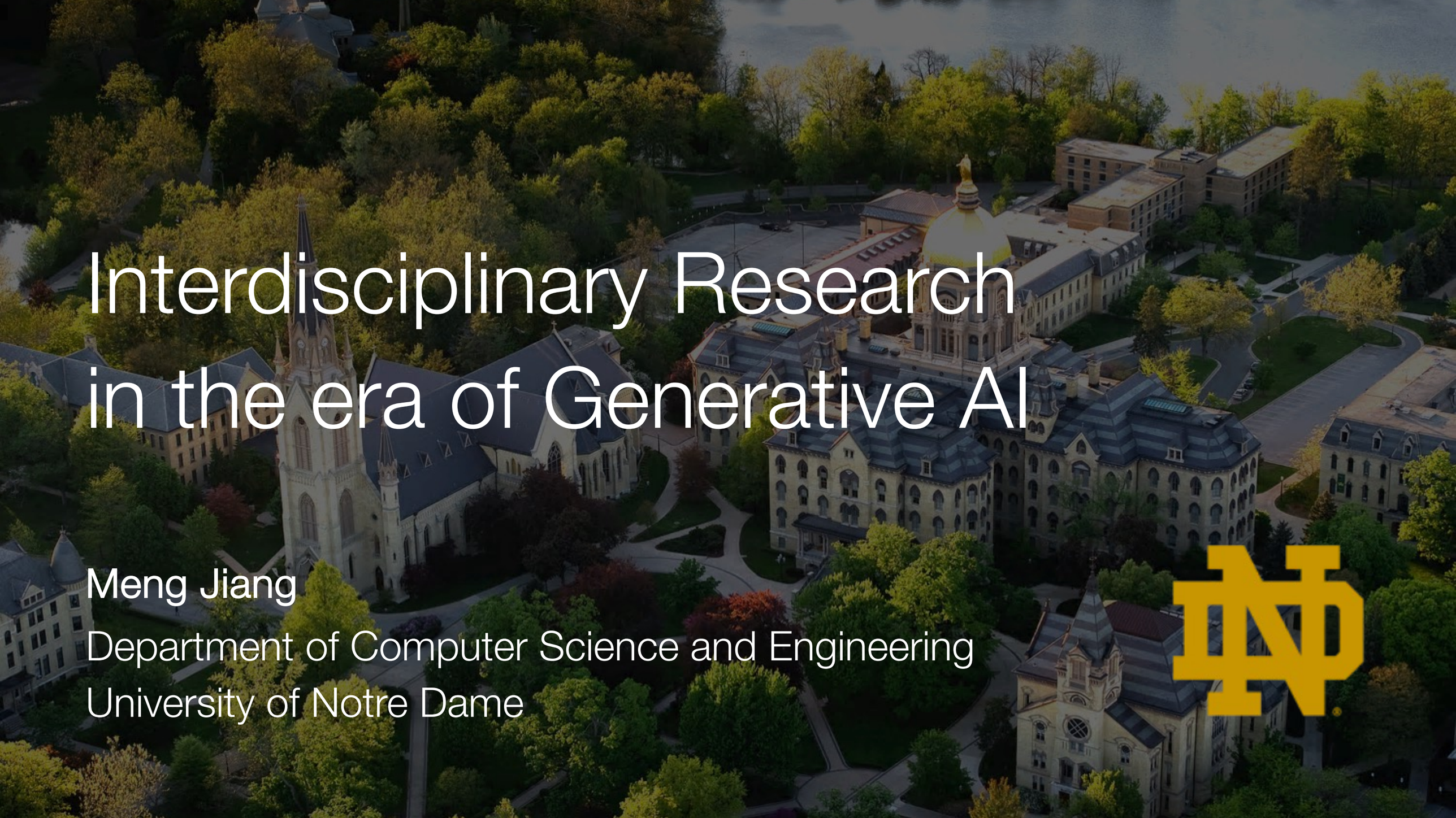


Interdisciplinary Research in the era of Generative AI

Meng Jiang
Department of Computer Science and Engineering
University of Notre Dame

My Background

*Graph and Text **Data Mining**: The process of knowledge discovery*

Tsinghua (2010-2015)

Given a very **large** text-attributed graph, how to predict links for low-degree nodes? (Social recommendation)

Jiang et al. Social context recommendation. CIKM 2012.
Jiang et al. FEMA: … for dynamic behavioral pattern discovery. KDD 2014.

CMU (2012-2013)

Given a very **large** graph, how to detect suspicious nodes? (Fake account detection)

Jiang et al. Catchsync: Catching synchronized behavior in large directed graphs. KDD 2014. (Best paper finalist)

UIUC (2015-2017)

Given a very **large** corpus, how to build a knowledge graph with minimal human effort? (Information extraction).

Jiang et al. MetaPAD: Meta pattern discovery from massive text corpora. KDD 2017.

Discover Knowledge for Broader Impact

Besides social network analysis and information systems

- KDD is mining useful and/or surprising patterns from big messy data.
  - Data processing, data engineering, data mining … on graph and text data
- Generative AI is bringing surprises:
  - A piece of accurate answer, a piece of nice story, an image or a 1-min video created from a few sentences…
  - Thanks to the big data and all data techs that make it less messy for exhaustive training.
  - Generative models are making **small data** useful and impactful.
- What can we do with graph and text generative models?
- This talk introduces a few interdisciplinary research projects at Notre Dame DM^2 Lab.

WARNING

This talk contains different types of techniques and has very limited detail. For the detail, you are welcome to ask questions at the end of the talk. You can find related papers cited at the bottom of slides.

Slide numbers are at the bottom right to help Q&A. Pages 23 and 34 may help you write a summary. Pages 7, 8, 11, 21, 32 may help too.

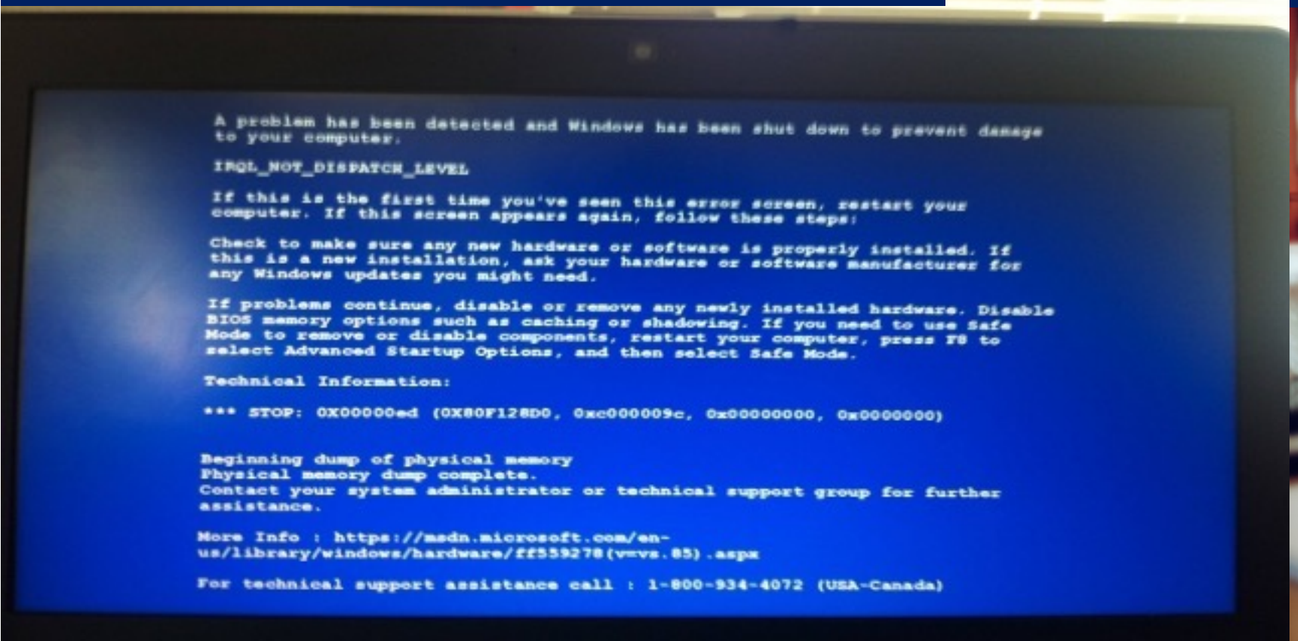
This talk would not be possible without our collaborators and amazing students that you will find at the top right.

[1] XX et al. (20XX). XXX, XXX.

