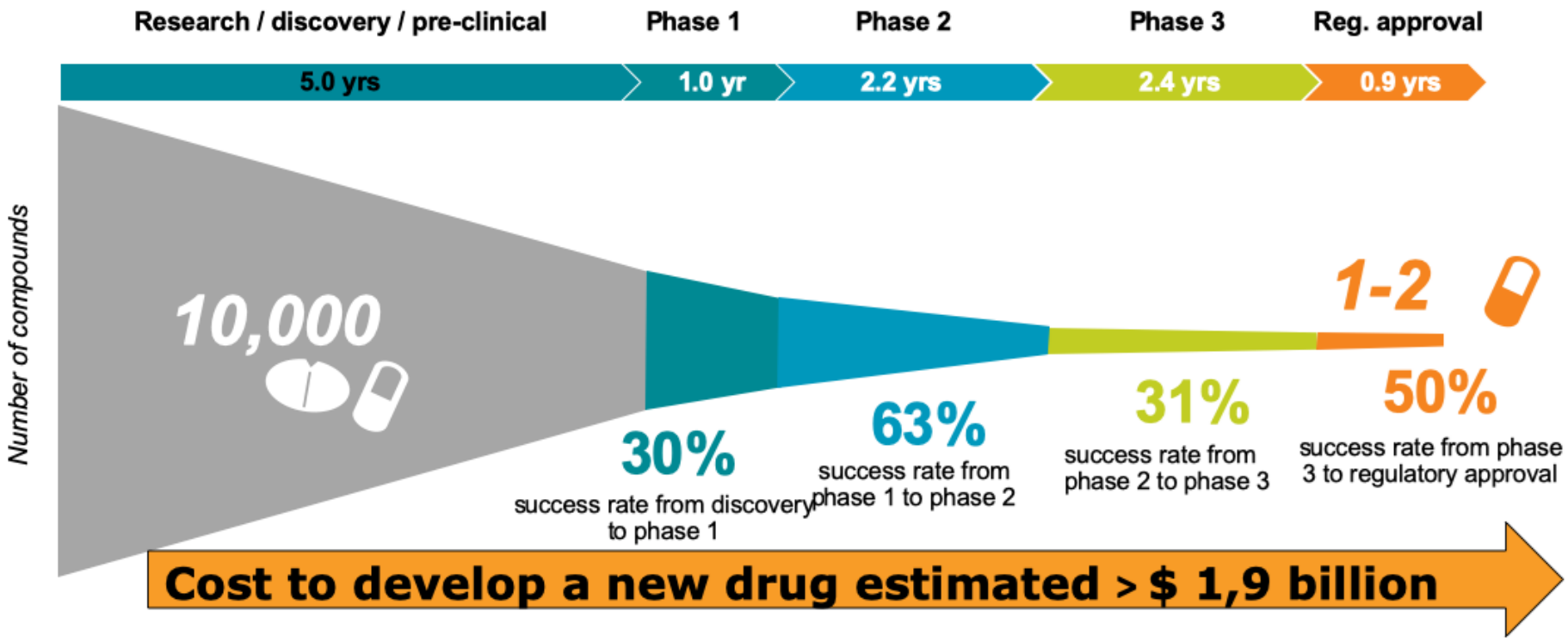
Advancing Drug Discovery with AI: An Integrated Fragment-Based Generative Model

Yuto Suzuki and Farnoush Banaei-Kashani
Department of Computer Science
University of Colorado Denver

Potential of Generative AI in Drug Discovery

- Drug discovery is a very time-consuming and expensive process
- Generative AI has been getting attention to accelerate drug discovery



https://www.efpia.eu/about-medicines/development-of-medicines/smes-in-europe/

Unique Challenges of Generative AI in Drug Discovery

- Extremely limited data and complex atom connections are the main challenges for AI to generate molecules

