

MultiSumm - Multimodal Summarisation Task at MediaEval 2025

Anastasiia Potyagalova, Gareth J. F. Jones

ADAPT Centre, School of Computing, Dublin City University, Ireland

Abstract

We describe the first edition of the MediaEval MultiSumm task. This task requests participants to use the online content relating to Food Sharing Initiative (FSIs) for each of a small number of cities accessed via a list of FSI URLs for each city. Evaluation of submissions explored the use of an “LLM-as-Judge” approach to assessment of the quality of submissions. Although this was the first edition of this task, participants’ submissions provided valuable initial insights into effective methodologies for the creation of multidocument summaries from diverse online content.

1. Introduction

Multidocument summarization for text documents has been a longstanding area of investigation [1], [2]. Traditionally this process has been complex and inflexible in terms of content style and test, requiring use of a wide variety of natural language processing (NLP) tools and detailed specification of the summarization process [3]. The emergence of large language models (LLM) technologies has many revolutionised NLP tasks including summarization [4]. The more recent arrival of multimodal LLMs is similarly impacting on topics relating to multimedia content [5].

While the MultiSumm tasks could be tackled using traditional NLP and multimedia processing tools, the expectation is that participants will tackle it using multimodal LLM methods,

To the best of our knowledge, this is the first benchmark task focusing on this topic and providing a potentially valuable venue for exploration of the potential and challenges of use of multimodal LLMs in tasks of this sort.


2. Task Definition


The MultiSumm task at MediaEval 2025 explores the generation of multimodal summaries from multiple heterogeneous web content sources. Specifically, participants are challenged to create concise, informative summaries that integrate textual and visual information extracted from websites describing Food Sharing Initiatives (FSIs) in various cities worldwide [6], [7]. The task builds on data provided by the H2020 CULTIVATE project ¹, which investigates and promotes urban and peri-urban food sharing through the creation of large-scale open resources [8].

A central resource underpinning the task is the ShareCity200 database, an automatically crawled and curated dataset of FSIs identified across 200 cities. The database extends the earlier ShareCity100 dataset ² by including both European and selected international cities [9], [10].

MediaEval’25: Multimedia Evaluation Workshop, October 25–26, 2025, Dublin, Ireland and Online

✉ anastasia.potyagalova2@mail.dcu.ie (A. Potyagalova); Gareth.Jones@dcu.ie (G. J. F. Jones)

 © 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)

¹<https://cultivate-project.eu/>

²<https://sharecity.ie/research/sharecity100-database/>

Each city-level subset of ShareCity200 contains verified web pages, metadata, and multimedia content relating to local food sharing activities.

Participants are provided with the URLs of FSIs identified within ShareCity200 for a small number of cities and are required to access the online content from these URLs produce a multimodal summary that effectively represents the landscape of FSIs in each city. The summary should reflect multiple aspects, including:

- Geographical distribution of initiatives by city district
- Categories and types of initiatives (e.g., sharing, swapping, gifting)
- Operational level (government-funded, district-supported, community-led)
- Popularity or activity level, as inferred from web and social signals
- Public sentiment or community feedback
- Representative visual content (e.g., photographs illustrating major FSIs)

The summaries must be produced in English and presented in a structured multimodal format, combining textual and visual elements to enhance comprehensibility and engagement. Participants will receive a reference schema specifying the expected summary structure, example outputs, and detailed evaluation criteria.

2.1. Research Questions

The task also aims to promote reflection on broader research issues related to multimodal summarization with LLMs, including:

- What are the key challenges in generating summaries from multi-source web content?
- How can LLMs be effectively applied in multimodal summarization?
- What open research problems remain in multidocument and multimodal summarization?
- How effective are LLM-based evaluation methods for assessing multimodal summaries?

By addressing these questions, the MultiSumm task seeks to advance understanding of LLM-driven summarization techniques [11], [12] and their potential for automatically generating rich, cross-media representations of real-world phenomena such as urban food sharing.

3. Dataset and Evaluation

The dataset provided for the MultiSumm 2025 task originates from the ShareCity200 collection developed within the H2020 CULTIVATE project. It consists of manually verified web links to online resources describing Food Sharing Initiatives (FSIs) in selected cities [9]. Each entry corresponds to an identified FSI website that forms part of the broader ShareCity200 database of global urban food sharing activity.

3.1. Data Splits

The MultiSumm datasets are organized by city and divided according to the main and subtask configurations. The dataset for Cork (Ireland) dataset is provided as the development and training material. It contains verified URLs of FSIs within City city and serves as an example for link structure, file format, and summary schema.

