.03.2021	Michael Kündig, Basil Philipp (GEN), Rasa Bocyte (NISV)	Input for 4u2 Messenger and Content sWitch
0.4	16.03.2021	Annette Wilson, Willy Lamm (RBB)	Input for 4u2 Smart Speaker Skill
0.41	20.03.2021	Arno Scharl (WLT)	Minor edits and revisions
0.5	22.03.2021	Michael Kündig (GEN), Willy Lamm (RBB)	Input for 4u2 Smart Speaker Skill
0.6	24.03.2021	Dimitra Atsidis (NISV), Krzysztof Ciesielski (GEN)	Input for 4u2 Messenger
0.7	25.03.2021	Annette Wilson, Karolin Krämer, Willy Lamm (RBB)	Input for Executive Summary, Introduction, 4u2 Smart Speaker Skill and Conclusion, minor edits
0.8	26.03.2021	Vasileios Mezaris, Kostas Apostolidis (CERTH)	QA review
0.9	29.03.2021	Rasa Bocyte (NISV), Annette Wilson, Karolin Krämer, Willy Lamm (RBB), Basil Philipp (GEN), Kostas Apostolidis (CERTH)	Post-QA corrections
1.0	30.03.2021	Lyndon Nixon (MOD)	Final check

Statement of Originality

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

This deliverable reflects only the authors' views and the European Union is not liable for any use that might be made of information contained therein.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Following a user-centric, flexible approach ReTV has developed a number of consumer applications and services using the capabilities of the Trans-Vector Platform (TVP)[1]. The aim was to offer consumers more relevant, engaging, tailor-made content created using novel personalisation and interaction functionalities on popular online platforms.

This deliverable describes how the initial prototypes within the project were further developed to create the **4u2 Messenger** application, the **4u2 Smart Speaker Skill** and the **Video Summarization Web Service**.

Both the 4u2 Messenger and the 4u2 Smart Speaker Skill were tested in longitudinal surveys with users in their home environment to ensure they had sufficient time to thoroughly test the applications under real world conditions. The evaluation indicates a high level of acceptance of the services and provided constructive feedback on possible improvements. The fact that there is also a demand for the Video Summarization Web Service on behalf of the users is shown by the usage statistics of the publicly available tool.

---

[1] The Trans-Vector Platform (TVP is an open architecture of components and subsystems developed within ReTV that, in flexible configurations, enable a content owner or creator to optimise their content's impact and reach through contextualisation, prediction, and adaptation mechanisms (see D4.3).

1 INTRODUCTION

The final WP6 deliverable of ReTV describes the work conducted in Task *6.3* (T6.3), *TVP Deployment and Testing with Consumers*. It consists of updated prototypes to evaluate and demonstrate the TVP functionality for consumers and reports the evaluation results. In the deliverable *D7.4 Business and Exploitation Plan*, ReTV partners described the steps planned for exploitation and commercialisation of ReTV results. The outcomes of the user evaluation reported here will underpin those efforts by providing feedback on user acceptance and guidance for optimization and further development. In this document we focus on three services:

- **4u2 Messenger** automatically distributes a personalized selection of video content directly to users via a personal messaging application. It was implemented for NISV and tested using archival content.

- **4u2 Smart Speaker Skill** is a skill (application) for smart speakers that enables users to create their own videos from existing archival material using voice commands. It was implemented and tested for the RBB children's programme "Unser Sandmännchen".

- **Video Summarization Web Service** is a public demo of an essential technology developed in the context of ReTV which was otherwise utilized by ReTV's TVP application *Content Wizard* (see Sections 3 and 5 of D5.3).

Deliverable *D6.2 First Validation of Personalization Prototype* reported on the first consumer prototypes developed and tested in ReTV. Task 6.3 continued to evolve the prototypes and related services, guided by both consumer feedback, as well as market developments and evolving requirements. This deliverable summarizes this work on the above-mentioned services.

Furthermore, we also explain the decision to discontinue the work on the **Content sWitch** which aimed to enable broadcasters to designate areas of their programming for on-demand personalization and thus enhance the linear TV experience by replacing certain parts with alternatives tailored for each viewer.

The Deliverable is structured as follows:

- **Section 2** focuses on the *4u2 Messenger*, presenting in detail its development and the setup of the final prototype as well as the evaluation methodology and results.

- Following the same structure, **Section 3** describes the *4u2 Smart Speaker Skill* scenario.

- The *Video Summarization Web Service* is presented in **Section 4**, describing the service itself, its user interface and its backend service integration. Insights from the evaluation of the summarization method as well as usage statistics are also presented.

- **Section 5** deals with the development of the *Content sWitch* and the rationale for the decision to not continue the research work on it within the project.

- **Section 6** summarizes the evaluation results and provides an outlook on the potential of the developed services with regard to future improvements and areas of use.

2 4U2 MESSENGER

The goal of the 4u2 scenario in ReTV is to develop solutions that personalize broadcaster content distribution and delivery, and propose novel ways for media archives and broadcasters to build connections with their audiences. Responding to this, the 4u2 Messenger application **automatically distributes a personalized selection of video content directly to users** via personal messaging applications. It largely builds on and improves the 4u2 Chatbot prototype presented in D6.2. 4u2 Messenger is envisioned as a flexible framework that could be adapted to work with widely available personalized messaging applications. During the project, a prototype of this application was implemented for NISV using Facebook Messenger.

Unlike the already common-used social media channels that address large audience segments, a messaging application allows broadcasters and media archives to reach individual users and tailor the video specifically to them. It actively pushes archival content into the information and entertainment stream of online audiences. The 4u2 Messenger was built with the specific goal in mind: to leverage the breadth of audiovisual archival content, especially the long tail of large-scale collections, in order to build long-lasting engagement between content owners and their audiences.

The following sections describe the development of the use case scenario since the previous reporting period and the functionalities of the final prototype, as well as its technical implementation. At the end of this chapter, we present the evaluation of the 4u2 Messenger application with end users.

2.1 SCENARIO DEVELOPMENT

In D6.2, we presented a concept and a first prototype for the 4u2 Chatbot application that was used as a foundation for further research. The first prototype of the chatbot was built on the Telegram mobile application but the evaluation results at the time revealed that to successfully exploit it and attract users, other messaging platforms would be more convenient, specifically WhatsApp, which was ranked as the most important during the user tests reported in D6.2. However, in December 2019 WhatsApp announced that it would no longer support any mass messaging via its main application to stop any companies or individuals who abuse this feature to spam users.[2] While the 4u2 Chatbot scenario does not violate these rules, still it was no longer possible to implement the scenario on this messaging application.

However, from a research standpoint, it was still interesting to investigate the potential of personalized messaging methods for organisations such as broadcasters and media archives. While the dependence on changes and regulations imposed by companies responsible for external applications and platforms limits technical possibilities and in some cases, present ethical concerns regarding data collection, the majority of organisations are not likely to leave these platforms until their audiences move away from them. Developing native platforms and applications is one of the possible alternatives. However, this is a more resource-demanding option that is not feasible for many organisations with limited resources. Also experiments with similar applications demonstrate that users tend to quickly lose interest in them unless they are

---

[2]

https://faq.whatsapp.com/general/security-and-privacy/unauthorized-use-of-automated-or-bulk-messaging-on-whatsapp

coupled with other services of interest to the user. Therefore, NISV decided to pursue the development of this use case and, in the process, address the following questions:

- How can media organizations use personal messaging applications to their advantage?
- What concerns should organisations address when developing and using such personalization-driven applications on external platforms?

During the scenario development and implementation process, NISV researchers forged a strong collaboration with their internal marketing department who showed high interest in the application during the first validation round and provided valuable input to guide the development of the application. In discussions with this team, it was decided to build the application for Facebook Messenger, primarily because NISV has a large number of users following on Facebook who are very actively engaging with audiovisual content from the archive shared on the platform. Furthermore, many other audiovisual archives and broadcasters use Facebook as a main external platform to publish audiovisual content. Curation and publication of this content is a time consuming task, and the COVID-19 pandemic further highlighted that the demand for content could not always be matched with the available time and resources to the team. Thus, this implementation for the Facebook Messenger would fit the business requirements for NISV as well as present a scenario applicable to other organisations in the field.

NISV proposed to use its Polygoon newsreel collection for the use case implementation.[3] The collection is composed of over 2.000 videos dating primarily between 1920-60s that need to be contextualized and presented in a new light to appeal to contemporary audiences. Therefore it was decided to build the application around the "This Week in History" concept popularized on social media platforms where content from the past is shared on a relevant date. Our initial proposal was to use the TVP's Prediction Service to determine how videos are selected since, during the first validation stage, the majority of users indicated that they would like to receive videos related to currently trending news stories. However, due to their age, a lot of the videos from the selected collection have little connection to contemporary news and would require further editorial efforts to contextualize them, and therefore the fully automated workflow of the 4u2 scenario would not bring satisfactory results.[4] With the "This Week in History" approach, the recommendation engine would look at the videos published in the same week across the years and use this content pool to curate 4-5 videos per week for each user. The process of narrowing down the collection to individual weeks leaves between 20 and 30 videos to choose from and contributes to a more meaningful automated curation process given the seasonality of topics represented in the collection (for example, holidays or historic events commemorated annually).

An important factor for NISV was working with a recommendation algorithm that does not place users in filter bubbles and instead tries to promote more diverse content, making full use of a large-scale audiovisual collection. While the algorithm should try to satisfy the diverse and niche interests of its users with highly related content, it was equally important to introduce a level of serendipity by recommending more unexpected videos. For example, a user who shows a high preference for sports-related content would be regularly nudged to watch videos about topics that might have some overlap with sports. User reaction to this content would help to

---

[3] https://openbeelden.nl/.en

[4] This is done in the Content Wizard application where the same collection is used to create social media posts. Here, users can manually create text to accompany the video and show why the historic archival content is relevant for today's new stories. See *D5.3 Second Validation of Engagement Monitoring Prototype* for more detail.

expand their regions of interest and create a more diverse profile. This element of surprise should also keep audiences engaged for longer and spark their curiosity to discover new things. This was achieved by balancing the relevance and diversity scores in the recommendation engine (see Section 3.2.1. of D3.3 for more details).

Development Process for Facebook Messenger

During the first validation of the 4u2 scenario reported in D6.2, end users provided the following feedback that went on to guide further development. In light of the switch to Facebook Messenger and in consultation with the NISV marketing department and ReTV partners, this feedback was re-evaluated. Table 1 summarizes how these user comments were taken into consideration to implement the first version of the prototype for Facebook Messenger (see Fig. 1).

User Comments	Implementation Approach
Users want to receive videos on request as well as at regular intervals.	Given that the focus of the application was a recommendation engine that tries to match users with content from the archive, we decided to focus on distribution of videos at regular intervals. At a larger stage, a hybrid approach could be introduced where content delivery is automatically scheduled but each user can also request a video at any moment (their requests could then be used to further optimise the recommendations).
Users want to explicitly provide feedback on their interests that would provide input for the personalization algorithm.	We decided to use Facebook Messenger's functionality where users can react to the messages with a number of emojis. These reactions were then used as input for the recommendation engine. Because of this change, we thought it was no longer accurate to call the application *4u2 Chatbot* as it does not support the communication between the user and the application that users might expect from chatbots. Hence, it was decided to call it *4u2 Messenger.*
Users want to engage with the application via buttons or by entering free text.	Given the choice of emojis, this requirement was no longer applicable at this stage. When the feature where users can request videos at any time is implemented, it could be facilitated via buttons or free text.
Users want to receive a gif preview of the video, rendered by CERTH's Video Summarization service.	In consultation with the NISV marketing team, it was decided that this preview was not necessary since the videos from the Polygoon newsreel collection used for the prototype are already relatively short.

Table 1: User comments on the 4u2 scenario from D6.2 and their implementation approach in the first version of the 4u2 Messenger.
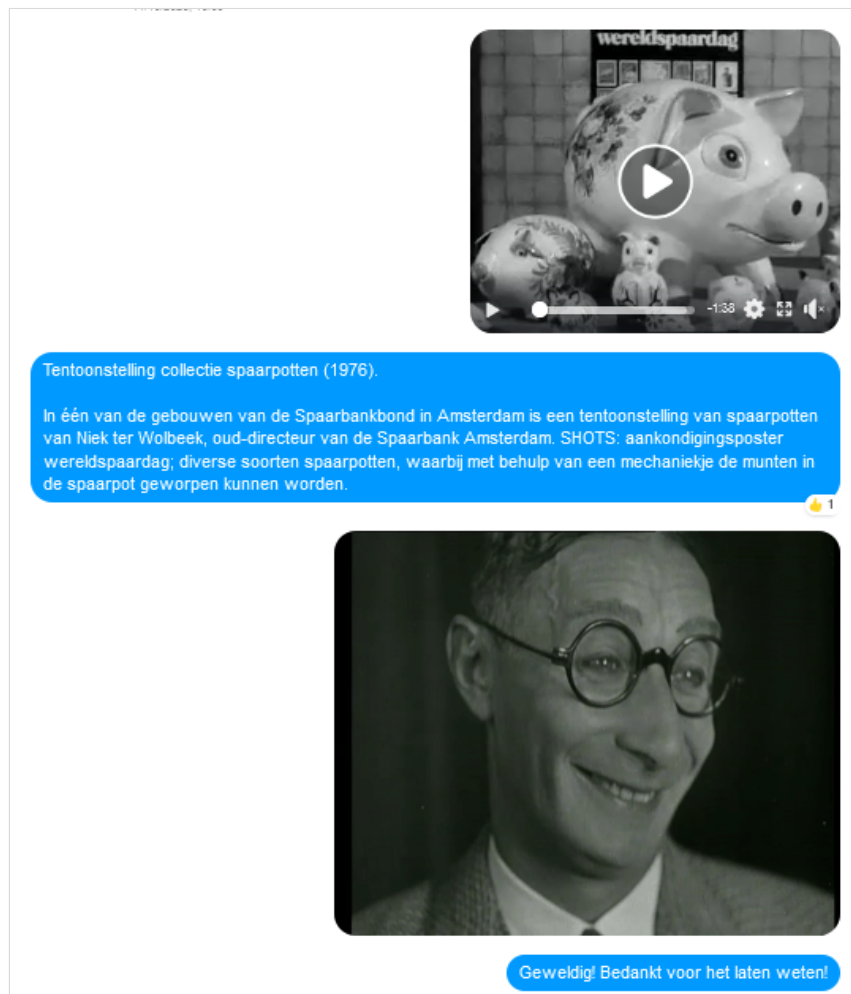
Figure 1: Screenshot of messages sent out by the first version of the 4u2 Messenger prototype on Facebook Messenger. A thank you message and a gif is sent to the users after they provide their reaction.

However, we encountered multiple obstacles that prevented us from releasing this version of the 4u2 Messenger as planned. In early 2021, Facebook announced two major changes that forced us to make fundamental changes to the prototype. First, it was announced that the functionality to directly send videos that could be watched within the Messenger application would be discontinued for EU-based pages due to the new privacy regulations.[5] Additionally, Facebook disabled the ability to see when a user reacted to a message. These features were turned off at a short notice, and might be reinstated at some point, but it is not clear when.
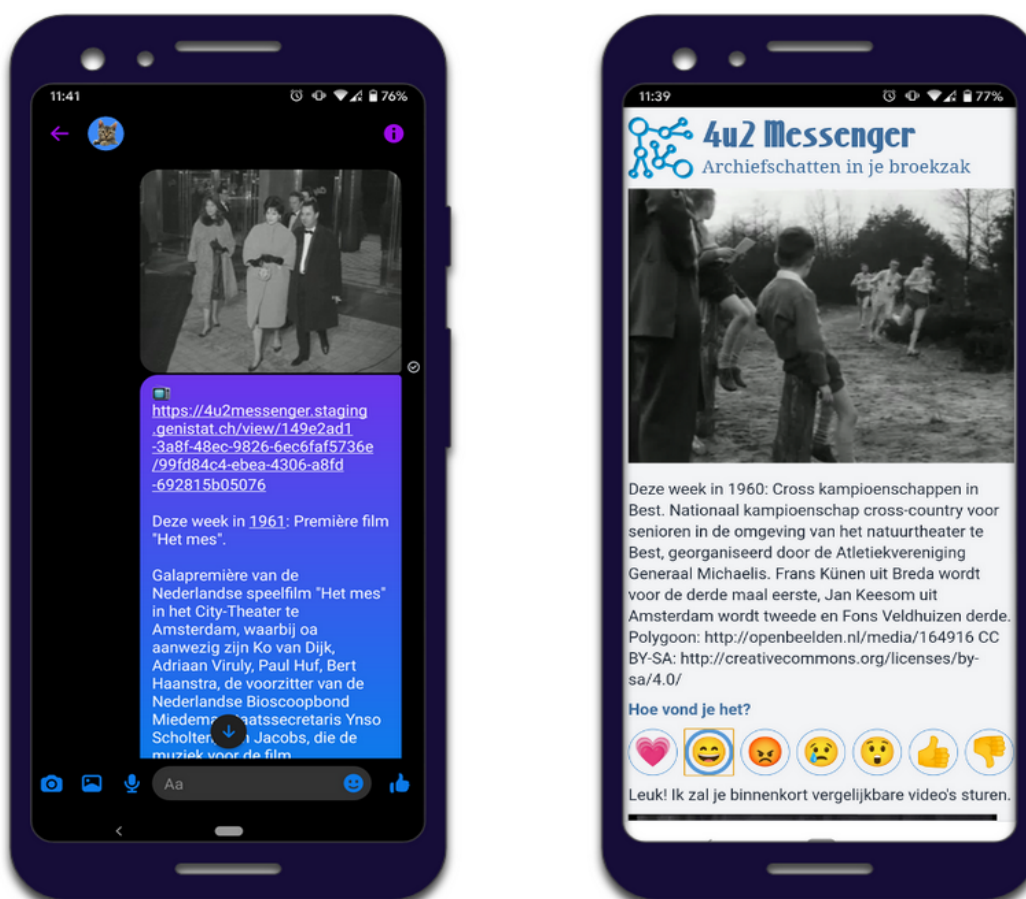
This change had significant consequences to the user experience we aimed to facilitate with the 4u2 Messenger, and we had to take swift measures to be able to test the application with users in Q1 of 2021. A decision was made to build a website where users could watch videos and react to them with the same emojis as they would have on the Facebook Messenger. This enabled us to recreate a very similar user experience to the one that was originally intended in a short period of time.

---

[5] https://developers.facebook.com/docs/messenger-platform/europe-updates

2.2 FINAL PROTOTYPE

The 4u2 Messenger prototype is connected to a dedicated Facebook page and users can start engaging with it by simply sending a message to the page.[6] From the moment the user signs up to use the 4u2 Messenger, the whole process is fully automated.[7] First, users are greeted by the messenger and asked to subscribe. After they do so, they are added to the list of active users and receive their first video within the next 24 hours.

Each week, users receive 4-5 messages from the 4u2 Messenger (see the left side of Fig. 2) that contains the video thumbnail, title and a short description of the video, copyright information, and most importantly, a link to the external browser where the video can be watched. In the external browser (see the right side of Fig. 2), the user can watch the full video and provide their feedback using one of the emoji reactions: heart/love, smile, angry, sad, surprised, thumbs up and thumbs down. Based on these explicit reactions, the recommendation algorithm is adjusted to better match their interests (see Section 3.2.1. of D3.3 for more details).

Figure 2: 4u2 Messenger. Left: example of a message that a user received in their Facebook Messenger application. Right: the external browser where the user can watch the video and add an emoji reaction.

Each user receives different videos sent at different times. The publication time is determined by monitoring each user's past behaviour and is iteratively optimised. The scheduled and distributed messages can be monitored on the editorial backend (see Fig. 3). If required, this interface could be also used to manually schedule and publish messages to all users.

