NewsImages in MediaEval 2025 – Comparing Image Retrieval and Generation for News Articles

Lucien Heitz[1,*], Luca Rossetto[2], Benjamin Kille[3], Andreas Lommatzsch[4], Mehdi Elahi[5] and Duc-Tien Dang-Nguyen[5]

[1]University of Zurich, Zurich, Switzerland
[2]Dublin City University, Dublin, Ireland
[3]Norwegian University of Science and Technology, Trondheim, Norway
[4]Technische Universität Berlin, Berlin, Germany
[5]University of Bergen, Bergen, Norway

Abstract

The NewsImages challenge in MediaEval 2025 explores matching news articles with fitting thumbnails. Participants receive a large set of English-language articles from international publishers. Given the text of a news article, the goal of the challenge is to (1) retrieve or (2) generate a fitting image. In two subtasks, participating teams create approaches to fully automate the retrieval/generation workflows as well as develop approaches that feature a human-in-the-loop. The overall quality of the image recommendation is evaluated in a crowd-sourced online event where image fit and relevance are assessed. This task overview paper provides details on the challenge setup, the crowdsourced online evaluation, and technical aspects of the group submissions.

Keywords

image retrieval, image generation, online news

1. Introduction

Images and visuals play an important role when it comes to news consumption, especially on online [1, 2, 3]. They receive special attention online platforms [4], given that they are a key factor that drives user engagement [5]. Editors and writers have different options available to help them with the image selection process. These options include automatically finding a matching image for a given news story (i.e., image retrieval, cf. [6]) or generating a new image instead (through generative visual AI, cf. [7]). The goal of NewsImages in MediaEval 2025 is to compare these two approaches across different news topics, identifying their advantages and disadvantages, in order to understand the implications that they may have for the areas of journalism and news personalization. The comparison is conducted at both qualitative and media policy levels. We acknowledge that this scope provides only limited insight into the complex areas of image retrieval/generation, as well as articles. Therefore, we created a complementary list of potential "Quest for Insight" topics for participants to investigate.[1]

MediaEval'25: Multimedia Evaluation Workshop, October 25–26, 2025, Dublin, Ireland and Online
*Corresponding author.
heitz@ifi.uzh.ch (L. Heitz); luca.rossetto@dcu.ie (L. Rossetto); benjamin.u.kille@ntnu.no (B. Kille); andreas.lommatzsch@dai-labor.de (A. Lommatzsch); mehdi.elahi@uib.no (M. Elahi); ductien.dangnguyen@uib.no (D. Dang-Nguyen)
0000-0001-7987-8446 (L. Heitz); 0000-0002-5389-9465 (L. Rossetto); 0000-0002-3206-5154 (B. Kille); 0009-0009-0532-7081 (A. Lommatzsch); 0000-0003-2203-9195 (M. Elahi); 0000-0002-2761-2213 (D. Dang-Nguyen)

[1]The topic list is available online: https://multimediaeval.github.io/editions/2025/tasks/newsimages

2. Related Work

Image captions are often assumed to describe the literally depicted content of the image [8]. The relationship becomes less clear in the news domain when images accompany news articles [9]. Since images are not always available for the most recent news stories, stock images or archived photos are frequently used as alternatives. Retrieval methods to query these collections often rely on the CLIP [6] or OpenCLIP model [10] (see previous challenge submissions [11, 12, 13, 14]). NewsImages at MediaEval 2023 [15] was the first iteration of the task that introduced diffusion-based AI models [7] for image generation as an alternative option for providing image recommendations. A user study conducted with this NewsImages dataset showed that, in fact, readers do perceive generated content as more fitting than editorially selected stock images for certain news topics [16], hinting at the future potential of this research direction.

While the automation of the image selection process is feasible from a technical standpoint, several open ethical issues remain, particularly with image generation [17]. News plays an important role in democratic societies [18, 19, 20], as they inform and enable citizens to form their own opinions [21, 22] and provide the foundation for public deliberation [23, 24, 25]. Therefore, it is of critical importance that the match between the news and the image is accurate. Images should not mislead or deceive readers into assuming they represent a real-life situation when they do not. This could otherwise lead to the spread of misinformation through fake news and images [26], which can have severe negative consequences for society.

