Huldra: A Framework for Collecting Crowdsourced Feedback on Multimedia Assets

Malek Hammou[1], Cise Midoglu[1], Steven A. Hicks[1], Andrea Storås[1,2], Saeed Shafiee Sabet[1], Inga Strümke[1], Michael A. Riegler[1,3], Pål Halvorsen[1,2]
[1]SimulaMet, Norway
[2]Oslo Metropolitan University, Norway
[3]UIT The Arctic University of Norway, Norway

ABSTRACT

Collecting crowdsourced feedback to evaluate, rank, or score multimedia content can be cumbersome and time-consuming. Most of the existing survey tools are complicated, hard to customize, or tailored for a specific asset type. In this paper, we present an open source framework called Huldra, designed explicitly to address the challenges associated with user studies involving crowdsourced feedback collection. The web-based framework is built in a modular and configurable fashion to allow for the easy adjustment of the user interface (UI) and the multimedia content, while providing integrations with reliable and stable backend solutions to facilitate the collection and analysis of responses. Our proposed framework can be used as an online survey tool by researchers working on different topics such as Machine Learning (ML), audio, image, and video quality assessment, Quality of Experience (QoE), and require user studies for the benchmarking of various types of multimedia content.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in interaction design**; **User studies**; • **General and reference** → **Empirical studies**.

KEYWORDS

crowdsourced feedback, multimedia content, open source, survey, UI, user study, web application

ACM Reference Format:
Malek Hammou[1], Cise Midoglu[1], Steven A. Hicks[1], Andrea Storås[1,2], Saeed Shafiee Sabet[1], Inga Strümke[1], Michael A. Riegler[1,3], Pål Halvorsen[1,2]. 2022. Huldra: A Framework for Collecting Crowdsourced Feedback on Multimedia Assets. In *13th ACM Multimedia Systems Conference (MMSys '22), June 14–17, 2022, Athlone, Ireland*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3524273.3532887

1 INTRODUCTION

Online systems for the generation and delivery of multimedia content have become prominent over the past years, and the trend is likely to keep increasing in the future. User studies are important to ensure the fidelity and usability of such systems that generate multimedia content, as the testing of how certain multimedia content is perceived by end-users and domain experts might provide crucial information, especially in high-risk and high-impact fields such as medicine, law, and industry [19, 25, 30]. Online surveys which allow for conducting user studies in a crowdsourced fashion over the Internet facilitate this process, and enable a large scale with little to no added effort. However, collecting crowdsourced feedback in any form, especially from experts, might be challenging due to several factors [13, 27]. Great care must be taken when designing the necessary survey tools, as researchers are likely to have a single chance to collect the required feedback from experts, and study participants must not be unduly influenced or biased by the survey design.

There are several challenges associated with study participants. **Experts are a sparse resource.** They usually have busy schedules and heavy workloads, and are therefore often unwilling to participate in studies that take a large amount of time, or in the same study more than once. Therefore, it is imperative that the survey framework provides a smooth experience without errors or interruptions, and facilitates fast completion. **Complexity reduces participation.** Faulty or confusing study design can make the experience cumbersome for participants and cause them to change their minds and exit while in the middle of providing feedback. **Increasing participant diversity requires universal accessibility.** Low diversity within the participant group (e.g., with respect to gender, native language, experience, and cultural background) might lead to biased results. Universal accessibility in design is very important to allow for participation from large and diverse groups of people. **Human feedback is subject to bias.** Any user study designed by humans runs the risk of involuntarily influencing the humans who are giving feedback. This is a particularly precarious issue if the study concerns subjective rather than objective feedback. The survey framework should avoid triggering bias in terms of style (e.g., colors and fonts), application flow (e.g., order of questions), and content placement. There are also a number of additional challenges associated with the development of such software frameworks. **Multi-disciplinary competence is necessary.** Developing a robust and adaptable survey framework to fit various scientific needs is time-consuming and requires specific skills such as systems programming or web development, as well as domain expertise. **Commercial tools are not easy to customize or use.** While there are several existing crowdsourcing platforms that provide survey templates and services, such as Microworkers [4] and Amazon Mechanical Turk [12], these frameworks come with challenges of their own. For example, the available templates

MMSys '22, June 14–17, 2022, Athlone, Ireland

ACM ISBN 978-1-4503-9283-9/22/06.
https://doi.org/10.1145/3524273.3532887

are often rudimentary and require extensive adjustments to fit the specific study, and participation itself involves extra effort such as registration on the platform. **Academic institutions might incur additional costs.** Research institutions often do not have the necessary competence in-house for high-end software development, resulting in either additional costs due to outsourcing, or compromises related to the quality of the study. **Privacy protection requires overall control.** The majority of commercial platforms come with potentially substantial privacy issues. This is especially pressing for studies presenting or collecting sensitive information, since the backends of commercial platforms are not available for modification, and they often store data on cloud services selected by the platform providers, instead of the survey designers. **Continuous development is challenging.** It is important that the software framework designed to collect feedback be as close to perfect as possible *before* data collection starts, in order to avoid disruptions in operation as well as maintain neutrality in the aggregation of results, by avoiding potential bias due to version differences. This speaks to the trade-off between development overhead and the quality of results. A framework that is easy to set up in a single iteration greatly improves launch and maintenance performance.

