Usage Agreement

for the MediaEval 2025 Research Collections

Please fill out this form and return it following the instructions that are given at the bottom of the last page. On page 1 (this page), mark the box next to the task or tasks for which you have registered. Then fill out page 2 with your team information. Sign on page 3, and then proceed to also sign any task-specific agreements related to the task or tasks for which you have registered.

Note: Please return one form per team, unless the team is composed of people from more than one organization. In that case, each organization (i.e., university or company) in the team should sign a separate form. We request that these multi-organization teams designate one person to collect and submit all forms from the team in a single email.

[ ] **Medico: VQA (with multimodal explanations) for gastrointestinal imaging:** The goal is to use Visual Question Answering (VQA) to interpret and answer questions based on gastrointestinal images, aiming to enhance decision support and improve AI-driven medical decision-making. We provide a gastrointestinal dataset containing images and videos with VQA labels and additional metadata.

[ ] **Memorability: Predicting movie and commercial memorability:** The goal of this task is to study the long-term memorability performance when recognising movie excerpts or commercial videos . We provide the videos, precomputed features or EEG features for the challenges proposed in the task such as *Will a person remember a video they have seen in the past? or Can you predict the brand memorability?*

[ ] **MultiSumm: Multimodal summarization of multiple topically related websites:** Participants are provided with multimodal web content from several cities listing food sharing initiatives (FSIs) in each city. For each city, participants are tasked with creating a multimodal summary of the FSI activities in the city which satisfy specified criteria. Evaluation will explore the use of emerging LLMs-based methods in automated assessment of multimodal multi-document summarization.

[ ] **NewsImages: Retrieval and generative AI for news thumbnails:** Participants receive a large set of articles (including the headline and article lead) in the English-language from international publishers. We offer two subtasks: retrieving an image for each article from a collection of images that can serve as a thumbnail, or generating an article thumbnail.

[ ] **Synthetic Images: Advancing detection of generative AI used in real-world online images:** The goal of this challenge is to develop AI models capable of detecting synthetic images and identifying the specific regions in the images that have been manipulated or synthesized. Approaches will be tested on images synthesized with state-of-the-art approaches and collected from real-world settings online.

Please follow these directions to submit this form:

- *Print, sign, and scan the whole form into a single .pdf file*
- *Please remember to sign both page 4 and also the appropriate task-specific sections (following pages).*
- ***Please name the file <teamname>_ME2025UA.pdf (add your organization name at the end of the filename for multi-organization teams)***
- *Return the form as an attachment to agree@mediaeven.org (do not use this email for any other purpose)*
- ***Give your email the subject line: <teamname> ME2025UA***

Team name used in MediaEval 2025 (as specified during registration): _____
Please note that it is important to provide the team name so that we are able to easily identify your team in the registration system. Thank you.

The _____ (the name of your organization, further referred to as "Organization") engages in research and development work in information retrieval, multimedia processing, music analysis, speech recognition or related areas.

Official mailing address: _____

_____

Telephone: _____

Contact person: _____

E-mail: _____

The Organization agrees to use the multimedia content and associated data including extracted features, automatically generated metadata, manually generated metadata, sensor readings, social metadata, and speech recognition transcripts (the "Information") under the following understandings, terms, and conditions. These understandings, terms, and conditions apply equally to all or to part of the Information, including any updates or new versions of the Information supplied under this agreement.

Copyright

1. This clause (points 1-4) applies to tasks that crawl content from the Internet. For content that is associated with a Creative Commons (cf. http://creativecommons.org) license, every possible measure has been taken to ensure that the association with a Creative Commons license is a valid one. However, the MediaEval 2025 organizers cannot fully guarantee that these collections contain absolutely no content without a Creative Commons license. Such content could potentially enter the collection if it was not correctly marked on the site from which it was collected.

2. The MediaEval 2025 organizers declare any metadata contained in the Information has been at some time made publicly available on the Internet.

3. Owners of copyright for elements contained in the Information may choose to request deletion of these elements from the Information.

4. The limitation on permitted use contained in the following section is intended to reduce the risk of any action being brought by copyright owners, but if this happens the Organization agrees to bear all associated liability.

Permitted Uses

1. The Information may only be used for research and development of multimedia and information retrieval systems.

2. Summaries, analyses and interpretations of the linguistic properties of the Information may be derived and published, provided it is not possible to reconstruct the Information from these summaries.

3. Small excerpts of the Information may be displayed to others or published in a scientific or technical context, solely for the purpose of describing the research and development carried out and related issues. The name of the Information's owner must be clearly identified in writing at the time of disclosure of the Information and/or in publication. In the case of the Creative Commons data, the "licensor" (cf. http://creativecommons.org/licenses) must be acknowledged.

Own Assessment of Information Use
The Organization must make its own assessment of the suitability of the Information for its research and development purposes under Permitted Uses.