These questions and challenges are at the core of the NewsImages 2025 challenge task. The goal of this task is to investigate these intricacies in more depth, in order to understand the implications that it may have for the areas of journalism and news personalization.

3. Task Description

This year's NewsImages challenge explores matching news articles with retrieved or generated images. Participants receive a large set of English-language articles from international news organizations and media outlets. Given the text of a news article, the goal of this task is to retrieve and/or generate a fitting image recommendation that can serve as an article thumbnail. In the Image Retrieval subtask, teams retrieve images from a pool of existing images. In the Image generation subtask, the teams use generative AI to produce new visuals instead. The main tasks of the challenge are as follows:

- **Image Retrieval**, where participants design and implement approaches to retrieve relevant images in an existing image pool that fits a given news headline and lead.
- **Image Generation**, where participants need to use and develop techniques for creating appropriate visuals for news articles.

Participating teams are encouraged to explore different image styles (e.g., use photo-realistic style or caricatures), image subjects (e.g., focus on people), and a combination thereof. Teams can take part in one or both of the main tasks. They are free to submit multiple runs.

3.1. Challenge Subtasks

The challenge of retrieving and/or generating fitting news images consists of two subtasks. Participants must make an image recommendation for all 8,500 articles. From among this article pool, a small one (30 images) and a large evaluation subset (20 images) are selected. The selection of the images for the large dataset is random and will occur after the submission deadline, whereas the image selection for the small dataset is predetermined and will be communicated beforehand. Teams are requested to provide at least one submission for each subtask.

The two subtasks allow participants to explore different retrieval and generation workflows. The large dataset presents a challenge that necessitates a fully automated and scalable workflow for image assignment. In contrast, the small subset focuses on creating a workflow where the journalist has an active role; this approach aims to be used as a tool to augment the editorial retrieval or generation capabilities for image assignments.

3.2. News Dataset and Resources

Participants are given a collection of 8,500 news articles with images (the article text is in English, collected during 2022–2023 by GDELT).[2] Each new item includes the following properties: *article_id* (internal ID of news article used for submission), *article_url* (URL of the original news article), *article_title* (title of the news article, may include lead), *article_tags* (automatically generated tags for the main article text/body), *image_id* (internal ID of original image for which a copy is shared), and *image_url* (URL of the original news image). The article text itself is not shared, but participating groups are free to retrieve it from the original source.

For the image retrieval task, we ask participants to use the *Yahoo-Flickr Creative Commons 100 Million* (YFCC100M) dataset as the source for the images.[3] Groups are free to select any publicly available model for image generation and/or create their own. The model and any resources used to retrieve/generate images must be shared as part of the final submission to verify the approaches. The image retrieval/generation pipeline must not rely on any closed-source APIs or resources that are not publicly available.

3.3. Paper and Image Submission

Staying true to the principles of MediaEval of promoting reproducible research, we ask participants to share their entire image retrieval pipelines and generation workflows (including all parameters and linking models to reproduce the image recommendations). For image generation, we recommend tools like ComfyUI[4] or Stable Diffusion Web UI[5] to automatically generate and embed the workflow data within the images. Images must be PNGs, with a target dimension of 460x260 pixels (i.e., landscape orientation).[6] These restrictions apply to both generated and retrieved images. Finally, document your work by preparing a working note paper that outlines your submission. Teams are free to submit multiple runs per task However, they must make sure that each run is properly documented within the paper.[7]

[2]GDELT website: https://www.gdeltproject.org
[3]YFCC100M website: https://www.multimediacommons.org
[4]ComfyUI on GitHub: https://github.com/comfyanonymous/ComfyUI
[5]Stable Diffusion Web UI on GitHub: https://github.com/AUTOMATIC1111/stable-diffusion-webui
[6]For more details, please see the official NewsImages GitHub page: https://github.com/Informfully/Challenges
[7]For instructions, please go to the official MediaEval 2025 website: https://multimediaeval.github.io/editions/2025

4. Evaluation

The quality of the retrieved and generated image will be evaluated during a crowd-sourced online event. The primary criterion for evaluating the submitted runs is the image fit (i.e., whether the images accurately capture key attributes of the text article, without depicting any important elements not present in the article). During the event, participants are presented with a news headline and a list of image recommendations. The fit of each image is rated on a 5-point Likert scale. The winning team is determined by the highest average image fit ratings achieved by their best submitted run. All participating teams must participate in this event to have a chance of winning the competition. We encourage participants to adhere to editorial standards and guidelines for image generation.[8] We are particularly interested in non-photorealistic images that do not suggest they accurately represent real events, in order to avoid misleading and/or deceiving readers.