The "This Week in History" concept chosen for the NISV prototype demonstrates how the application could be customized to support the distribution of a particular type of content. Other media organisations interested in using the application for their purposes would have to define this concept and after that, the whole process is fully automated. This would enable organisations to sustainably engage with their audiences with little to no edit efforts needed.
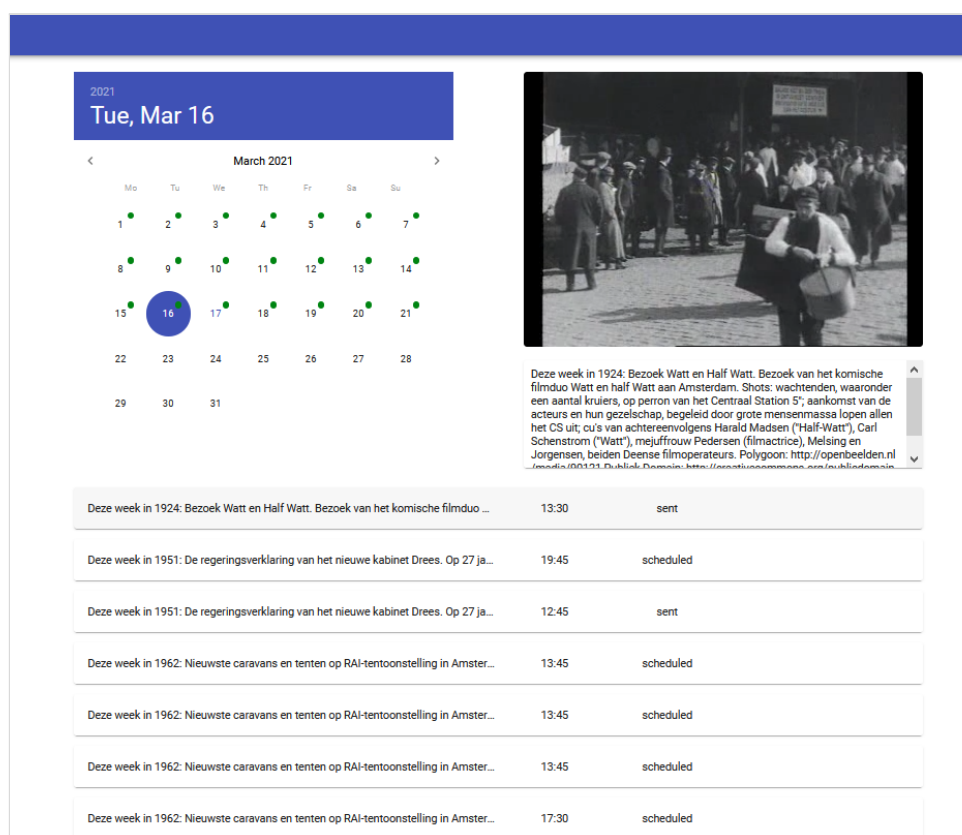
Figure 3: The editorial backend with an overview of scheduled and already distributed messages.

Technical implementation

The description of the recommendation algorithm was presented in Section 3 of D3.3. Its main features are:

- video similarity is based on text2video embeddings developed in WP3;
- user profile is constructed as an aggregated, weighted "embedding", based on individual video/scenes embeddings. The user embedding is close to the videos that the user perceived positively and further away from the videos that user perceived negatively;

- the mapping from Facebook emojis to a positive/negative sentiment (numerical weight) is learned automatically from behavioral data (the optimized metric used is the user retention);
- the optimal scheduling time is learned automatically from behavioral data such that the distance between message sent time and opening time is diminishing over time;
- the scheduling date is chosen based on the "This Week in History" concept, described in the previous section. The algorithm balances the distance from the exact anniversary of the event presented in the video and the uniform distribution of videos scheduled for a given user in the current week.

The messenger consists of multiple components we developed. They all interface with the user through a Facebook app. Users grant our Facebook app permission to connect with their Facebook account[8]. After this explicit approval, we received the public User ID of the user, allowing us to send them messages.

The system architecture depicted in Fig. 4 consists of the following components:

- fetcher
- hooker
- react-app frontend
- Airflow scheduler DAG[9]

___
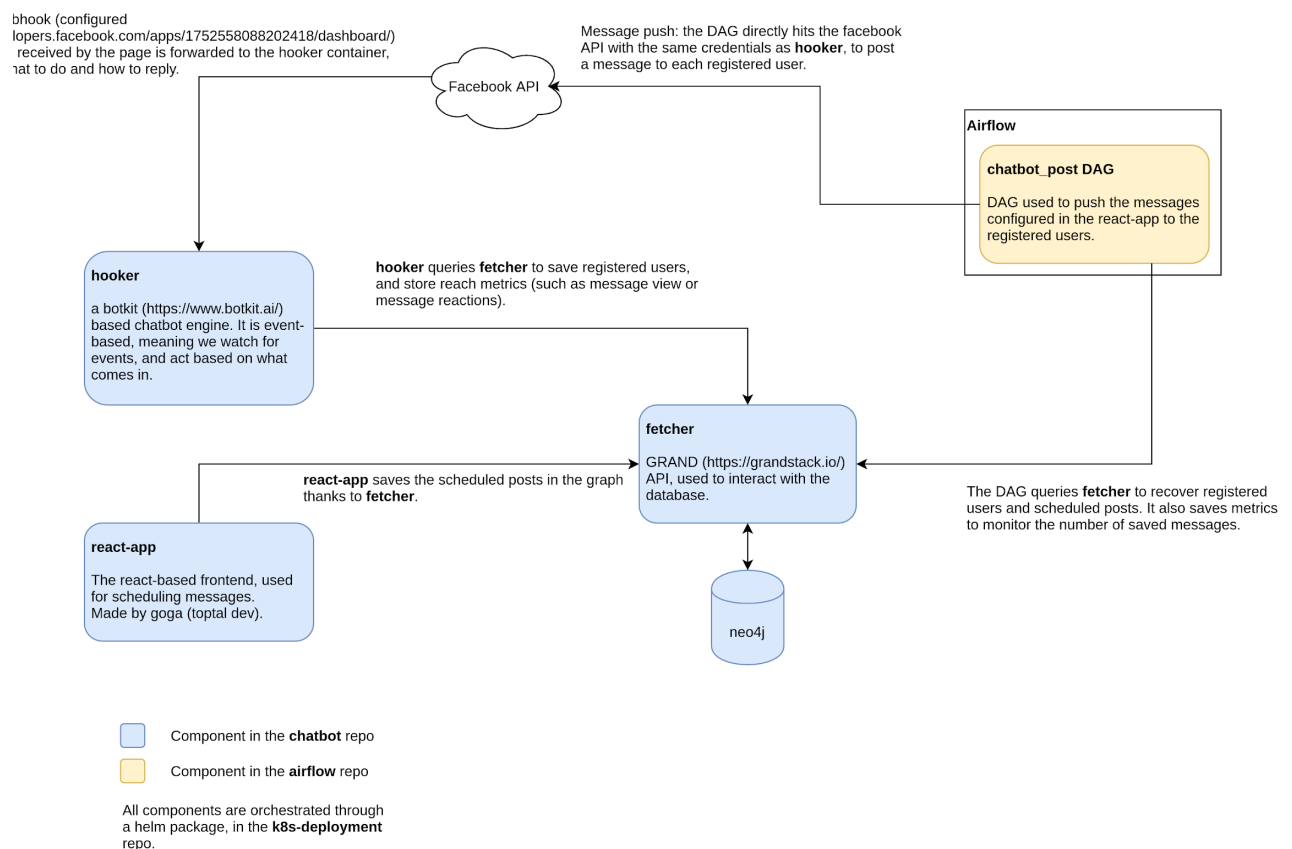
[9] DAG stands for Directed Acyclic Graph. The tasks are organized in a graph, where edges between the tasks represent dependencies. In other words, DAG represents the sequence of the execution of individual, smaller tasks.

Figure 4: 4u2 Messenger architecture.

Fetcher is the interface with the graph database. The database schema is depicted in Fig. 5:

- User: represents a user and stores their public Facebook User ID. We also document whether the user is registered to the message notification. This will be used by the DAG to know who to message. Please note that we do not store any "personal" information: we have no idea what the name, age or gender of the person we are talking to is;
- UserMessage: even though the bot cannot answer "any" message from the user, and gives a generic answer, we still store what the users say. This will be useful in the future to ask qualitative feedback about the chatbot, and the content it pushes;
- Post: a message we schedule in the web interface. It holds all the information needed to send a message on Facebook;
- PostMessage: object linking a Post to a User, it basically represents the actual single message sent to a User for a given Post. This is where we store the reach metrics, such as the seen events and the user reaction;
- NISVVideo: these are the only objects present in the graph after a "fresh" install. These represent all the videos available on NISV, properly parsed in both English and Dutch.
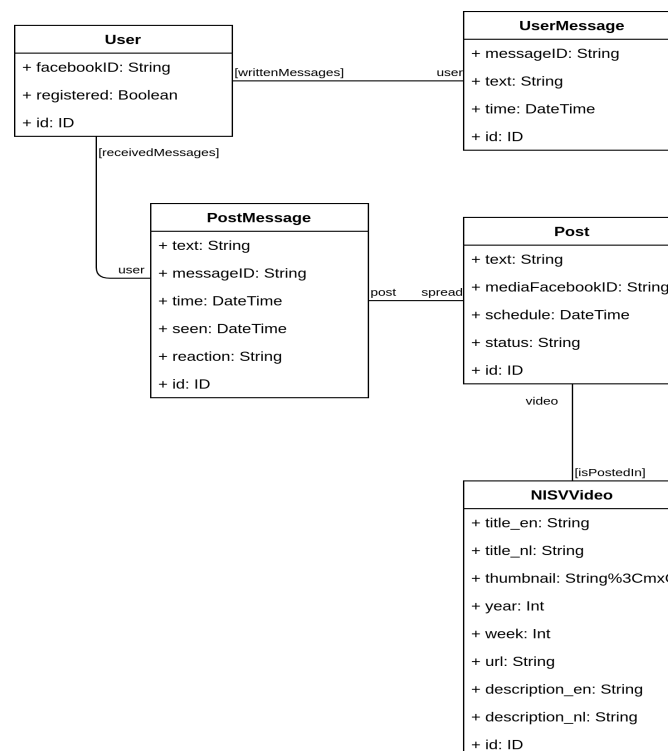

Figure 5: 4u2 Messenger database schema.

Hooker is the part responsible for the direct communication with the Facebook API. It is webhook-based, meaning the registered endpoint will be called by Facebook for each message (or event) the chatbot receives.

Facebook let us register for more than incoming messages: on this page, we can configure the webhook to be called in all the following cases:

- messages: a message was sent to the bot.
- messaging_postbacks: event fired in case of button interaction, for instance with the "get started" initial message.
- message_reads: an event fired when the user reads a message.
- message_reactions: someone reacted on a message.

In the first stage of the project, we built a frontend interface that allows editors to manage the users, monitor the messages that are being scheduled, and to manually create and schedule posts if necessary (see Fig. 6).
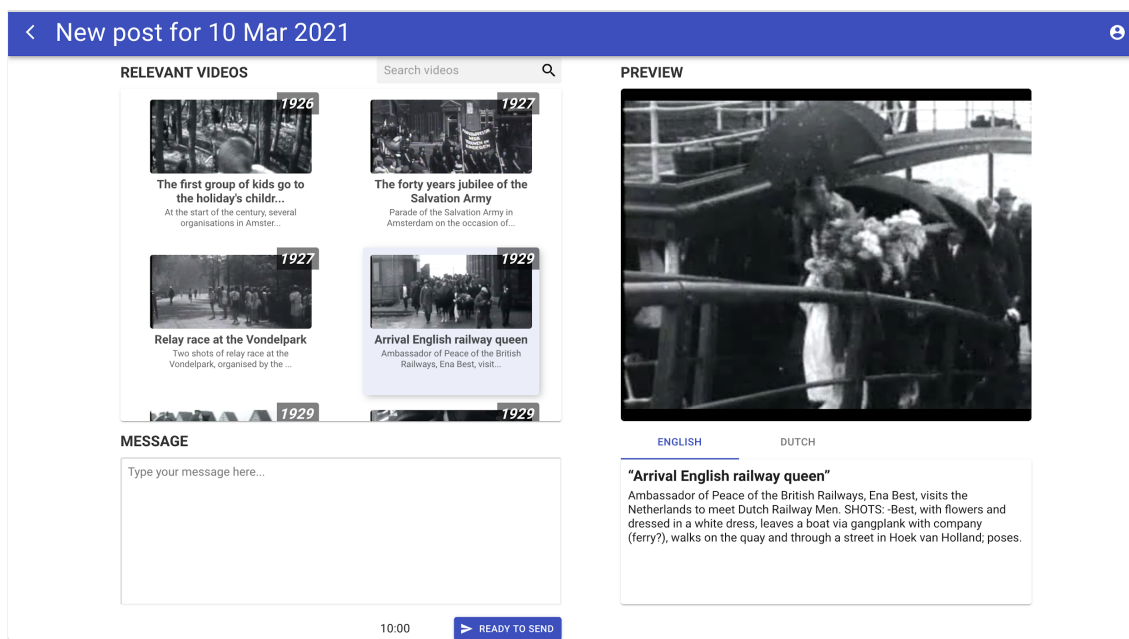
Figure 6: Manual creation of a post for the chatbot.

Message scheduling is managed by an Airflow scheduler and a daily running *create_personal_recommendations* task. The idea behind Airflow scheduler DAG is to actually send the messages that were returned by the recommendation algorithm (so called personalized messages) or manually configured in the admin interface (so called broadcast messages) to the registered users. PostMessage objects are stored in the database to keep track of them (opening times and reactions in particular).

The DAG runs every 5 minutes, with a scope window of 5 minutes. It means that a post can be, "worse case scenario" 5 minutes late compared to the time scheduled on the interface.

The tasks (steps) executed by DAG are the following:

- *load_posts_task*: recover the posts scheduled on the interface for the last 5 minutes.
- *exit_if_no_post_task*: exit if no post was identified (this will mark the rest of the tasks as "skipped").
- *upload_media_to_facebook_task*: as mentioned earlier, not all the information on a post is set from the beginning. One example being the mediaFacebookID, which is an

identifier of the video on Facebook. Indeed, any media sent on Facebook first needs to be saved on the platform itself (presumably for validation purposes, and to make sure it complies with Facebook guidelines). This task basically takes the URL from each Post object, and uploads it to Facebook, while finally saving the ID in the Post object itself.

- *send_posts_task*: We loop through the posts and registered users and individually send them a message.
- *set_status_to_sent_task*: finally, we mark the Post as "sent", so that it appears as such on the admin interface.

There is a different DAG, running daily that consists of just one task: create_personal_recommendations. This task triggers the recommendation algorithm and turns results into database entries (namely, PostMessage entries discussed earlier).

2.3 METHODOLOGY FOR EVALUATION

The evaluation goals for the 4u2 Messenger application were the following:

- Measure the **acceptance** of the 4u2 Messenger as a service for end user;
- Evaluate the **performance of the personalization engine**;
- Evaluate whether the application can **retain user engagement** over a prolonged period of time.

The testing with end users was planned to take place over a period of three weeks in February-March 2021.[10] NISV invited Dutch-speaking Facebook users to participate in the evaluation. The invitation was shared via NISV's social media channels and a paid advertisement was promoted on Facebook. Users who agreed to participate were invited to use the application for three weeks. It was decided intentionally not to offer any direct financial rewards to users participating in the process in order to more accurately monitor their engagement with the application or lack thereof. Intrinsic motivation (curiosity, interest in heritage, etc.) was the driving factor for users to participate. At the end of this testing period, participants were asked to fill out a short questionnaire, focusing on their acceptance of the service and the quality of recommendations (see Appendix A). The survey was conducted using the Typeform survey tool approved by NISV. It was selected as it is fully compliant with GDPR regulations (no personal data, including IP address, are collected) and it offers an easily customizable and user-friendly interface needed for evaluation with end-users.

The backend database of the 4u2 Messenger stores the following information about user activity:

- emoji reaction to each message (can be null if user did not react);
- timestamp of the message opening event (can be null if the user did not open the message).

Unfortunately, due to the limitation of data that is made available in the Facebook API, we cannot observe for how long a user has watched the video, or if they have watched it to the end. That also limits the options to evaluate user engagement, as well as the options to optimize the personalization algorithm itself.

---

[10] Originally, the testing was foreseen for a longer period of time, but due to the need to adapt the prototype to the new Facebook regulations described in Section 2.1, the release had to be delayed.

EVALUATION RESULTS

Quantitative Evaluation

The analysis is based on the behavioral data collected between the 3rd of March and 21st of March 2021 (19 days of data). During this period, we observed the behavior of 16 subscribed users. Due to the very short observation period, we decided to send as many as five personalized messages to each user per week. In the end, we scheduled and sent 198 messages (some users have registered during the test period, so they got less than 5 messages in some weeks).

Overall, users actively reacted to the messages. Their reactions were quite diversified and they were using all emojis that are available in the 4u2 Messenger (see the list below). In total, we collected 96 reactions (out of 198 messages, i.e. on average users reacted to every second message). The distribution of emoji reactions was the following:

- like - 37 reactions
- dislike - 20 reactions
- love - 17 reactions
- smile - 10 reactions
- wow - 8 reactions
- sad - 3 reactions
- angry -1 reactions

If we consider "like", "love", "smile" and "wow" reactions as positive, and "dislike" and "angry" reactions as negative, then in total we observed 72 positive reactions (77.4%) and 21 negative reactions (22.5%). We intentionally did not include "sad" reaction in any of the groups, as it may be interpreted differently, depending on the context (for example, a user may be engaged with the video, but the video content depicts sad events which call for a sad reaction). However, it should be noted that the algorithm learns the emoji interpretation automatically from the data (by adjusting positive/negative weights, see Section 3.2.2 of D3.3 for the detailed discussion).

Going deeper into user reactions over the testing period, we can observe the proportion of positive and negative reactions in the three weeks of the test period. It is worth stressing that during the first two weeks most of the users were served with the personalized messages selected by the cold-start mode of the algorithm (see Section 3.2.1 of D3.3 for the detailed description). Cold-start mode goal was to diversify the topical spread of the videos, in order to quickly learn user profile/interests. Only then, in the last week, the algorithm switched into the full-fledged personalization mode, based on the learnt profile. For some users that subscribed later in the testing period, they were still in the cold-start mode in the last week, so the numbers can be slightly biased and should be observed for a much longer period. Still, overall proportion of the positive reactions grew over time (from 77% in the first two weeks to 81% in the last week), so we may cautiously conclude that the switch from the more randomized cold-start mode to the targeted personalization mode increased user satisfaction.

Last but not least, it should be stressed that the behavior of individual users varied significantly and can also bias the conclusions based on such a small sample and limited observation period. If we look at the reactions of each user, it is visible that negative reactions come mostly for just two users (out of 16) - 12 out of 21 negative reactions. Only one of the 16 users had a prevalence of negative reactions over positive ones (7 negative reactions, 5 positive reactions). Finally, some users (including the two users mentioned above) reacted negatively even without

playing the video, just by seeing a textual description. In such cases, a qualitative user questionnaire is a better way to interpret and investigate the motivation behind some of the reactions. From a statistical point of view, we can conclude that such a behavior is rare, so the majority of the test population seemed to be interested in the personalized content.

Finally, one of the aspects of personalization implemented in WP3 was the data-driven optimization of the scheduling time. We observed the time difference between the scheduled time and the time when the user has actually opened the message. We assumed that the shorter this time difference is, the more accurate is the scheduling. It should be noted that the algorithm (see Section 3.2 of D3.3 for details) learns from the user behavior (i.e. collected past data on when they typically open messages) and adapts the scheduling accordingly. So, the more time (and data) we provide, the better the end result should be. Still, we can already observe that the median difference between the scheduling time and the opening time is 1 hour 50 minutes, and 90% of the differences were below 6 hours. Over the testing period, the mean difference decreased from 4 hours in the first two weeks to 1 hour 55 minutes in the last week (median difference from 1 hour 54 minutes to 1 hour 24 minutes, and lastly - the maximum difference to 6 hours 16 minutes).