Material Science

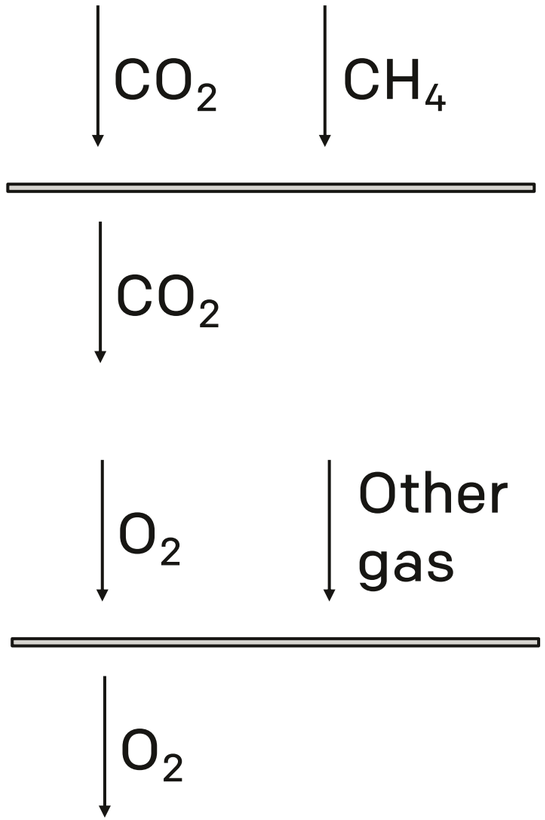
Online Education

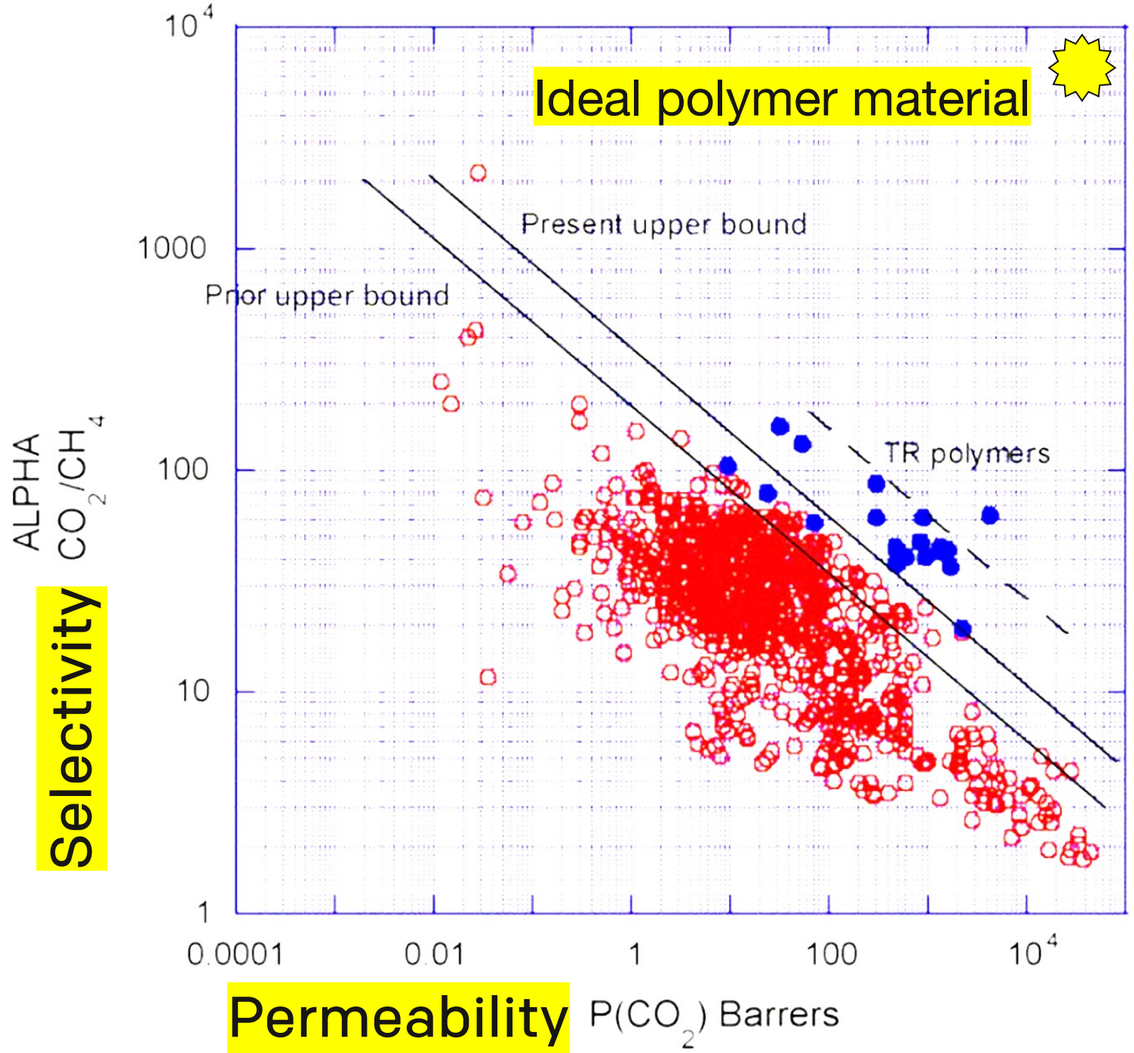
Software & Network Security



Mental Health

Gas-separation Polymer Material [1]

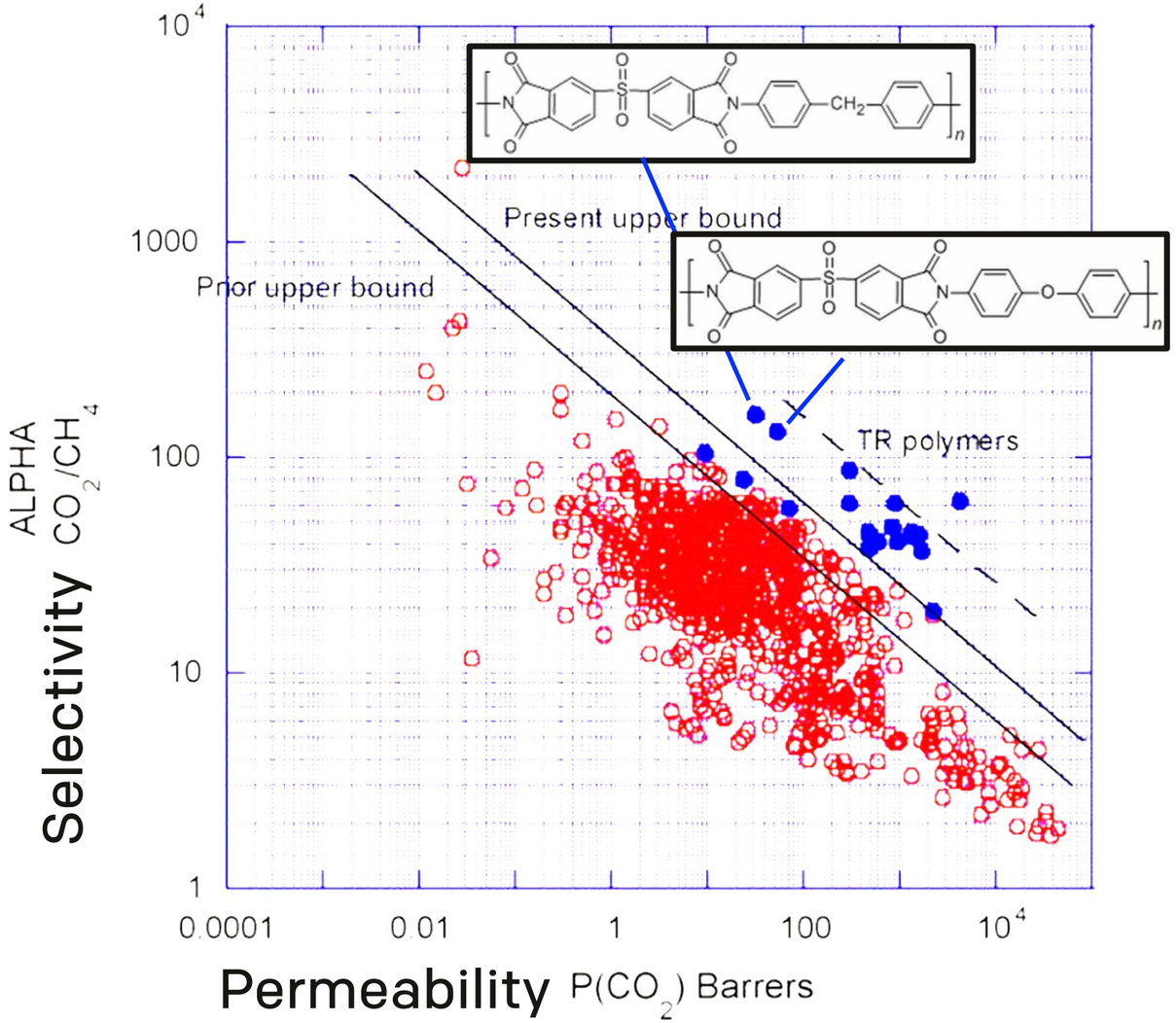




[1] Barnett et al. (2020). Designing exceptional gas-separation polymer membranes using machine learning. Science Advances, 6(20), eaaz4301.

Research on Polymer Discovery

- Red points: ~**700** polymeric constructs that have been measured to date (60+ years)
- **1991** Robeson upper bound [1]
- Blue points: “Thermally rearranged (TR) polymers, which are considered the **next-generation of membrane materials** because of their excellent properties”
- **2008** Robeson upper bound



[1] Tena et al. (2016). Claisen thermally rearranged polymers. Science Advances, 2(7), e1501859.

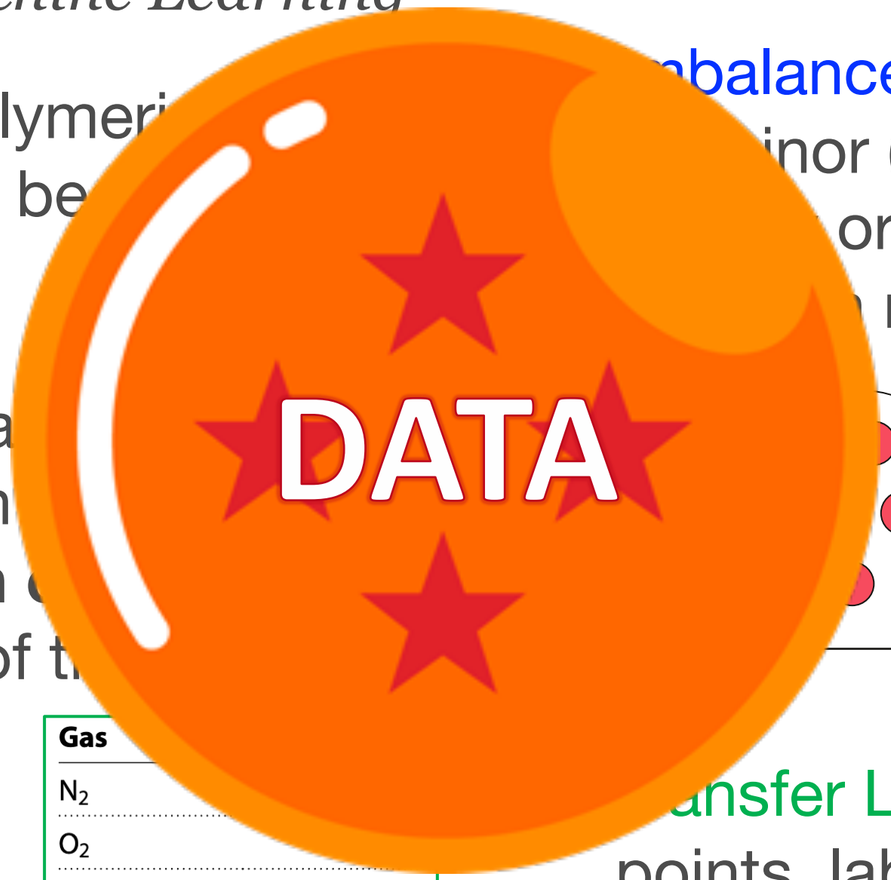
Graph Regression

Problems in Graph Machine Learning

- Red points: ~**700** polymeri
  constructs that have be
  to date (60+ years)

- Blue points: "Therma
  (TR) polymers, which
  the **next-generation**
  materials because of t
  properties"

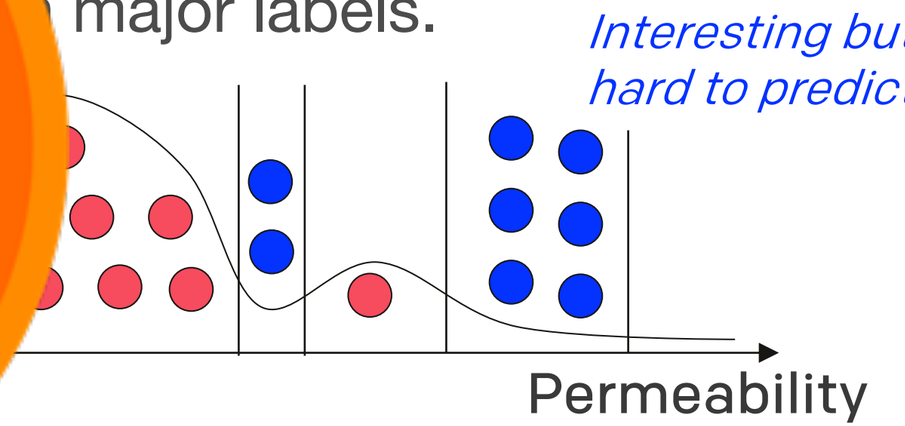
- Over 1.2 million unlabeled points
- Six different gases



Gas	
N_2	
O_2	
H_2	324
He	282
CH_4	420
CO_2	471

- Supervised Learning: Training sets of ~500 examples are too limited.

- balanced Learning: Predictions
  inor (continuous) labels are
  or more important than
  major labels.



- ansfer Learning: Extra data points, labeled or unlabeled, the same domain or different domains, could be useful.

[1] Tena et al. (2016). Claisen thermally rearranged polymers. Science Advances, 2(7), e1501859.

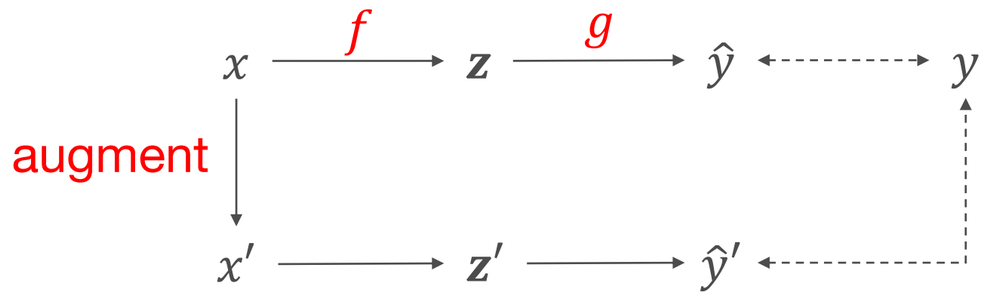
Data Augmentation

in Machine Learning

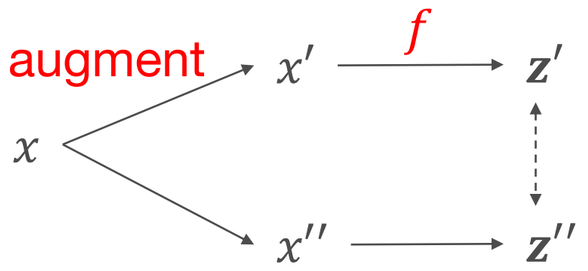
What is data augmentation?