In this paper, we present an open source framework called Huldra, which is designed to address the above listed challenges, and to provide the research community with the necessary software baseline to implement their own survey solutions efficiently. Our focus during development has been on accessibility, modular/reusable component design, support for various multimedia types, and flexibility in configuration. To the best of our knowledge, no open source framework with similar functionality is currently available. The contributions of this paper are two-fold. First, we present the open source framework Huldra which allows for conducting large-scale crowdsourced user studies about multimedia content through a web-based survey. Second, we present 2 active use-cases, where the proposed framework has been customized for: (1) a study on eXplainable Artificial Intelligence (XAI) methods which aim to retrieve feedback from medical experts on the automatic identification of pathological findings in the Gastrointestinal (GI) track, called HOST-XAI, and (2) a study on automatic thumbnail selection for soccer videos, which aims to retrieve feedback from the general public on alternative thumbnail candidates for a given soccer video clip, called HOST-ATS.

The remainder of this paper is structured as follows. We provide background information and an overview of related work in Section 2, followed by a detailed description of our proposed framework in Section 3. We present example use cases of Huldra, including HOST-XAI and HOST-ATS, in Section 4. We discuss the ability of Huldra in addressing the above listed challenges in Section 5. We touch upon aspects related to a live demonstration in Section **??**. We conclude the paper with perspectives for future work in Section 6.

2 BACKGROUND AND RELATED WORK

Subjective assessment of multimedia content has traditionally been conducted in controlled laboratory settings, where invited participants have a physical presence in the laboratory and express their opinion about the stimulus on a given scale. However, laboratory studies have a limited diversity of participants and are often time-consuming and expensive. Crowdsourcing has gained attention during recent years, especially during the COVID-19 pandemic, as an alternative method for collecting feedback. Crowdsourcing platforms usually provide access to a pool of globally distributed users, facilitating possible to access a larger and demographically balanced group of users at lower costs. Participants in crowdsourced studies can take part in experiments remotely from their own working environment, using their own hardware.

Although crowdsourced studies are not conducted in controlled environments unlike laboratory studies, research has shown that if best practices are followed, crowdsourcing can also produce valid and reliable results [16, 28]. The ITU-T P.808 standard [20] provides recommendations on conducting crowdsourced studies, specifically for audio listening, in order to ensure the collection of valid results. The standard recommends using qualification and training jobs in addition to the rating job. In a qualification job, crowdworkers who are eligible to participate in the study are selected, this job could be performing a simple task or answering a number of questions. Crowdsourced studies with and without domain experts are used within the field of biomedicine, where Amazon Mechanical Turk and Crowdflower are the most popular platforms [22]. Studies have in fact shown that crowdsourcing tasks to a large number of non-experts can provide quite accurate results as compared to domain experts [14, 22].

There are a number of available templates and frameworks for conducting crowdsourced studies, such as BeaqleJS [23], a JavaScript-based framework designed for subjective assessment of audio signals, crowdMOS [26], an open source project designed for video and image quality assessment, QualityCrowd [21], a web-based framework designed for video quality assessment, and Crowdgaming [29], a JavaScript-based framework for conducting crowdsourced studies in cloud gaming. In addition, there are a number of crowdsourcing studies that use cloud-based survey alternatives such as Google Forms [8], SurveyMonkey [11], Doodle [7], and similar online services. However, most of these templates and frameworks are difficult to customize for a specific study, as they are designed to assess a certain type of content (e.g., audio, video) in a certain quality scale (e.g,. likert scale, multiple choice). Furthermore, some of the platforms require that the multimedia content is made publicly available (e.g., videos to be uploaded to YouTube), and it is often not clear where the participant data and responses are stored. This makes it challenging to keep control over privacy and security-related aspects, and distribution rights are further complicated by additional legal challenges. Some of the cloud-based alternatives are not well-suited for multimedia content at all, do not support live updates, or do not facilitate the reuse of established template designs with new multimedia content.