The main evaluation focuses on Dublin (Ireland) and Brighton and Hove (U.K.), representing English-speaking cities. Participants are required to use the verified URLs to collect content and generate multimodal summaries describing the food sharing landscape in each city.

The subtask expands the evaluation to London (U.K.), Milan (Italy), and Barcelona (Spain). These datasets introduce the challenges of a much larger city and additional linguistic and cultural variation, enabling the assessment of summarization systems under multilingual and cross-domain conditions.

Participants are responsible for retrieving relevant textual and visual content from the listed web resources in accordance with the summary schema and ethical web use guidelines. All datasets are released under the CULTIVATE research data sharing policy, and may be expanded with additional city links on request to support multilingual or region-specific experimentation.

3.2. Evaluation Methodology

Systems are assessed using a hybrid automatic–LLM judging framework. The textual component of each submission is evaluated by large language models acting as evaluators (LLM-as-Judge), following recent methods proposed for generative IR and summarization quality assessment. The visual component (images accompanying summaries) is evaluated using a combination of automatic metadata verification and LLM-based multimodal reasoning (via GPT-4V-like models), allowing both factual and contextual assessment of relevance.

3.3. LLM-as-Judge Implementation

Evaluation is conducted using prompting templates in which the LLM acts as an expert assessor [13], [14]. A representative prompt is structured as follows:

“You are an expert evaluator reviewing a multimodal report about Food Sharing Initiatives (FSIs) in [City]. Evaluate the report using the following criteria—Informational Coverage, Accuracy and Factual Consistency, Clarity and Structure, Use of Visuals, Local Relevance, and Practical Usefulness—assigning a score from 1 to 5 for each and providing justifications for scores.”

This framework enables consistent large-scale evaluation with manual scoring, while maintaining transparency and flexibility for future iterations. The methodology also facilitates later fine-tuning of smaller evaluators using the collected LLM-as-Judge data [15].

3.4. Evaluation Output and Reporting

Each evaluated submission produces a structured JSON report containing per-dimension scores and textual explanations. These can be aggregated to compute mean performance across cities and dimension-specific leaderboards. The evaluation schema also supports inclusion of human reviewer annotations for calibration and cross-validation of LLM-based judgments.

4. Discussion and Insights

The comparative analysis across teams reveals three central insights regarding how participants approached the FSI overview task.

Structure and methodological maturity. The structure and methodological clarity of the reports had a strong positive influence on overall quality. Participants demonstrated a thoughtful approach to organization, producing coherent narratives that effectively combined quantitative and qualitative perspectives. Submissions provide valuable foundations for future refinement and analysis. Collectively, participant efforts highlight the growing methodological awareness within the community and reinforce the importance of a consistent analytical framework for documenting diverse, community-driven and government-driven initiatives.

Visual and spatial representation. The integration of images and spatial context played a crucial role in enhancing both credibility and engagement. Visual content that is meaningfully linked to specific locations or activities (such as community gardens, food redistribution hubs, and cooperative fridges) helped ground written descriptions in reality. When effectively employed, photographs were used as illustrations of local initiatives and events. It is necessary to mention that many participants detected and provided accurate images for their summarisation reports.

Localization and practical impact. Reports that incorporated references to municipal networks, community organizations, and regionally specific initiatives (e.g., Cork Food Policy Council, FareShare London, Espigoladors Barcelona) demonstrated a strong grasp of contextual dynamics and community priorities. By connecting observed activities with broader social and policy frameworks, participants translated descriptive data into practical insights. These locally informed perspectives enriched the narrative depth of each report and introduced elements of practical applicability, positioning the outputs as valuable resources for city councils, community initiatives, and local policy planning.

In summary, the comparative evaluation underscores that report quality depends not solely on data completeness, but on methodological consistency, contextual accuracy, and the ability to translate findings into locally actionable insight.

5. Conclusions and Future Directions

The assessment of participant submissions demonstrates a clear learning trajectory across teams, evolving from loosely structured datasets to comprehensive analytical overviews that integrate verified evidence, coherent reasoning, and authentic visual documentation.

Future iterations of this task should emphasize:

- **Source transparency and verification:** participants should include explicit data references (URLs, publication dates, or metadata) for every listed initiative.
- **Quantitative synthesis:** each report should provide numerical summaries (e.g., initiative counts by type or funding structure) and geospatial statistics to enhance comparability and reproducibility.
- **Standardized visual documentation:** images must include captions, organization names, and geographic coordinates to improve traceability and reuse.
- **Cross-city comparability:** adoption of consistent taxonomies and scoring metrics would facilitate benchmarking across cities and time periods.
- **Open data integration:** validated outputs could feed into a central repository, such as the *ShareCity200* or related datasets, supporting long-term monitoring of food-sharing activity.