Extremely Limited Data

Complex Atom Connections

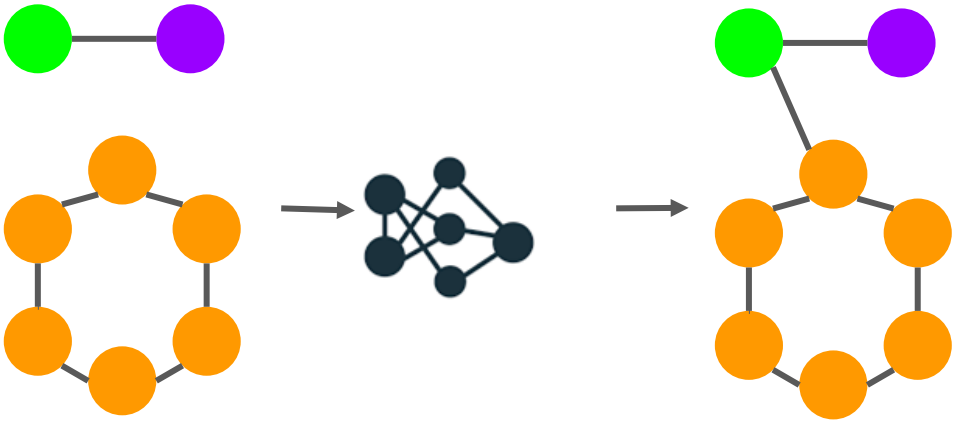
Types of Generative AI for Molecules

- Fragment-based generative model can generate more realistic structures

Atom-based generative AI

× Complex structures

Fragment-based generative AI



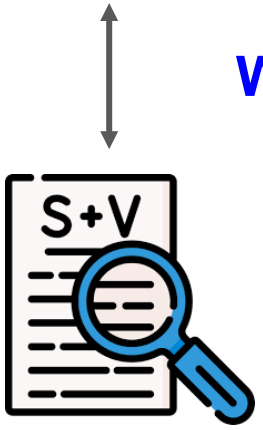
○ Complex structures

The Selection of Useful Fragments is Crucial

- Fragments are akin to words
- Understanding sentences requires knowledge of words and grammar

Human Languages

This | is | a | pen.