Our proposed framework Huldra aims to address these shortcomings. Huldra is easily adjusted to the study of interest, requiring minimal to no programming skills. It is easily replicated and reused as a new survey, or augmented within the same survey, with new content. It allows controlled access to all data, including multimedia assets as well as participant responses, which can be stored publicly or privately. Live updates are supported, which enables agile reconfiguration. Huldra is particularly targeted for studies that require participant feedback on the ranking or benchmarking of

multimedia assets, e.g., asking participants to select the best audio or video clip, or providing feedback in easily configurable discrete (likert scale, multiple choice) and continuous (text field) formats. As a web-based framework, it supports global participation from experts and non-experts alike.

3 FRAMEWORK DESCRIPTION

Huldra is a React-based framework for collecting crowdsourced feedback on various types of multimedia content. It is built in a modular and configurable fashion to allow for the easy adjustment of the User Interface (UI) and the multimedia content, while providing integrations with reliable and stable backend solutions to facilitate the collection and analysis of responses. Huldra provides a series of React components that can be (re-)used to construct a custom survey with a simple UI, and can integrate different types of multimedia assets such as audio, video, and images. Components are grouped into 3 categories (major components, minor components, and utilities), and the overall layout is designed to support 8 types of pages. Several minor components facilitate the display of multimedia content on these pages. The major components make use of the app context to pass data across the component tree. Huldra can be hosted locally or publicly, and supports 3 cloud-based third-party integrations for web deployment, multimedia asset management, and response collection. The framework is designed to facilitate easy setup with minimal programming knowledge. Despite the use of the React library as a basis for the framework, little to no knowledge of this library is needed to build a custom survey. Huldra is provided as an open source software under [3].

3.1 React Library

We implement Huldra using React [10], a popular JavaScript library for building user interfaces, originally developed by Facebook. There are a number of reasons for our choice. **React is lightweight and flexible.** For instance, compared to Angular[1] which provides additional routing functionalities, React only renders UI components. It is very easy to use third-party libraries with React. **React is easy to customize.** It is component-based and declarative, which allows for building complex UIs from encapsulated components and a codebase that is easy to manage, while still providing a lot of features out-of-the-box. For instance, new page templates can be added to the Huldra framework by re-using existing minor components from the codebase. **React is cross-platform compatible.** It can be used across different platforms, as it can render on any server using Node[2] and also power mobile apps using React Native. **React is in widespread use**, with a community of millions of developers, and consequently very well-documented and supported.

3.2 Overall Functionality

Huldra is designed to serve the following functionality.

Registration and Login: Registration functionality for the participants to enter their information and agree to any terms and conditions, in order to get a new participant ID, and login functionality for participants with an existing participant ID to continue the survey from where they left off. The primary goal of this step is to retrieve information *from* the participants. Relevant page types: `Home`, `Registration`.

Introduction: Introduction functionality refers to dedicated pages where the participants are informed about what the study is about, and what they can expect from the rest of the survey. These can include background information, textual instructions, sound and video checks, demonstrations and/or tutorials. The primary goal of this step is to display information *to* the participants. Relevant page types: `Background`, `Demonstration`.

Questionnaire: Questionnaire functionality refers to the series of survey questions, or "case"s, where participants are asked to perform the duties explained in the previous step, such as evaluation, ranking, selection, etc., for various multimedia content. The primary goal of this step is to retrieve information *from* the participants. Relevant page types: `CaseAudio`, `CaseHybrid`, `CaseImage`, `CaseVideo`.

Summary and Feedback: Summary and feedback functionality refers to the presentation of a summary of the survey questions ("case"s), along with a form for collecting participant feedback. The feedback can be about the content of the particular study, and/or the survey framework itself. The primary goal of this step is to retrieve information *from* the participants (in the form of user feedback), but this functionality is supported by the display of information *to* the participants (in the form of a summary), in order for them to be able to provide more informed comments. Relevant page types: `SummaryAndFeedback`.

3.3 Application Flow

Figure 1 presents an overview of the application flow, including the pages that were designed and implemented to support the target functionality described above. Huldra supports 8 types of pages.

Warning Page: This is a warning/error page template that communicates to users that there is a problem with the rendering of the application, and serves as a redirect. There can be multiple warning pages for different error warnings. By default, a single warning page is used. The default warning page is displayed if the resolution of a client browser window is too low, and contains a warning message asking participants to view the survey in a window with a minimum resolution of 1200×800 pixels.

Home Page: This is the landing page under normal operation (i.e., no warnings), once the survey link is clicked. It allows participants to select between logging in with an already registered participant ID, or generating a new participant ID to start the survey anew. There is a single `Home` page. Under normal operation, it is followed by the `Registration` page by default.

Registration Page: This page allows participants to enter required and optional pieces of information for registration, agree to terms and conditions if any, and get a new participant ID. There is a single `Registration` page. Under normal operation, it is followed by the `Background` page by default.

