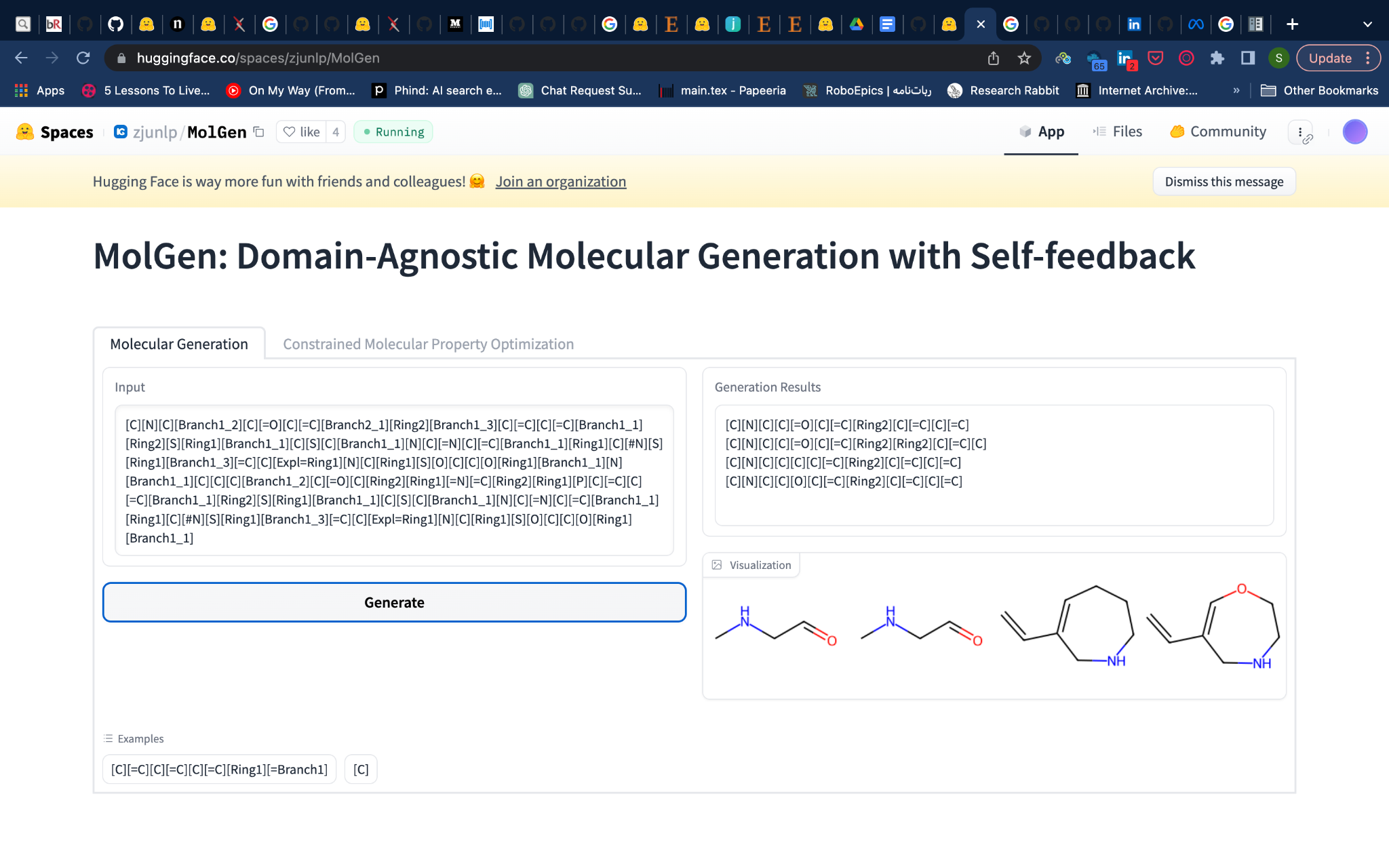