> Data augmentation techniques are used to increase the amount of data by adding slightly modified copies of existing data or newly created synthetic data from existing data. It helps reduce overfitting when training machine learning models. — Wikipedia [1]

Supervised learning



Self-supervised learning

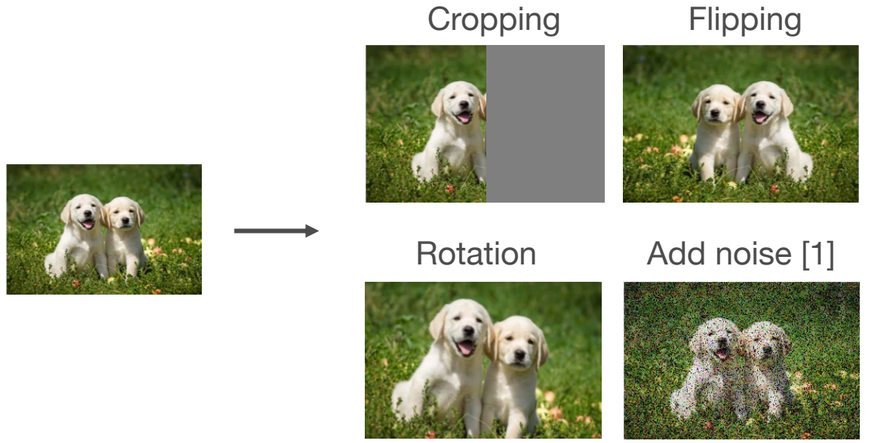


[1] https://en.wikipedia.org/wiki/Data_augmentation

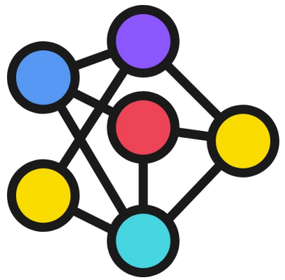
Data Augmentation

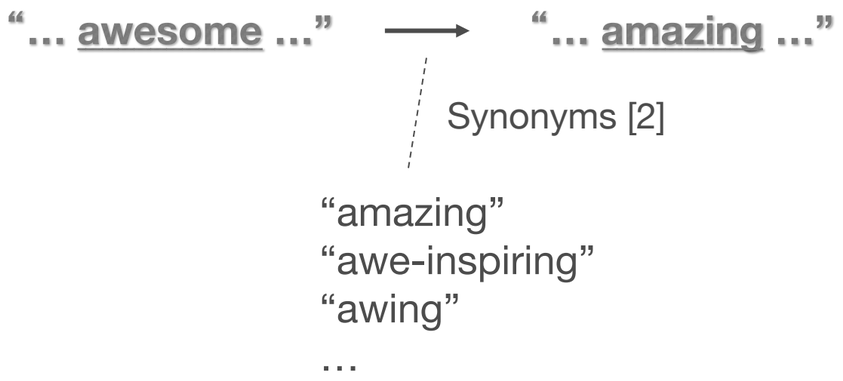
in Computer Vision and NLP

- Data augmentation improved learning.
- Image and text data could be augmented.
- Heuristics were effective for augmentation.



Synthetic nodes / links in a large graph?





- Dependencies have to be learned to augment data.

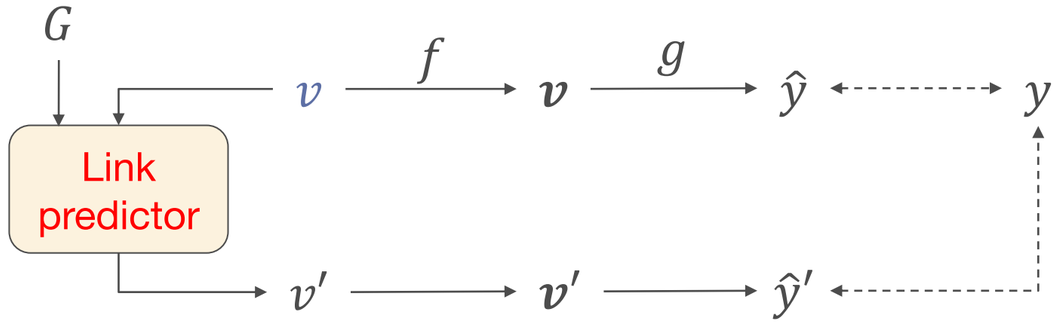
[1] Zhong et al. “Random erasing data augmentation.” AAAI 2020.
[2] Wei et al. “EDA: Easy data augmentation techniques for boosting performance on text classification tasks.” EMNLP 2019.

Data Augmentation for GNNs

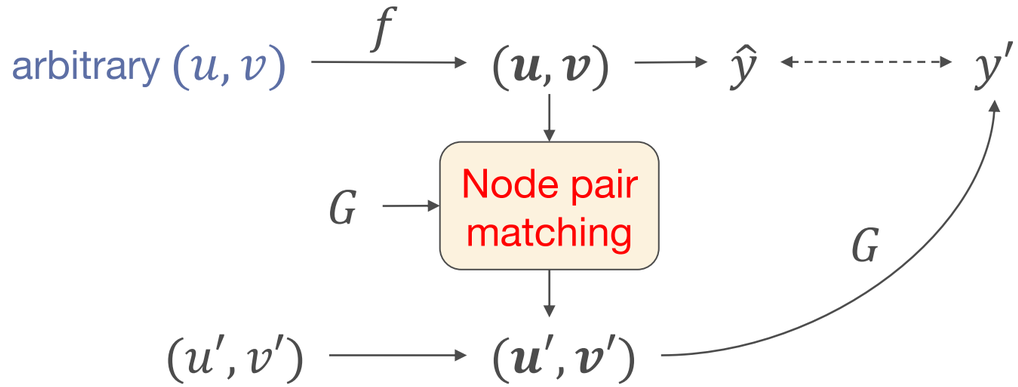
AAAI'21, ICML'22, LoG'22

Tong Zhao (Snap)

- Link prediction (addition and removal) created synthetic nodes in a large graph.
- Improved node classification accuracy by relatively +17% [1].



- Counterfactual link labels improved link prediction Hits@20 relatively by +16.4% [2].



- Learning to augment is effective for graphs.

[1] Zhao et al. "Data augmentation for graph neural networks." AAAI 2021.
[2] Zhao et al. "Learning from counterfactual links for link prediction." ICML 2022.

Learning to Augment Graph Data

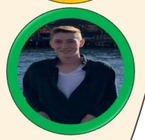
from Node- and Link-level tasks to Graph-level

IIS-2146761

III: Small:
Comprehensive Methods to Learn to Augment Graph Data
(2022-2025, $515,999)

PI: 

GRA:  

REU:  

- Supervised Learning

  Given only 500 labeled graphs, can we create synthetic labeled examples?

- Imbalanced Learning

  Given imbalanced numeric labels where we are interested in minor areas, can we create synthetic labeled examples to make data balanced?

- Transfer Learning

  Given 1.2 million unlabeled graphs and diverse downstream tasks (~500 labeled graphs each), can we guarantee positive transfer from unlabeled data?

Augment by Rationale-Environment Separation

KDD'22

Gang Liu

The separation was used in data augmentation for sentiment analysis [1].

i: I've just got home and one of my burgers was stone cold. → negative

"Environments" "Rationales"

j: I find that my French toast has a nice texture! → positive

Can you make a new example of a negative label using i and j?

Rationale of i + Environment of j: "I find that my French toast was stone cold!"

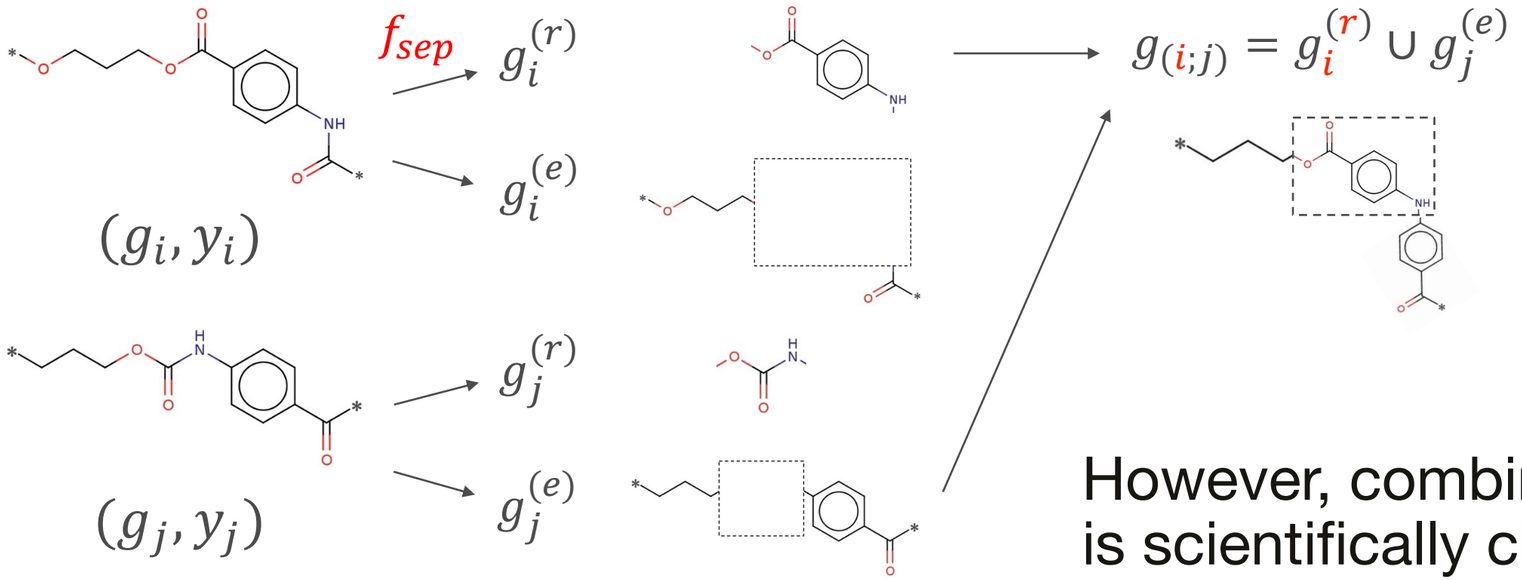
[1] Abonizio et al. "Toward text data augmentation for sentiment analysis." TAI 2021.

Augment by Rationale-Environment Separation

KDD'22

Gang Liu

Environment replacement on molecule/polymer graphs [1]:



What is likely its label?