Background Page: This is the first of the 2 types of introductory pages, where the participants can be informed about what the study is about. There is a single `Background` page (with no limits on

[1] Angular is a JavasSript framework built using Typescript, while React is a Javascript library and built using JavaScript XML (JSX). Angular is mostly used to build complex enterprise-grade apps like single-page apps and progressive web apps, while React is used to build UI components in any app with frequently variable data.

[2] Node.js is an open-source, cross-platform, backend JavaScript runtime environment that runs on the V8 engine and executes JavaScript code outside a web browser.

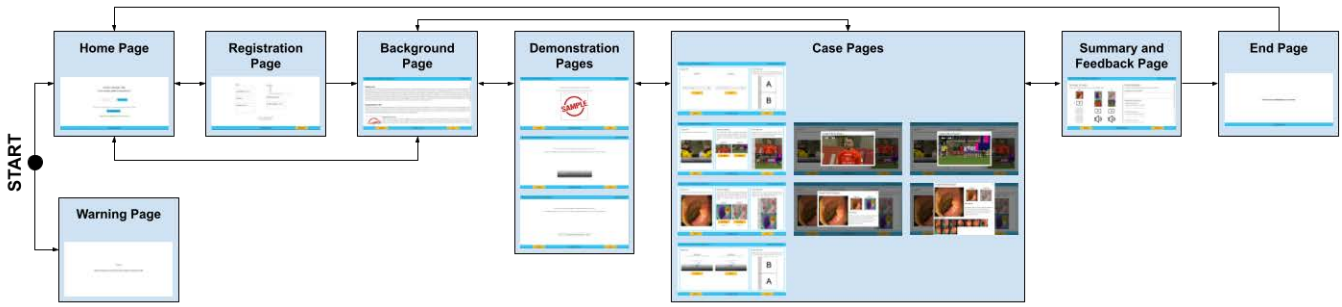


Figure 1: Overview of the `Huldra` application flow. From the Home Page, participants can proceed to either the Registration Page or the Background Page. This is optionally followed by a variable number of Demonstration Page(s), each of which can include an image, video clip, or audio clip. Next are a variable number of Case Pages, each of which can be of 4 different types (Audio Case, Hybrid Case, Image Case, or Video Case), with different layout and popup window options. Once the participants have gone through all Case Pages, they are presented with the Summary and Feedback page. Upon submitting their responses, participants are taken to the End Page, which automatically redirects back to the Home Page after a configurable amount of time.

downscrolling). Under normal operation, it is followed by the (first) `Demonstration` page if one is configured, or by the (first) `Case` page if no `Demonstration` page has been configured.

Demonstration Page(s): This is the second of the 2 types of introductory pages, where the participants can be informed about additional details related to the survey content (e.g., the format of the questions), as well as the application (e.g., what to click and when). These pages can also be used to conduct sound and/or video checks. There can be between 0-3 `Demonstration` pages in a Huldra-based survey. Under normal operation, a `Demonstration` page is followed by the next[3] `Demonstration` page if one is configured, or the (first) `Case` page.

Case Page(s): This page presents a survey question and has 4 alternative formats. There are typically multiple `Case` pages in a survey, Huldra allows for a minimum of 1, and there is no limit on the maximum. Under normal operation, a `Case` page is followed by the next[4] `Case` page until the last case, after which it is followed by the `Feedback and Summary` page.

- **Audio Case:** Targeted for audio assets, this type of `Case` page is structured into 2 columns: the left-middle column for a textual case description and 2 audio assets as answer options, and the right column for displaying the participant answer.
- **Hybrid Case:** Targeted for a mix of video and image assets, this type of `Case` page is structured into 3 columns: the left column for the case description including a single video asset, the middle column for 2 image assets as answer options, and the right column for displaying the participant answer.
- **Image Case:** Targeted for image assets, this type of `Case` page is structured into 3 columns: the left column for the case description including a single image asset, the middle column for 2 image assets as answer options, and the right column for displaying the participant answer.
- **Video Case:** Targeted for video assets, this type of `Case` page is structured into 2 columns: the left-middle column for a textual case description and 2 video assets as answer options, and the right column for displaying the participant answer.

Summary and Feedback Page: This page consists of a left column displaying the summary of the questions and answer options in the survey, and a right column displaying a feedback form to be filled by participants. There is a single `Feedback and Summary` page in a survey. Under normal operation, it is followed by the `End` page, provided that participants give explicit approval via a confirmation dialog box to end the survey and submit their responses.

End Page: This is the final page of the survey, which is displayed to let the participants know that the survey has ended, and that their responses have been submitted. There is a single `End` page in a survey. Under normal operation, this page automatically redirects to the `Home` page after a configurable amount of time.