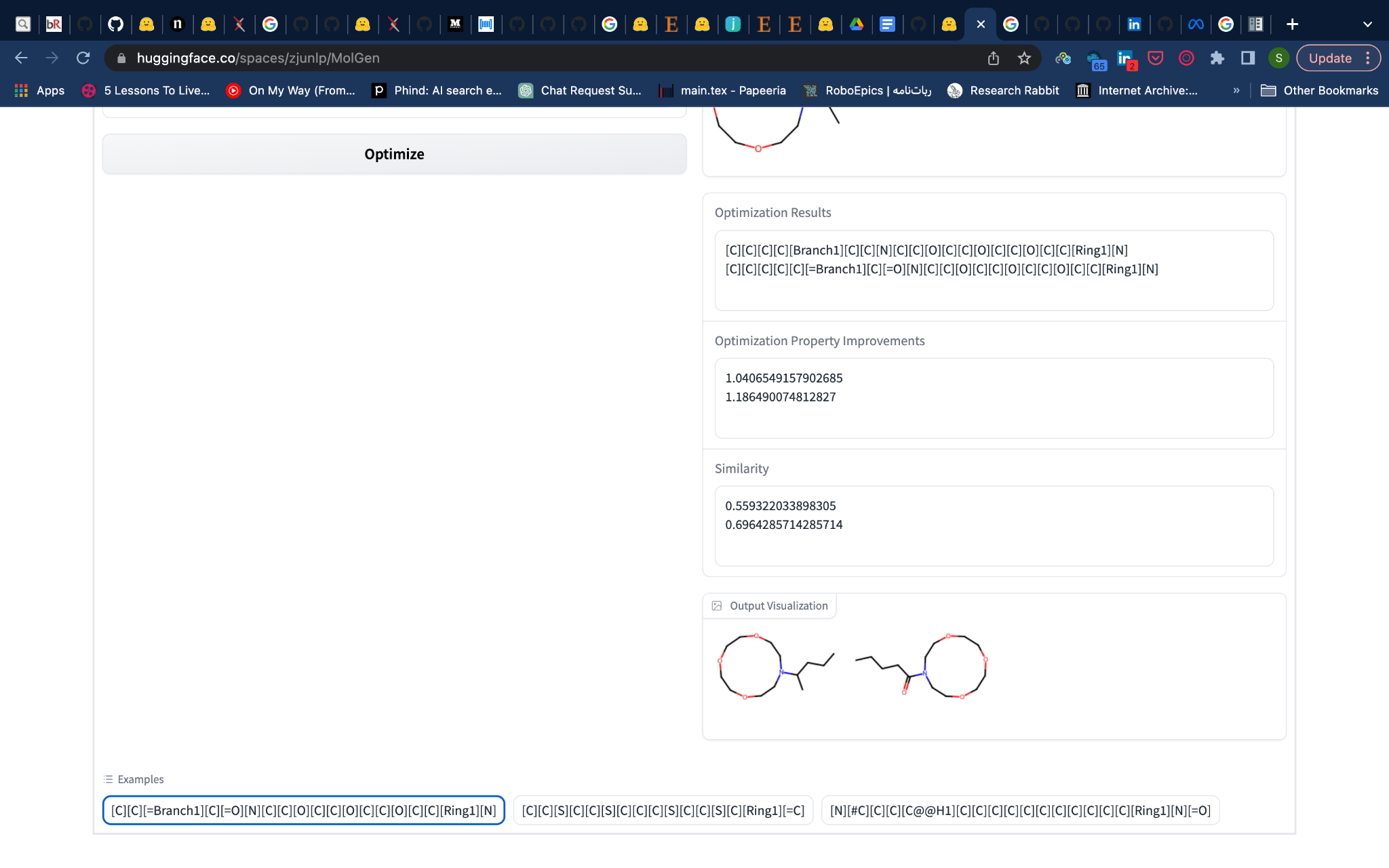