User testing evaluation

At the end of the testing period, users were asked to answer a survey with eight questions focusing on their acceptance of the service and the quality of recommendations. Out of the 16 testers who signed up, 11 filled out the survey.

First, users were asked to rate their overall experience with the 4u2 Messenger on a 1-10 scale (see Fig. 7). 91% of the testers indicated they had a positive experience with a rating of 6 and higher. Most users rated their experience with a 7. The highest rating was 8 with 37%, the lowest rating was 2 with only 9%. The different approach taken to develop the 4u2 Messenger application, as explained in the previous sections, did not hinder an overall satisfactory user participation for most testers.
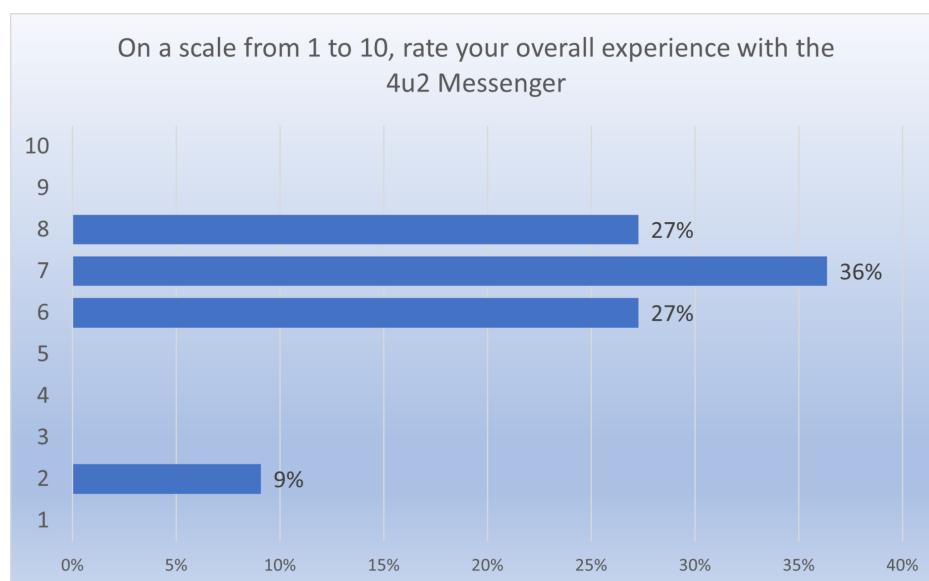

Figure 7: The rating of the overall experience with the 4u2 Messenger, on a scale from 1-10.

Regarding the quality of the recommendation engine, 73% of the users responded positively when asked if they liked the videos curated for them (see Fig. 8). When asked if the videos matched their interests the responses were more negative with 27% disagreeing or strongly

disagreeing with the statement. Nonetheless, 45% of the participants found that the videos did match their interests. One explanation can be found in the Quantitative Evaluation section on the behavioral data; out of the three weeks of testing, only in the last week the algorithm switched into the full-fledged personalization mode after the cold-start status. These survey results might have differed if the testing period lasted for a longer period. Nevertheless, 90% of the testers responded positively, of which 45% very positively, on the statement whether they enjoyed receiving videos on a wide variety of topics. Users liked the serendipity of the recommendation engine providing unexpected videos, even if these videos did not necessarily match their personal interests.
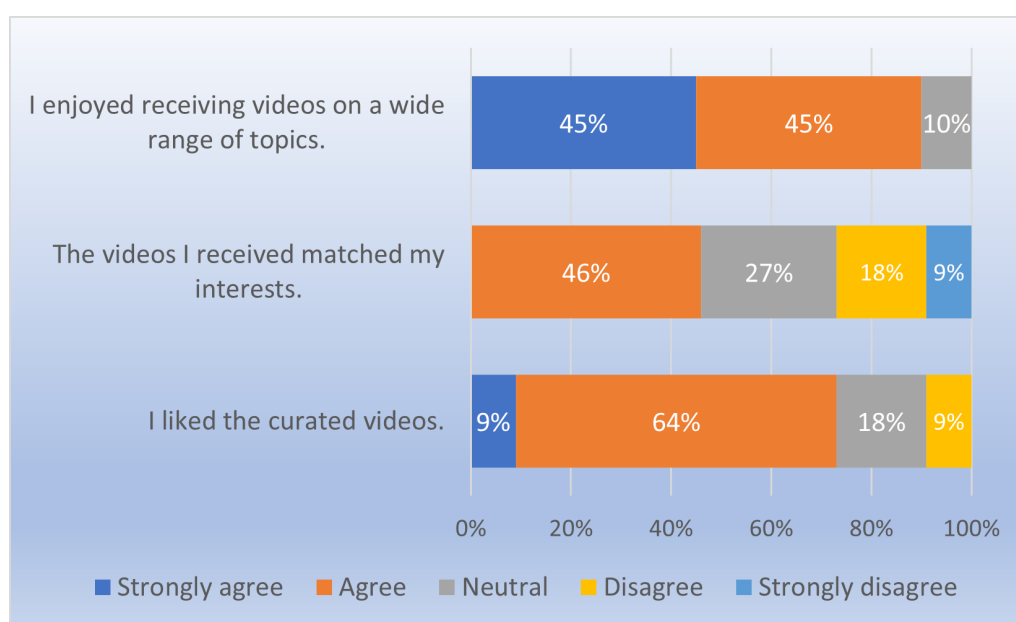
Figure 8: Statements about the selected videos the participants received.

We presented two statements concerned with the timing at which the participants received the videos (see Fig. 9). With 64%, most users responded neutral to the statement "I thought most of the messages were sent out at a time that suited me". 36% of the users agreed or strongly agreed with the statement, there was no disagreement. When asked whether users would prefer to receive messages at the same time each day, the responses were diverse. 36% objected, or strongly objected to receiving messages at the same time each day, while 55% agreed or strongly agreed with the statement. Comparing the outcomes of these two statements can suggest that the data-driven optimization of the scheduling time, described in the Quantitative Evaluation section, became more precise during the testing period matching the time when the user has actually interacted with the 4u2 Messenger, which can be stable every day or more varied from time to time. This is substantiated with the observation that while some testers objected to receiving messages each day at the same time, there was no disagreement when asked if the actual time the messages were sent suited them.
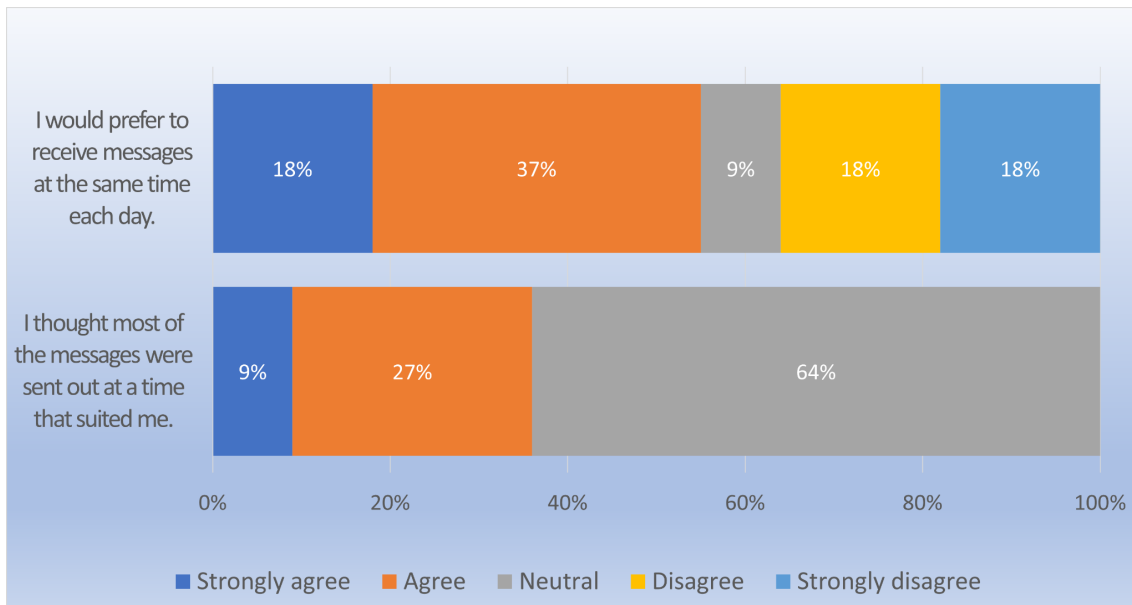
Figure 9: Statements about the timing that the participants received the videos.

A multiple choice question asked users if they would like to keep using the application in the future (see Fig. 10). 27% of the respondents answered Yes, and 64% indicated that they want to keep using the 4u2 Messenger in the future but after some adjustments to the application. Only one user would prefer not to use the application anymore. The open question that concluded the survey can point to possible suggestions for alterations that would convince the participants to keep using the application.
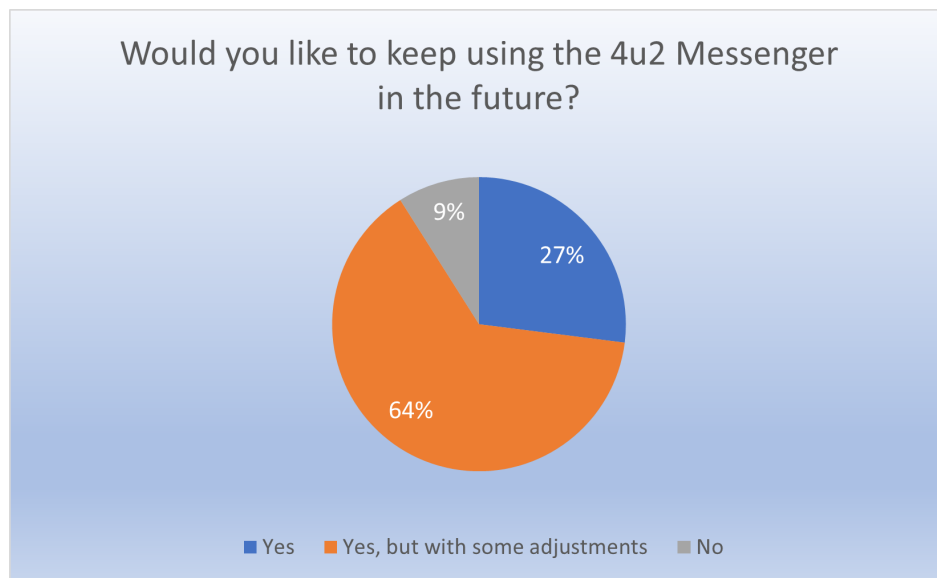
Figure 10: Would participants like to keep using the 4u2 Messenger in the future?

We ended the questionnaire with an open question, asking whether there were additional comments or suggestions for improvements. Interestingly, most of the comments and suggestions in the open questions could either be related back to the constraints by Facebook to develop the application that was intended, or the initial requirements of the 4u2 Messenger

described in the Development Process for Facebook Messenger of Section 2.1, that were not or could not be implemented at this stage for several reasons:

- It was suggested that too many videos were sent during the testing period. Less videos per week could enhance the experience since this way it remains special if you receive a video (one user called it a "welcome interruption"). Due to the circumstances we could only test the application for three weeks, which led to as many as five personalized messages per week to collect the behavioral data. For the future use of the 4u2 Messenger, it can be considered to lower the amount of videos users receive per week to retain user engagement over a longer period of time;
- It was noted that the application did not always meet the high expectations of testers. Respondents mentioned that more interaction with the 4u2 Messenger was expected and missing. Also, the benefit of the functionalities of the Messenger compared to the regular videos that NISV publishes on their own Facebook account were not clear to all users. For the future use of the 4u2 Messengers these kinds of concerns can be addressed by performing more qualitative evaluations with users. Also revisiting the initial requirements that address more interaction with the 4u2 Messenger, like providing feedback on interests, or engaging with the application via buttons or text could be an option;
- Respondents also gave feedback on the personalization engine and the use of emojis to react to videos; users did not always understand whether their responses with certain emojis eventually led to video content on different topics. Since out of the three weeks of testing, only in the last week the algorithm switched into the full-fledged personalization mode, this would have been difficult for the users to notice and a longer testing period could give users more insight on this feature. It was also mentioned that the meaning of the emojis was not always clear and that different types of categories instead of reacting with emojis could lead users intuitively to more personalized content as well, like a button with "want more like this or less like it". For the future use of the 4u2 Messenger, the initial requirements could be revisited that address the engagement with the application via buttons with specific text or by entering free text.

3   4u2 Smart Speaker Skill

In late 2019, as a result of changes in WhatsApp terms of use (see Section 2.1), RBB decided to cease all current and future planned messaging activities aimed at contacting consumers as a company, e.g. by sending text messages, audios or videos, by the end of the year. The main reason for this decision was that the user base of RBB's existing messenger offerings at the time was predominantly made up of WhatsApp users.

However, since WhatsApp was no longer an option for the 4u2 Messenger scenario as described in Section 2.1, RBB decided, after consultation with the other ReTV partners, not to continue its work on it and instead to develop a completely new scenario together with the consortium. This should be an approach in which, on the one hand, the knowledge gained and developments made in Tasks T6.1 and T6.2 could be incorporated and, on the other hand, with which the requirements of T6.3 could also be realized. In addition, the new scenario had to be compatible with the core objective of WP6 "to develop prototypical applications to evaluate consumer benefits from TVP data-driven media content placement and distribution" and had to have a clear future potential.

All this applies to the 4u2 Smart Speaker Skill. It is an application ("Skill") for Amazon Smart Speakers with displays that **enables users to create their own videos from existing archive material using voice commands**. This scenario was designed, developed and evaluated by RBB together with the project partners in the third year of ReTV. The process is described and explained in detail in the following sub-sections.

3.1 SCENARIO DEVELOPMENT

Already in mid-2019, ReTV partners recognized Smart Speakers as a great additional opportunity for content owners, e.g. broadcasters and media archives, to offer consumers video content tailored to their needs. The TVP to be developed within the project, in particular its Recommendation and Scheduling as well as its Video Adaptation and Repurposing components, was the ideal prerequisite for this. Current statistics also confirm that it is worthwhile for content owners to deal with smart speakers as a further distribution channel. According to these, their worldwide sales amount to 30 million units per quarter[11]. In Germany, every fourth person owns at least one smart speaker[12] and 32% of Germans who are active online intend to buy one[13]. Among the latter, smart speaker usage has increased by 61% between 2018 and 2020[14]. Appropriately, RBB was able to quickly contribute a suitable scenario: enabling consumers to create their own episodes of "Unser Sandmännchen" using a smart speaker application.

"Unser Sandmännchen" is a long established, very popular animated children's programme from RBB[15]. Targeted at pre-school children, "Unser Sandmännchen", a seven-minute show broadcasted daily at 18:00, accompanies children to bed with a bedtime story called "Abendgruß".

Sandmännchen episodes are simply structured: there is always a framing story, consisting of an intro and an outro (the Sandmännchen arrives and leaves a selected setting in a specified way), which surrounds a main story, usually an adventure of one or more of his friends. There is a continuously growing amount of archival material from which these structural elements can be drawn.

Currently, Sandmännchen episodes are created primarily for TV broadcasting. This means that the selection of production elements (i.e., which framing story to use, which main story elements to combine with which other) depends strongly on the requirements of RBB's daily (live) TV programme. Live broadcasting requires that Sandmännchen elements are limited by available broadcast time, and RBB's on-demand platforms, such as the Mediathek[16], host only these TV versions of the programme.

Here is where the application for smart speakers comes into play. The core idea is the following: there are unrestricted possibilities for the creation of personalised Sandmännchen episodes by combining your favorite framing and main story using voice commands. So these can be created by the application according to user preference, and are no longer defined by the requirements of live TV broadcast. Users can create their own Sandmännchen episode completely free from editorial restrictions, with their choices being driven only by their

---

[11] https://de.statista.com/themen/4662/smart-speakers/
[12] https://www.infosat.de/digitale-welt/studie-jeder-vierte-deutsche-hat-einen-smart-speaker
[13] https://rms.de/audio-und-radiowerbung/studien/smart_speaker_studien/smart_speaker
[14] https://rms.de/audio-und-radiowerbung/studien/smart_speaker_studien/smart_speaker
[15] https://www.sandmann.de/
[16] https://www.ardmediathek.de/rbb/

personal preferences. To achieve that goal, the application had to be supported by functions for video adaptation and re-purposing developed within ReTV that are based on Artificial Intelligence (AI) techniques, most notably deep neural networks.

Following this idea, RBB, CERTH and GENISTAT developed a first prototype and called it "Abendgruß"[17]. This was an application ("Action") for the Google Assistant, designed primarily for the Nest Hub, Google's smart speaker with display. The app was equipped with a limited range of functions and video elements that could be combined with each other. There were no connections to the TVP, all potential episodes were created manually.

The "Abendgruß" prototype was presented to the RBB departments "Familie & Kinder" (family & children), responsible for RBB's complete content for children including "Unser Sandmännchen", "Online Koordination" (online coordination), controlling the product development of all internet and app products of RBB as well as the voice applications, and "Programmbegleitende Dienste" (TV online), responsible for the whole online representation of the the brand "RBB Fernsehen" (RBB TV). In addition, the app was showcased to project stakeholders and a wider audience at various fairs such as the IFA[18]. The feedback was very positive throughout. In particular, the idea of enabling consumers to interact directly with the content of a broadcaster in order to personalize it was met with great approval. The smart speaker as the device of choice was considered reasonable and in line with market trends. The application's simple, child-friendly operation appealed to the "Abendgruß" target group, i.e., children aged four to six years or parents with children of that age. Moreover, the approach of using AI techniques to realize the concept generated great interest and was considered highly innovative. In addition, the RBB departments "Familie & Kinder", "Online Koordination" and "Programmbegleitende Dienste" confirmed that the idea of the "Abendgruß" for smart speakers fits perfectly into RBB's plans to 1) open up new distribution channels and 2) expand its digital offer for the Sandmännchen.

Encouraged by this feedback, RBB, in close cooperation with the above-mentioned RBB departments and GENISTAT, began working on a detailed content concept for the final prototype in the first quarter of 2020. This formed the basis for the technical implementation, for which GENISTAT and CERTH were responsible. The focus lay on the following points:

- the identification of the relevant TVP components
- the specific development of the relevant TVP components in the sense of the scenario
- the integration of the relevant TVP components into the existing "Abendgruß" application

The aim was to be able to present the focus group with a stable "Abendgruß" prototype for the longitudinal tests at the end of the project, which, thanks to the TVP components, could be used to reliably create personalized Sandmännchen episodes.

To validate the idea and the concept of the "Abendgruß", RBB conducted a survey at the beginning of the development process among parents with children in preschool age, i.e. the intended target group. The survey ran from 22nd June to 6th July 2020 in using an online questionnaire. The tool LamaPoll was used for this purpose. A total of 49 participants completed the questionnaire.