5. Conclusion

Images remain a crucial factor in capturing the reader's attention on online news platforms. To explore the available options and opportunities for assigning visuals to articles, this year's iteration of the NewsImages challenge task examines both the technological and user aspects of incorporating both retrieved and generated images for news stories.

From a technological perspective, the task focuses on comparing traditional retrieval approaches with new AI-based image generation capabilities. The goal here is to identify the strengths and weaknesses of both options with different stories and topics. For the large challenge subtask, this is achieved through a fully automated workflow. And with the small subtask, this challenge examines situations where retrieval/generation systems play a supporting role and work hand in hand with writers or editors to select images.

From the user's point of view, the goal is to conduct a qualitative analysis of both image recommendation approaches to identify existing preferences in terms of image style, subject, and other factors. By focusing on existing editorial guidelines regarding AI, this task empirically validates how these policies are perceived by readers. Is there a preference for more photo-realistic images, which could potentially be used to mislead people, or are they more in favor of highly stylized pictures that adhere to most existing AI policies?

Understanding in what scenarios retrieval or generation performs best, and how audiences perceive image recommendations, helps online platforms develop better news offers, increasing reader satisfaction and engagement with news.

Acknowledgments

We thank Martha Larson, Steven Hicks, and Mihai Gabriel Constantin for their feedback and support through organizing this challenge.

Declaration on Generative AI

During the preparation of this task overview paper, the authors used Grammarly for grammar and spelling checks. After using this service, the authors reviewed and edited the content as needed. They take full responsibility for the publication's content.

[8] For an overview of existing policies, please see: https://www.ebu.ch/groups/ai-ethics

References

[1] H. Caple, Visual media: The importance of visuals as partners in the news, in: The Routledge Handbook of Language and Media, Routledge, 2017, pp. 230–243.

[2] Y. De Haan, S. Kruikemeier, S. Lecheler, G. Smit, R. Van der Nat, When does an infographic say more than a thousand words? audience evaluations of news visualizations, Journalism Studies 19 (2018) 1293–1312.

[3] H. Stöckl, Image-centricity–when visuals take center stage: Analyses and interpretations of a current (news) media practice, in: Shifts towards image-centricity in contemporary multimodal practices, Routledge, 2020, pp. 19–41.

[4] T. Thomson, To see and be seen: The environments, interactions and identities behind news images, Rowman & Littlefield, 2019.

[5] J. R. Collier, J. Dunaway, N. J. Stroud, Pathways to deeper news engagement: Factors influencing click behaviors on news sites, Journal of Computer-Mediated Communication 26 (2021) 265–283.

[6] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, I. Sutskever, Learning transferable visual models from natural language supervision, in: M. Meila, T. Zhang (Eds.), Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event, volume 139 of *Proceedings of Machine Learning Research*, PMLR, 2021, pp. 8748–8763.

[7] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, B. Ommer, High-resolution image synthesis with latent diffusion models, in: IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022, IEEE, 2022, pp. 10674–10685.

[8] M. Z. Hossain, F. Sohel, M. F. Shiratuddin, H. Laga, A comprehensive survey of deep learning for image captioning, ACM Computing Surveys (CsUR) 51 (2019) 1–36.

[9] N. Oostdijk, H. v. Halteren, E. Basar, M. A. Larson, The connection between the text and images of news articles: New insights for multimedia analysis (2020).

[10] M. Cherti, R. Beaumont, R. Wightman, M. Wortsman, G. Ilharco, C. Gordon, C. Schuhmann, L. Schmidt, J. Jitsev, Reproducible scaling laws for contrastive language-image learning, in: IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023, Vancouver, BC, Canada, June 17-24, 2023, IEEE, 2023, pp. 2818–2829.