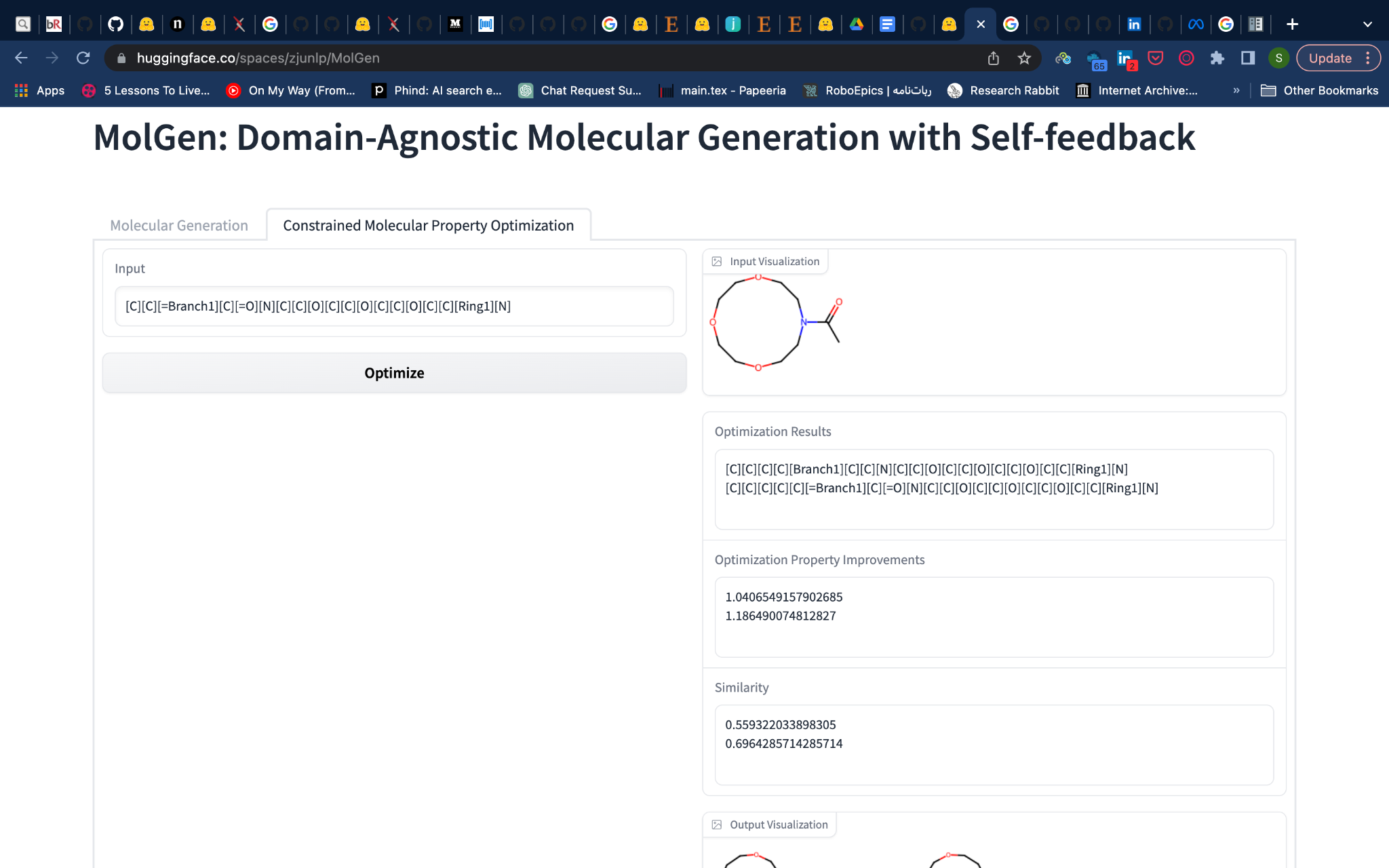
# Sources:

MolGen: <https://github.com/zjunlp/MolGen>

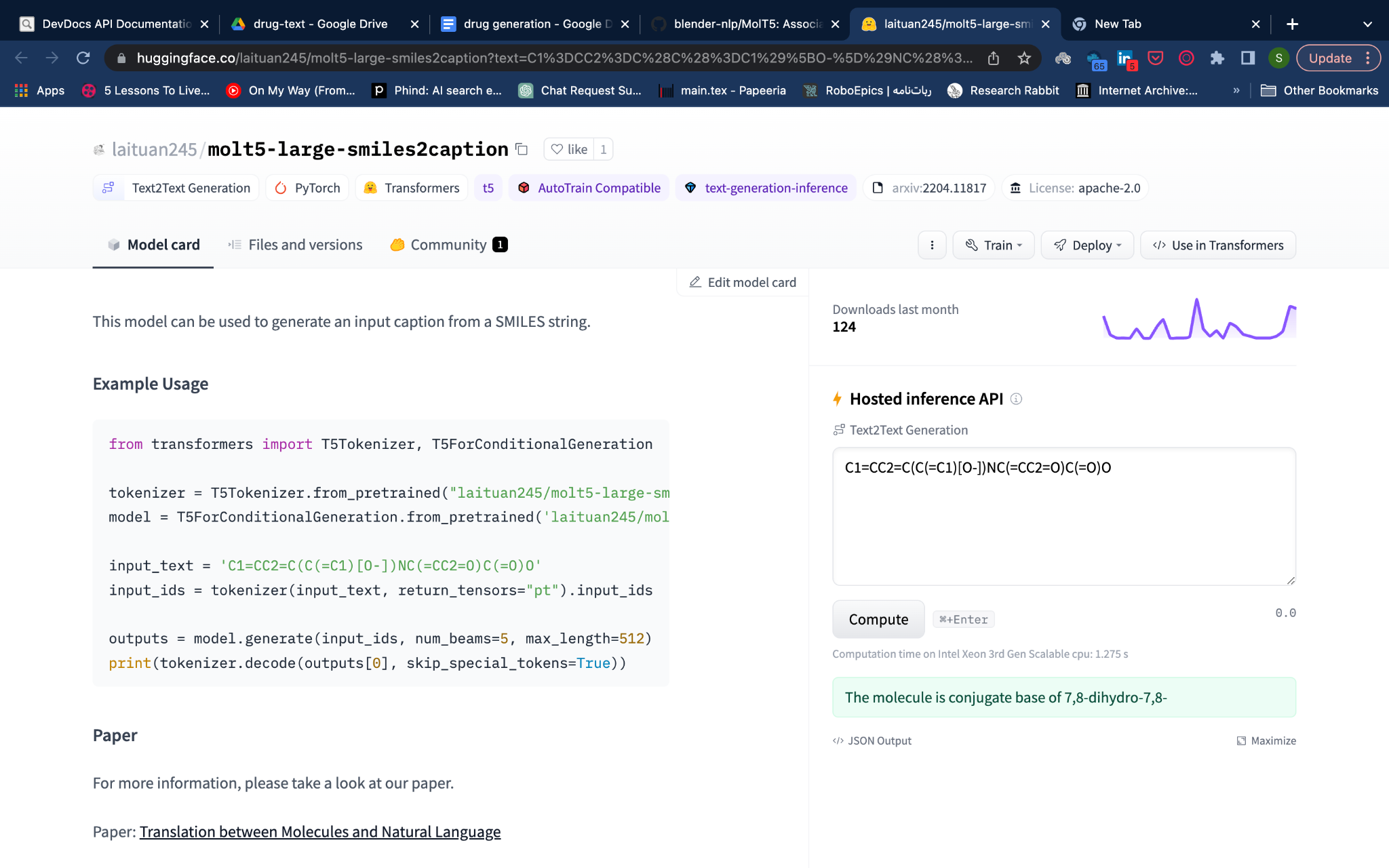
* Molecule generation using SELFIES
* Weights are available





