Overview of The MediaEval 2026 Predicting Movie and Commercial Memorability Task

Iván Martín-Fernández[1], Aashutosh Ganesh[2], Mihai Gabriel Constantin[3], Claire-Hélène Demarty[4], Manuel Gil-Martín[1], Sebastian Halder[5], Bogdan Ionescu[3], Rukiye Savran Kiziltepe[6], Ana Matran-Fernandez[5] and Alba G. Seco de Herrera[7]

[1]*Grupo de Tecnología del Habla y Aprendizaje Automático (THAU), Information Processing and Telecommunications Center, E.T.S.I. de Telecomunicación, Universidad Politécnica de Madrid (UPM), Madrid, 28040, Spain*

[2]*Department of Advanced Computing Sciences, Maastricht University, The Netherlands*

[3]*AI Multimedia Lab, National University of Science and Technology Politehnica Bucharest, Bucharest, Romania*

[4]*InterDigital, R&I France, 35510 Cesson-Sévigné, France*

[5]*School of Computer Science and Electronic Engineering, University of Essex, Colchester, CO4 3SQ, United Kingdom*

[6]*Department of Software Engineering, Ankara University, Golbasi, Ankara, 06830, Ankara, Türkiye*

[7]*ETSI Informática, UNED, Madrid, Spain*

Abstract

This paper presents the 2026 edition of the MediaEval Predicting Movie and Commercial Memorability Task. Much like the previous edition, the benchmark is structured around four different subtasks: (1.1) predicting long-term memorability using a movie clip, (1.2) classifying successful recognition using participants EEG signals, (2.1) predicting long-term commercial video memorability and (2.2) predicting the brand memorability associated with a video clip. This edition builds upon previous years, introducing an expanded feature set that integrates colour palettes and movie metadata to enable modelling strategies that go beyond video characteristics.

1. Introduction

This paper presents the 2026 edition of the MediaEval Predicting Movie and Commercial Memorability Task. Media memorability is defined as the likelihood that a piece of multimedia, such as images and videos, will be recalled after a period of time. Approaching this concept from a computational standpoint improves our understanding of how humans perceive stimuli and informs potential applications in content creation, marketing, and media literacy. The successful 2025 edition of the challenge shifted the focus from short, out-of-context videos to longer formats in the film and advertising industries [1]. This iteration builds upon lessons provided by the previous iteration and further explores the intrinsic factors that make movie clips and promotional material, such as advertising and commercials, memorable. The 2026 edition intentionally retains the 2025 benchmarks to establish a longitudinal baseline for long-form media analysis. By maintaining the same four sub-tasks and datasets, this iteration facilitates a direct comparative analysis of algorithmic evolution. The main novelty of the 2026 edition is the expanded feature set, which now includes movie metadata and colour palettes. This enables studies on how aesthetic decisions by creative studios and aspects beyond the visual information of the video itself contribute to the memorability of the underlying media. This paper details the tasks, the datasets and the evaluation methodology.

MediaEval'26: Multimedia Evaluation Workshop, June 15–16, 2026, Amsterdam, Netherlands and Online
ivan.martinf@upm.es (I. Martín-Fernández)
0009-0004-2769-9752 (I. Martín-Fernández)

CEUR Workshop Proceedings (CEUR-WS.org)

2. Related Work

The interaction between audiovisual stimuli and human memory has been largely studied, both from the neuroscience and psychology fields and from the computer science and artificial intelligence standpoints. Earlier works explored human visual memory, its consistency across observers and its variability with the intrinsic characteristics of the provided stimuli [2, 3, 4, 5]. Recently, the research community has converged by developing memorability systems based on the interaction of semantic content and high level visual attributes through Transformer-based modality encoders and generative approaches, with special attention to multimodal large language models [6, 7, 8, 9, 10]. In parallel, research in functional Magnetic Resonance Imaging (fMRI) and EEG has identified several brain processes and areas involved in the memory process, providing consistent neural dynamics that motivate the study of subject-independent recall prediction [11, 12].

Despite this progress in neural and semantic modelling, a gap remains in understanding how artistic elements—independent of pure signal or plot—drive these processes. While the seminal work from Isola et al. [5] suggested that colours and image memorability are weakly correlated at a general level, it is well-established that film studios use photography and palette design to leave lasting impressions on viewers [13]. This opens the door to studying the impact of colour-based aesthetic choices on multimedia memorability. Additionally, since no clear correlations have been found between aesthetics [5], interestingness [14] or popularity [15] and memorability, it remains an open question whether clips from more popular movies or brands are more memorable. The 2026 edition addresses these gaps by providing colour palettes and movie metadata. By maintaining the 2025 benchmarks, this iteration enables a longitudinal study of how intentional aesthetic choices and industry context—beyond raw visual signals—drive long-term human recall.