[11] T.-H. Nguyen, H.-L. Nguyen-Huu, T.-D. Le, H.-L. Tran, Q.-K. Le-Tran, H.-B. Ngo, M.-H. An, Q.-V. Dinh, Multimodal fusion in newsimages 2023: Evaluating translators, keyphrase extraction, and clip pre-training., in: MediaEval, 2023.

[12] A. Leventakis, D. Galanopoulos, V. Mezaris, Cross-modal networks, fine-tuning, data augmentation and dual softmax operation for mediaeval newsimages 2023., in: MediaEval, 2023.

[13] X. Wang, M. Liang, M. A. Larson, Investigating the performance of the clip model and concept matching in text-image retrieval systems., in: MediaEval, 2023.

[14] L. Heitz, Y. K. Chan, H. Li, K. Zeng, A. Bernstein, L. Rossetto, Prompt-based alignment of headlines and images using openclip, in: Working Notes Proceedings of the MediaEval 2023 Workshop, 2024. URL: http://ceur-ws.org/Vol-3658/paper7.pdf.

[15] A. Lommatzsch, B. Kille, Ö. Özgöbek, M. Elahi, D.-T. Dang-Nguyen, News images in mediaeval 2023, Proceedings of the MediaEval Benchmarking Initiative (2023).

[16] L. Heitz, A. Bernstein, L. Rossetto, An empirical exploration of perceived similarity between news article texts and images, in: Working Notes Proceedings of the MediaEval 2023 Workshop, 2024. URL: http://ceur-ws.org/Vol-3658/paper8.pdf.

[17] T. Thomson, R. J. Thomas, P. Matich, Generative visual ai in news organizations: Challenges, opportunities, perceptions, and policies, Digital Journalism (2024) 1–22.
[18] A. Bernstein, C. De Vreese, N. Helberger, W. Schulz, K. Zweig, L. Heitz, S. Tolmeijer, et al., Diversity in news recommendation, Dagstuhl Manifestos 9 (2021) 43–61. URL: https://doi.org/10.4230/DagMan.9.1.43.
[19] L. Heitz, J. A. Lischka, A. Birrer, B. Paudel, S. Tolmeijer, L. Laugwitz, A. Bernstein, Benefits of diverse news recommendations for democracy: A user study, Digital Journalism 10 (2022) 1710–1730. URL: https://doi.org/10.1080/21670811.2021.2021804.
[20] N. Mattis, L. Heitz, P. K. Masur, J. Moeller, W. van Atteveldt, Nudges for news recommenders: Prominent article positioning increases selection, engagement, and recall of environmental news, but reducing complexity does not, Journal of Communication (2025) jqaf019. URL: https://doi.org/10.1093/joc/jqaf019.
[21] N. Helberger, On the democratic role of news recommenders, Digital Journalism 7 (2019) 993–1012.
[22] L. Heitz, J. A. Lischka, R. Abdullah, L. Laugwitz, H. Meyer, A. Bernstein, Deliberative diversity for news recommendations: Operationalization and experimental user study, in: Proceedings of the 17th ACM Conference on Recommender Systems, ACM New York, NY, USA, 2023, pp. 813–819. URL: https://doi.org/10.1145/3604915.3608834.
[23] J. Habermas, The structural transformation of the public sphere, Polity Press, 1989.
[24] L. Heitz, K. Rozgonyi, B. Kostic, D. Wagner, J. Haas, AI in Content Curation and Media Pluralism, Organization for Security and Co-operation in Europe, Austria, 2021, pp. 56–70. URL: https://www.zora.uzh.ch/id/eprint/213723.
[25] L. Heitz, O. Inel, S. Vrijenhoek, Recommendations for the recommenders: Reflections on prioritizing diversity in the recsys challenge, in: Proceedings of the Recommender Systems Challenge 2024, ACM New York, NY, USA, 2024, pp. 22–26. URL: https://doi.org/10.1145/3687151.3687155.
[26] T.-P. Chang, T.-C. Hsiao, T.-L. Chen, T.-C. Lo, A framework for detecting fake news by identifying fake text messages and forgery images, Enterprise Information Systems 19 (2025).