MolT5: <https://github.com/blender-nlp/MolT5>

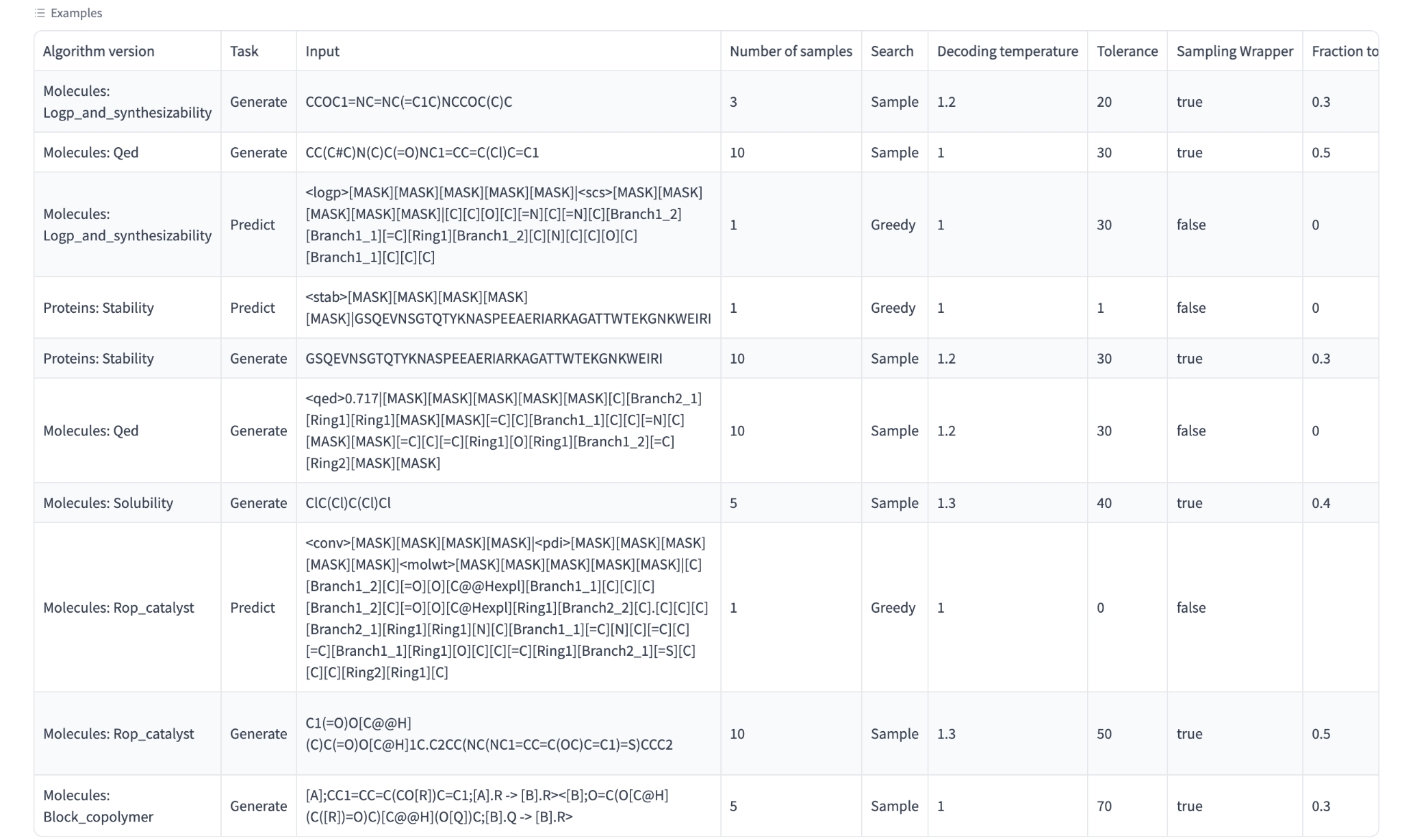
* Smiles2caption and Caption2Smiles
* Weights are avalible



# [Multitask Text and Chemistry T5](https://arxiv.org/abs/2301.12586): <https://github.com/GT4SD/multitask_text_and_chemistry_t5>



# Regression Transformer (generate and predict): <https://github.com/IBM/regression-transformer>



# Molecular Property Prediction using a Smiles Tokenization Strategy: <https://github.com/deepchem/deepchem/blob/master/examples/tutorials/Transfer_Learning_With_ChemBERTa_Transformers.ipynb>

RNN-based SMILE generation: <https://github.com/grisoniFr/de_novo_design_RNN/tree/main>

# GPTs:

* MolGPT: Molecular Generation Using a Transformer-Decoder Model. [[PAPER]](https://chemrxiv.org/engage/chemrxiv/article-details/60c7588e469df48597f456ae) [[REPO]](https://github.com/devalab/molgpt)
* ChemGPT: Neural Scaling of Deep Chemical Models. [[PAPER]](https://chemrxiv.org/engage/chemrxiv/article-details/627bddd544bdd532395fb4b5) [[REPO]](https://github.com/ncfrey/litmatter)
* Material transformers: deep learning language models for generative materials design. [[PAPER]](https://iopscience.iop.org/article/10.1088/2632-2153/acadcd/meta) [[REPO]](https://github.com/usccolumbia/MTransformer)
* Galactica: A Large Language Model for Science. [[PAPER]](https://arxiv.org/abs/2211.09085) [[REPO]](https://www.youtube.com/watch?v=dQw4w9WgXcQ)
* cMolGPT: A Conditional Generative Pre-Trained Transformer for Target-Specific De Novo Molecular Generation. [[PAPER]](https://www.mdpi.com/1420-3049/28/11/4430) [[REPO]](https://github.com/VV123/cMolGPT)
* SmileGPT: <https://github.com/sanjaradylov/smiles-gpt>
* ChatDrug: <https://github.com/chao1224/ChatDrug>