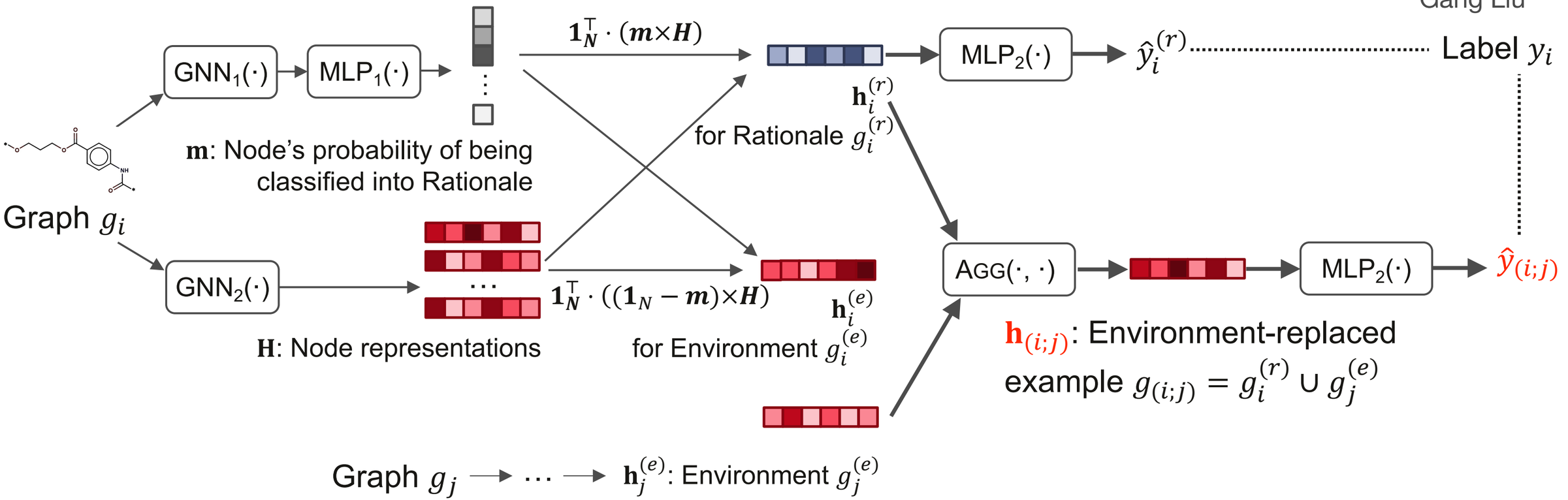
However, combining two chemical subgraphs is scientifically challenging!

[1] Liu et al. "Graph rationalization with environment-based augmentations." KDD 2022.

Augmentation in Latent Space

KDD'22

Gang Liu



- Increased molecule classification AUC from .738 to .779 on HIV, from .766 to .819 on BASE (drug).
- Reduced polymer regression MAE from 60.6 to 42.6 on MeltingTemp, from 770 to 524 on O_2Perm.
- Graph data augmentation can be performed in latent space!

[1] Liu et al. "Graph rationalization with environment-based augmentations." KDD 2022.

Learning to Augment Graph Data

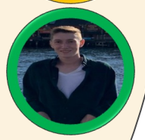
from Node- and Link-level tasks to Graph-level

IIS-2146761

III: Small:
Comprehensive Methods to Learn to Augment Graph Data
(2022-2025, $515,999)

PI: 

GRA:  

REU:  

- Supervised Learning

  Given only 500 labeled graphs, can we create synthetic labeled examples?

- Imbalanced Learning

  Given imbalanced numeric labels where we are interested in minor areas, can we create synthetic labeled examples to make data balanced?

- Transfer Learning

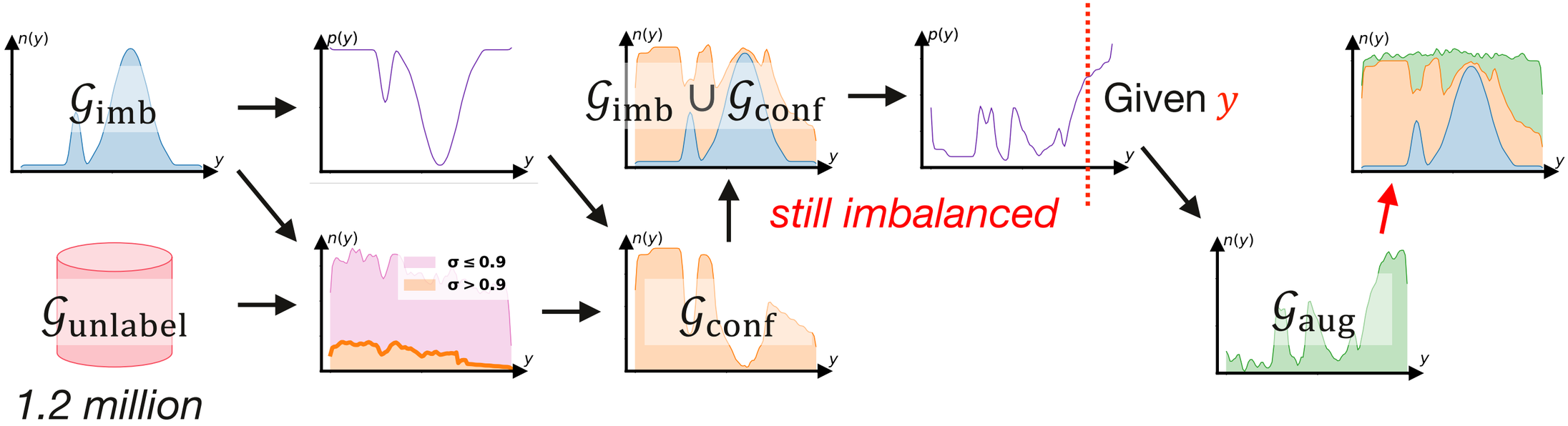
  Given 1.2 million unlabeled graphs and diverse downstream tasks (~500 labeled graphs each), can we guarantee positive transfer from unlabeled data?

Augment for Label Balance in Regression

KDD'23

Gang Liu

Can self-training *fully* address label imbalance in graph regression? [1]



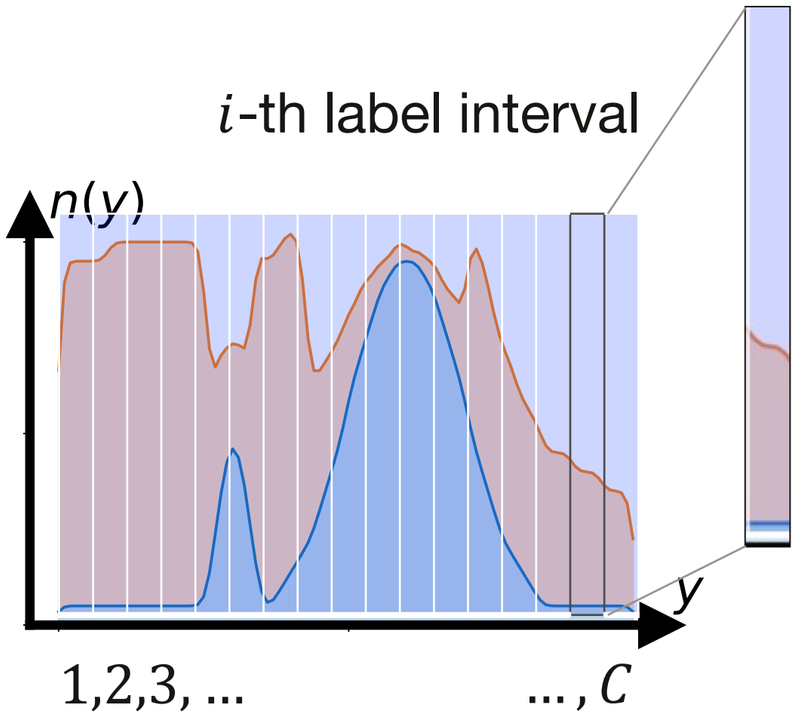
Explicit label-anchored graph decoding is technically challenging!

[1] Liu et al. "Semi-supervised graph imbalanced regression." KDD 2023.

In Latent Space: Label-anchored Mix-up

KDD'23

Gang Liu



$$\boldsymbol{z}_i = \text{mean}(\{\boldsymbol{h}'\})$$
$$a_i = \text{mean}(\{y'\})$$

Given a target label value y:

1. Sample a labeled example
$$(G_j \in \mathcal{G}_{\text{imb}} \cup \mathcal{G}_{\text{conf}}, y_j)$$

2. Get $\boldsymbol{h}_j = \text{GNN}(G_j)$

3. Sample a label interval $(\boldsymbol{z}_i, a_i)$

4. Constrain the sampling processes:
$$\boldsymbol{h} = \lambda \cdot \boldsymbol{z}_i + (1-\lambda) \cdot \boldsymbol{h}_j$$
$$s.t. \quad y = \lambda \cdot a_i + (1-\lambda) \cdot y_j$$

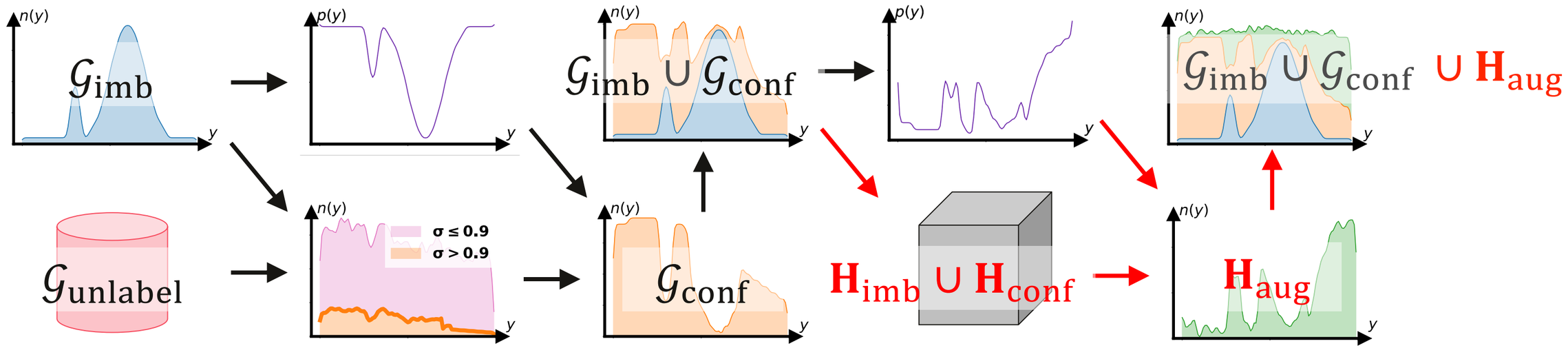
5. Augment with $(\boldsymbol{h}, y)$

Augment by Label-anchored Mix-up

KDD'23

Gang Liu

Semi-supervised graph imbalanced regression (SGIR):



- Reduced molecule regression MAE on FreeSolv from 1.154 to .777 in few-shot area, from .726 to .563 in all areas. Reduced polymer regression MAE on MeltingTemp from 54.7 to 51.4 in few-shot, from 41.8 to 38.9 in all areas.
- Again, graph data augmentation can be performed in latent space!

Exciting Results and Effort

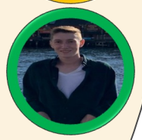
from Graph Regression and Data Augmentation to Polymer Informatics

IIS-2146761

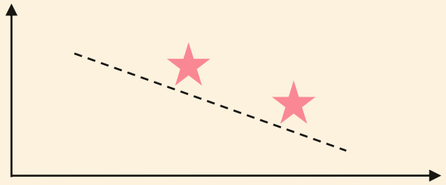
III: Small:
Comprehensive Methods to Learn to Augment Graph Data
(2022-2025, $515,999)

PI: 

GRA:  

REU:  

Found two novel polymers in gas separation. Patents under review.



NeurIPS 2024 competition: Open Polymer Challenge. Under review.

A data-centric cyber platform for polymer-by-design. Under construction.