Overall, the study confirms that combining automated discovery with structured human interpretation yields high-quality, context-sensitive documentation of food-sharing ecosystems. The best-performing submissions exemplify how rigorous analysis, grounded localization, and visual verification can transform dispersed community data into actionable urban knowledge.

6. Acknowledgement

This research received support from the SFI ADAPT II, and European Union's Horizon Europe Research and Innovation Programme under Grant Agreement No 101083377.

References

- [1] Anonymous, Survey on multi-document summarization: Systematic literature review, arXiv preprint arXiv:2312.12915 (2023). URL: <https://arxiv.org/abs/2312.12915>.
- [2] Supriyono, A. P. Wibawa, Suyono, F. Kurniawan, A survey of text summarization: Techniques, evaluation and challenges, *Natural Language Processing Journal* 7 (2024) 100070. URL: <https://www.sciencedirect.com/science/article/pii/S2949719124000189>. doi:<https://doi.org/10.1016/j.nlp.2024.100070>.
- [3] Z. Sheng, K. Yang, et al., Multi-document summarization via deep learning techniques, *ACM Transactions on Information Systems* (2021). doi:[10.1145/3529754](https://doi.org/10.1145/3529754).
- [4] H. Naveed, A. U. Khan, S. Qiu, M. Saqib, S. Anwar, M. Usman, N. Akhtar, N. Barnes, A. Mian, A comprehensive overview of large language models, 2024. URL: <https://arxiv.org/abs/2307.06435>. arXiv:2307.06435.
- [5] C. X. Liang, P. Tian, C. H. Yin, Y. Yua, W. An-Hou, L. Ming, T. Wang, Z. Bi, M. Liu, A comprehensive survey and guide to multimodal large language models in vision-language tasks, 2024. URL: <https://arxiv.org/abs/2411.06284>. arXiv:2411.06284.
- [6] A. R. Davies, *Urban food sharing: Rules, tools and networks*, Policy Press, 2019.
- [7] A. R. Davies, A. Cretella, V. Franck, Food sharing initiatives and food democracy: Practice and policy in three european cities, *Politics and Governance* 7 (2019) 8–20.
- [8] H. Wu, H. Cho, A. R. Davies, G. J. F. Jones, Llm-based automated web retrieval and text classification of food sharing initiatives, in: *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management, CIKM '24*, ACM, 2024, p. 4983–4990. URL: <https://doi.org/10.1145/3627673.3680090>. doi:[10.1145/3627673.3680090](https://doi.org/10.1145/3627673.3680090).
- [9] A. Davies, H. Cho, A.-M. Gatejel, R. Martinez Varderi, M. Vedo, CULTIVATE Briefing note - Food sharing landscapes in Hub city locations, 2024. URL: <https://doi.org/10.5281/zenodo.11030355>. doi:[10.5281/zenodo.11030355](https://doi.org/10.5281/zenodo.11030355).
- [10] D. Phelan, A. Davies, N. Gomboli, *The european food sharing dictionary*, 2023. URL: <https://doi.org/10.5281/zenodo.10160274>. doi:[10.5281/zenodo.10160274](https://doi.org/10.5281/zenodo.10160274).
- [11] Y. Zhang, M. Wang, C. Ren, Q. Li, P. Tiwari, B. Wang, J. Qin, Pushing the limit of llm capacity for text classification, arXiv preprint arXiv:2402.07470 (2024).
- [12] X. Sun, X. Li, J. Li, F. Wu, S. Guo, T. Zhang, G. Wang, Text classification via large language models, arXiv preprint arXiv:2305.08377 (2023).
- [13] Y. Lu, X. Yang, X. Li, et al., Llmscore: Unveiling the power of large language models in text-to-image synthesis evaluation, arXiv preprint arXiv:2305.11116 (2023). URL: <https://arxiv.org/abs/2305.11116>.
- [14] J. Gu, X. Jiang, Z. Shi, et al., A survey on llm-as-a-judge, arXiv preprint arXiv:2411.15594 (2024). URL: <https://arxiv.org/html/2411.15594v1>.
- [15] H. Wei, S. He, T. Xia, F. Liu, A. Wong, J. Lin, M. Han, Systematic evaluation of llm-as-a-judge in llm alignment tasks: Explainable metrics and diverse prompt templates, 2025. URL: <https://arxiv.org/abs/2408.13006>. arXiv:2408.13006.