# GPT-GNN: Generative Pre-Training of Graph Neural Networks: <https://github.com/UCLA-DM/GPT-GNN>

<https://huggingface.co/ncfrey/ChemGPT-4.7M>

<https://huggingface.co/seyonec/ChemBERTa-zinc-base-v1>

<https://huggingface.co/recobo/chemical-bert-uncased>

<https://huggingface.co/recobo/chemical-bert-uncased-pharmaceutical-chemical-classifier>

<https://huggingface.co/mrm8488/chEMBL_smiles_v1>

<https://huggingface.co/microsoft/biogpt?text=what+is+smile+format+of+asprin>

gptChem: <https://github.com/kjappelbaum/gptchem>

* Need openAI key (api)

<https://huggingface.co/liyuesen/druggpt>

# Data Augmentation:

# Data:

<https://huggingface.co/datasets/haitengzhao/molecule_property_instruction>

# Tools:

Map the fingerprints: <https://github.com/rxn4chemistry/rxnfp>

* generates chemical reaction fingerprints from reaction SMILES

## Uni-Mol: <https://github.com/dptech-corp/Uni-Mol>

* Has some functionality like binding pose prediction

T5 Chem: Unified Deep Learning Model for Multitask Reaction Predictions with Explanation. [[REPO]](https://yzhang.hpc.nyu.edu/T5Chem)

KV-PLM: <https://github.com/thunlp/KV-PLM>

# ChemBERTa: <https://github.com/seyonechithrananda/bert-loves-chemistry>

MolBERT: <https://github.com/BenevolentAI/MolBERT>

MoLFormer: <https://github.com/IBM/molformer>

MatSciBERT: <https://github.com/M3RG-IITD/MatSciBERT>

Molecule Attention Transformer: <https://github.com/ardigen/MAT>, <https://github.com/gnina/SolTranNet>

GIMLET: <https://github.com/zhao-ht/GIMLET>

CDDD: <https://github.com/jrwnter/cddd>

<https://github.com/Kohulan/Smiles-TO-iUpac-Translator>

CPA: <https://www.linkedin.com/posts/visionarynet_ai-artificialintelligence-machinelearning-activity-6788813800876380160-kJS9/>

# SELFormer: Molecular Representation Learning via SELFIES Language Models: <https://github.com/HUBioDataLab/SELFormer>

Pretrained SMILES transformation model for fine tuning for diverse molecular tasks: [https://github.com/MolecularAI/MolBART](https://github.com/MolecularAI/MolBART#2)

ModBERT: <https://github.com/junxia97/Mole-BERT>

# Molecular Geometry Pretraining with SE(3)-Invariant Denoising Distance Matching: <https://github.com/chao1224/GeoSSL>

# Pre-training via Denoising for Molecular Property Prediction: <https://github.com/shehzaidi/pre-training-via-denoising>

semi-supervised learning for molecular property prediction: <https://github.com/zhang-xuan1314/Molecular-graph-BERT>

MolecularTransformer: <https://github.com/pschwllr/MolecularTransformer>

TransformerMolecules: <https://github.com/pfizer-opensource/transform-molecules>

[SMILES Transformer](http://arxiv.org/abs/1911.04738): <https://github.com/DSPsleeporg/smiles-transformer>

MolBART: <https://github.com/MolecularAI/Chemformer>

### IUPAC2Struct: <https://github.com/sergsb/IUPAC2Struct>

CLAMP: <https://github.com/ml-jku/clamp>

<https://github.com/chemprop/chemprop>

<https://huggingface.co/google/flan-t5-xl?text=Premise%3A++At+my+age+you+will+probably+have+learnt+one+lesson.+Hypothesis%3A++It%27s+not+certain+how+many+lessons+you%27ll+learn+by+your+thirties.+Does+the+premise+entail+the+hypothesis%3F>

# References:

Chemical language models for de novo drug design - Challenges and opportunities

<https://github.com/junxia97/awesome-pretrain-on-molecules>

<https://github.com/alxfgh/Large-Language-Models-in-Chemistry>

<https://github.com/AspirinCode/papers-for-molecular-design-using-DL>

<https://github.com/HHW-zhou/LLM4Mol>