- **Polymer property prediction**
- **Polymer inverse design**

CBET-2332270

CBET: Developing and Understanding Thermally Conductive Polymers by Combining Molecular Simulation, Machine Learning and Experiment
(2024-2026, $405,726)

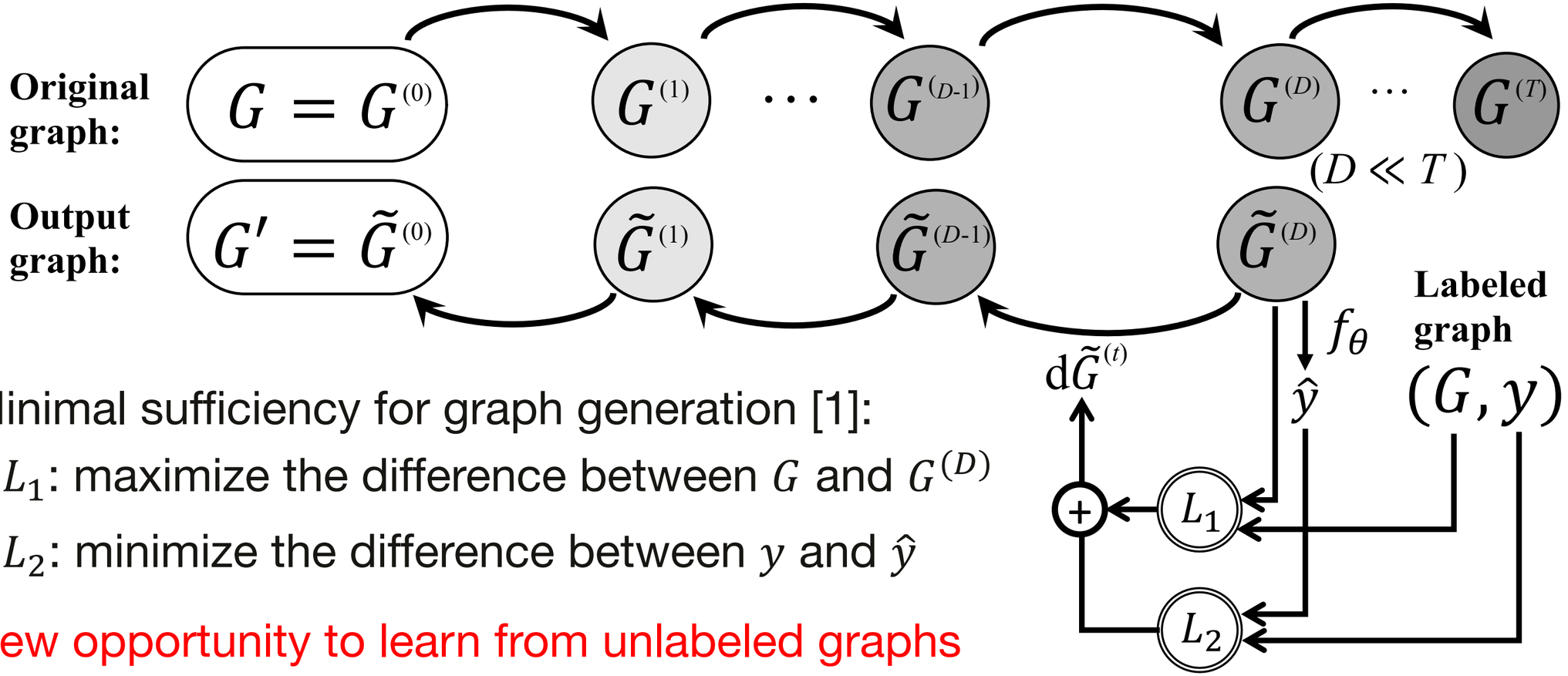
PI: 

Co-PI: 

Task data guide Graph Diffusion Transformer

NeurIPS'23

Gang Liu



- Minimal sufficiency for graph generation [1]:
  - L_1: maximize the difference between G and $G^{(D)}$
  - L_2: minimize the difference between y and $\hat{y}$
- New opportunity to learn from unlabeled graphs

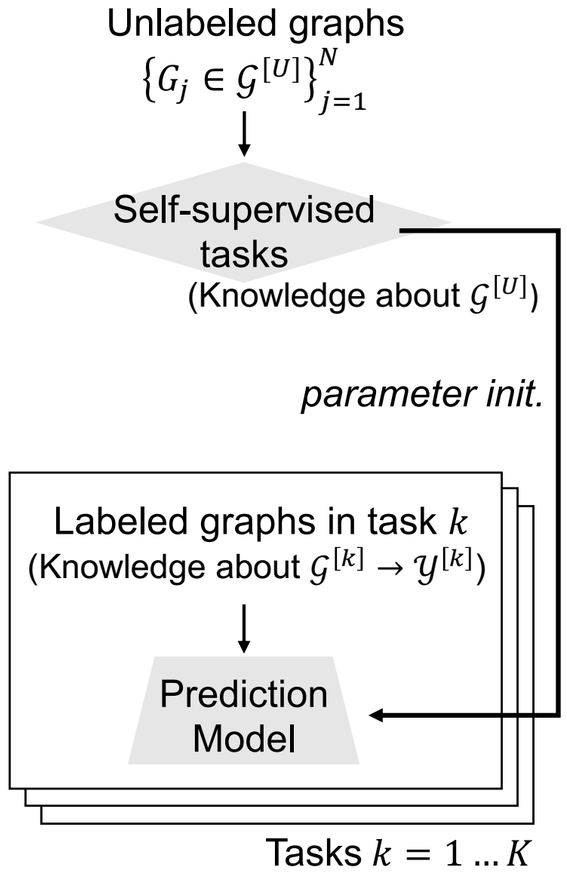
[1] Liu et al. "Data-centric learning from unlabeled graphs with diffusion model." NeurIPS 2023.

Data-Centric Transfer

NeurIPS'23

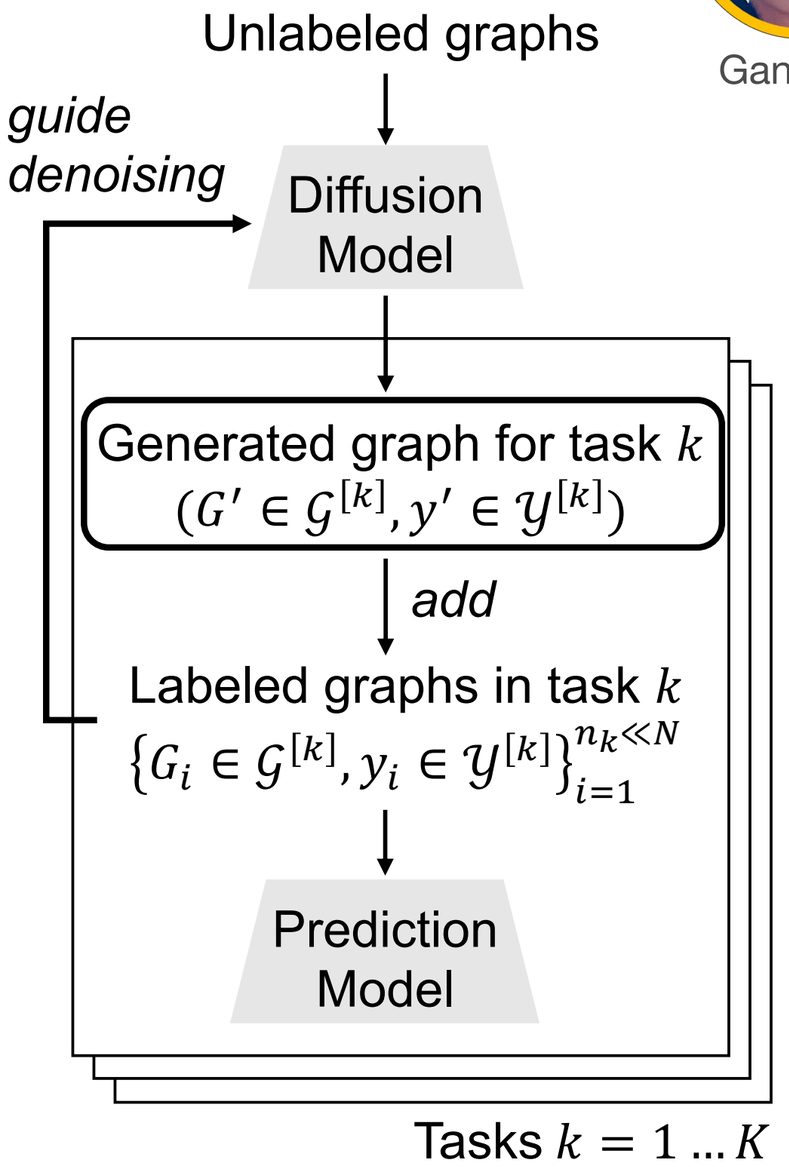
Gang Liu

Parameter-centric transfer:



Patterns learned from **self-supervised** tasks (e.g., reconstruction, perturbation) **might not be aligned** or even conflicting with **downstream** predictions.

Parameter initialization could **hardly interpret** how unlabeled graphs were able to improve prediction models.



Data-Centric Transfer

NeurIPS'23

Gang Liu

		Molecule Classification: AUC (%) ↑						
	# Training Graphs	ogbg-HIV 32,901	ogbg-ToxCast 6,860	ogbg-Tox21 6,264	ogbg-BBBP 1,631	ogbg-BACE 1,210	ogbg-ClinTox 1,181	ogbg-SIDER 1,141
	GIN	77.4(1.2)	66.9(0.2)	76.0(0.6)	67.5(2.7)	77.5(2.8)	88.8(3.8)	58.1(0.9)
Self-Supervised	EDGEPRED	78.1(1.3)	63.9(0.4)	75.5(0.4)	69.9(0.5)	79.5(1.0)	62.9(2.3)	59.7(0.8)
	ATTRMASK	77.1(1.7)	64.2(0.5)	76.6(0.4)	63.9(1.2)	79.3(0.7)	70.4(1.1)	60.7(0.4)
	CONTEXTPRED	78.4(0.1)	63.7(0.3)	75.0(0.1)	68.8(1.6)	75.7(1.0)	63.2(6.5)	60.7(0.8)
	INFOMAX	75.4(1.8)	61.7(1.0)	75.5(0.4)	69.2(0.5)	76.8(0.2)	73.0(0.2)	58.6(0.5)
	JOAO	76.2(0.2)	64.8(0.3)	74.8(0.5)	69.3(2.5)	75.9(3.9)	69.4(4.5)	60.8(0.6)
	GRAPHLOG	74.8(1.1)	63.2(0.8)	75.4(0.8)	67.5(2.3)	80.4(3.6)	69.0(6.6)	57.0(0.9)
	D-SLA	76.9(0.9)	60.8(1.2)	76.1(0.1)	62.6(1.0)	80.3(0.6)	78.3(2.4)	55.1(1.0)
Semi-SL	INFOGRAPH	73.3(0.7)	61.5(1.1)	67.6(0.9)	61.6(4.4)	75.9(1.8)	62.2(5.5)	56.3(2.3)
	ST-REAL	78.3(0.6)	64.5(1.0)	76.2(0.5)	66.7(1.9)	77.4(1.8)	82.2(2.4)	60.8(1.2)
	ST-GEN	77.9(1.6)	65.1(1.0)	75.8(0.9)	66.3(1.5)	78.4(3.0)	87.3(1.3)	59.3(1.3)
GDA	FLAG	74.6(1.7)	59.9(1.6)	76.9(0.7)	66.6(1.0)	79.1(1.2)	85.1(3.4)	57.6(2.3)
	GREA	79.3(0.9)	67.5(0.7)	77.2(1.2)	69.7(1.3)	82.4(2.4)	87.9(3.7)	60.1(2.0)
	G-MIXUP	77.1(1.1)	55.6(1.1)	64.6(0.4)	70.2(1.0)	77.8(3.3)	60.2(7.5)	56.8(3.5)
	DCT (Ours)	**79.5**(1.0)	**68.1**(0.2)	**78.2**(0.2)	**70.8**(0.5)	**85.6**(0.6)	**92.1**(0.8)	**63.9**(0.3)

- SSL may have negative impact on downstream.
- Augment downstream datasets by *graph generation*.
- Experts can evaluate and control the transfer.