All pages after the `Registration` page are subject to route protection[5]. In order to ensure complete responses overall, Huldra makes it mandatory for participants to submit their responses to a `Case` page before proceeding to the next `Case` page.[6] All pages are configurable via the global configuration file `config.json`.

3.4 Third-Party Integrations

Huldra uses 3 third-party integrations.

Google Cloud Platform (GCP) is a suite of cloud computing services that runs on Google infrastructure [5]. Alongside a set of management tools, it provides a series of modular cloud services

[3]Demonstration pages appear in the order specified in the `config.json` file. This fixed order applies to all participants.

[4]Case pages can be completely shuffled, shuffled per asset category (image cases shuffled within themselves, followed by video cases shuffled within themselves, followed by audio cases shuffled within themselves, followed by hybrid cases shuffled within themselves), or displayed in a particular fixed order, depending on the specification in the `config.json` file. Fixed order applies to all participants, whereas shuffling is undertaken at the time of participant ID generation individually for each participant, therefore yielding a random personalized case order for each participant, ensuring that each participant will see the same cases overall but with a different order.

[5]Protected routes are routes that can only be accessed if a condition is met. Huldra only allows browser sessions with a valid participant ID to proceed to protected routes.

[6]Participants can later go backwards to revisit and/or change their answers.

including computing, data storage, data analytics and machine learning. HULDRA uses a GCP S3 bucket for cloud storage, both for storing the multimedia assets and the participant responses. Access to the bucket is maintained through Firebase integration, where credentials can be specified as environment variables in the Heroku configuration (configuration of access to the storage bucket can also be undertaken using the `config.json` file in the codebase, but this is not advisable due to privacy reasons, unless a public bucket is used on purpose).

Firebase is a platform for creating mobile and web applications [6]. Used in connection with the GCP S3 bucket, Firebase allows HULDRA to fetch the multimedia assets to be used in the study seamlessly from, and write participant responses to, configurable locations in this bucket. Details about the folder structure and file naming conventions expected by the default HULDRA implementation are provided in the open GitHub repository [3].

HULDRA uses **Heroku** as a cloud Platform as a Service (PaaS) for web deployment, which supports the triggering of automatic deployments from a GitHub repository. Applications deployed on the web via Heroku can be accessed via a URL of the form: `https://<application-name>.herokuapp.com`.

All of the HULDRA third-party integration functionalities can be used with free plans as well as paid plans: minimum requirements are a free Google account for GCP and Firebase, and free hobby dynos for a Heroku personal application.

3.5 Configuration and Outputs

HULDRA incorporates a multitude of configuration parameters, which allow survey designers to configure everything from the UI (e.g., page layout), to content (e.g., multimedia assets and text), to file operations and storage (e.g., the structure of the participant response files). Configuration parameters are organized into dictionaries per page type. The values for all configuration parameters can be specified in the `config.json` file, or set as environment variables. The full list of parameters can be found in the HULDRA open GitHub repository [3].

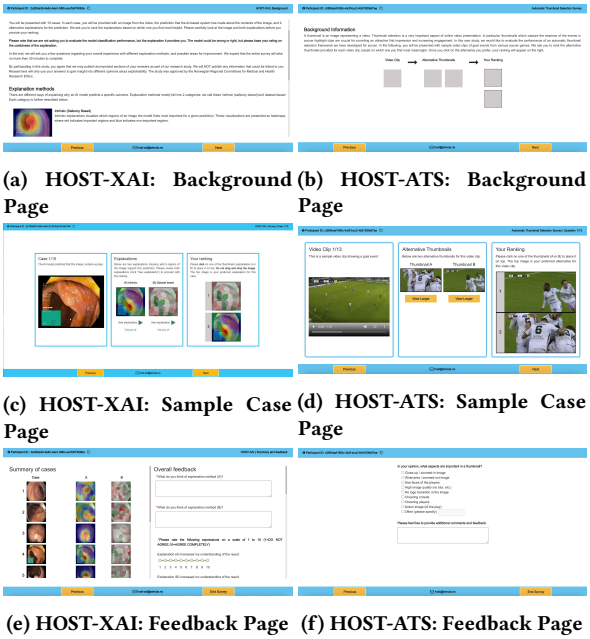
HULDRA uses browser session storage to track participant responses locally[7]. At certain points in the survey (e.g., clicking of the "Next" button) information from the session storage is written to the cloud bucket, where responses from each participant are stored in separate JavaScript Object Notation (JSON) files, identifiable by the participant ID. File structure, dictionary, field names, storage location, etc. are all configurable.

3.6 Instructions to Use the Codebase

The HULDRA open source software repository can be found under [3]. Assuming that the third-party service accounts are already established, the below steps can be used to run a HULDRA instance:

1. **Assets:** Set up the necessary folder structure in the S3 bucket, prepare and upload the multimedia assets corresponding to your desired "case"s.
2. **Codebase:** Clone HULDRA repository.