Words and grammar should be learned together

Our approach

fragment | fragment

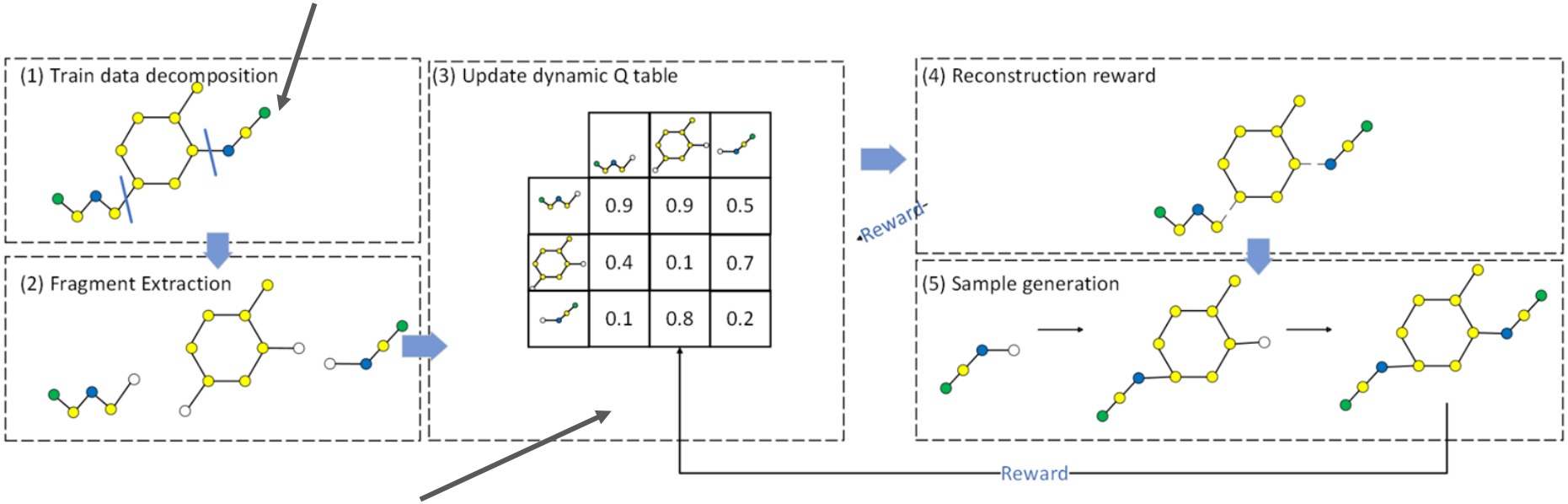
Generative AI

Fragments and generative AI are learned together to find best combination

Model Overview

AI model predicts molecule decomposition (extracts words)

Words

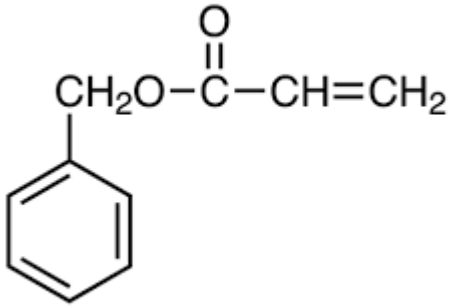


AI model which combines fragments (grammar)

Grammar

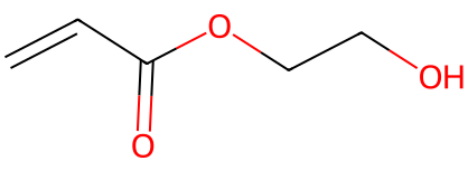
Experimental Methodology

32 known Acrylates



Train generative AI

Generate new synthesizable Acrylates



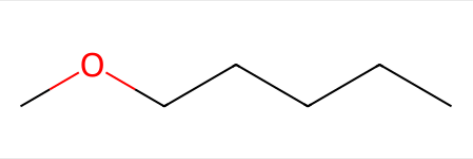
Experimental Results

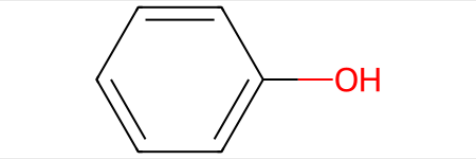
- Our AI generated 64 new and synthesizable Acrylates out of 1,000 samples, whereas the state-of-the-art model produced only 39. (**60% increase**)

Results with 32 Acrylates

Model	Dis w/	Dis w/o	Valid	RS	Unique	Novel	Cham.	Div.	Mem.
GraphNVP	-	-	0	-	-	-	-	-	-
HierVAE (w/ft)	0	2.6%	100%	**0.98**	3%	**100%**	0.44	0.67	0%
HierVAE (w.ft)	0	7.8%	100%	0.56	32%	95%	0.20	0.77	0%
DEG	3.9%	13.6%	100%	0.33	70%	**100%**	**0.63**	**0.87**	45%
LVSEF(ran)	6.2%	19.2%	100%	0.35	**72%**	**100%**	0.60	0.84	**51%**
LVSEF (bal)	**6.4%**	**20.3%**	100%	0.44	63%	**100%**	0.52	0.84	**51%**

Sample fragments





Future Work

- Apply this AI model to real-world drug discovery (KDM4 inhibitors)
- Develop an AI model that can modify molecules while preserving their specific functionalities

Thank you very much.

Acknowledgement

- This work used the computing resources at the Center for Computational Mathematics, University of Colorado Denver, including the Alderaan cluster, supported by the National Science Foundation award OAC-2019089.
- We would like to express our deepest gratitude to Professor Daniel LaBarbera, PhD, Director of the CU AMC Center for Drug Discovery and Co-Director of the Drug Discovery and Development Shared Resource (D3SR), for his invaluable collaboration on this research. We are also thankful to his dedicated team for their unwavering support and contributions.