

1st Workshop on Language + Molecules (Lang + Mol 2024)

Bangkok, Thailand
15 August 2024

ISBN: 979-8-3313-0167-5

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© (2024) 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>L+M-24: Building a Dataset for Language+Molecules @ ACL 2024</i> Carl Edwards, Qingyun Wang, Lawrence Zhao and Heng Ji	1
<i>Could Chemical Language Models benefit from Message Passing</i> Jiaqing Xie and Ziheng Chi	11
<i>ALMol: Aligned Language-Molecule Translation LLMs through Offline Preference Contrastive Optimisation</i> Dimitris Gkoumas	22
<i>Evaluating Extrapolation Ability of Large Language Model in Chemical Domain</i> Taehun Cha and Donghun Lee	29
<i>Design Proteins Using Large Language Models: Enhancements and Comparative Analyses</i> Kamyar Zeinalipour, Neda Jamshidi, Monica Bianchini, Marco Maggini and Marco Gori	35
<i>Enhanced BioT5+ for Molecule-Text Translation: A Three-Stage Approach with Data Distillation, Diverse Training, and Voting Ensemble</i> Qizhi Pei, Lijun Wu, Kaiyuan Gao, Jinhua Zhu and Rui Yan	49
<i>ChatMol Copilot: An Agent for Molecular Modeling and Computation Powered by LLMs</i> Jinyuan Sun, Auston Li, Yifan Deng and Jiabo Li	56
<i>SciMind: A Multimodal Mixture-of-Experts Model for Advancing Pharmaceutical Sciences</i> Zhaoping Xiong, Xintao Fang, Haotian Chu, Xiaozhe Wan, Liwei Liu, Yameng Li, Wenkai Xiang and Mingyue Zheng	67
<i>Knowledge Graph Extraction from Total Synthesis Documents</i> Andres M Bran, Zlatko Jončev and Philippe Schwaller	75
<i>NLPeople at L+M-24 Shared Task: An Ensembled Approach for Molecule Captioning from SMILES</i> Shinnosuke Tanaka, Carol Mak, Flaviu Cipcigan, James Barry, Mohab Elkaref, Movina Moses, Vishnudev Kuruvanthodi and Geeth De Mel	86
<i>Knowlab's Submission to L+M Shared Task: All you need is continued pretraining of chemistry texts even for molecule captioning</i> Yunsoo Kim and Honghan Wu	92
<i>Mol2Lang-VLM: Vision- and Text-Guided Generative Pre-trained Language Models for Advancing Molecule Captioning through Multimodal Fusion</i> Duong Thanh Tran, Nhat Truong Pham, Nguyen Doan Hieu Nguyen and Balachandran Manavalan	98
<i>DNA Language Model and Interpretable Graph Neural Network Identify Genes and Pathways Involved in Rare Diseases</i> Ali Saadat and Jacques Fellay	104
<i>Repurformer: Transformers for Repurposing-Aware Molecule Generation</i> Changhun Lee and Gyumin Lee	117
<i>Lang2Mol-Diff: A Diffusion-Based Generative Model for Language-to-Molecule Translation Leveraging SELFIES Representation</i> Nguyen Doan Hieu Nguyen, Nhat Truong Pham, Duong Thanh Tran and Balachandran Manavalan	129