[7]Session storage enables apps to store data on a browser, different values can then be saved and retrieved. Unlike local storage, session storage only keeps data for a particular session. The data is cleared once the user closes their browser window.



(a) HOST-XAI: Background Page **(b) HOST-ATS: Background Page**

(c) HOST-XAI: Sample Case Page **(d) HOST-ATS: Sample Case Page**

(e) HOST-XAI: Feedback Page **(f) HOST-ATS: Feedback Page**

Figure 2: Sample pages from 2 actively deployed HULDRA use cases.

3. **Configuration:** Update configuration parameters in the `config.json` file as needed, to customize your instance.

4a. **Deployment (Heroku):** Enter the Firebase connection parameters in the Heroku configuration for your app. Deploy the relevant branch of your repository. The survey will be accessible at `https://<application-name>.herokuapp.com` by default.

4b. **Deployment (local):** Enter the Firebase connection parameters in your local environment variables. Run `npm install` and `npm start` to start your local server. The survey will be accessible at `localhost:3000` by default.

5 **Outputs:** Retrieve participant response files from the bucket at your convenience.

4 USE CASES

In this section, we present actively deployed and potential use cases for the HULDRA framework, in order to exemplify how it can apply to real-world scenarios. Figure 2 presents screenshots of sample pages from the 2 actively deployed use cases: HOST-XAI and HOST-ATS.

4.1 Image Use Case: HOST-XAI

HOST-XAI [15] is a study of XAI methods, which aims to retrieve feedback from medical experts on Artificial Intelligence (AI)-based automatic identification of pathological findings in the GI track. More specifically, medical experts are asked to compare different XAI methods, which explain the predictions of a Convolutional Neural Network (CNN) used to find abnormalities in the GI tract. The goal is to find out which XAI method is preferred by doctors as an understandable and trust-building visual explanation of the image

classifications: *intrinsic* (saliency-based) or *extrinsic* (dataset-based). The HOST-XAI online survey, deployed as a publicly accessible Heroku application under [1] is a customized version of Huldra, generated only by changing the `config.json` file from the Huldra codebase [3]. For instance, HOST-XAI has been configured to display information about the 2 families of XAI methods (saliency based and dataset based), as well as a category of relevant pathological findings in the GI tract (polyps) in the Background Page (Figure 2a). It uses the `Image Case` type of pages for the questionnaire (Figure 2c). A tutorial on how to participate in the HOST-XAI survey can be found under [2].

4.2 Hybrid Use Case: HOST-ATS

HOST-ATS [17, 18] is a study of automatic thumbnail selection for soccer videos, which aims to retrieve feedback from the general public on alternative thumbnail candidates for a given soccer video clip. In each question, participants are asked to compare thumbnail proposals from 2 different thumbnail selection methods (different pairwise combinations of: manual, automatic, static). Participant responses allow for the benchmarking of alternative thumbnail selection methods, as well as provide deeper and more general insights into viewer expectations from thumbnails. The HOST-ATS online survey, deployed as a publicly accessible Heroku application under [9] is a customized version of Huldra, generated only by changing the `config.json` file from the Huldra codebase [3]. For instance, HOST-ATS has been configured to present a checklist of various aspects viewers might deem important in a thumbnail, as well as a free form input field for additional feedback and suggestions, in the `Summary and Feedback Page` (Figure 2f). It uses the `Hybrid Case` type of pages for the questionnaire (Figure 2d), with 1 video and 2 image assets per question.

4.3 Audio Use Cases

A potential audio use case for Huldra is a user study to compare the subjective quality of various audio codecs. Using `Audio Case` type of pages, participants can be asked to listen to 2 alternative audio clips which are encoded with different codecs, in each question, and pick the "best" according to their personal opinion. As disinformation campaigns are increasingly powered by advanced AI techniques, and a lot of effort is put into the detection of fake content by researchers, an audio clip comparison survey can also be helpful as a part of an investigation into fake multimedia detection. For instance, a user study asking participants to compare audio clips and to pick which was real and which was artificially generated, can yield insights about how to combat the spread of fake multimedia in society.

4.4 Video Use Cases

Similar to the audio use cases described above, Huldra has various video use cases. These include video codec benchmarking, fake multimedia detection, and video clip comparisons. For instance, evaluation of the submissions for the "event clipping" task[8] of the MMSys'22 Grand Challenge on AI-based Video Production for Soccer [24] could easily be undertaken using a deployment based on Huldra.