3. Task Description

Due to the novelty of the tasks and data introduced in the previous edition and the shorter cycle of this edition, the 2026 task intentionally retains the 2025 format and subtasks: Movie Memorability (Subtask 1) and Commercial Memorability (Subtask 2). A detailed description of each is provided below.

3.1. Subtask 1: Movie Memorability

The aim of this subtask is to examine the long-term memorability of movie clips. In 2026 the subtask maintains two challenges introduced in 2025:

1.1 *How memorable is this movie excerpt? (Prediction).* Participants must predict the long-term memorability score of a movie clip from the Movie Memorability Dataset. The participants can use features derived from the raw video, such as the pre-computed feature embeddings and colour palettes, or metadata to estimate memorability.

1.2 *Is this person familiar with this video? (EEG-based recall detection).* Participants must predict whether a subject correctly recalls a clip from a previously watched movie, using EEG signals provided by the Essex EEG Movie Memory Dataset (EEMMD). The participants must use the raw EEG signal to derive features and predict recall.

3.2. Subtask 2: Commercial Memorability

This subtask focuses on assessing how well commercial videos are remembered over the long-term. This edition continues the two challenges introduced in 2025:

2.1 *How memorable is this commercial video? (Prediction).* Participants must predict the long-term memorability score of a commercial video from the Video Effectiveness and Memorability Dataset (VIDEM). The participants can use pre-computed feature embeddings, colour palettes and metadata for their proposed models.

2.2 *Can you predict the brand memorability? (Prediction-Brand).* Using the same inputs as in challenge 2.1, participants must predict the memorability score of the brand associated with a given commercial. The participants can use the same features as those used for video memorability.

4. Datasets

To ensure longitudinal consistency, the 2026 edition retains the Movie Memorability Dataset [16] and the Essex EEG Movie Memory Dataset (EEMMD) [17], and the Video Effectiveness and Memorability Dataset (VIDEM) [18]. A set of pre-computed features is also provided to the participants: (i) image-level features (AlexNetFC7, HOG, HSVHist, RGBHist, LBP, VGGFC7, DenseNet121, ResNet50, and EfficientNetB3); (ii) video-level features (C3D); and (iii) EEG-based representations: ERPs (event-related potentials, i.e., EEG amplitudes measured at the onset of each video) and ERSPs (event-related spectral perturbations, representing time-frequency features across the full duration of the video). The primary advancement in this edition is the inclusion of contextual and aesthetic descriptors. For Challenge 1.1, participants are provided with a subset of the MovieLens dataset containing user ratings, genres and tags. Furthermore, we introduce colour palettes extracted via the Binned-KMeans method, which focuses on extracting "memorable colours" from videos [19]. We hypothesise that these palettes offer a more detailed representation of intentional colour usage in videos than raw RGB values, enabling teams to explore the interplay between artistic decisions, industry context, and long-term human recall.

4.1. Movie Video Memorability

The Movie Memorability Dataset [16] will be employed in Subtask 1. It consists of 660 short movie segments taken from 100 occidental films, each paired with long-term memorability annotations. These labels capture the extent to which a viewer can recognise a clip after a long duration (ranging between few days to several months) between subsequent viewings. As in the 2025 edition, we provide the same 521 videos for the development set and the same 139 videos as the official test set. We refer to the dataset paper [16] for more information regarding the annotation protocol.

4.2. Essex EEG Movie Memory Dataset

The EEMMD [17] consist of 32 channel EEG recordings derived from the Movie Memorability Dataset. The EEG signals were recorded from 27 participants while they viewed a set of clips drawn from both movies they had watched previously and movies they had never watched. The resulting dataset includes 3484 epochs of EEG data (2122 unrecognised and 1362 remembered). The data from each participant are divided at the epoch level into the exact same development (80%) and testing (20%) subsets used in the 2025 edition.

4.3. Video Effectiveness and Memorability Dataset

The VIDEM dataset [18] has collected video and brand memorability annotations for advertisements, educational, and explanatory content. It comprises 424 video advertisements released on YouTube between June 2018 and June 2021, with lengths spanning from 7 seconds to 94 minutes. The 1403 annotators were allowed to watch up to a minute of each video. However, each annotator was allowed to select which parts of the video to watch.