17

https://retv-project.eu/2020/11/12/ok-google-mit-abendgrus-sprechen-our-smart-speaker-application-for-children/

[18] https://b2b.ifa-berlin.com/en/

The survey focused on three main areas:

- children living in the household and their use of technical devices
- familiarity with "Unser Sandmännchen" and its media channels
- experiences with voice-controlled applications for children and wishes for a corresponding Sandmännchen application

The core results were the following:

- Only 14% of children were allowed to use smart speaker applications alone. They were usually assisted by an adult.
- 82% of the participants knew "Unser Sandmännchen".
- 70% of the participants watched the episodes of "Unser Sandmännchen" on demand via the official Sandmännchen channels, e.g. in the ARD Mediathek (22%), the smartphone app (18%) and on YouTube (12%).
- 45% of participants watched the episodes for entertainment purposes, more than for going to sleep (38%), but 70% watched the episodes nevertheless in the evening (17:00 to 21:00).
- 33% of the participants who knew "Unser Sandmännchen" and 100% of those who already used voice-controlled applications for children would like to see a Sandmännchen application for a smart speaker.
- Regarding potential functions of a voice-controlled Sandmännchen application for children, 50% of the participants would like to be able to watch the current TV episode. 46% would like to choose from different main stories of "Unser Sandmännchen" and 31% would like to put together and play an episode of "Unser Sandmännchen" in a playful way (e.g. choose between two main stories).

Thus, ReTV partners regarded the "Abendgruß" scenario as validated and RBB, GENISTAT and CERTH proceeded to develop its final prototype.

3.2 FINAL PROTOTYPE

Switch from Google Action to Amazon Skill

In contrast to the first prototype, the final "Abendgruß" prototype was developed for smart speakers with displays from Amazon. There were two reasons for this:

- Amazon is the leading company in smart speaker sales, with 15.8 million units sold per quarter worldwide (in Germany, the proportion of smart speaker owners who own an Amazon Echo is 81%).[19] In addition, statistics on the use of already published rbb/ARD smart speaker applications revealed that this is predominantly done via Amazon devices or the Amazon voice service Alexa integrated into other devices. Therefore, the ReTV partners, also in consultation with the online coordination of RBB, saw a significantly greater potential in the development of an Amazon skill with regard to the market introduction and establishment of the "Abendgruß".

- The core of the developed "Abendgruß" concept, i.e. the personalized creation of Sandmännchen episodes and their direct playback on the Smart Speaker, can currently only be realized via Amazon Skills. In Google Actions, only YouTube videos can be integrated. Google confirmed this in response to a RBB enquiry. The ReTV partners were confronted with this fact during the development of the first prototype, which

---

[19] https://de.statista.com/themen/4662/smart-speakers/

meant that a workaround had to be developed to present the functionality (scanning a QR code using an additional device in order to then watch the episode created). Amazon is much more open in this respect and allows the integration/playing of external video URLs in its skills. This option is essential for the technical workflow of the "Abendgruß" (see "Implementation").

Functionality

The "Abendgruß" skill includes two functions:

- showing the current Sandmännchen episode from the linear TV programme
- creating an own Sandmännchen episode

The former was the function most desired by the participants in the parent questionnaire (see Section 3.1) for a voice-controlled Sandmännchen application for children. Its integration therefore represents an ideal control mechanism in terms of both user acceptance and technical functionality with regard to the second function, i.e. the main function of the skill from ReTV's point of view.

The functions are realized as shown in Table 2.

Showing the current Sandmännchen episode from the linear TV programme	Creating your own Sandmännchen episode
To start the "Abendgruß" skill, the user has to say "Alexa, hallo Abendgruß".	
The skill then opens and the start screen is displayed. At the same time, the skill welcomes the user by offering him or her two options for interaction: "Hallo, was möchtest Du Dir heute angucken? Die neueste Folge oder eine eigene Folge? (Hello, what would you like to watch today? The newest episode or an episode of your own?)" The two options can also be read on the screen (see Fig. 11).	
The user answers: "Die neueste Folge. (The newest episode.)"	The user answers: "Eine eigene Folge. (My own episode.)"
	After saying that, the user is offered two sub-options in succession.
-	Firstly, the user can choose how the Sandmännchen should arrive, in other words, in this step the framing story is defined. The Skill asks the respective question ("Wie soll das Sandmännchen heute zu uns kommen?/How should the Sandmännchen arrive today?") and offers two possibilities, e.g., "Mit dem Pferd? (By horse?)" and "Mit dem Fahrrad? (By bike?)". In addition, a third variant is available to the user: "Oder mit etwas anderem? (Or with something else?)". If he or she selects this, he or she is offered two new arrival types

	from which to choose. The two arrival types available for choice are also always shown on the display by means of thumbnails and a descriptive sub-headline (see Fig. 12).
-	Secondly, the user can determine the main story by answering the question "Und wen möchtest Du heute sehen? (And who would you like to see today?)". Again, two variants are presented, e.g., "Rita und und das Krokodil oder Jan und Henry? (Rita and the crocodile or Jan and Henry?)", completed by a third to change the offer ("Oder jemand anderes?/Or someone else?") (see Fig. 13).
The skill shows the current episode from the linear TV programme.	After choosing the desired character, the created episode is shown.
Afterwards, the skill says goodbye to the user with "Gute Nacht, bis zum nächsten Mal. (Good night, see you next time.)" At the same time, "Gute Nacht" (good night) is shown on the screen.	
Following this, the skill closes itself.	

Table 2: The functionality of the "Abendgruß" skill.

Figure 11: The start screen of the "Abendgruß" skill.

Figure 12: Example for the sub-option "Wie soll das Sandmännchen heute zu uns kommen? (How should the Sandmännchen arrive today?)" of the "Abendgruß" skill.

Figure 13: Example for the sub-option "Und wen möchtest Du heute sehen? (And who would you like to see today?)" of the "Abendgruß" skill.

Technical Implementation
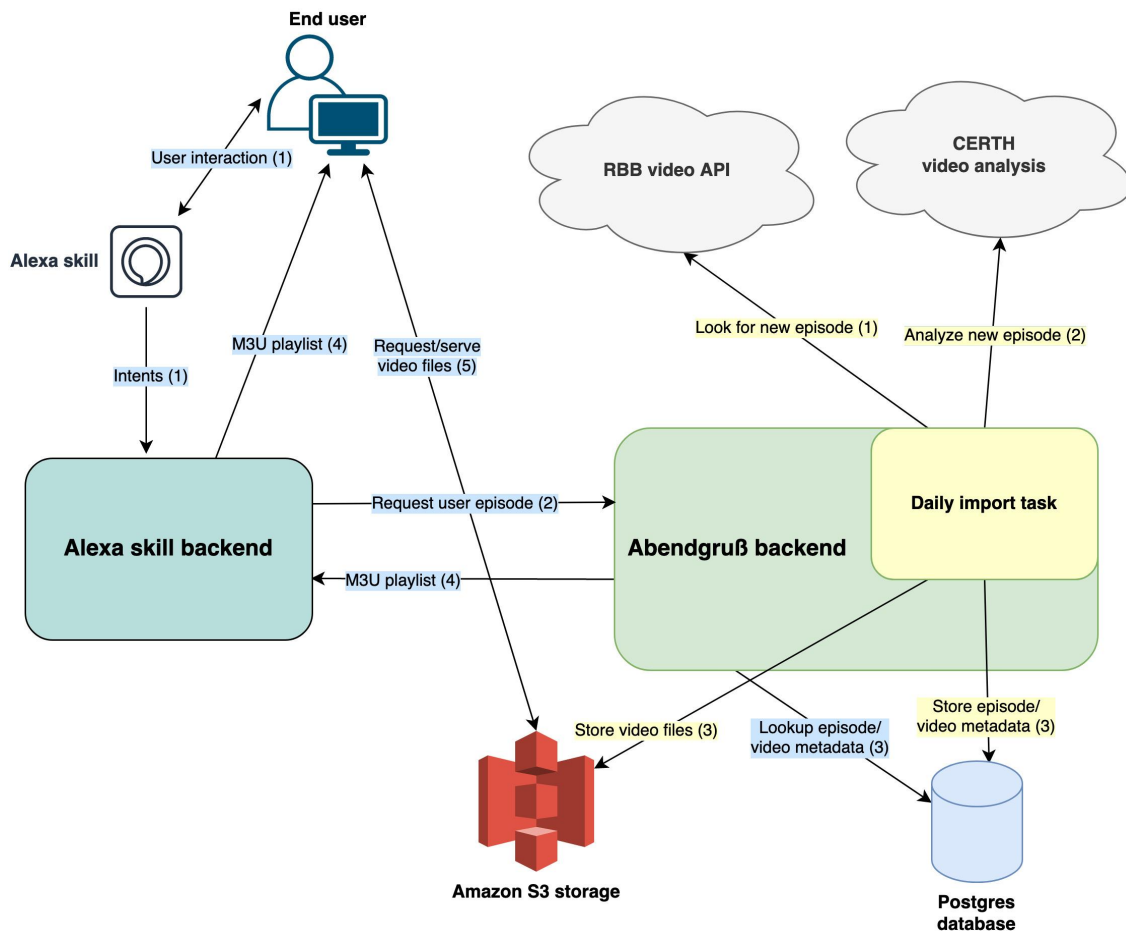
Figure 14: "Abendgruß" system architecture.

Workflow - Adding new Episodes

Every day, the process checks for new Sandmännchen episodes in the RBB video API (1) (see Fig. 14). If a new episode exists, the episode is downloaded and a request to the CERTH API is sent to analyze the new episode (2). As soon as the CERTH analysis is available, the video is cut into an intro, main, and outro part. The timings of the video cuts are taken from the CERTH video analysis. Also, the information about the Sandmännchen character is extracted from the CERTH video analysis and stored in the GENISTAT backend.

Next, the three video parts are converted into so-called fragmented MP4 video files. This is necessary, so the video parts can later be independently used in a video stream. After the video files have been converted into fragmented MP4 files, they are stored in Amazon's object storage (3). The metadata that is needed for creating m3u video playlists later is saved in a Postgres database (3).

New episodes can also be added manually through the admin web interface of the "Abendgruß" backend. In this case, the video files need to be provided manually.

Workflow - Requesting a custom Episode

When a user interacts with his Amazon Alexa device, he/she indirectly interacts with the "Abendgruß" Alexa skill. The Alexa skill on Amazon communicates with GENISTAT's Alexa skill backend, which provides intro and character options that the user can choose from (1). As soon as a user has selected an intro option (e.g. "Mit dem Fahrrad.") and a character (e.g. "Rita und das Krokodil"), GENISTAT's Alexa skill backend sends a request to the "Abendgruß" backend (2). The request simply contains the identifier of the episode that belongs to the intro option and the identifier of the episode that belongs to the selected character. Based on the selected episode identifiers, the GENISTAT backend looks up the metadata of the intro, main and outro video and creates a M3U playlist file (3). The resulting M3U playlist file consists of video chunks pointing to three different video files. The M3U playlist file is returned to the Alexa backend (4), which forwards it to the Amazon Alexa device where it is directly interpreted and played (5).

Administration - Alexa Skill

The interaction model of the "Abendgruß" skill, which refers to an "app" on Amazon Alexa devices, is kept very simple. There are four user intents defined: play the latest episode, create a custom episode, select an intro option, and select a character. For each intent, a list of utterances, which trigger the intent, are defined. Utterances like "newest episode", "latest episode", or "play latest episode" will trigger the intent that plays the latest episode. Utterances like "own episode" or "custom episode" will trigger the custom episode intent, which will then ask for an intro option and a character.

Both, the intro and character options, are defined as so-called slot types in the skill. A slot type is a list of utterances. Each utterance has to have at least one value, e.g. the name of the character, but can have additional synonyms, i.e. other ways a user might express the canonical value. The value is used to identify the framing story option or character in the backend. In case a new character or intro option is added to the skill, it has to be defined as a slot type value in the Alexa skill and as a character or intro option in the Abendgruß backend. Only this way, it can be identified across the two systems.

Administration - "Abendgruß" Backend

The "Abendgruß" backend has an integrated administration web interface that offers the following functionalities (see Fig. 15):

- monitor automatically added episodes
- link intro option and character to episode
- publish/unpublish episode
- choose thumbnail for intro option and character
- manually upload new episode
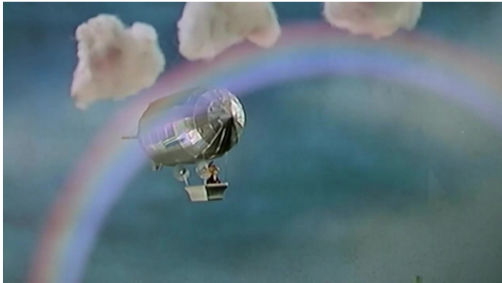
Figure 15: Example view of the administration web interface of the "Abendgruß" backend.

As new episodes are automatically added to the backend every day, it's important that the quality of the video cuts is monitored. This can be done through watching the video parts in the admin interface. Also, an intro option type needs to be assigned to a new episode before it can be published and used in custom episodes.

Furthermore, it's possible to change the thumbnail of the intro option or character that the user sees on the Alexa device and add a new episode manually by uploading the three required video files directly.

Supporting Features Extraction

The necessary features to provide the aforementioned functionalities are extracted by CERTH's Video Analysis Service. Specifically, the support signals to detect the structure of a Sandmännchen episode are discussed in Section 4.1.1 of D1.3, while the employed concept detection pool for the main character identification is detailed in Section 4.2.4 of D1.3. We

presented a publication[20] and demo[21] of how these features are extracted and utilized at the DataTV 2020 workshop (Zwicklbauer et al. 2020).

3.3 METHODOLOGY FOR EVALUATION

Once the prototype was successfully implemented we wanted to validate its use by families in realistic, day-to-day use. To achieve this we developed a test methodology, concentrating on usability of the service, over an extended time period by members of the target group in their home environment.

The target group was defined as German speaking families with preschool children aged 4 to 6 years. Ideally they knew the Sandmännchen brand, were technically savvy but were not required to own a smart speaker. We decided to provide the participating families with smart speakers (Amazon Echo 8 model) to ensure consistency of use and allow comparability of experience across respondents. This also opened up a potentially larger pool of testers and provided motivation to participate. In total 13 families were recruited to take part in the test. Additionally, four RBB colleagues participated with their children.

We wanted to investigate how families used the skill on a day-to-day level on how this usage progressed or changed over time. To achieve this we developed a range of questionnaires, an initial long questionnaire at the start of the test, followed by shorter weekly questionnaires and at the end of the test period the families were asked to complete a final questionnaire:

- **The initial questionnaire** concentrated on questions related to media usage and experience with voice controlled applications for children (see Appendix B). The answers provided formed the basis for questions in the final questionnaire.
- **The weekly questionnaires** were designed to be completed in five minutes and concentrated on questions directly related to the usage of the skill. These concerned frequency of use, time of day, understandability, general feedback and development over time (see Appendix B).
- **The final questionnaire** was about evaluating the skill (see Appendix B).

For all questionnaires the online survey tool LamaPoll was used as it complies to data protection regulations in Germany. The usage of LamaPoll and the test design and methodology was reviewed and approved by RBB data protection officer Anke Naujock-Simon[22] . The timeframe for the test was planned for February to March 2021. The actual testing took place from 15th February to 17th March.

3.4 EVALUATION RESULTS

Initial Situation

Two children live in the households of the majority of our test families (85%). In addition, there was one family with one child and one family with three children. With regard to the age of the children, the desired target group was also represented in all households: age 4 (38%), age 5 (38%) and age 6 (31%).

---

[20] https://dl.acm.org/doi/10.1145/3422839.3423061
[21] https://memad.eu/wp-content/uploads/AI4TV2020-4-Apostolidis-Sandmannchen.pdf
[22] https://www.rbb-online.de/datenschutz/datenschutzerklaerung.html

The No. 1 device in the households of the test families is the smartphone, all of them own (at least) one. 92% each stated the television and the tablet. **69% own a smart speaker with a display**, 38% a smart speaker without a display. 62% own streaming devices (see Fig. 16).

Which of the following devices are in your household?

Device	Percentage
smartphone	100,00%
TV	92,31%
tablet	92,31%
smart speaker with display (e.g. Echo Show, Nest Hub)	69,23%
streaming devices (e.g. Fire TV Stick, Apple TV, Chromecast)	61,54%
smart speaker without display (e.g. Echo Dot, Home Mini,…)	
none of the above	
other	
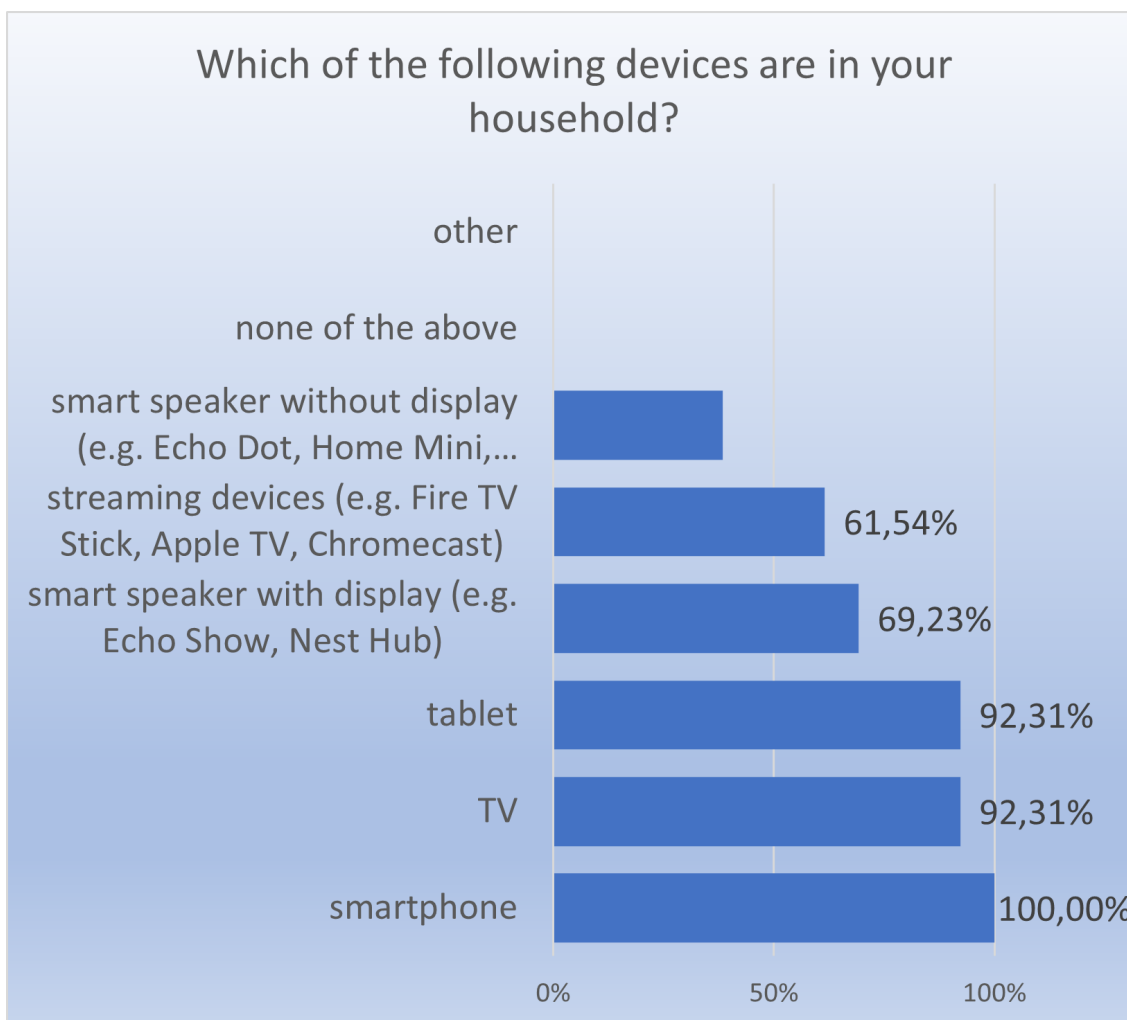
Figure 16: Devices used per household.