5 DISCUSSION

With Huldra, we have tried to address the shortcomings listed in Section 1 in a number of ways. Huldra: (1) uses a simple, easy to follow interface to reduce complexity, (2) facilitates easy completion by experts and non-experts alike, including people who are not familiar with advanced software, so that each participant can complete the survey in a single attempt, not exacerbating the scarcity of time, (3) strives to remove potential human bias by, e.g., shuffling case orders for each participant, and allowing for an overview of all cases at the end of the survey, so that participants can see if they have changed their mind, (4) is highly configurable, and supports different types of user input (multiple choice, likert, free form, radio button, tickbox, etc.), (5) supports privacy protection through various schemes related to the handling of multimedia assets and participant responses, including bucket authentication.

We provide Huldra as an open source framework to help the research community against increased commercialization. It can easily be integrated with existing crowdsourcing platforms to recruit crowdworkers. For instance, the link to a Huldra-based web application can be distributed to crowdworkers in platforms such as MTurk, and they can be asked to simply copy their Huldra participant ID in their MTurk submission.

One of the biggest shortcomings of Huldra is its current dependency on specific third-party integrations, namely GCP, Firebase, and Heroku. We plan to alleviate this limitation by extending asset management and survey deployment functionalities in two directions: (1) supporting a fully local solution including local storage and self-hosting, and (2) supporting integrations with additional cloud service providers (e.g., Amazon Web Services) to serve as alternatives. In its current form, Huldra can already be deployed locally (without Heroku) for testing[9] as well as production[10].

6 CONCLUSIONS AND FUTURE WORK

In this paper, we present Huldra, an open source framework which can be used to design web-based surveys for large-scale crowdsourced feedback collection on multimedia assets. The framework is built using the React library in a fully configurable fashion, and provides various reusable components for future extensions. We present sample use cases of Huldra which also serve as guidelines for future customizations of the basic version. We hope that this framework will be useful for researchers from various fields to conduct user studies efficiently and in a less time-consuming manner. As future work, we plan to increase the number of service and utility components to make Huldra even more configurable, support alternative storage and hosting solutions to reduce its dependency on specific third-party integrations and make the framework simpler to deploy, support multi-asset ranking and more (heterogeneous) multimedia modalities, as well as provide an out-of-the box functionality for the automatic analysis of survey results.

[8] Task 1 of the grand challenge focuses on the automatic generation of short highlight clips from full soccer game videos, based on important events such as goals, bookings and substitutions.

