

Workshop on Multimodal Augmented Generation via Multimodal Retrieval (MAGMaR 2025)

Vienna, Austria
1 August 2025

ISBN: 979-8-3313-2413-1

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571



Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2025) by the Association for Computational Linguistics
All rights reserved.

Printed with permission by Curran Associates, Inc. (2025)

For permission requests, please contact the Association for Computational Linguistics
at the address below.

Association for Computational Linguistics
209 N. Eighth Street
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006
Fax: 1-570-476-0860

acl@aclweb.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2633
Email: curran@proceedings.com
Web: www.proceedings.com

Table of Contents

<i>MultiReflect: Multimodal Self-Reflective RAG-based Automated Fact-Checking</i>	
Uku Kangur, Krish Agrawal, Yashashvi Singh, Ahmed Sabir and Rajesh Sharma	1
<i>ColLEX – A Multimodal Agentic RAG System Enabling Interactive Exploration of Scientific Collections</i>	
Florian Schneider, Narges Baba Ahmadi, Niloufar Baba Ahmadi, Iris Vogel, Martin Semmann and Chris Biemann	18
<i>VoxRAG: A Step Toward Transcription-Free RAG Systems in Spoken Question Answering</i>	
Zackary Rackauckas and Julia Hirschberg	40
<i>Cross-modal Clustering-based Retrieval for Scalable and Robust Image Captioning</i>	
Jingyi You, Hiroshi Sasaki and Kazuma Kadowaki	47
<i>Multimodal Retrieval-Augmented Generation: Unified Information Processing Across Text, Image, Table, and Video Modalities</i>	
Nazarii Drushchak, Nataliya Polyakovska, Maryna Bautina, Taras Semchenko, Jakub Koscielicki, Wojciech Sykala and Michal Wegrzynowski	59
<i>Making LVLMS Look Twice: Contrastive Decoding with Contrast Images</i>	
Avshalom Manevich and Reut Tsarfaty	65
<i>MT2ST: Adaptive Multi-Task to Single-Task Learning</i>	
Dong Liu and Yanxuan Yu	79
<i>Cross-Modal Augmentation for Low-Resource Language Understanding and Generation</i>	
Zichao Li and Zong Ke	90
<i>FORTIFY: Generative Model Fine-tuning with ORPO for ReTrieval Expansion of InFormal NoisY Text</i>	
Dan DeGenaro, Eugene Yang, David Etter, Cameron Carpenter, Kate Sanders, Alexander Martin, Kenton Murray and Reno Kriz	100