Each clip in the dataset includes features such as video metadata, including the title, description, view count, and duration, as well as behavioural metrics including the numbers of likes and dislikes, views, and the engagement rate. We provide the exact same 339 samples for the development set and the identical 85 samples for the test set as used in the 2025 edition.

5. Evaluation

To maintain consistency with the benchmarking approach established in MediaEval 2025, Challenges 1.1, 2.1 and 2.2 employ the Spearman Rank Correlation Coefficient (SRCC, ρ) as the primary evaluation metric. Challenge 1.2, the EEG-based recall classification task, uses the Area Under the Receiver Operating Characteristic Curve (AUC).

For each sub-task, teams may submit a maximum of five runs. One run must use datasets provided by the task organisers;for this specific run, participants must not use supplemental videos/EEG signals. However, they are free to apply any feature extraction procedure they deem appropriate for their modelling strategy. The remaining four runs may incorporate additional datasets for training or validation purposes.

6. Conclusions

This paper presents the 2026 edition of the MediaEval Predicting Movie and Commercial Memorability task, continuing the exploration of computational modelling of memorability across movie clips, marketing-related videos and neural-based recall. The findings from this task will deepen our understanding of human memory, shape technologies for media literacy, and provide a collaborative framework for research across multimedia, cognitive science, and neuroscience.

Acknowledgements

The research of Iván Martín-Fernández was supported by the Universidad Politécnica de Madrid (Programa Propio I+D+i). This work was funded by the Spanish Ministry of Science and Innovation through the project TRUSTBOOST (PID2023-150584OB-C21), funded by MCIN/AEI/10.13039/501100011033 and by the European Union “NextGenerationEU/PRTR”. This work was partly supported by the projects GRESEL-UNED (PID2023-151280OB-C22) funded by MICIU/AEI/ AEI 501100011033 and ANNOTATE (PID2024-156022OB-C31) funded by MICIU/AEI/10.13039/501100011033 and the European Social Fund Plus (ESF+). This work was partly supported by TÜBİTAK under the 2209-A (Project No: 1919B012468481). This work was funded by UEFISCDI Romania, grant PN-IV-P7-7.1-PTE-2024-0382, agreement 12PTE/2025. This work is partially supported by project ROBUST: Trustworthy AI-based Systems for Sustainable Growth with project number KICH3.LTP.20.006, which is (partly) financed by the Dutch Research Council (NWO), RTL, and the Dutch Ministry of Economic Affairs and Climate Policy (EZK) under the program LTP KIC 2020-2023.

Declaration on Generative AI

During the preparation of this work, the authors used DeepL Write and Writefull in order to: Grammar and spelling check.

References

[1] I. Martín-Fernández, M. G. Constantin, C.-H. Demarty, M. Gil-Martín, S. Halder, B. Ionescu, A. Matran-Fernandez, R. Savran Kiziltepe, A. García Seco de Herrera, Overview of the mediaeval 2025 predicting movie and commercial memorability task, in: Proc. of the MediaEval 2025 Workshop, Dublin, Ireland and Online, 2025.

[2] R. N. Shepard, Recognition memory for words, sentences, and pictures 6 (1967) 156–163. URL: https://linkinghub.elsevier.com/retrieve/pii/S0022537167800677. doi:10.1016/S0022-5371(67)80067-7.

[3] L. Standing, Learning 10000 pictures 25 (1973) 207–222. URL: https://journals.sagepub.com/doi/10.1080/14640747308400340. doi:10.1080/14640747308400340.

[4] A. Jaegle, V. Mehrpour, Y. Mohsenzadeh, T. Meyer, A. Oliva, N. Rust, Population response magnitude variation in inferotemporal cortex predicts image memorability 8 (2019) e47596. URL: https://elifesciences.org/articles/47596. doi:10.7554/eLife.47596.

[5] P. Isola, J. Xiao, A. Torralba, A. Oliva, What makes an image memorable?, in: CVPR 2011, IEEE, 2011, pp. 145–152. URL: http://ieeexplore.ieee.org/document/5995721/. doi:10.1109/CVPR.2011.5995721.

[6] P. Kumar, E. Khandelwal, M. Tapaswi, V. Sreekumar, Seeing eye to ai: Comparing human gaze and model attention in video memorability, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2025, pp. 2082–2091. doi:10.1109/WACV61041.2025.00209.