When Material Science meets Generative AI

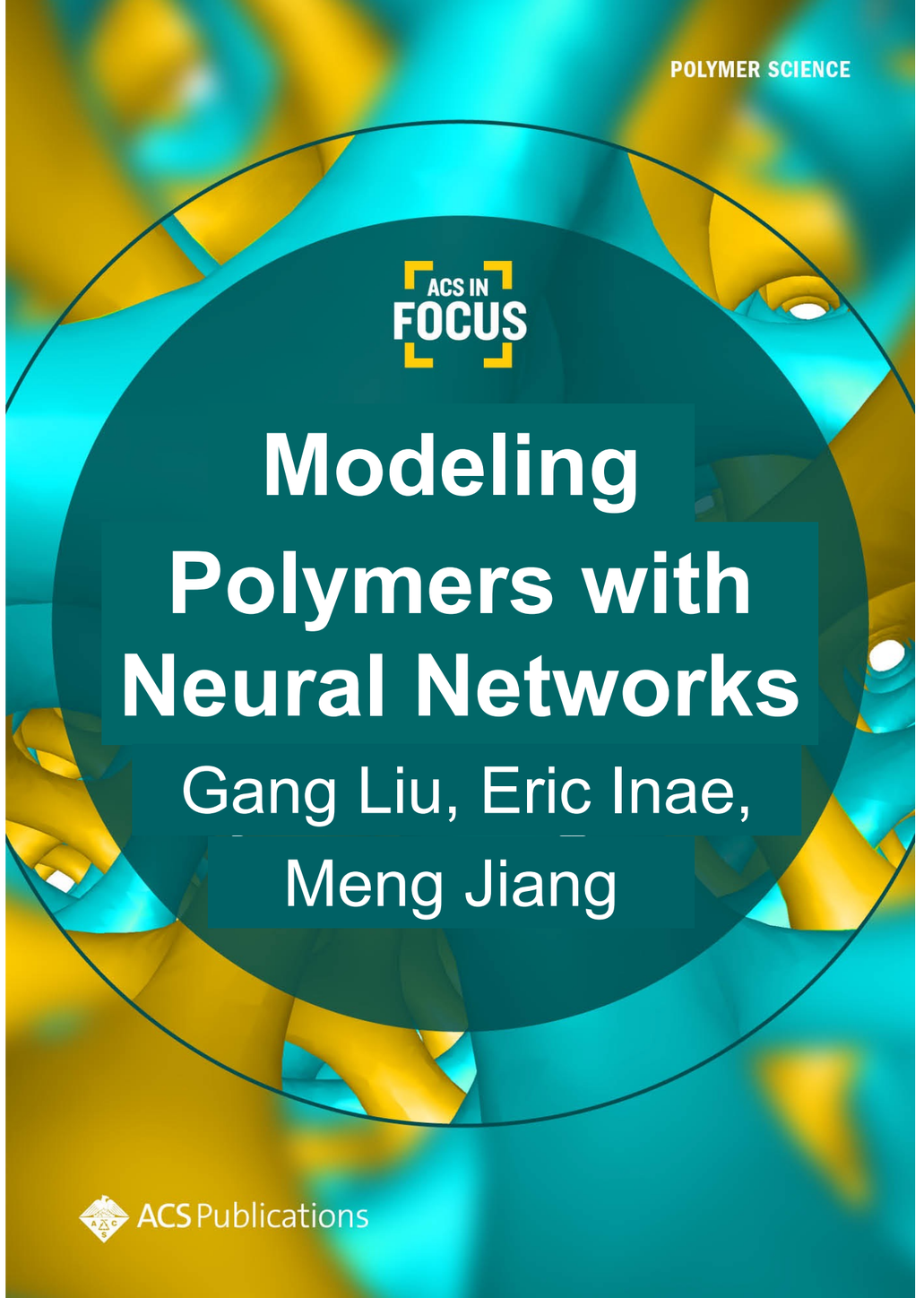
Take Away

- Graph data augmentation can improve property prediction. It can be implicit or explicit.
- Graph data augmentation can address labeled data scarcity and imbalance.
- Diffusion transformer works for not only video creation but also material science.

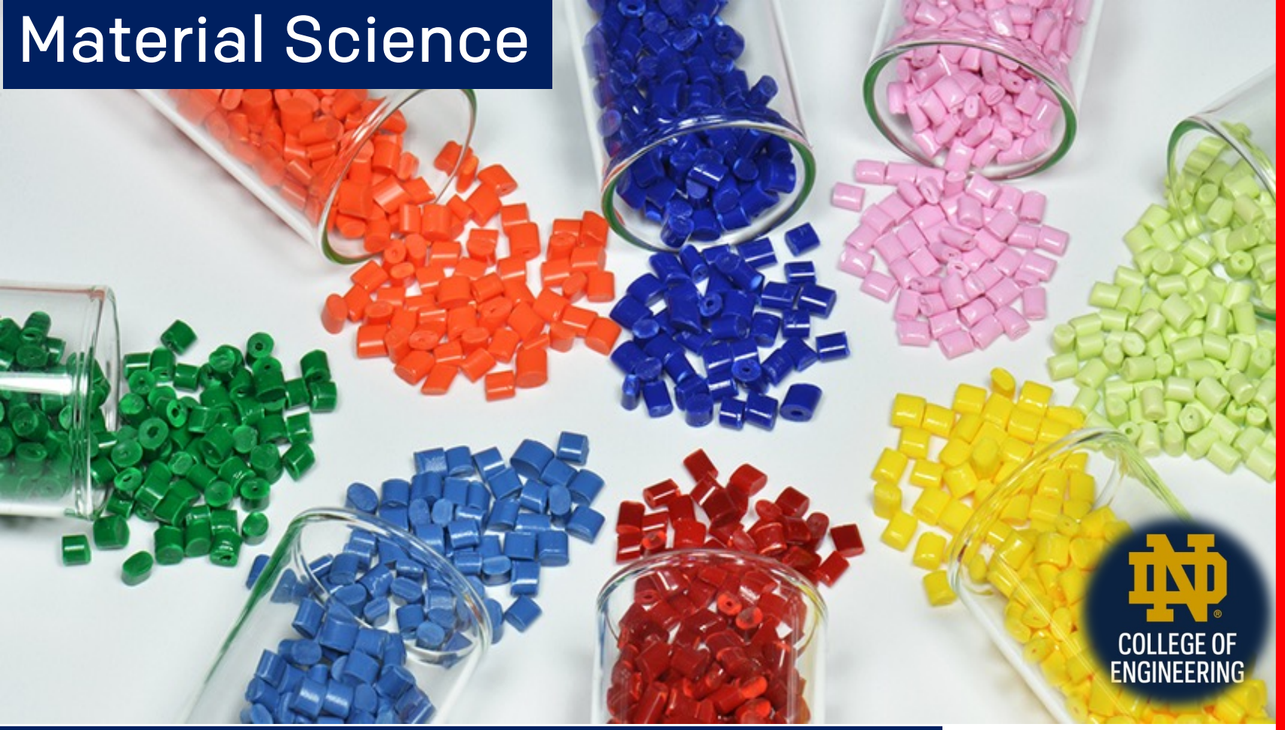
- Interesting directions:
  - Guide generative models with multiple desired properties [1]
  - Guide with domain knowledge such as motifs [2]
  - Guide with knowledge from material science literature and large language models (LLMs)

[1] Liu et al. “Inverse molecular design with multi-conditional diffusion guidance.” https://arxiv.org/abs/2401.13858
[2] Inae et al. “Motif-aware attribute masking for molecular graph pre-training.” https://arxiv.org/abs/2309.04589

Polymer Science

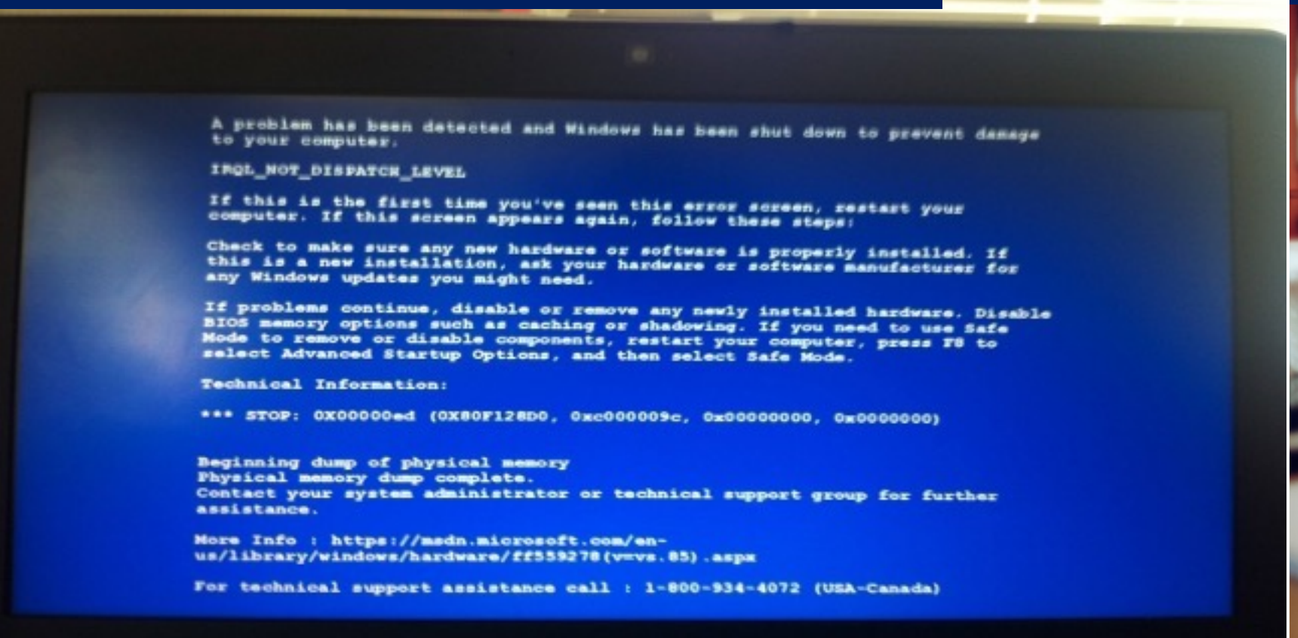


Material Science



Online Education

Software & Network Security



Mental Health

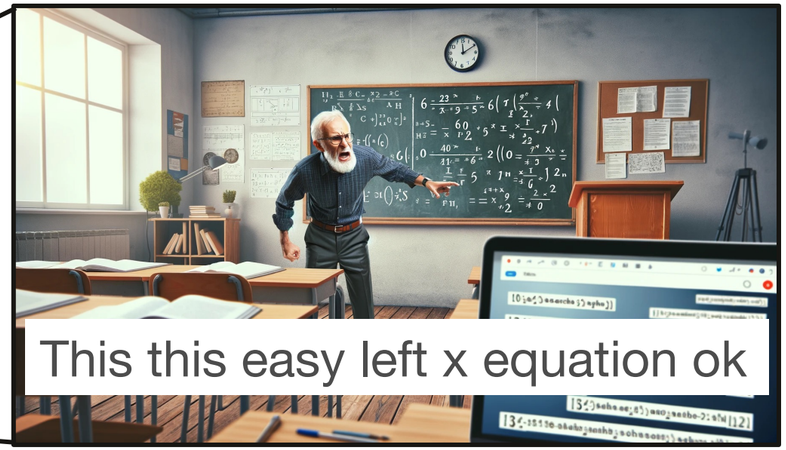
When Online Learning meets Generative AI

*Collaboration with UIUC and Gallaudet University - private university in Washington, D.C., for the education of the deaf and hard of hearing (**DHH**)*

IIS-2119531

RETTL: Collaborative Research: Advancing STEM Online Learning by Augmenting Accessibility with Explanatory Captions and AI (2021-2024, $849,994)

PIs:



Improving Learning Effectiveness

1

Quiz Generation (QG)

Question Answering (QA)

LMs

Knowledge

Reasoning Abilities

2

Instructor's Learning Goals

Student / Classroom Simulation

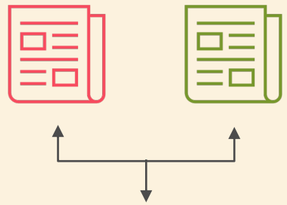
Reasoning in Natural Language

EMNLP'23 (Outstanding Paper Award), EMNLP'22, ACL'22


Mengxia Yu


Wenhao Yu (Tencent)

Comparative Reasoning [1]



- Created numerous training pairs with open (un-)structured data
- Pre-trained LMs continually for zero/few-shot comparative QA/QG

Counterfactual Reasoning [2]

IfQA

- Collected 3,800+ questions that have counterfactual presuppositions
- Evaluated supervised retrieve-then-read and GPT prompting

Commonsense Reasoning [3]

Abductive Reasoning [4]

[1] Yu et al. "Pre-training language models for comparative reasoning." EMNLP 2023.
[2] Yu et al. "IfQA: Open-domain question answering under counterfactual presuppositions." EMNLP 2023. (Outstanding paper award)
[3] Yu et al. "Retrieval augmentation for commonsense reasoning: a unified approach." EMNLP 2022.
[4] Yu et al. "Diversifying content generation with mixture of knowledge graph experts." ACL 2022.