However, the device that children are most allowed to use independently in the household is the tablet (67%). In contrast, the smartphone may only be used by children alone in just under a third of families. **In households with smart speakers with displays, children can use them on their own in 56%.**

Regarding the Sandmännchen, there was a high level of familiarity among the test families. 90% of them know the Sandmännchen and 75% watch its episodes. The main channels used for that are: the ARD Mediathek[23] (50%), the linear TV broadcast (40%) and the Sandmännchen app for tablets (30%). Since there is not much difference between non-linear (Mediathek) and linear (TV) use, the preference for a fixed reception time is likely. Consequently, it is not surprising that the Sandmännchen episodes are watched exclusively in the evenings between 5:01 to 9 pm (see Fig. 17).

---

[23] https://www.ardmediathek.de/ard/
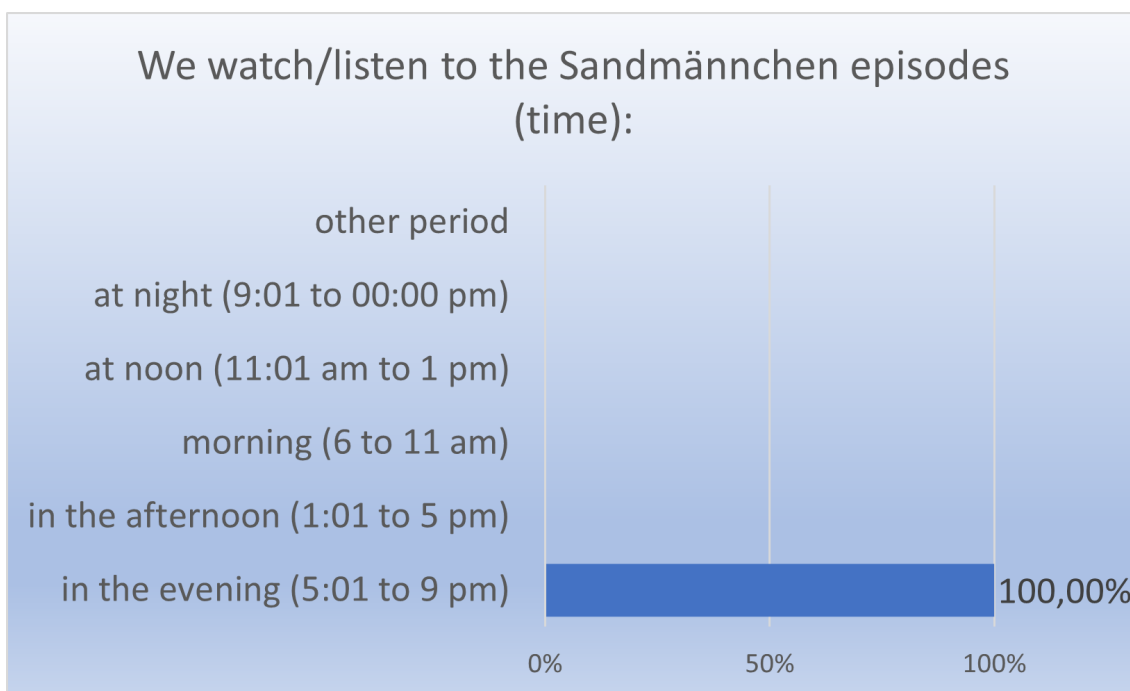
Figure 17: Preferred time for watching or listening to the Sandmännchen episodes.

The purpose for watching them is mainly "for entertainment" (70%), followed by "to fall asleep" (50%). In 80% of cases, **one Sandmännchen episode is watched at a time**.

Finally, **69%** of the test families said that they had **not yet used voice-controlled applications** (see Fig. 18).
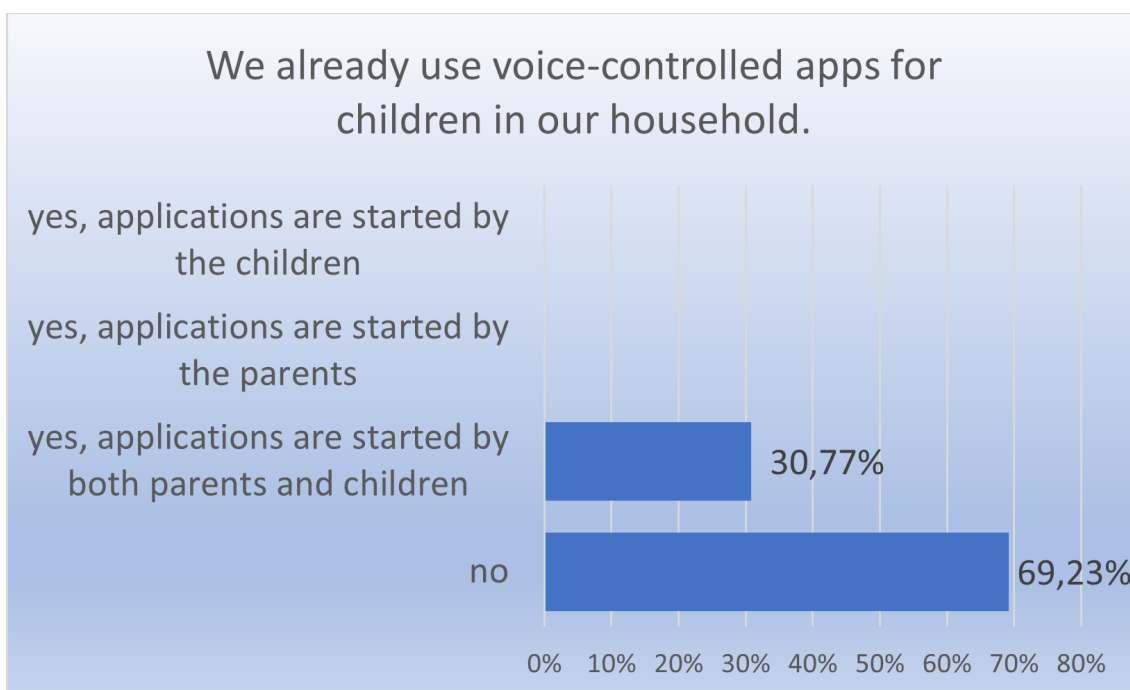
Figure 18: Statements on whether voice-controlled applications are used and if so, by whom.

In summary, it can be said that our test families completely corresponded to the target group we defined for the testings. The fact that the majority of them had no previous experiences

with voice-controlled applications for children provided an ideal opportunity to validate the implementation of the "Abendgruß" skill, particularly with regard to acceptance and usability.

Usage and Feedback during the Testing Phase

Over the weeks, the "Abendgruß" skill was constantly used between 3 to 4 times and 7 times or more. Over time, there was a clear tendency to use it three to four times a week (see Fig. 19).
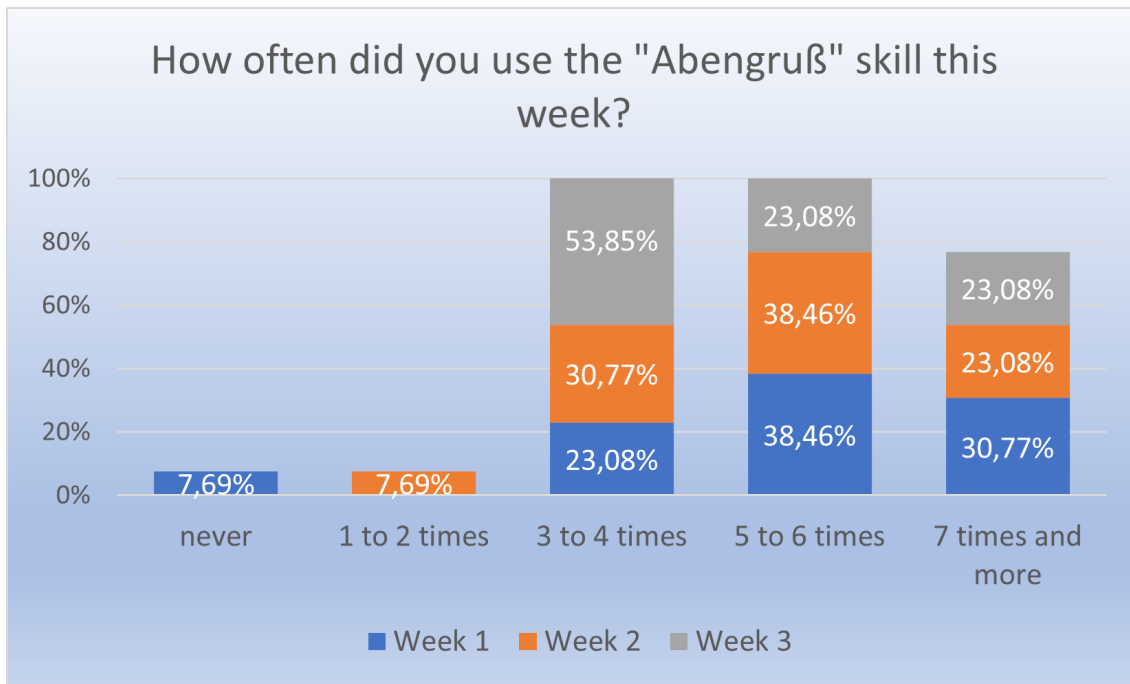
Figure 19: Frequency of use of the "Abendgruß" skill during the weeks.

All test families used the "Abendgruß" skill consistently in the evening between 5:01 pm and 9 pm, so completely in keeping with their previous habits for watching the Sandmännchen episodes (see previous paragraph).

The main users of the "Abendgruß" skill were children under the age of 6. In the first test week, 75% of the users were parents/adults. In the second and third test week, children up to 6 years of age predominated with 85% each, with the number of parents/adults using the skill increasing again in the third test week (see Fig. 20).
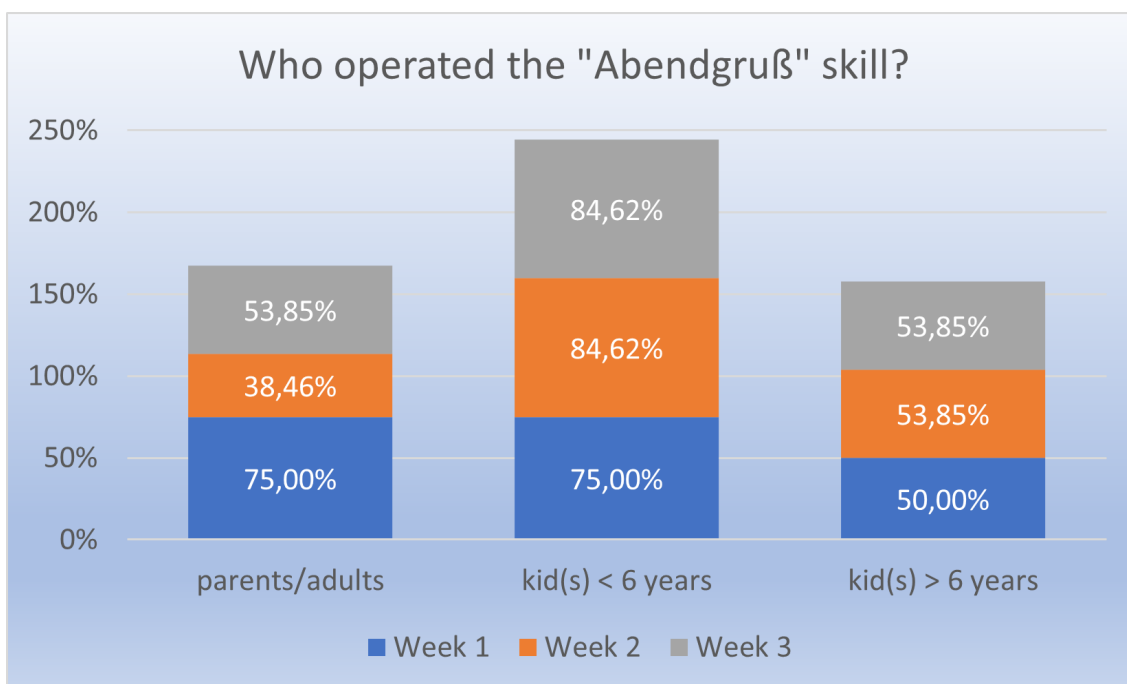
Figure 20: Details of who operated the "Abendgruß".

Of the two options "Die neueste Folge. (The newest episode.)" and "Eine eigene Folge. (My own episode.)" offered, the latter, i.e. the main function from ReTV's point of view, was clearly the most preferred. While the vast majority of the test families used "Die neueste Folge. (The newest episode.)" never or 1 to 2 times a week, "Eine eigene Folge. (My own episode.)" was much more popular. All test families mostly created their own episode at least seven times or more. In the first week, all of them created their own episode, while in the second week 15% and in the third week 8% never did it (see Fig. 21).
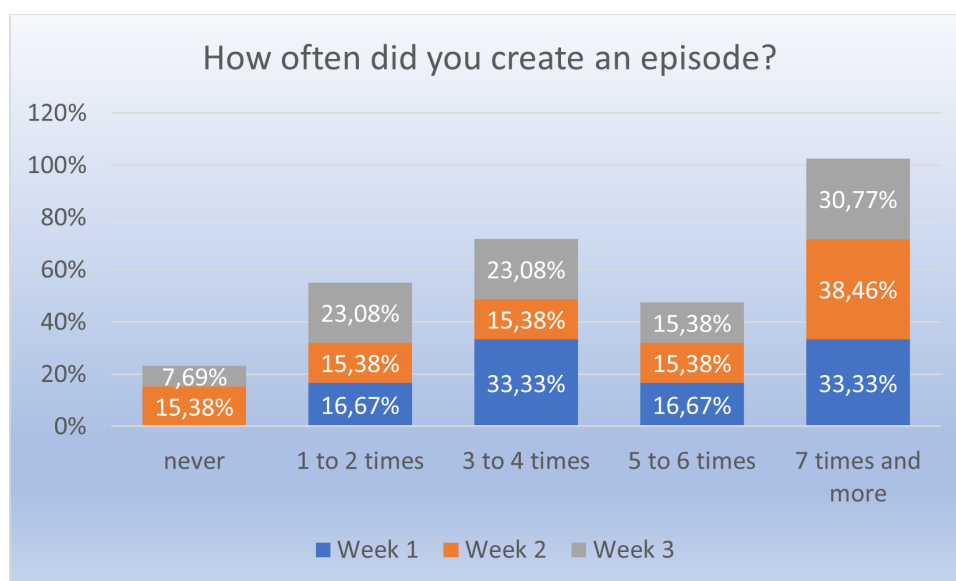
Figure 21: Frequency of using "Die neueste Folge (The newest episode.)" during the weeks.

Even though the function itself was popular, "Die neueste Folge. (The newest episode.)" obviously needs to become more user-friendly. This is indicated by the test families' answers to the question "How easy has it been for you creating an episode?". Most of them responded

that it was sometimes (very) easy and sometimes (very) hard. For weeks 1 (42%) and 3 (58%), this was the dominating statement. In these weeks, possible reasons for this are listed in Tables 3 and 4. However, it can be observed that the test families found it increasingly easy to create their own episode as the weeks went (see Fig. 22).
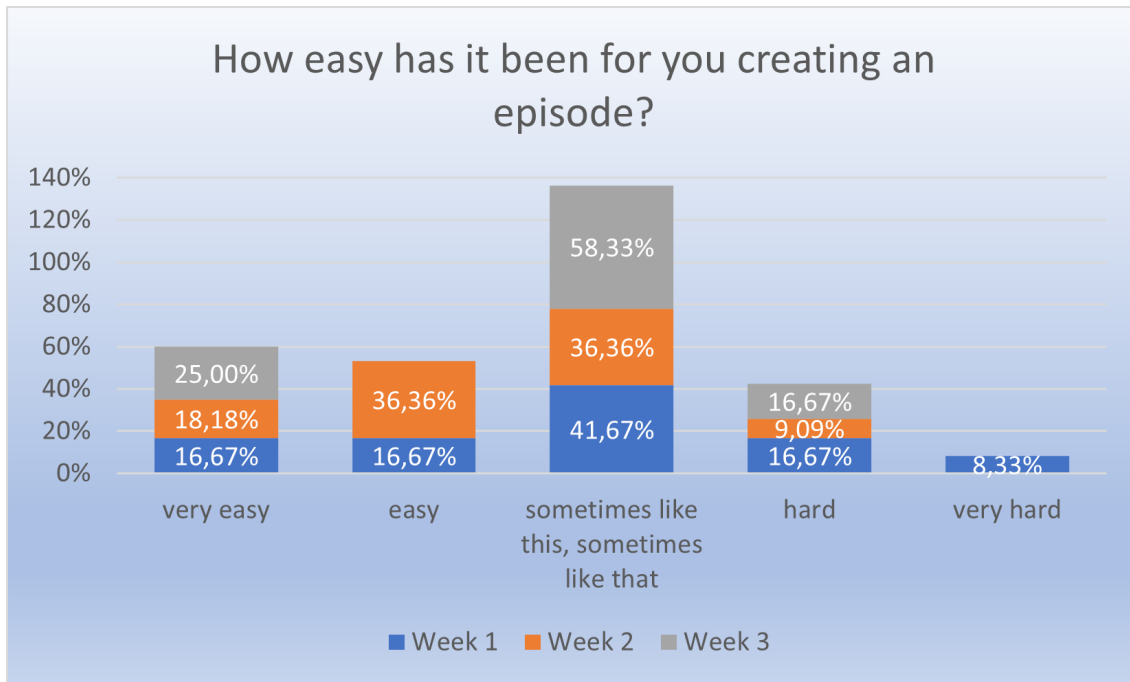
Figure 22: Degree of challenge with which an episode could be created.

Most of the test families said that the "Abendgruß" skill often understood them correctly: the highest value for this was in the second week (54%), the lowest in the third week (46%). The second most frequent response was that the skill "sometimes" understood them. There was a slight increase in the selection "always". In the first week, 8% of the test families indicated this, in the third week it was already 15%.

In each weekly questionnaire, the test families were also given the opportunity to share further experiences with the use of the skill via an open text field. Fortunately, this option was actively used. The main findings are summarized in the following.

With regard to the "Abendgruß" skill **in general**, the test families mentioned these points:

Topic	Description
Invocation and missing commands	Regarding the invocation, i.e. the command to start the skill, it had been difficult for the children of the test families to meet or remember the right wording. On the other hand, the right command ("hallo Abendgruß") was often not understood well enough by Alexa so it did not start the skill. Commands for pause were missed, the command "Stop" made Alexa stop the application. The skill was stopped as well, only by the word "Alexa".

Crashes/Errors	Concerning the technical workflow**,** the test families noticed crashes in the first and third week. The skill mainly crashed as a result of commands regarding the volume control as well as without no apparent reason. Some of the participants had to start the skill again. The lack of volume control was found to be annoying. It was reported that it had not been possible to stop the skill completely by the command "Alexa, stop" as the video stopped while the sound continued.
Ritual and habit	Using the "Abendgruß" skill was described as a ritual and it was mentioned that the children were able to operate the skill on their own.
Design	The design was found to be improvable regarding the visual representation and little design errors, related to font size and contrasts on the start screen (see Fig. 11.) .

Table 3: The summarised and categorised feedback of the test families from the weekly questionnaires on the "Abendgruß" skill.