The MediaEval 2025 organizers do not make any warranties or representations, whether expressed or implied or statutory, of any kind with respect to their Information, including without limitation:

1. that the Information is suitable for any particular purpose
2. regarding the results of any use of the whole or part of the Information
3. as to the accuracy, reliability or content of the Information
4. of the absence of any infringement of any proprietary right (including, without limitation, IPRs, trade secret rights and right over confidential information) of third parties by the use of such Information

The Organization shall in any case bear the entire risk of any consequences that may arise from the use to which it, or to which any person that it directly or indirectly permits or allows to use such Information, puts such Information.

The Information's owner shall not have any liability in respect of the infringement of any patent or other right of any third party resulting from any other Organization exercising any of the permitted uses granted under this agreement.

No Information owner makes any representation or warranty, express or implied, other than as expressly stated in this Agreement.

The Organization agrees and acknowledges that the Information's owners shall not be held responsible, alone or jointly and severally, for any loss, damage or injury resulting from the use made by the Organization of their respective Information.

Agreement to Delete Data on Request

The Organization undertakes to delete within thirty days of receiving notice all copies of any named document that is part of the Information whenever requested to do so by any one of:

1. The MediaEval Organizers
2. the owner of copyright for a particular element

Access to the Information by Individuals:

The Organization:

1. must control access to the Information by individuals and may only grant access to people working under its control, i.e., its own members, consultants to the Organization, or individuals providing service to the Organization.
2. remains responsible for any breach of this access restriction by individuals under its control.

Termination

Either party may terminate the Agreement at any time by notifying the other party in writing. On termination, the Organization must a) destroy all copies of the Information and b) notify the MediaEval 2025 organizers in writing of the action taken.

Applicable Law This Agreement is governed by the laws of the Netherlands. Signed by the Organization:

Signature: _____ Date:_____

Name (please print): _____

Position/Organizational Role: _____

E-mail _____
(if different from contact person above)

Medico: VQA (with multimodal explanations) for gastrointestinal imaging

(Data: Kvasir-VQA (https://huggingface.co/datasets/SimulaMet-HOST/Kvasir-VQA), Hyper-Kvasir (https://datasets.simula.no/hyper-kvasir)

The data is free to use for academic purposes. No patient-identifiable data is included. All study participants gave consent to use the data for research. Data is anonymized following the Norwegian and European data protection regulations (fully anonymized, no key lists are available).

Any use of the data will need to cite the following papers:

Sushant Gautam, Andrea M. Storås, Cise Midoglu, Steven A. Hicks, Vajira Thambawita, Pål Halvorsen, and Michael A. Riegler. 2024. Kvasir-VQA: A Text-Image Pair GI Tract Dataset. In Proceedings of the First International Workshop on Vision-Language Models for Biomedical Applications (VLM4Bio'24). Association for Computing Machinery, New York, NY, USA, 3–12. https://doi.org/10.1145/3689096.3689458

Borgli, H., Thambawita, V., Smedsrud, P.H. et al. HyperKvasir, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy. Sci Data 7, 283 (2020). https://doi.org/10.1038/s41597-020-00622-y

Signature_____

(sign here if participating in the **Medico: VQA (with multimodal explanations) for gastrointestinal imaging** to indicate you have read and accepted the task specific conditions)

NewsImages: Retrieval and generative AI for news thumbnails
(Data: We have collected a set of news articles from GDELT. The dataset contains both textual and visual information in the form of images.)

The participating organization guarantees to uphold and comply with the MediaEval terms that restrict the usage of data for research purposes. The data must not be used commercially. The organization confirms that they will delete instances (article texts or images) upon request by the copyright holder. The organization acknowledges that data must not be shared with third parties or non-registered users. The organization will respect the naming rights of copyright holders. In particular, the data coming from GDELT can be used when naming the data source.

By participating in this task, you grant the NewsImages and MediaEval organizers a non-exclusive, non-transferable license to use any of the submitted material. You also confirm that any content generation was done in accordance with the license and usage restriction of all the tools and models used as part of your workflow.

Signature_____

(sign here if participating in the **NewsImages** to indicate you have read and accepted the task specific conditions)

Memorability: Predicting movie and commercial memorability:
(Data: Movie Memorability dataset and VIDEM dataset. The Movie Memorability dataset contains video excerpts together with human scores of memorability. EEG signals dataset. The EEG signals dataset contains EEG signals (preprocessed and sets of features) from participants watching videoclips from the Movie Memorability dataset together with a label reflecting whether the video was successfully recalled in memory or not.)

The Movie Memorability dataset (https://www.interdigital.com/data_sets/movie-memorability-dataset) is publicly available under an R&D license. Part of the dataset was derived from movie excerpts extracted from 100 Hollywood-like movies. Non-commercial entities are granted access to this part of the dataset under the herein license. The use of such excerpt for any other use than research and/or the redistribution to any third party of such excerpt is strictly prohibited.

The VIDEM (VIDeo Effectiveness and Memorability) dataset consists of commercial videos annotated with long-term video and brand memorability scores and is accompanied by metadata. The dataset is developed by the University of Essex and Hub Productions with support from Innovate UK's Knowledge Transfer Partnership under grant agreement No. 11071 and provided strictly for non-commercial research use only.