Multi-choice Commonsense Question Answering

Where can I stand on a river to see water falling without getting wet?

A) waterfall
B) bridge
C) valley
D) stream
E) bottom

Datasets:
- OpenBookQA: Clark et al. arxiv 2018. (5,957 **questions**)
- CommonsenseQA 1.0: Talmor et al. NAACL 2019. (12,102)

Evaluation: Accuracy.

Commonsense Fact Verification

A pound of cotton has the same weight as a pound of steel.

True
False

Datasets:
- CommonsenseQA 2.0: Talmor et al. (14,343 **assertions**)
- CREAK: Onoe et al. NeurIPS 2021 Benchmark Track. (13,000)

Evaluation: Accuracy.

Constrained Commonsense Generation

Constraints: dog, frisbee, catch, throw

Description: A girl throws a frisbee and her dog catches it.

Datasets:
- CommonGen: Lin et al. Findings of EMNLP 2020. (79,000+ **commonsense descriptions**)

Evaluation: SPICE, BLEU-4, ROUGE-L, CIDEr.

Counterfactual Explanation Generation

I am hungry for water.

Outputs:
1) Water is used to quench thirsty.
2) You are thirsty for water, not hungry.
3) Water is not food.

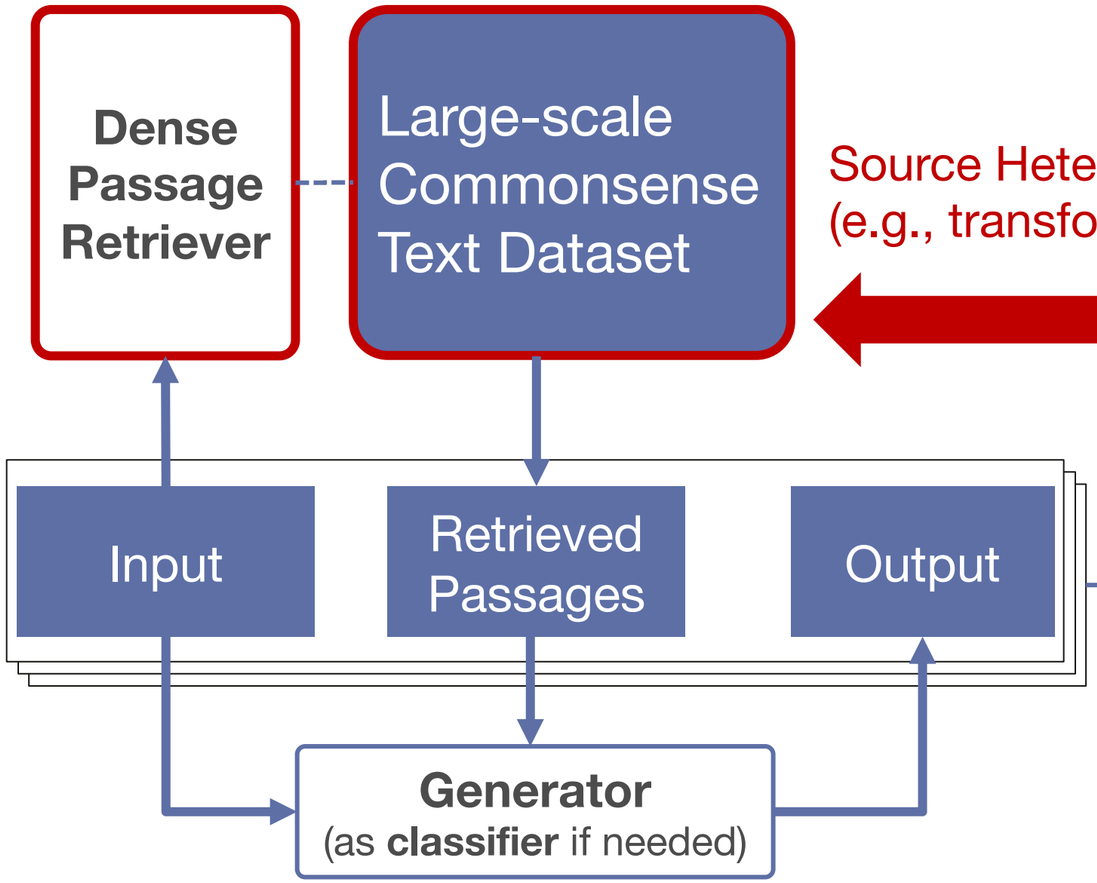
Datasets:
- ComVE: Wang et al. SemEval-2020 Task 4 (11,997 **examples**)

Evaluation: SPICE, BLEU-4, ROUGE-L, CIDEr.

Retrieval Augmentation

Unified Approach for Commonsense Reasoning (RACo) [1]

Wenhao Yu (Tencent)



Source Heterogeneity
(e.g., transform KG / KB)

- Commonsense Corpora
- Commonsense Knowledge Graphs
  (from human annotated relations)
- Commonsense Knowledge Bases
  (from a large collection of facts and events)

- Multi-choice Commonsense Question Answering
- Commonsense Fact Verification
- Constrained Commonsense Generation
- Counterfactual Explanation Generation

[1] Yu et al. "Retrieval augmentation for commonsense reasoning: a unified approach." EMNLP 2022.

Multi-choice Commonsense Question Answering

(Accuracy)	CSQA1.0	OBQA
RACo	75.76	71.25
GreaseLM	74.20	66.90
QA-GNN	73.40	67.80
UNICORN	71.60	70.02
T5-Large	70.14	66.02
KAGNet	69.00	-

Commonsense Fact Verification

(Accuracy)	CSQA2.0	CREAK
RACo	61.75	84.17
UNICORN	54.90	79.51
GreaseLM	-	77.51
T5-Large	54.60	77.32

Constrained Commonsense Generation

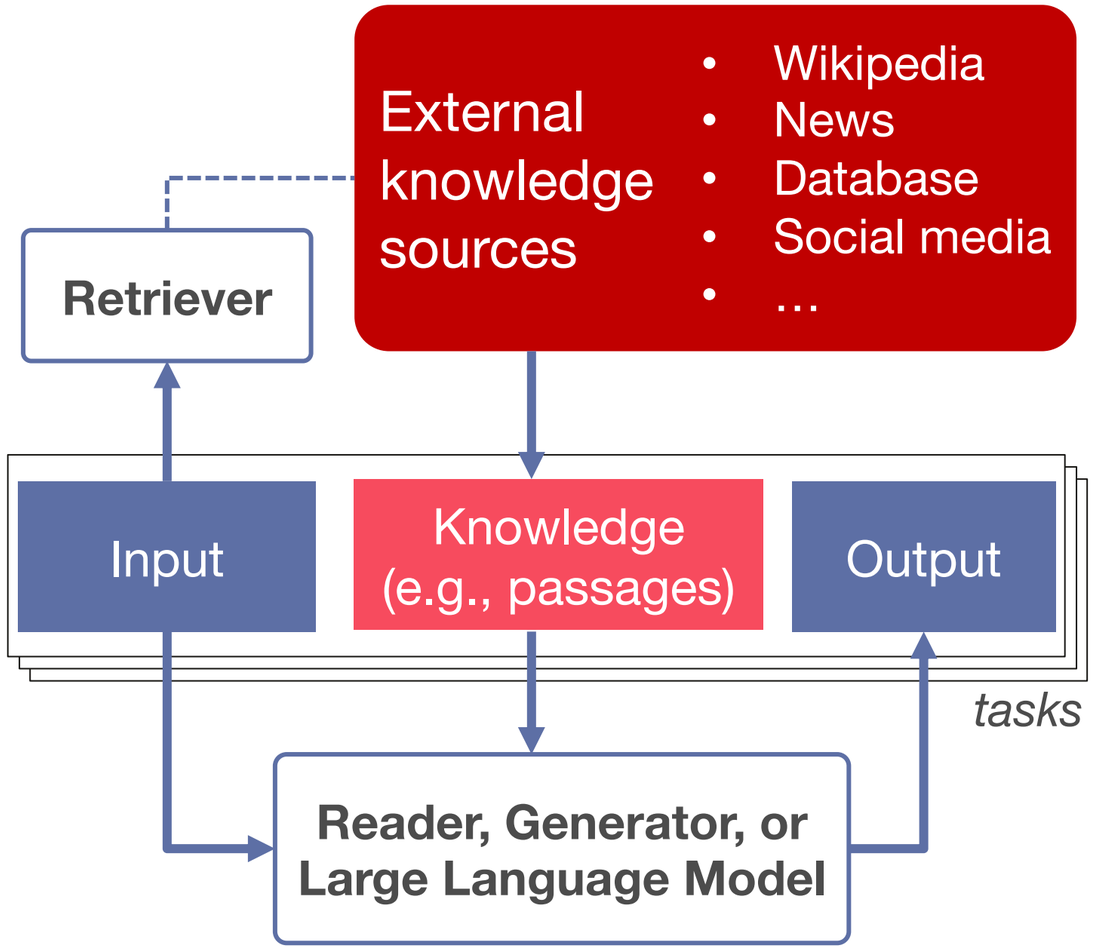
CommonGen	(BL-4)	(SPICE)
RACo	42.76	33.89
KFCNet	41.97	33.11
UNICORN	39.86	33.20
KG-BART	30.90	32.70
CALM	29.50	30.20
T5-Large	28.60	31.60

Counterfactual Explanation Generation

ComVE	(BL-4)	(SPICE)
RACo	25.30	36.37
UNICORN	24.46	35.79
CALM	23.50	35.23
MoKGE	22.87	34.88
T5-Large	22.77	34.62
GraphRF	22.07	33.09

Knowledge-Augmented NLP

"Open-Book" and Knowledge Heterogeneity



Survey paper:

- A Survey of Knowledge-Enhanced Text Generation. ACM Computing Surveys, 2022.

Tutorials:

- Knowledge-Enriched Natural Language Generation. EMNLP 2021.
- Knowledge-Augmented Methods for NLP. ACL 2022.
- Knowledge-Augmented Methods for NLP. WSDM 2023.