With regard to the **two offered functions, i.e. to watch the current episode and to create an own episode**, the test families reported these issues:

Topic	Description
Wording	Regarding the creation of an own episode**,** it was stated that the wording "Oder mit etwas anderem? (Or with something else?)"/"Oder jemand anderes? (Or someone else?)" was found to be too difficult, but the option itself was appreciated. In addition, some of the terms of the options offered under "Eine eigene Folge. (My own episode.)", e.g. the names of the Sandmännchen friends "Internationale Kinderlieder" ("International children's songs") or "Impressionen" ("Impressions") were indicated as being too difficult for children (pronunciation and comprehension). Hence, it was considered helpful that even the partial naming of an option was sufficient to select it, e.g. "Fuchs (Fox)" instead of "Fuchs und Elster ("Fox and Magpie)".
Time for selection	Some test families mentioned that the time to decide on one of the options offered under "Eine eigene Folge. (My own episode.)" was too short. On the other hand, it was reported that Alexa's questions took too long and there was no way to interrupt it if users had already made their choice.
Crashes/Errors	In the first week the current episode was reported to have crashed during the first minute. It was also told that the skill crashed if "Oder mit etwas anderem? (Or with something else?)"/"Oder jemand anderes? (Or someone else?)" had been chosen. Another reason for crashes during the process of creating an episode was that Alexa did not stop saying that it had not understood. In some cases, the skill crashed almost daily after the main story had been selected. In this case it jumped back and

	asked again which vehicle the Sandmännchen should use to come in. Those errors were mainly reported in week 2 and 3. Another error linked to the process of creating an episode was reported to have happened after a story had been composed. The video could not be seen or crashed, and the screen appeared in white whereas the audio continued. Ending the skill was not possible then. Crashes that seemed to happen suddenly had been reported in week 2.
Repetitions and duplications	Repetitions and duplications of options were reported in all three weeks, concerning the vehicle of the Sandmännchen as well as the main story. The selection sometimes did not differentiate but offered identical episodes several times. Some test families wished for more variety regarding the stories whereas others rated the limited variety of stories as positive.
Understanding	Regarding the workflow of creating an episode, problems of understanding were only reported in case the child wanted to choose a story. As a solution it was suggested to link the options to numbers. It also was reported that after a while the understanding improved. Since no changes have been made to the skill in this regard, this can have various reasons. The most likely are two: (I) the Alexa language model has adapted to the users and learned to interpret the user commands correctly or (II) the users have (unconsciously) adapted their way of interacting with the skill. It may also be a combination of the two.
Default answer	Alexa's default answer "Sorry, I did not understand you." was found to be too monotonous and other answers for that case were proposed.
No touchscreen	The lack of using the skill via touchscreen was also mentioned.

Table 4: The summarized and categorized feedback of the test families from the weekly questionnaires on the "Abendgruß" skill functions.

With regard to the **ending** of the skill, the volume differences between the watched current or own episode and the closing words "Gute Nacht, bis zum nächsten Mal. (Good night, see you next time.)" were felt to be too striking. It was suggested that the closing words be whispered.

In summary, it can be said that the "Abendgruß" skill was used intensively by the test families over the entire test period, i.e. at least 3 to 4 times a week. The main users were children from the desired target group up to the age of six. The core function of creating an episode, which was developed within the project on the basis of the TVP and its components, was clearly preferred. It was also possible to identify clear points to improve the application:

- identify and fix the reasons for crashes/errors
- facilitate the communication with the skill in general
- facilitate the process of creating an episode
- provide a greater variety of options to select from to create an episode

Final Evaluation

The majority of the test families were satisfied with the "Abendgruß" skill overall. About 62% of them rather agreed or fully agreed to the respective statement. No one voted for "rather

disagree" or "disagree at all". 92% of the test families also stated that the skill can be used without any prior knowledge of the Sandmännchen (see Fig. 23).
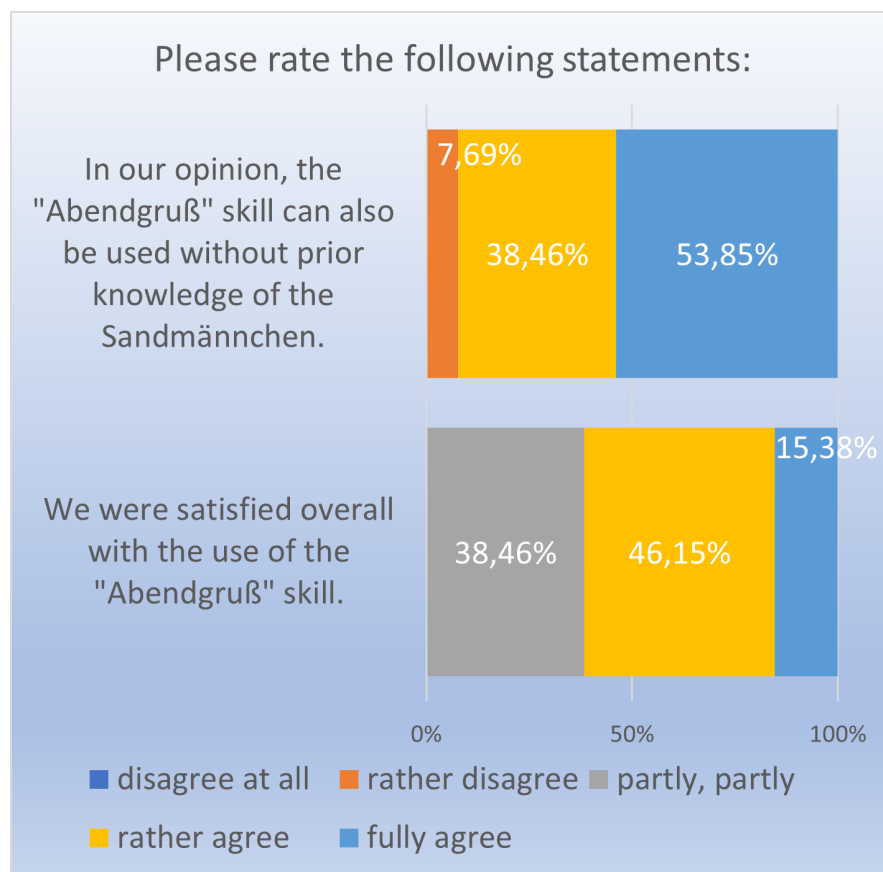
Figure 23: Ratings on the use of the "Abendgruß" skill without prior knowledge as well as on the general satisfaction with it.

Regarding the Net Promoter Score[24] question "How likely is it that you will recommend the "Abendgruß" skill to others (e.g. family, friends, acquaintances)?", the application achieved a value of -7.69. For a prototype, this is quite a good and encouraging result, especially if you take a closer look at the details: we got four "Detractors", six "Passives" and three "Promoters". That totally matches the previous results.

These ratings are all the more remarkable because most of the test families were undecided about the statement that smart speaker devices in general are suitable for offering media content to children (62%).

---

[24] The Net Promoter Score (NPS) is a metric that measures the extent to which consumers would recommend a company, a product, or a service to a friend or colleague. To calculate it, from the percentage of "Promoters", i.e. customers rating their likelihood as 9 or 10, the percentage rating this at 6 or low ("Detractors") on a scale from 0 to 10 is subtracted. Respondents who provide a score of 7 or 8 are referred to as "Passives" and do indeed enter into the overall percentage calculation. The Net Promoter Score can range between -100 and 100 and is indicated without a percentage sign: https://www.netpromotersystem.com/about/measuring-your-net-promoter-score/.

With regard to its usability the "Abendgruß" skill achieved an average System Usability Score[25] of 81.5 out of 100. The highest single value was 95, the lowest 65. Based on the facts that the average overall SUS score is 68 and that systems with a score of 80.3 or higher are awarded an A grade for high usability, the "Abendgruß" performed excellently.

Both functions of the "Abendgruß" skill as well as the two options offered for "Eine eigene Folge. (My own episode.)" were rated as useful (see Fig. 24). This is a clear vote, especially with regard to the core function of creating a personalised episode developed within ReTV, including the concept behind it.

Please rate the following statements:

Statement	disagree at all	rather disagree	partly, partly	rather agree	fully agree
We think the option "Oder mit etwas anderem / oder mit jemand anderem" is useful.				30,77%	69,23%
We think the option "Und wen möchtest du heute sehen" is useful.				23,08%	76,92%
We think the option "Wie soll das Sandmännchen heute zu uns kommen" is useful.				23,08%	76,92%
We think the option "Eine eigene Folge erstellen" is useful.			7,69%	15,38%	76,92%
We think the option "Die neueste Folge ansehen" is useful.		7,69%	7,69%	15,38%	69,23%

Figure 24: Evaluation of the usefulness of the different options of the "Abendgruß" skill.

The test families were also asked if they noticed any differences in the transitions between the intro/outro and main story elements when watching a created episode. The background to the question is that every Sandmännchen episode is split into its individual elements by the CERTH's Video Analysis Service by means of hard cuts. One consequence of this can be that the elements strung together in the created episodes are perceived as not coordinated/not fitting

---

[25] The System Usability Score (SUS)is a simple, ten-item attitude Likert scale giving a global view of subjective assessments of usability. SUS is used to provide an overall usability assessment measurement, as defined by ISO 9241–11, which is made up of the following characteristics: (I) effectiveness (Can users successfully achieve their objectives?), (II) efficiency (How much effort and resource is expended in achieving those objectives?) and (II) satisfaction (Was the experience satisfactory?): https://medium.com/thinking-design/the-system-usability-scale-how-its-used-in-ux-b823045270b7.

together. 54% of the test families did not notice any differences. Almost a third (31%) perceived picture and sound differences, while 15% noticed sound differences only.

Based on the feedback we received through both the weekly questionnaires and the parents' questionnaire on the idea and the concept of the "Abendgruß" skill, we concluded by asking the test families what specific improvements are needed to ensure that they continue to use (I) "Abendgruß" skill in general (see Fig. 25) and (II) the function of creating an episode in particular (see Fig. 26).

For (I), the results are as follows:

- It must be operable via touchscreen as well.
- The frequency of technical problems must be significantly reduced.
- Common functions such as volume control and pausing the playback of episodes must be integrated.
- It has to understand the users better.

For (II), the following two improvements are absolutely necessary:
- It must be avoided that the same option is suggested twice for "Wie soll das Sandmännchen heute zu uns kommen?(How should the Sandmännchen arrive today?)" or "Und wen möchtest Du heute sehen? (And who would you like to see today?)".
- It must be avoided after selecting "Oder mit etwas anderem? (Or with something else?)" or "Oder jemand anderes? (Or someone else?)" options already suggested are offered again.

To increase the chances of the "Abendgruß" skill being used in the future, …

Statement	disagree at all	rather disagree	partly, partly	rather agree	fully agree
… it has to greet us/ say goodbye to us personally (e.g. "Hallo, Martina." or "Gute Nacht, Martin.").	15,38%	53,85%	15,38%	15,38%	
… it must speak to us in a different voice (e.g. child, narrator, Sandmännchen friend).	7,69%	53,85%	15,38%	23,08%	
… the frequency of technical problems must be significantly reduced (e.g. crashes, getting stuck in the "Entschuldige, ich habe dich nicht verstanden"…		15,38%	7,69%	23,08%	53,85%
… it must also be operable via tapping on the display.	7,69%		7,69%	61,54%	23,08%
… more options than "Die neueste Folge ansehen" and "Eine eigene Folge erstellen" must be offered.	30,77%	23,08%	7,69%	30,77%	7,69%
… common functions such as controlling the volume and pausing the playback of episodes must be integrated.	7,69%	7,69%	15,38%	23,08%	46,15%
… the invocation (= command to start the skill, in our case "Hallo Abendgruß") must be changed.	30,77%	30,77%	15,38%	23,08%	
… it must understand us better.	7,69%	23,08%	7,69%	38,46%	23,08%

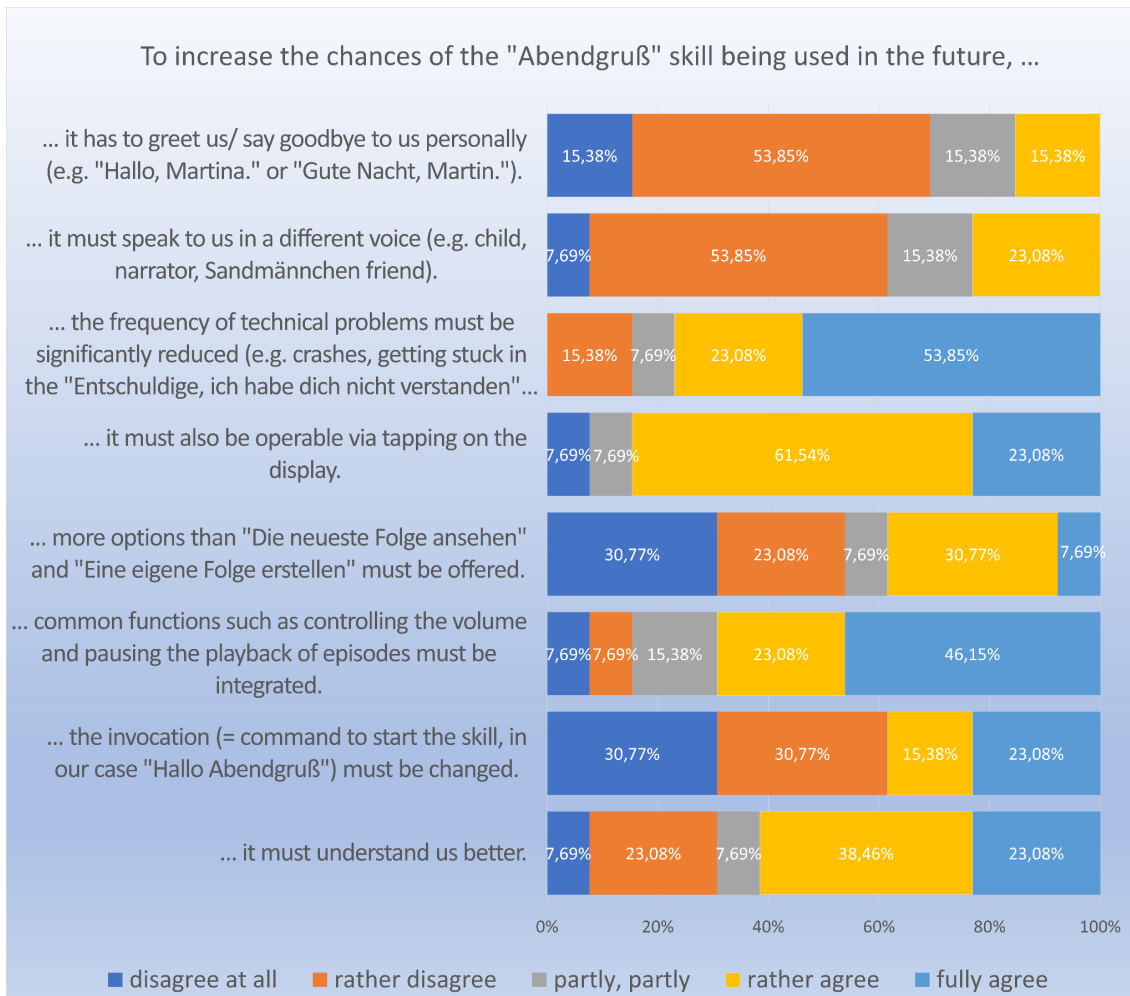*Figure 25: Requirements to increase the chances of the "Abendgruß" skill being used in the future.*

To increase the chances of "Eine eigene Folge erstellen" being used in the future, …

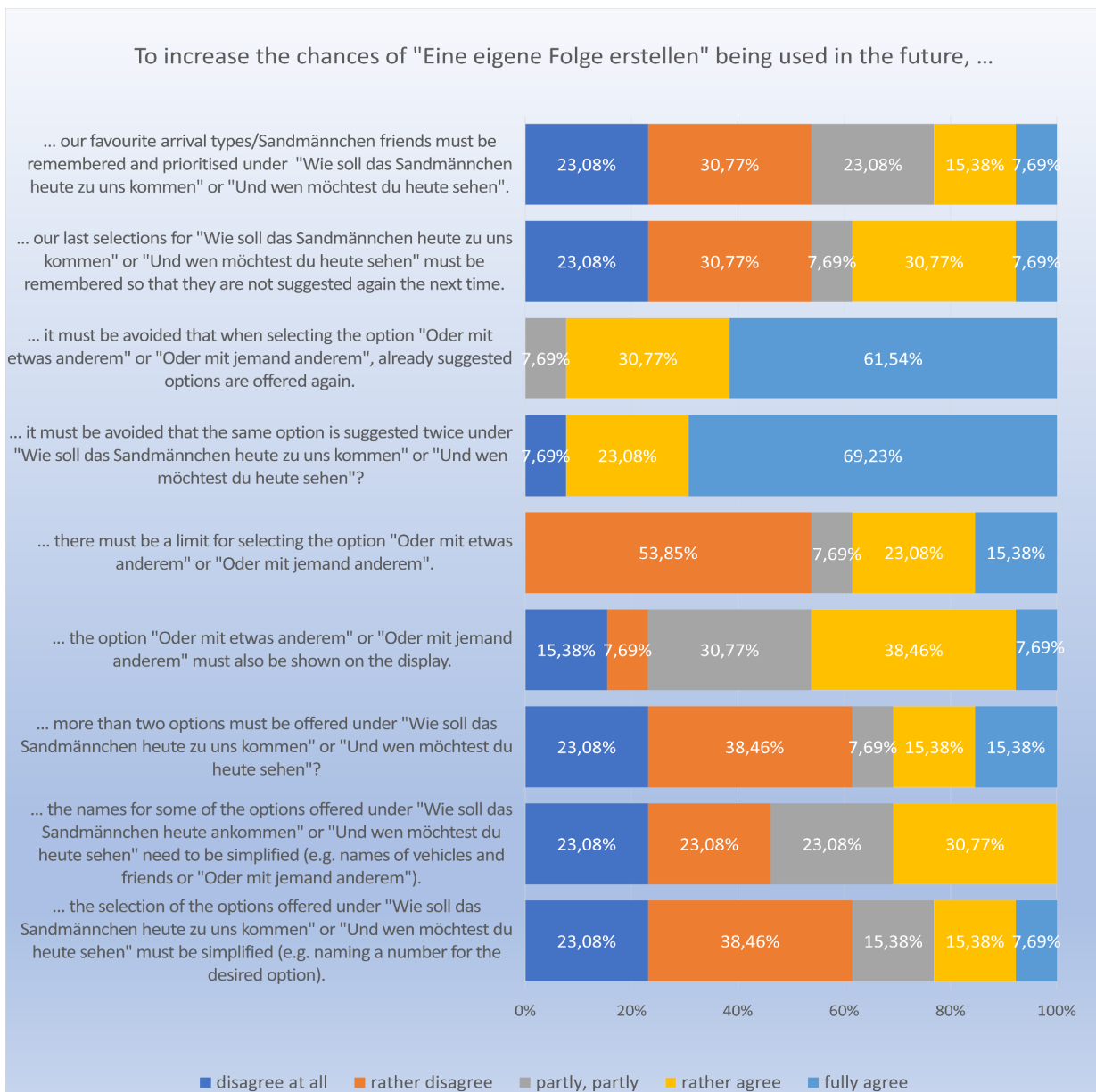| Question	disagree at all	rather disagree	partly, partly	rather agree	fully agree
… our favourite arrival types/Sandmännchen friends must be remembered and prioritised under "Wie soll das Sandmännchen heute zu uns kommen" or "Und wen möchtest du heute sehen".	23,08%	30,77%	23,08%	15,38%	7,69%
… our last selections for "Wie soll das Sandmännchen heute zu uns kommen" or "Und wen möchtest du heute sehen" must be remembered so that they are not suggested again the next time.	23,08%	30,77%	7,69%	30,77%	7,69%
… it must be avoided that when selecting the option "Oder mit etwas anderem" or "Oder mit jemand anderem", already suggested options are offered again.			7,69%	30,77%	61,54%
… it must be avoided that the same option is suggested twice under "Wie soll das Sandmännchen heute zu uns kommen" or "Und wen möchtest du heute sehen"?	7,69%			23,08%	69,23%
… there must be a limit for selecting the option "Oder mit etwas anderem" or "Oder mit jemand anderem".		53,85%	7,69%	23,08%	15,38%
… the option "Oder mit etwas anderem" or "Oder mit jemand anderem" must also be shown on the display.	15,38%	7,69%	30,77%	38,46%	7,69%
… more than two options must be offered under "Wie soll das Sandmännchen heute zu uns kommen" or "Und wen möchtest du heute sehen"?	23,08%	38,46%	7,69%	15,38%	15,38%
… the names for some of the options offered under "Wie soll das Sandmännchen heute ankommen" or "Und wen möchtest du heute sehen" need to be simplified (e.g. names of vehicles and friends or "Oder mit jemand anderem").	23,08%	23,08%	23,08%	30,77%	
… the selection of the options offered under "Wie soll das Sandmännchen heute zu uns kommen" or "Und wen möchtest du heute sehen" must be simplified (e.g. naming a number for the desired option).	23,08%	38,46%	15,38%	15,38%	7,69%

Figure 26: Requirements to increase the chance of "Eine eigene Folge. (My own episode.)" being used in the future.