Any use of the Movie Memorability dataset will be accompanied by the citation of the following paper:

R. Cohendet, K. Yadati, N. Q. Duong and C.-H. Demarty. Annotating, understanding, and predicting long-term video memorability. In Proceedings of the ICMR 2018 Conference, Yokohama, Japan, June 11-14, 2018.

You understand that the video media used for this task must be used in accordance with the terms of release of the original dataset.

By downloading the EEG signals dataset, you agree to the following terms:

1. You will use the data only for non-commercial research and educational purposes.

2. You will NOT distribute the Datasets or any parts thereof, nor copy any of the images, videos, tags or text onto a public site or social media of any kind.

3. Contributors to the dataset make no representations or warranties regarding the dataset, including but not limited to warranties of non-infringement or fitness for a particular purpose.

4. You accept full responsibility for your use of the datasets and accept all liability and risks associated with its use.

5. This data comes with no warranty or guarantee of any kind, and you accept full liability.

Features: The dataset for this task might be accompanied by automatically extracted features from the EEG recordings to facilitate use by those without a background in signal processing and/or dealing with neural recordings. These features must be used in compliance with the usage conditions set out in the main usage agreement (above). Features are provided on an as-is basis with no guarantee of any kind.

The VIDEM dataset https://github.com/multimediaeval/2025-Predicting-Media-Memorability-Task) is made available under an R&D License for non-commercial research purposes only. Redistribution of the dataset or any portion thereof to third parties is strictly prohibited. By downloading and using the dataset (including videos, metadata, extracted features, and associated materials), you agree to the following terms:

1. You will use the data only for non-commercial research and educational purposes.

2. You will NOT distribute the dataset or any of its parts (videos, metadata, features, or annotations) to any third party or publicly share them in any form, including online platforms or social media.

3. Hub Productions and the contributors to the VIDEM dataset provide the dataset "as is" and make no representations or warranties of any kind, including, but not limited to, warranties of accuracy, non-infringement, or fitness for a particular purpose.

4. You accept full responsibility for your use of the datasets and accept all liability and risks associated with its use, including but not limited to your use of any copies of copyrighted videos or images that you may create from the datasets.

5. You will treat people and animals appearing in this data with respect and dignity.

6. This data comes with no warranty or guarantee of any kind, and you accept full liability.

7. Cite our paper if you use our dataset: Kiziltepe, R.S., Sahab, S., Valladares Santana, R., Doctor, F., Paterson, K., Hunstone, D., & García Seco de Herrera, A. (2025). VIDEM: VIDeo Effectiveness and Memorability Dataset. (Under review).

Features: The dataset for this task might be accompanied by automatically extracted low-level features. These features must be used in compliance with the usage conditions set out in the main usage agreement (above). Features are provided on an as-is basis with no guarantee of any kind.

Signature_____

(sign here if participating in the **Memorability: Predicting movie and commercial memorability** to indicate you have read and accepted the task specific conditions)

MultiSumm: Multimodal summarization of multiple topically related websites

(*Data: URLs of the websites from the target cities)*

No additional specific terms. Please stay in close communication with the task organizers.

Signature_____

(sign here if participating in the **MultiSumm: Multimodal summarization of multiple topically related websites** to indicate you have read and accepted the task specific conditions)

Synthetic Images: Advancing detection of generative AI used in real-world online images

The participating organization hereby guarantees to uphold and comply with the MediaEval terms and conditions, which restrict the use of the provided data strictly for research purposes. The organization affirms that:

- **Commercial Use Prohibited**: The data will not be used for any commercial purposes.
- **Deletion Upon Request**: Any instance of the data (including article texts or images) will be deleted upon request by the copyright holder.
- **No Redistribution**: The data will not be shared with third parties or individuals who are not registered users under the MediaEval terms.

Furthermore, any use of the data in research outputs, publications, or presentations will properly cite the following papers:

For **subtask A (**Synthetic Image Detection)

- Guillaro, F., Zingarini, G., Usman, B., Sud, A., Cozzolino, D., & Verdoliva, L. (2024). A Bias-Free Training Paradigm for More General AI-generated Image Detection. arXiv preprint arXiv:2412.17671.
- Bammey, Q. (2023). Synthbuster: Towards detection of diffusion model generated images. *IEEE Open Journal of Signal Processing*, *5*, 1-9.

For **subtask B** (Manipulated Region Localization)

- Mareen, H., Karageorgiou, D., Van Wallendael, G., Lambert, P., & Papadopoulos, S. (2024, December). TGIF: Text-guided inpainting forgery dataset. In 2024 IEEE International Workshop on Information Forensics and Security (WIFS) (pp. 1–6). IEEE.
- Giakoumoglou, P., Karageorgiou, D., Papadopoulos, S., & Petrantonakis, P. C. (2025). A Large-scale AI-generated Image Inpainting Benchmark. arXiv preprint arXiv:2502.06593.

Signature_____

(sign here if participating in the **Synthetic Images: Advancing detection of generative AI used in real-world online images** to indicate you have read and accepted the task specific conditions)