[9] https://github.com/simula/huldra#4b-deployment-local

[10] https://create-react-app.dev/docs/deployment

REFERENCES

[1] 2021. HOST-XAI. https://host-xai.herokuapp.com.
[2] 2021. HOST-XAI Tutorial. https://www.youtube.com/watch?v=JJ8uc5gReko/.
[3] 2021. Huldra. https://github.com/simula/huldra.
[4] 2021. Templates - Microworkers - work & earn or offer a micro job. https://www.microworkers.com/.
[5] 2022. Cloud Computing Services - Google Cloud. https://cloud.google.com/.
[6] 2022. Firebase. https://firebase.google.com/.
[7] 2022. Free online meeting scheduling tool - Doodle. https://doodle.com/.
[8] 2022. Google Forms. https://docs.google.com/forms.
[9] 2022. HOST-ATS. https://host-ats.herokuapp.com.
[10] 2022. React - A JavaScript library for building user interfaces. https://reactjs.org/.
[11] 2022. SurveyMonkey - Free online survey software and questionnaire tool. https://www.surveymonkey.com/.
[12] Amazon. 2021. *Amazon Mechanical Turk.* https://www.mturk.com
[13] Sandy J. J. Gould, Anna L. Cox, and Duncan P. Brumby. 2018. *Influencing and Measuring Behaviour in Crowdsourced Activities.* Springer International Publishing, Cham, 103–130. https://doi.org/10.1007/978-3-319-73374-6_7
[14] Danna Gurari, Diane Theriault, Mehrnoosh Sameki, Brett Isenberg, Tuan A. Pham, Alberto Purwada, Patricia Solski, Matthew Walker, Chentian Zhang, Joyce Y. Wong, and Margrit Betke. 2015. How to Collect Segmentations for Biomedical Images? A Benchmark Evaluating the Performance of Experts, Crowdsourced Non-experts, and Algorithms. In *2015 IEEE Winter Conference on Applications of Computer Vision*. 1169–1176. https://doi.org/10.1109/WACV.2015.160
[15] Steven Hicks, Andrea Storås, Michael Riegler, Cise Midoglu, Malek Hammou, Thomas de Lange, Sravanthi Parasa, Pål Halvorsen, and Inga Strümke. 2022. Visual explanations for polyp detection: How medical doctors assess intrinsic versus extrinsic explanations. https://doi.org/10.48550/ARXIV.2204.00617
[16] Tobias Hoßfeld, Matthias Hirth, Judith Redi, Filippo Mazza, Pavel Korshunov, Babak Naderi, Michael Seufert, Bruno Gardlo, Sebastian Egger, and Christian Keimel. 2014. Best Practices and Recommendations for Crowdsourced QoE-Lessons learned from the Qualinet Task Force" Crowdsourcing". (2014).
[17] Andreas Husa, Cise Midoglu, Malek Hammou, Pål Halvorsen, and Michael A. Riegler. 2022. HOST-ATS: Automatic Thumbnail Selection with Dashboard-Controlled ML Pipeline and Dynamic User Survey. In *13th ACM Multimedia Systems Conference (MMSys '22), June 14–17, 2022, Athlone, Ireland.* ACM, New York, NY, USA. https://doi.org/10.1145/3524273.3532908
[18] Andreas Husa, Cise Midoglu, Malek Hammou, Steven A. Hicks, Dag Johansen, Tomas Kupka, Michael A. Riegler, and Pål Halvorsen. 2022. Automatic Thumbnail Selection for Soccer Videos using Machine Learning. In *13th ACM Multimedia Systems Conference (MMSys '22), June 14–17, 2022, Athlone, Ireland.* ACM, New York, NY, USA. https://doi.org/10.1145/3524273.3528182
[19] O. Iida, M. Urakami, and T. Iwamura. 1993. Applications and evaluation of AI technology in the steel industry. In *Proceedings of IEEE 2nd International Workshop on Emerging Technologies and Factory Automation (ETFA '93).* 156–163. https://doi.org/10.1109/ETFA.1993.396416
[20] ITU-T Recommendation P.808. 2018. *Subjective evaluation of speech quality with a crowdsourcing approach.* International Telecommunication Union, Geneva.
[21] Christian Keimel, Julian Habigt, Clemens Horch, and Klaus Diepold. 2012. QualityCrowd — A framework for crowd-based quality evaluation. In *2012 Picture Coding Symposium.* 245–248. https://doi.org/10.1109/PCS.2012.6213338
[22] Ritu Khare, Benjamin M. Good, Robert Leaman, Andrew I. Su, and Zhiyong Lu. 2016. Crowdsourcing in biomedicine: challenges and opportunities. *Briefings in Bioinformatics* 17, 1 (01 2016), 23–32. https://doi.org/10.1093/bib/bbv021 arXiv:https://academic.oup.com/bib/article-pdf/17/1/23/6684984/bbv021.pdf
[23] Sebastian Kraft and Udo Zölzer. 2014. BeaqleJS: HTML5 and JavaScript based Framework for the Subjective Evaluation of Audio Quality.
[24] Cise Midoglu, Steven A. Hicks, Vajira Thambawita, Tomas Kupka, and Pål Halvorsen. 2022. MMSys'22 Grand Challenge on AI-based Video Production for Soccer. arXiv:2202.01031 [cs.CV]
[25] Perry L Miller. 1986. The evaluation of artificial intelligence systems in medicine. *Computer methods and programs in biomedicine* 22, 1 (1986), 3–11.
[26] Flavio Protasio Ribeiro, Dinei Florencio, Cha Zhang, and Mike Seltzer. 2011. CROWDMOS: An Approach for Crowdsourcing Mean Opinion Score Studies. In *ICASSP* (icassp ed.). IEEE. https://www.microsoft.com/en-us/research/publication/crowdmos-an-approach-for-crowdsourcing-mean-opinion-score-studies/
[27] Michael Riegler, Vamsidhar Reddy Gaddam, Martha Larson, Ragnhild Eg, Pål Halvorsen, and Carsten Griwodz. 2016. Crowdsourcing as self-fulfilling prophecy: Influence of discarding workers in subjective assessment tasks. In *2016 14th International Workshop on Content-Based Multimedia Indexing (CBMI).* 1–6. https://doi.org/10.1109/CBMI.2016.7500256
[28] Dietmar Saupe, Franz Hahn, Vlad Hosu, Igor Zingman, Masud Rana, and Shujun Li. 2016. Crowd workers proven useful: A comparative study of subjective video quality assessment. In *QoMEX 2016: 8th International Conference on Quality of Multimedia Experience.*
[29] Steven Schmidt, Babak Naderi, Saeed Shafiee Sabet, Saman Zadtootaghaj, and Sebastian Möller. 2020. Assessing Interactive Gaming Quality of Experience Using a Crowdsourcing Approach. In *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX).* IEEE, 1–6.
[30] Jeremy Wyatt and David Spiegelhalter. 1991. Evaluating Medical Expert Systems: What To Test, And How ?. In *Knowledge Based Systems in Medicine: Methods, Applications and Evaluation*, Jan L. Talmon and John Fox (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 274–290.