4 VIDEO SUMMARIZATION WEB SERVICE

Video summarization is an essential technology developed in the context of ReTV and is utilized by specific components of ReTV's TVP, such as the 4u2 Messenger (see Section 2) as well as the Content Wizard (see Section 3 of D5.3). Additionally, to the best of our knowledge there are no freely available web-based video summarization tools. Of course, a plethora of scientific papers exists; some of them provide source code that can be used for video summarization - e.g. (Apostolidis et al., 2020), (Apostolidis et al., 2019), (Otani et al., 2019), (Shemer et al., 2019), (Zhou et al., 2018), (Fajtl et al., 2018). However, this requires that a user is well-informed on machine learning and proficient in computer programming, in order to use such code. For an
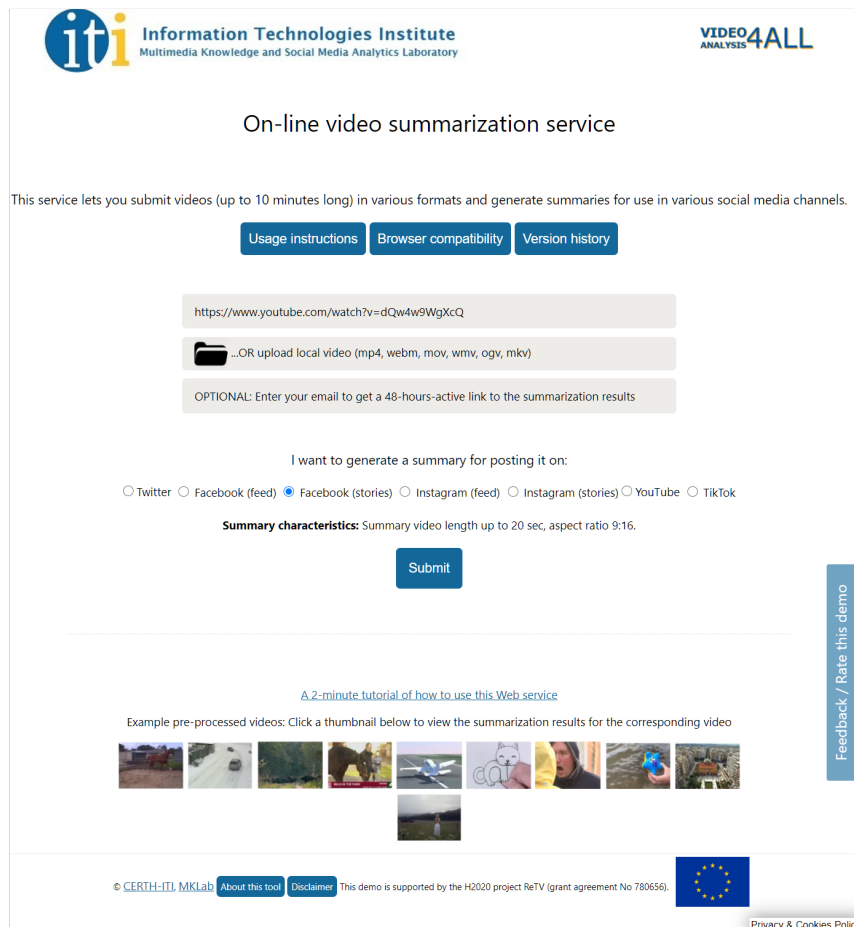
extensive analysis of related video summarization works, the interested reader is referred to (Apostolidis et al., 2020) and Section 2.1. of D3.3.

To showcase this particular ReTV technology to a wider audience of both professional and non-professional users, and motivated by the lack of web-based video summarization tools, we built a Web service that enables anyone to submit locally-stored or on-line available videos and automatically generate shorter versions of them. Our Web service takes as input a video and produces a video summary that encapsulates the flow of the story and the essential parts of the full-length video, adapting the length and format of the produced summary for publication on social media platforms, thus easing the creation of engaging video stories for online audiences. The Web service is based on the same core technologies detailed in Section 2.1 of D3.3, employed by the 4u2 Messenger and Content Wizard components of ReTV's TVP, albeit a simpler version, and the smart-cropping method discussed in Section 2.2 of D3.3.

The designed framework consists of: a) an interactive user interface (UI) that allows the user to exploit the functionality of our Web service and b) a REST service that hosts the developed technologies for video summarization (backend). In the sequel of this document we present the user interface in Section 4.1, the backend in Section 4.2, and finally the results and usage statistics of the implemented Web service in Section 4.3.

4.1 DESCRIPTION OF USER INTERFACE

The designed interactive user interface (see Fig. 27) allows the user to: a) submit a video for summarization (either available online or locally stored in the user's device), b) select the specifications of the generated summary from a list of predefined configurations that were properly adjusted for the most common social media and video sharing platforms, and c) get the created summary in a way that enables both immediate online inspection through the UI of our tool and the downloading of the video file in the user's device. Video submission is performed on a one-by-one basis (i.e., no video collection analysis is supported) and, for demonstration purposes, the submitted videos are allowed to be up to 10 minutes long and 200MB in file size.

Figure 27: The landing page of the frontend.

To submit a video for summarization the user can either provide its URL or upload a local copy of it from his/her device. The supported online video sources include YouTube, Facebook, Twitter, Instagram, Vimeo, DailyMotion, LiveLeak and Dropbox. Regarding local files, the service can handle videos in mp4, webm, avi, mov, wmv, ogv, mpg, flv, and mkv formats. Regardless of the mode of submission (local file or URL), it is the responsibility of any person submitting a video to this service to make sure that he or she has the right to do so without violating any copyrights associated with the submitted video. After submitting a video, the user can monitor the progress of the summarization, and is also able to submit additional requests while the previous ones are being analyzed as shown in Fig. 28. The submitted video and the summarization results are cached in a server for 24 hours, and after this time period, the local copy or the video URL, the summarization results and the user's email address (if provided) are automatically deleted from the server. When the analysis is completed, and after an automatic refresh of the user interface, the generated summary is presented to the user through the user interface presented in Fig. 29.
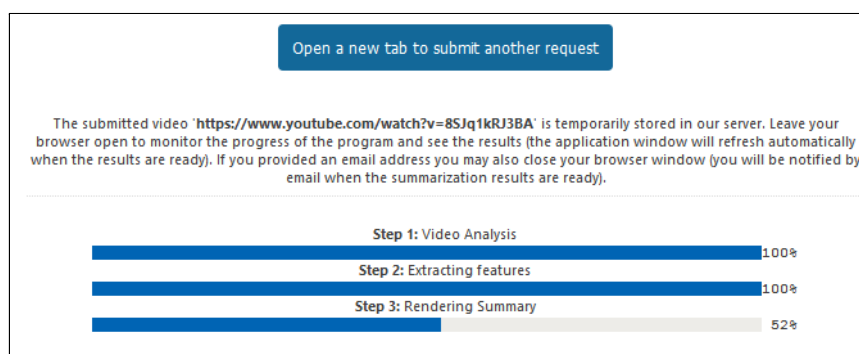
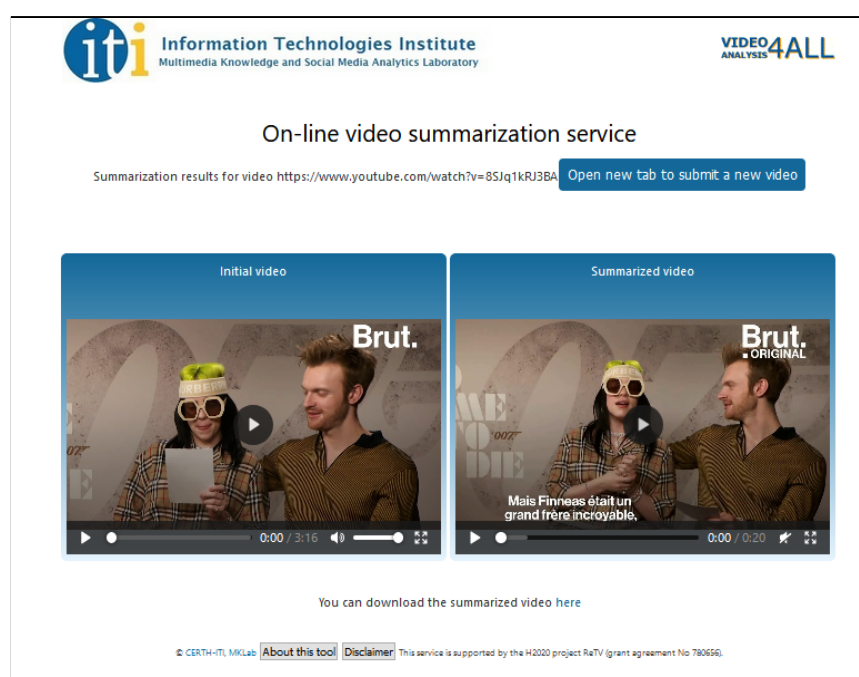*Figure 28: The summarization progress bars of the frontend.*

Figure 29: The results page of the frontend.

Optionally, if the user provided an email address she/he may close the Web browser and be notified by e-mail when the summary result is ready. The results page is an interactive webpage with two video players (for viewing the original and summarized videos) implemented using the HTML5 video tag. The players support all standard functionalities such as play/pause the video and toggle the video in full screen mode. Furthermore, the user is able to download the produced summary. The developed Web service for video summarization is fully compatible with Mozilla Firefox (>41.0), Chrome (>45.0), Opera (>32.0), Microsoft Edge (>77.0.200.1), and IE (>11.0).

Aiming to help the user and relieve him/her from deciding the appropriate target duration and aspect ratio of the produced video summary, we have created a list of configurations for the most common social media and video sharing platforms. The user can select one using the respective option buttons found in the service's landing page (see Fig. 27). This action configures the service to produce a video summary that fully meets the prerequisites of the target platform, based on information gathered from the internet and the works of (Moldovan, 2019) and (Dimopoulos, 2013). The specifications for five widely used social media and video sharing platforms are shown in Table 5, while an example of applying two different presets on

the same input video is shown in Fig. 30. To meet these specifications, we automatically adjust the duration of the video summary and employ the smart-cropping technique to retarget the video summary to the required aspect ratio. The parameters presented in Table 5 aim to maximize user engagement and experience. Nevertheless, it is important to stress that these are best practices and they are subject to change based on various trends.

Video sharing platform	Optimal summary length (seconds)	Video length hard limit (seconds)	Aspect ratio
Twitter	30	140	2:1
Facebook (feed)	120	120	16:9
Facebook (stories)	20	20	9:16
Instagram (feed)	30	30	4:5
Instagram (stories)	15	15	9:16
YouTube	120	900 / unlimited*	original
TikTok	15	60	9:16

* This depends on the user account type; and additional limitations based on file size apply.

Table 5: Configuration settings for each social media platform. The developed service generates summaries that conform to the optimal summary length and aspect ratio listed above.

3' 25'' long
(19 shots)

(a)

1'59'' long
(13 shots)

(b)

15'' long
(3 shots)

(c)

Figure 30: Key-frames of shots for the: (a) original video with 19 shots and 205 seconds total length, (b) video summary for Facebook feed (16:9 aspect ratio) with 13 shots and 119 seconds total length, (c) video summary for TikTok (9:16 aspect ratio) with 3 shots and 15 seconds total length.

4.2 DESCRIPTION OF BACKEND

The backend process for video summarization starts by segmenting the video into shots. A shot is an elementary structural unit of the video, that is composed of a set of consecutive frames captured by a single camera without interruption (Apostolidis et al., 2014). For the video shot segmentation we trained a 3D CNN with a similar architecture to (Gygli, 2017) using the BBC Planet Earth dataset (Baraldi et al., 2015), which contains ground truth annotations for training a shot segmentation method. Each detected shot is ranked according to a value denoting its suitability to be part of the video summary, calculated as the mean of the estimated importance scores of the shot's frames, using the method of (Apostolidis et al. 2020), thus producing a ranked list of shots. The shot with the highest rank is selected as a candidate to be included in the summary and is removed from the ranked list. Then, two empirically-set thresholds, *min_segment_duration* and *max_segment_duration* are utilized, with the intent to impose bounds on the duration of the selected part from each shot, i.e., to avoid the inclusion of very short or very long segments in the summary. Specifically, if the selected shot's duration is greater than *min_segment_duration* and lower than *max_segment_duration* then the whole shot is included in the summary. If this duration is greater than *max_segment_duration* then we select a part of the shot (of *max_segment_duration* seconds) for which its frames exhibit the maximum sum of importance scores. If this duration is lower than *min_segment_duration* then the shot is discarded. The procedure is repeated until the video summary has a duration that is very close to the target duration, and results in an array containing the start and end time of each selected segment. The array is sorted based on the time of appearance of each segment in the original video and is subsequently fed to a separate module that is responsible for decoding the original video, finding the respective selected segments, transforming them to fit the target aspect ratio and rendering them to a final summary video file.

The above-described processing pipeline is deployed as a REST service that: a) retrieves a video file, b) analyzes the video using the method of (Apostolidis et al., 2020) to estimate frame-level importance scores, c) performs temporal segmentation of the video to shots, d) ranks the shots and selects a part from each of the top-ranked shots until the summary's specified time-budget is filled based on the determined thresholds and the selection process described above, e) transforms the video frames to the target aspect ratio, and f) renders the video summary.

The frontend and backend components of our Web service are deployed as independent modules. Once a video is submitted to the frontend, a call to the backend service is initiated, which includes the specifications of the video summary to be generated. In turn, the backend instantiates a processing session, returns a unique session ID for reference to the frontend and inserts the session in the processing queue. The frontend periodically queries about the status of the backend session process. The response of the backend contains information about the position of the queried session in the queue or the progress, so that the user interface can provide the corresponding visual feedback to the user. Once the backend session is completed its status is set accordingly. The frontend can then retrieve the video summary as well as the original full length using another call, display the two videos side-by-side in the results interface and notify the user via email that the processing has been completed (if the user had opted to use this feature by submitting his/her email along with the original video).

4.3 RESULTS AND USAGE STATISTICS

The Web service for video summarization was first made publicly available on 13th of March, 2020. It can be accessed and tested online[26]. After the initial release there were several updates to the Web service, to improve its results, make adjustments to the text/messages in the user interface to make it more informative, and integrate new functionalities as these were made available (e.g. smart-cropping).

The learning-based summarization method of (Apostolidis et al., 2020) that the Web service utilizes, was evaluated on the SumMe (Gygli et al., 2014) and TVSum (Song et al., 2015) video summarization standard benchmark datasets. The findings show that the utilized method performs consistently well in both datasets, and is the most competitive one among the literature approaches. Details about that evaluation and an extensive analysis of the results can be found in Section 2.1.5 of D3.3. Also, the component testing and software quality assessment procedure carried out for the video summarization and smart cropping components, as discussed in Section 2.4.3 of D3.3, was extended to include the implementation used in the Video Summarization Web Service. The employed video summarization technology has been evaluated, as part of the testing with ReTV users, and results are reported in Section 5 of D5.3.

The backend service, discussed in Section 4.2, is deployed on a PC with an Intel i7-4770K at 3.50 GHz, 32GB of RAM and a NVIDIA GeForce GTX 1660 graphics card. By exploiting the multi-thread and multi-core processing capabilities of the available CPU and GPU, the analysis is faster than real-time video, taking approximately 45% of the original video's playback time; though of course delays may be noticed if multiple analysis requests are submitted to the service, since these requests are processed on a one-by-one basis.

There is no particular focus group that we considered during the development of the Web service; video creators, journalists, broadcasters, as well as simple users who regularly create video content to be published, can freely use our Web service to help themselves in the tedious task of generating tailored versions of video content for publication in multiple platforms. To gather usage statistics we employed a logging mechanism to register for each call the following information:

- Call date and time
- Caller's IP (just for the purpose of counting unique and returning users)
- Country and city of the caller (again, just for statistical purposes)

The Web service is GDPR-compliant, clearly displaying a Privacy Policy that explains the above collection of the least-possible amount of data that is necessary purely for statistical and reporting purposes.

Based on the above-mentioned usage statistics, during the period between 13th of March 2020 and 10th of March 2021 (excluding calls from IPs affiliated with ReTV partners) there were:

- >900 videos submitted to the Web service
- 200 unique users
- 75 returning users, i.e. the same IP using the Web service more than once
- Calls to the service were submitted from 29 countries and 73 different cities

In addition to these usage statistics, we should highlight that we presented a publication & demo of this Web service at ACM IMX 2020 (Collyda et al. 2020), with an estimated audience of

---

[26] http://multimedia2.iti.gr/videosummarization/service/start.html

more than 200 people coming from both academia and industry in the area of interactive media.

5   CONTENT SWITCH

The provision of linear TV on the internet - either through broadcasters' websites and apps or internet-TV services such as Zattoo - makes it possible to take advantage of content personalization. In ReTV, we believe that is an unexplored opportunity that could benefit both the broadcasters and the advertisers.

The idea behind the Content sWitch was to allow broadcasters to designate areas of their programming for on-demand personalisation. It **enhances the linear TV experience by replacing certain parts with alternatives tailored for each viewer**. The possible applications are:

- replacing ads with targeted ads
- replacing ads with promos trailers for shows
- replacing ads with other, relevant content like the weather forecast

After developing and testing the first prototype of Content sWitch (see D6.2) we started to see a strategic and technical misalignment between the Content sWitch and the other three use-cases.

The Content Wizard (see Section 3 of D5.3), the Topics Compass (see Section 4 of D5.3), and the 4u2 Messenger (see Section 2) are all SaaS products built on top of the ReTV backend services. They can be sold without further customization and are stand-alone products. The Content Switch, on the other hand, would need to be tightly integrated with the encoding pipeline of an OTT provider.

The Content Wizard, the Topics Compass, and the 4u2 Messenger can be sold directly to their actual customers. For the Content sWitch, sales are more complicated: the practical technical implementation needs to be done by the OTT provider, but they, in turn, would be paid by the broadcasters who benefit from the dynamically replaced content. Proving to a broadcaster how much revenue they gained through the Content sWitch would be tricky in practice: It requires a homogeneous measurement of user-engagement across all the OTT providers that broadcast a particular channel.

If we wanted to allow, for example, RBB to use the Content sWitch in practice, we would have to:

- Get all of the large OTT providers like Deutsche Telekom, Waipu, and Zattoo to integrate with the Content sWitch.
- Deploy a homogenous measurement at all the participating OTT providers across all of their apps on different platforms.
- Persuade RBB that they should pay ReTV and the OTT providers for the increased user-engagement.

Additionally, after our first round of user tests, we realized that advertising didn't play as large a role we had initially envisioned, neither for end users nor for the professional users. This sets the Content sWitch apart from the three other use case scenarios, which are purely content focused. Ultimately, trading off overall strategic fit with the project goals and required engineering effort, we decided to focus our efforts on the three remaining use case scenarios and therefore increase the potential impact of ReTV on the broadcaster and archive landscape.