[7] T. Dumont, J. S. Hevia, C. L. Fosco, Modular memorability: Tiered representations for video memorability prediction, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023, pp. 10751–10760. doi:10.1109/CVPR52729.2023.01035.

[8] L. Sweeney, G. Healy, A. F. Smeaton, Diffusing surrogate dreams of video scenes to predict video memorability, in: Working Notes Proceedings of the MediaEval 2022 Workshop, volume 3583 of *CEUR Workshop Proceedings*, CEUR-WS.org, Bergen, Norway, 2023. URL: https://ceur-ws.org/Vol-3583/paper52.pdf.

[9] H. Si, S. Singh, Y. K. Singla, A. Bhattacharyya, V. Baths, C. Chen, R. R. Shah, B. Krishnamurthy, Long-term ad memorability: Understanding & generating memorable ads, in: Proceedings of the Winter Conference on Applications of Computer Vision (WACV), 2025, pp. 5707–5718. doi:10.1109/WACV61041.2025.00557.

[10] I. Martín-Fernández, S. Esteban-Romero, F. Fernández-Martínez, M. Gil-Martín, Parameter-efficient adaptation of large vision—language models for video memorability prediction, Sensors 25 (2025). URL: https://www.mdpi.com/1424-8220/25/6/1661. doi:10.3390/s25061661.

[11] J. Han, C. Chen, L. Shao, X. Hu, J. Han, T. Liu, Learning computational models of video memorability from fMRI brain imaging, IEEE Transactions on Cybernetics 45 (2014) 1692–1703.

[12] W. Xie, W. A. Bainbridge, S. K. Inati, C. I. Baker, K. A. Zaghloul, Memorability of words in arbitrary verbal associations modulates memory retrieval in the anterior temporal lobe, Nature Human Behaviour 4 (2020) 937–948. URL: https://www.nature.com/articles/s41562-020-0901-2. doi:10.1038/s41562-020-0901-2.

[13] M. Risk, How to Use Color in Film: 50+ Examples of Movie Color Palettes, 2024. URL: https://www.studiobinder.com/blog/how-to-use-color-in-film-50-examples-of-movie-color-palettes/, section: Film Theory.

[14] M. Gygli, H. Grabner, H. Riemenschneider, F. Nater, L. V. Gool, The Interestingness of Images , in: 2013 IEEE International Conference on Computer Vision (ICCV), IEEE Computer Society, Los Alamitos, CA, USA, 2013, pp. 1633–1640. URL: https://doi.ieeecomputersociety.org/10.1109/ICCV.2013.205. doi:10.1109/ICCV.2013.205.

[15] A. Khosla, A. S. Raju, A. Torralba, A. Oliva, Understanding and predicting image memorability at a

large scale, in: 2015 IEEE International Conference on Computer Vision (ICCV), IEEE, 2015, pp. 2390–2398. URL: https://ieeexplore.ieee.org/document/7410632/. doi:10.1109/ICCV.2015.275.

[16] R. Cohendet, K. Yadati, N. Q. K. Duong, C.-H. Demarty, Annotating, understanding, and predicting long-term video memorability, in: Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval, ICMR '18, Association for Computing Machinery, New York, NY, USA, 2018, p. 178–186. URL: https://doi.org/10.1145/3206025.3206056. doi:10.1145/3206025.3206056.

[17] A. Matran-Fernandez, S. Halder, An EEG dataset to study neural correlates of audiovisual long-term memory retrieval, Scientific Data 12 (2025) 1933. URL: https://www.nature.com/articles/s41597-025-06203-1. doi:10.1038/s41597-025-06203-1.

[18] R. S. Kiziltepe, S. Sahab, R. V. Santana, F. Doctor, K. Paterson, D. Hunstone, A. G. Seco de Herrera, Videm: Video effectiveness and memorability dataset, in: I. Rojas, G. Joya, A. Catala (Eds.), Advances in Computational Intelligence, Springer Nature Switzerland, Cham, 2026, pp. 41–54.

[19] A. Pocol, L. Istead, C. Kaplan, Binned-kmeans: Extracting memorable, explainable, weighted and sorted colour palettes from videos, in: 2025 International Conference on Content-Based Multimedia Indexing (CBMI), 2025, pp. 1–5. doi:10.1109/CBMI66578.2025.11339288.