

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

Lucien Heitz<sup>1,†</sup>, Luca Rossetto<sup>2</sup>, Benjamin Kille<sup>3</sup>, Andreas Lommatzsch<sup>4</sup>, Mehdi Elahi<sup>5</sup> and Duc-Tien Dang-Nguyen<sup>5</sup>

<sup>1</sup>University of Zurich, Zurich, Switzerland

<sup>2</sup>Dublin City University, Dublin, Ireland

<sup>3</sup>Norwegian University of Science and Technology, Trondheim, Norway

<sup>4</sup>Technische Universität Berlin, Berlin, Germany

<sup>5</sup>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 crowdsourced online event where image fit and relevance are assessed. This task overview paper goes into the details of the challenge setup, the evaluation of the images, and the technical details 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 it may have for the areas of journalism and news personalization. The comparison is conducted at both a qualitative and a media policy level. We acknowledge that this scope allows only for a limited insight into the complex area of image retrieval/generation and news articles. Therefore, we created a complementary list of potential “Quest for Insight” topics for participants to investigate.<sup>1</sup>

---

*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); 0000-0002-5389-9465 (A. Lommatzsch); 0000-0003-2203-9195 (M. Elahi); 0000-0002-2761-2213 (D. Dang-Nguyen)

© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

<sup>1</sup>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 there are always images available for the most recent news stories, stock images or archived photos are frequently used instead. Retrieval methods to query these collections often rely on the CLIP [6] or OpenCLIP model [10] (see previous NewsImages challenge submissions [11, 12, 13, 14]). The previous NewsImages challenge at MediaEval 2023 [15] was the first iteration to introduce diffusion-based AI models [7] for image generation as an alternative option for providing image recommendations. And a user study conducted within the scope of this challenge showed that, in fact, readers do perceive generated content as more fitting than editorially selected stock images for certain news topics [16].

While the automation of the image selection process is possible from a technical point of view, there are a number of open ethical issues, especially with image generation [17]. News plays an important role in democratic societies [18, 19], as they inform and enable citizens to form their own opinions [20] and provide the foundation for public deliberation [21]. Therefore, it is of critical importance that the match between the news and the image is accurate. And images should not mislead and/or deceive readers into assuming they represent a real-life situation when it does not. This could otherwise lead to the spread of misinformation through fake news and images [22], 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 main task 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 large evaluation subset (1,000 images) and a small one (50 images) are selected. The selection of the images for the large dataset will happen after the submission deadline, whereas the image selection of the small dataset is predetermined and will be communicated beforehand. Teams are requested to provide at least one submission for each of the subtasks.

The two subtasks allow participants to explore different retrieval and generation workflows. The large set provides a challenge that requires a fully automated and scalable workflow for image assignment. In contrast to that, the small subset focuses on creating a workflow wherein 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).<sup>2</sup> 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 to source the images for the retrieval task.<sup>3</sup> Groups are free to select any publicly available model for image generation and/or create their own. The model and any resource used to retrieve/generate images must be shared as part of the final submission in order 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<sup>4</sup> or Stable Diffusion Web UI<sup>5</sup> 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).<sup>6</sup> 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.<sup>7</sup>

<sup>2</sup>GDELT website: <https://www.gdeltproject.org>

<sup>3</sup>YFCC100M website: <https://www.multimediacommons.org>

<sup>4</sup>ComfyUI on GitHub: <https://github.com/comfyanonymous/ComfyUI>

<sup>5</sup>Stable Diffusion Web UI on GitHub: <https://github.com/AUTOMATIC1111/stable-diffusion-webui>

<sup>6</sup>Detailed instructions are available online: <https://github.com/Informally/Challenges>

<sup>7</sup>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 main criterion for evaluation of the submitted runs is the image fit and relevance (i.e., whether the images capture key attributes of the text article, but do not depict any important elements not present in the article). All participating teams must take part in the event, where they rate image fit and relevance of the submissions of their peers on a 7-point Likert scale. The winning team is determined by the highest average image fit ratings achieved by their best submissions across both the large and small challenge subtasks.

We encourage participants to adhere to editorial standards and guidelines for image generation.<sup>8</sup> We are particularly interested in non-photorealistic images that do not suggest that they accurately represent real events so as not to mislead and/or deceive readers.

## 5. Conclusion

Images and visuals remain an important factor when it comes to capturing the reader's attention on an online platform. To explore the available options and opportunities, this year's iteration of the NewsImages challenge task looks at both the technological side as well as the user side when it comes to adding images and visuals to the online news stories.

From a technological point of view, 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. With the large challenge subtask, this is done for a fully automated workflow. And with the small subtask, this challenge looks at situations where retrieval/generation systems have more of a supporting role and work hand in hand with the writers or editors for selecting images.

From the user's point of view, the goal is to have a qualitative analysis of both image recommendation approaches to see what existing preferences are in terms of image style, subject, and more. By focusing on existing editorial guidelines when it comes to AI, this task empirically validates how the policies are seen by readers. Is there a preference towards more photo-realistic images, that could potentially be used to mislead people, or are they more in favor of highly stylized pictures that follow most existing AI policies?

Understanding in what scenario retrieval or generation performs best, and how audiences perceive the image recommendations, helps online platforms to develop better news offers. Offers that increase 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.

---

<sup>8</sup>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.
- [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.

- [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.
- [19] H. Sargeant, E. Pirkova, M. C. Kettemann, M. Wisniak, M. Scheinin, E. Bevensee, K. Pentney, L. Woods, L. Heitz, B. Kostic, et al., *Spotlight on Artificial Intelligence and Freedom of Expression: A Policy Manual*, Organization for Security and Co-operation in Europe, 2022.
- [20] N. Helberger, On the democratic role of news recommenders, *Digital Journalism* 7 (2019) 993–1012.
- [21] J. Habermas, *The structural transformation of the public sphere*, Polity Press, 1989.
- [22] 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) 2436495.