6 CONCLUSION AND OUTLOOK

In T6.3 we developed prototypical applications based on the TVP that demonstrate clear consumer benefits. The two consumer prototypes, the 4u2 Messenger and 4u2 Smart Speaker Skill "Abendgruß" each show in their own right a high level of potential. The level of engagement with both of the services in the longitudinal tests indicate a demand for and acceptance of personalized services. They both reflect the TVP ability to offer users innovative personalized content services. The technology developed is transferable to other services beyond the prototypes tested.

As mentioned in the introduction of this document, ReTV partners have outlined their plans to exploit and commercialise the TVP and ReTV services in D7.4 Business and Exploitation Plan. The plans outlined in that document combined with the results reported here, provide the partners with input to further develop the services and engage in further steps towards commercialization. Both the "Abendgruß" skill and the 4u2 Messenger prototypes used in the longitudinal consumer tests offer concrete exploitation opportunities.

At RBB, the editorial department "Familie & Kinder" responsible for Sandmännchen has indicated a willingness and interest to look into incorporating the interactive personalization function into its official Sandmännchen Skill.[27] The results of the survey provide a clear indication of improvements required before this would happen. A further avenue of investigation is to look at the suitability of the ReTV technology to create the actual Sandmännchen episode broadcast on TV. Each TV episode is currently created by an editor using professional video editing software and includes transition effects between the individual sections of the show. To replicate these effects with ReTV technology would require further development. However, the high level of acceptance of the skill is encouraging and merits considering this type of personalization and interaction for other content and target audiences.

Equally, after the positive evaluation results, NISV is eager to officially release the 4u2 Messenger and invest in its further development, in particular, extending the engagement possibilities between the users and the messenger and making use of TVP's predictive capabilities. However, this is subject to the pending permissions from Facebook and an alternative messaging platform could be selected to release the application.

The exploitation agreement and business plan detailed in D7.4, ensure that the TVP services required for the RBB and NISV consumer applications will remain available beyond the end of the ReTV project. To support this cooperation partners have agreed to sign a Memorandum of Understanding, which sets out the basis for this continued cooperation.

The Video Summarization Web Service offers an easy to use tool to anyone interested in creating summaries of existing videos for publication on multiple social media platforms. CERTH will continue to keep the website online even after the project ends, utilizing the feedback collected in order to improve the employed video summarization core technologies.

---

[27] https://www.amazon.de/dp/B08QSG6XM3

REFERENCES

Apostolidis, E., Adamantidou, E., Metsai, A., Mezaris, V., & Patras, I. (2020). Unsupervised Video Summarization via Attention-Driven Adversarial Learning. In International Conference on Multimedia Modeling (pp. 492–504).

Apostolidis, E., Metsai, A., Adamantidou, E., Mezaris, V., & Patras, I. (2019). A stepwise, label-based approach for improving the adversarial training in unsupervised video summarization. In Proceedings of the 1st International Workshop on AI for Smart TV Content Production, Access and Delivery (pp. 17–25).

Apostolidis, E., & Mezaris, V. (2014). Fast shot segmentation combining global and local visual descriptors. In 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 6583–6587).

Collyda, C., Apostolidis, K., Apostolidis, E., Adamantidou, E., Metsai, A. I., & Mezaris, V. (2020, June). A Web Service for Video Summarization. In ACM International Conference on Interactive Media Experiences (pp. 148-153).

Gygli, M. (2017). Ridiculously fast shot boundary detection with fully convolutional neural networks. arXiv preprint arXiv:1705.08214.

Baraldi, L., Grana, C., & Cucchiara, R. (2015). A deep siamese network for scene detection in broadcast videos. In Proceedings of the 23rd ACM international conference on Multimedia (pp. 1199–1202).

Song, Y., Vallmitjana, J., Stent, A., & Jaimes, A. (2015). Tvsum: Summarizing web videos using titles. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 5179–5187).

Gygli, M., Grabner, H., Riemenschneider, H., & Van Gool, L. (2014). Creating summaries from user videos. In European conference on computer vision (pp. 505–520).

Moldovan, C., Wamser, F., & Hossfeld, T. (2019). User Behavior and Engagement of a Mobile Video Streaming User from Crowdsourced Measurements. In 2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX) (pp. 1–3).

Dimopoulos, G., Barlet-Ros, P., & Sanjuas-Cuxart, J. (2013). Analysis of YouTube user experience from passive measurements. In Proceedings of the 9th International Conference on Network and Service Management (CNSM 2013) (pp. 260–267).

Zwicklbauer, M., Lamm, W., Gordon, M., Apostolidis, K., Philipp, B., & Mezaris, V. (2020, October). Video Analysis for Interactive Story Creation: The Sandmännchen Showcase. In Proceedings of the 2nd International Workshop on AI for Smart TV Content Production, Access and Delivery (pp. 17-24).

Zhou, K., Qiao, Y., & Xiang, T. (2018). Deep reinforcement learning for unsupervised video summarization with diversity-representativeness reward. In Thirty-Second AAAI Conference on Artificial Intelligence.

Fajtl, J., Sokeh, H., Argyriou, V., Monekosso, D., & Remagnino, P. (2018). Summarizing videos with attention. In Asian Conference on Computer Vision (pp. 39–54).

Otani, M., Nakashima, Y., Rahtu, E., & Heikkila, J. (2019). Rethinking the evaluation of video summaries. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 7596–7604).

Shemer, Y., Rotman, D., & Shimkin, N. (2019). ILS-SUMM: Iterated Local Search for Unsupervised Video Summarization. arXiv preprint arXiv:1912.03650.

APPENDIX A: EVALUATION QUESTIONNAIRE FOR 4U2 MESSENGER

Question	Question Type	Possible Answers
On a scale from 1 to 10, rate your overall experience with the 4u2 Messenger.	Scale	Choice from 1-10
Did you like the videos curated for you by the 4u2 Messenger?	Single choice	- Strongly disagree - Disagree - Neutral - Agree - Strongly Agree
The videos I received matched my interests.	Scale	- Strongly disagree - Disagree - Neutral - Agree - Strongly Agree
I enjoyed receiving videos on a wide range of topics.	Scale	- Strongly disagree - Disagree - Neutral - Agree - Strongly Agree
I thought most of the messages were sent out at a time that suited me.	Scale	- Strongly disagree - Disagree - Neutral - Agree - Strongly Agree
I would prefer to receive messages at the same time each day.	Scale	- Strongly disagree - Disagree - Neutral - Agree - Strongly Agree
Would you like to keep using the 4u2 Messenger in the future?	Single choice	- Yes - Yes, but with some adjustments - No
Do you have additional comments or suggestions for improvement?	Open Textfield	

APPENDIX B: EVALUATION QUESTIONNAIRES FOR 4U2 SMART SPEAKER SKILL

PARENTS' QUESTIONNAIRE ON THE IDEA AND THE CONCEPT

Question	Question Type	Possible Answers
Children in the Household		
How many children aged 6 and under live in your household?	Single Choice	- 1 - 2 - 3 - 4 - 5 - 6 - more than 6 - I have no children in my household
How old are the children living in your household?	Multiple Choice	- under 1 year - 1 year - 2 years - 3 years - 4 years - 5 years - 6 years - older than 6 years
Which of these devices may your children use independently?	Multiple Choice	- TV - smartphone - tablet - smart speaker without display (e.g. Amazon Echo Dot, Google Home, Sonos One, Bose Home Speaker) - smart speaker with display (e.g. Amazon Echo Show, Google Nest Hub) - streaming devices (e.g. Fire TV Stick, Google Chromecast, Apple TV) - none of the above - other [with open textfield]
Unser Sandmännchen		
Do you know "Unser Sandmännchen"?	Yes-No question	- yes - no

We watch the episodes of "Unser Sandmännchen" on the following channels:	Multiple Choice	- daily broadcast on TV (rbb/MDR/KiKA) - ARD-Mediathek - website - app for HbbTV - app(s) for smartphones - app(s) for tablets - YouTube - We don't watch the episodes of "Unser Sandmännchen" - other [with open text field]
We watch the episodes of "Unser Sandmännchen":	Multiple Choice	- to fall asleep - to play - for learning - for entertainment - for distraction (e.g. long car rides, comforting) - only on certain days to see a certain Sandmännchen friend - other [with open textfield]
We watch the episodes of "Unser Sandmännchen":	Multiple Choice	- in the morning (6 to 11 am) - at noon (11:01 am to 1 pm) - in the afternoon (1:01 to 5 pm) - in the evening (5:01 to 9 pm) - at night (9:01 to 00:00 pm) - other period [with open textfield]
Voice-controlled Applications for Children		
In our household, voice-controlled applications for children are already in use.	Single Choice	- yes - application is started by the parents - yes - application is started by the children - yes application is started by both parents and children - no

We already use the following voice-controlled applications for children:	Open Textfield	
We would like to use a voice-controlled Sandmännchen application.	Yes/No Question	- yes - no
We would want to start a voice-controlled Sandmännchen application in combination with the following term, e.g. "Alexa, open …":	Multiple Choice	- Abendgruß - Sandmann - Sandmännchen - Abendgruß-Geschichte - Sandmann-Geschichte - Sandmännchen-Geschichte - Freunde-Geschichte - Traumsand - other [with open textfield]
We would like to see the following functions for a voice-controlled Sandmännchen application:	Multiple Choice	- play the current episode of "Unser Sandmännchen" - choose from different friends stories of "Unser Sandmännchen" - put together and play an episode of "Unser Sandmännchen" in a playful way (e.g. choose between two friends stories) - play several episodes of "Unser Sandmännchen" in a row - be guided through the app by a child's voice - also play other content, e.g. Ohrenbär or Zappelduster - use the app and its content as an audio-only version
We would like to see the following additional features for a voice-controlled Sandmännchen application:	Open Textfield	

INTAKE QUESTIONNAIRE

Question	Question Type	Possible Answers
Log in		
Please identify yourself with your test pseudonym.	Dropdown	- Der kleine König - Der kleine Rabe Socke - Die Moffels - Dr. Brumm - Ferdinand und Paula - Fuchs und Elster - Jan & Henry - Kalli - Meine Schmusedecke - Pittiplatsch - Plumps - Pondorondo - Rita und das Krokodil
Children in the Household		
How many children live in your household?	Single Choice	- 1 - 2 - 3 - 4 - 5 - 6 - more than 6
How old are the children?	Multiple Choice	- under 1 year - 1 year - 2 years - 3 years - 4 years - 5 years - 6 years - older than 6 years
Which of the following devices are in your household?	Multiple Choice	- TV - smartphone - tablet - smart speaker without display (e.g. Echo Dot, Home Mini, Sonos One, Bose Home Speaker) - smart speaker with display (e.g. Echo Show, Nest Hub) - streaming devices (e.g.

		Fire TV Stick, Apple TV, Chromecast) - none of the above - other [with open textfield]
Which of the following devices are the children allowed to use independently?	Multiple Choice	- TV - smartphone - tablet - smart speaker without display (e.g. Echo Dot, Home Mini, Sonos One, Bose Home Speaker) - smart speaker with display (e.g. Echo Show, Nest Hub) - streaming devices (e.g. Fire TV Stick, Apple TV, Chromecast) - none of the above - other [with open textfield]
Unser Sandmännchen		
Do you know the Sandmännchen?	Yes-No Question	- yes - no
Do you prefer to watch or to listen to the Sandmännchen episodes?	Single Choice	- We watch them. - We listen to them. - We neither watch nor listen to them. - We don't watch and listen to them.
We watch/listen to the Sandmännchen episodes via the following channels:	Multiple Choice	- daily broadcast on television (rbb/MDR/KiKA) - website - ARD-Mediathek - app for HbbTV - app for smartphone - app for tablet - app for streaming Devices (Apple TV, Fire TV) - app for voice assistants (Amazon Alexa, Google Assistant) - YouTube

		- other [with open textfield]
We watch/listen to the Sandmännchen episodes:	Multiple Choice	- to fall asleep - to play - for learning - for entertainment - for distraction (e.g. long car journeys, comforting) - only on certain days to see a certain Sandmännchen friend (e.g. Pittiplatsch or Jan & Henry) - other [with open textfield]
We watch/listen to the Sandmännchen episodes:	Multiple Choice	- in the morning (6 to 11 am) - at noon (11:01 am to 1 pm) - in the afternoon (1:01 to 5 pm) - in the evening (5:01 to 9 pm) - at night (9:01 to 00:00 pm) - other period [with open textfield]
We usually watch/listen to the Sandmännchen episodes (frequency):	Single Choice	- 1 at a time - 2 to 3 at a time - 4 to 5 at a time - 6 and more at a time
Voice-controlled Applications for Children		
We already use voice-controlled apps for children in our household.	Single Choice	- yes, applications are started by the parents - yes, applications are started by the children - yes, applications are started by both parents and children - no
We use the following voice applications for children:	Open Textfield	
Do you already know the new	Yes-No Question	- yes

Sandmännchen application for voice assistants (Amazon Alexa, Google Assistant)?		- no

WEEKLY QUESTIONNAIRE

Question	Question Type	Possible Answers
Log in		
Please identify yourself with your test pseudonym.	Dropdown	- Der kleine König - Der kleine Rabe Socke - Die Moffels - Dr. Brumm - Ferdinand und Paula - Fuchs und Elster - Jan & Henry - Kalli - Meine Schmusedecke - Pittiplatsch - Plumps - Pondorondo - Rita und das Krokodil
Our Questions to You		
How often did you use the "Abendgruß" skill this week?	Single Choice	- never - 1 to 2 times - 3 to 4 times - 5 to 6 times - 7 times and more
Why did you never use the "Abendgruß" skill this week? [only shown if "never" was selected in the previous question]	Open Textfield	
When did you use the "Abendgruß" skill?	Multiple Choice	- in the morning (6 to 11 am) - at noon (11:01 am to 1 pm) - in the afternoon (1:01 to 5 pm) - in the evening (5:01 to 9 pm) - at night (9:01 to 00:00 pm) - other period [with open

		textfield]
Who operated the "Abendgruß" skill?	Multiple Choice	- parents/adults - children up to six years - children from six years
How often did you watch the current episode?	Single Choice	- never - 1 to 2 times - 3 to 4 times - 5 to 6 times - 7 times and more
How often did you create an episode?	Single Choice	- never - 1 to 2 times - 3 to 4 times - 5 to 6 times - 7 times and more
How easy has it been for you creating an episode?	Single Choice	- very easy - easy - sometimes like this, sometimes like that - hard - very hard
Did the "Abendgruß" skill get you right?	Single Choice	- always - often - sometimes - rarely - never
Are there any other experiences with the "Abendgruß" skill that you would like to share with us?	Open Textfield	

FINAL QUESTIONNAIRE

Question	Question Type	Possible Answers
Log in		
Please identify yourself with your test pseudonym.	Dropdown	- Der kleine König - Der kleine Rabe Socke - Die Moffels - Dr. Brumm - Ferdinand und Paula

		- Fuchs und Elster - Jan & Henry - Kalli - Meine Schmusedecke - Pittiplatsch - Plumps - Pondorondo - Rita und das Krokodil
General		
Please rate the following statements: - Overall, we were satisfied with the use of the "Abendgruß" skill. - In our opinion, the "Abendgruß" skill can also be used without prior knowledge of the Sandmännchen.	Matrix	- Strongly disagree - Disagree - Neutral - Agree - Strongly agree
How likely is it that you will recommend the "Abendgruß" skill to others (e.g. family, friends, acquaintances)?	Net Promoter score	Choice from 1-10
Please rate the following statements: - In our opinion, the smart speaker device is suitable for offering media content to children. - In our opinion, voice-controlled applications are suitable for children. - Regardless of the application used (e.g. smart speaker skill, smartphone app, media library website), the ability to personalise media content is generally important to us.	Matrix	- Strongly disagree - Disagree - Neutral - Agree - Strongly agree
Usability		
Please rate the following statements:	Matrix	- Strongly disagree - Disagree

- We can very well imagine using the "Abendgruß" skill regularly. - We find the "Abendgruß" skill unnecessarily complex. - We find the "Abendgruß" skill easy to use. - We think that we would need technical support to use the "Abendgruß" skill. - We find that the different functions of the "Abendgruß" skill are well integrated. - We find that there are too many inconsistencies in the "Abendgruß" skill (e.g. when selecting "Eine eigene Folge erstellen" the latest episode is displayed or when selecting "Wie soll das Sandmännchen heute zu uns kommen?" Sandmännchen friends are suggested). - We can imagine that most people will quickly learn to master the "Abendgruß" skill. - We find the "Abendgruß" skill very cumbersome to use. - We felt very confident using the "Abendgruß" skill. - We had to learn a lot of things before we could use the "Abendgruß" skill.		- Neutral - Agree - Strongly agree
Skill now and in the future		
Please rate the following statements: - We think the option "Die neueste Folge ansehen" is useful. - We think the option "Eine eigene Folge erstellen" is useful. - We think the option "Wie	Matrix	- Strongly disagree - Disagree - Neutral - Agree - Strongly agree

soll das Sandmännchen heute zu uns kommen? is useful. - We think the option "Und wen möchtest du heute sehen?" is useful. - We think the option "Oder mit etwas anderem?"/"Oder mit jemand anderem?" is useful.		
When watching our own episodes, we noticed in the selected elements (arrival/disappearance of the Sandmännchen and friend story)...	Single Choice	- ... picture differences. - ... differences in sound. - ... picture and sound differences. - ... no differences.
To increase the chances of the "Abendgruß" skill being used in the future, … - … it must understand us better. - ... the invocation (= command to start the skill, in our case "hallo Abendgruß") must be changed. - ... common functions such as controlling the volume and pausing the playback of episodes must be integrated. - ... more options than "Die neueste Folge ansehen" and "Eine eigene Folge erstellen" must be offered. - ... it must also be operable via tapping on the display. - ... the frequency of technical problems must be significantly reduced (e.g. crashes, getting stuck in the "Entschuldige, ich habe dich nicht verstanden" loop). - ... it must speak to us in a	Matrix	- Strongly disagree - Disagree - Neutral - Agree - Strongly agree

different voice (e.g. child, narrator, Sandmännchen friend). - ... it has to greet us/say goodbye to us personally (e.g. "Hallo, Martina." or "Gute Nacht, Martin.").		
To increase the chances of "Eine eigene Folge erstellen" being used in the future, … - ... the selection of the options offered under "Wie soll das Sandmännchen heute zu uns kommen?" or "Und wen möchtest du heute sehen?" must be simplified (e.g. naming a number for the desired option). - ... the names for some of the options offered under "Wie soll das Sandmännchen heute ankommen?" or "Und wen möchtest du heute sehen?" need to be simplified (e.g. names of vehicles and friends or "Oder mit jemand anderem?"). - ... more than two options must be offered under "Wie soll das Sandmännchen heute zu uns kommen?" or "Und wen möchtest du heute sehen"? - ... the option "Oder mit etwas anderem?" or "Oder mit jemand anderem?" must also be shown on the display. - ... there must be a limit for selecting the option "Oder mit etwas anderem?" or "Oder mit jemand anderem?". - ... it must be avoided that the same option is suggested twice under	Single Choice	- Strongly disagree - Disagree - Neutral - Agree - Strongly agree

"Wie soll das Sandmännchen heute zu uns kommen?" or "Und wen möchtest du heute sehen?". - ... it must be avoided that when selecting the option "Oder mit etwas anderem?" or "Oder mit jemand anderem?", already suggested options are offered again. - ... our last selections for "Wie soll das Sandmännchen heute zu uns kommen?" or "Und wen möchtest du heute sehen?" must be remembered so that they are not suggested again the next time. - ... our favourite arrival types/Sandmännchen friends must be remembered and prioritised under "Wie soll das Sandmännchen heute zu uns kommen?" or "Und wen möchtest du heute sehen?".		
Final Feedback		
Is there any feedback on the "Abendgruß" skill that you would like to share with us?	Open Textfield	