Workshops:

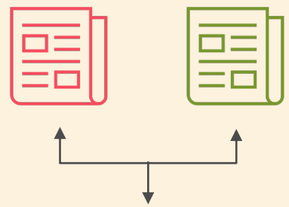
- KnowledgeNLP-AAAI 2023. https://knowledge-nlp.github.io/aaai2023/
- KnowledgeNLP-KDD 2023. https://knowledge-nlp.github.io/kdd2023/
- KnowledgeNLP-**ACL 2024**. https://knowledge-nlp.github.io/acl2024/ (August 11-16 in Bangkok, Thailand)

Reasoning in Natural Language

EMNLP'23 (Outstanding Paper Award), EMNLP'22, ACL'22


Mengxia Yu Wenhao Yu (Tencent)

Comparative Reasoning [1]

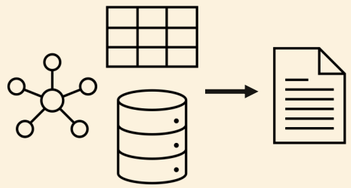


- Created numerous training pairs with open (un-)structured data
- Pre-trained LMs continually for zero/few-shot comparative QA/QG

Counterfactual Reasoning [2]

- Collected 3,800+ questions that have counterfactual presuppositions
- Evaluated supervised retrieve-then-read and GPT prompting

Commonsense Reasoning [3]



- Integrated heterogeneous knowledge into a corpus by data-to-text
- Developed retrieval augmentation as a unified (best) approach for 4 tasks

Abductive Reasoning [4]

Went to work,
leaving windows open

Back home
Found a mess

[1] Yu et al. "Pre-training language models for comparative reasoning." EMNLP 2023.
[2] Yu et al. "IfQA: Open-domain question answering under counterfactual presuppositions." EMNLP 2023. (Outstanding paper award)
[3] Yu et al. "Retrieval augmentation for commonsense reasoning: a unified approach." EMNLP 2022.
[4] Yu et al. "Diversifying content generation with mixture of knowledge graph experts." ACL 2022.

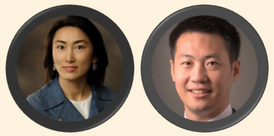
Scientific Text Intelligence

Knowledge-augmented Comparative and Abductive Reasoning

IIS-2119531

RETTL: Collaborative Research: Advancing STEM Online Learning by Augmenting Accessibility with Explanatory Captions and AI (2021-2024, $849,994)

PIs:




IIS-2234058

III: Small: Intelligent Scientific Text Analytics with Knowledge-augmented Abductive Reasoning (2023-2026, $599,962)

PI:

GRA:


When Software & Network Security meets Generative AI

Human-oriented Knowledge Representation with Text and Graph

IIS-1849816

CRII: III: Beyond Similarity Learning: Complementarity Learning for Contextual Behavior Modeling
(2019-2021, $174,910)

PI:

NSF

- *Text generation on graph:* To build a taxonomy, concepts can be generated word by word to address low recall [1].

Qingkai Zeng

- *Text generation from tree:* LLMs can interpret decision tree predictions with domain knowledge to experts who know little about machine learning [2, 3].

Noah Ziems

CCF-1901059

SHF: Medium: Collaborative Research: Semantically-Enhanced Software Traceability for Supporting Human-Centric Tasks
(2019-2024, $1,035,030)

NSF

PI:



N00014-22-1-2507

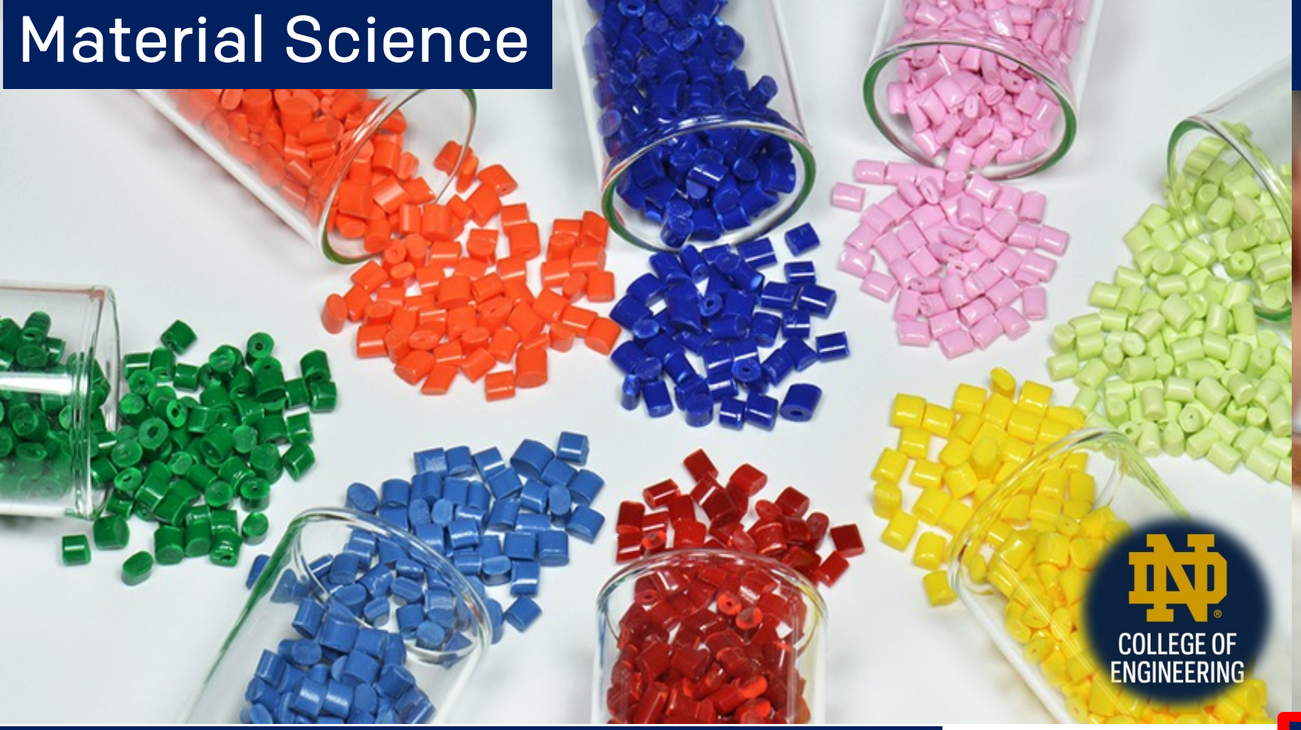
"Towards Third Generation Network Intrusion Detection and Prevention Systems
(2022-2025, $459,251)

Office of Naval Research ONR Science & Technology

PI:

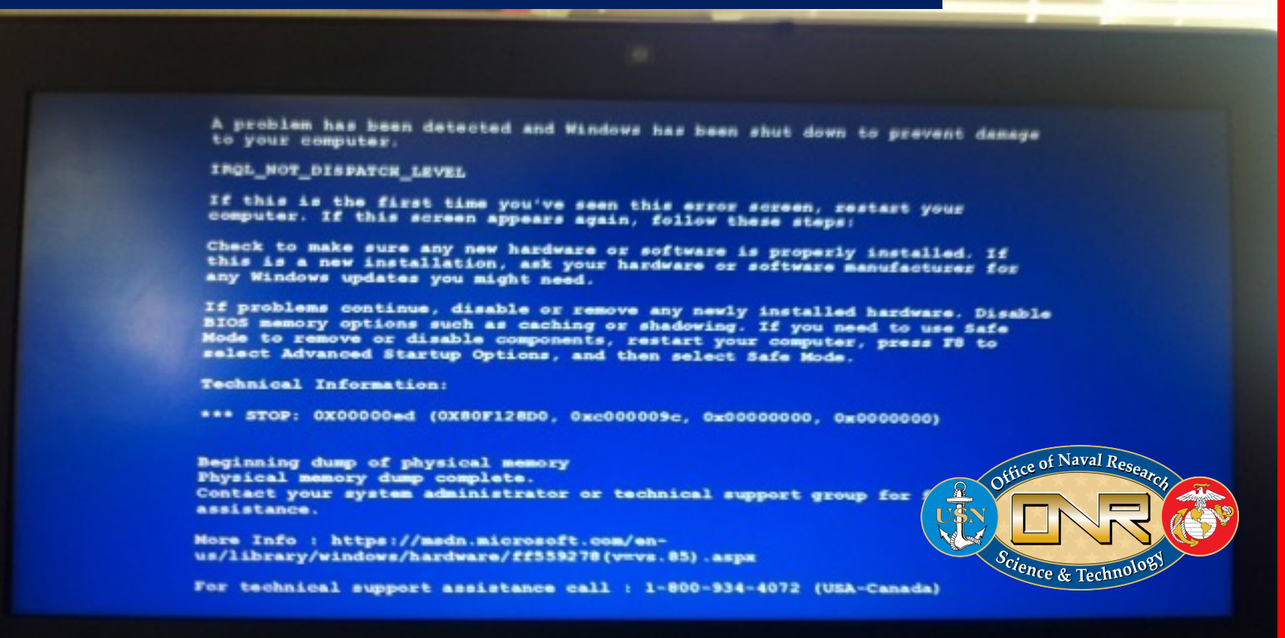
[1] Zeng et al. "Enhancing taxonomy completion with concept generation via fusing relational representations." KDD 2021.
[2] Ziems et al. "Explaining tree model decisions in natural language for network intrusion detection." NeurIPS-XAIA 2023.
[3] Ziems et al. "Large language models are built-in autoregressive search engines." ACL 2023.

Material Science



Online Education

Software & Network Security



Mental Health

When Suicidal Ideation meets Generative AI

Specialized Generative AI for Specialized IA

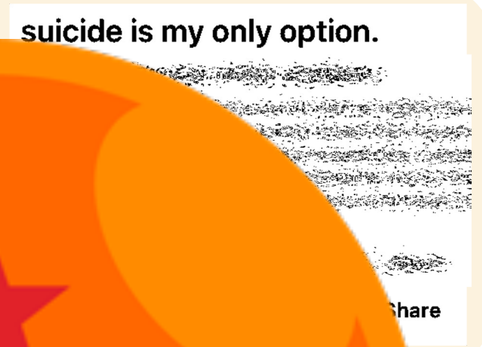
IIS-2142827

CAREER: Synergistic Approaches for Specialized Intelligent Assistance
(2022-2027, $549,969)

PI:

GRA:

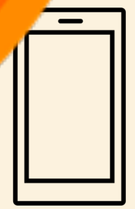
r/SuicideWatch [1, 2]
Peer support
anyone s
with su
467K
Me



Jo
NL
Sc

two p
active ri
in suicidal c
psychiatric hospitalization [3]

- Screenomics screenshot capture
- Ecological momentary assessment (EMA)



GPT-5?

1R21MH129688

Advancing Real-Time Suicide Risk Detection through Digital Phenotyping
(2022-2024, $430,375)

PI:

NIMH
National Institute of Mental Health

[1] Jiang et al. "Phrase-level pairwise topic modeling to uncover helpful peer responses to online suicidal crises." Nature Humanities and Social Sciences Communications 2020.

[2] Dang et al. "Embedding mental health discourse for community recommendation." ACL-CODI 2023.

[3] Jacobucci et al. "A comparative case analysis of passively collected smartphone-based data in the days prior to psychiatric hospitalization for a suicidal crisis." Open Science Framework 2024.

Conclusions

Doing interdisciplinary research with Generative AI is lots of fun!

- **Material Science | Polymer Informatics**
  - Graph data augmentation
    - Supervised learning; Imbalanced learning; Transfer learning
    - Implicit aug.: Environment (non-Rationale) replacement; Label-anchored mix-up
    - Explicit aug.: Parameter- or data-centric transfer; Graph diffusion transformer
- **Online Education** (for DHH)
  - Knowledge-augmented NLP
    - Question answering; Question generation
    - Knowledge heterogeneity; Multiple types of Reasoning (e.g., Comparative, Abductive)
- **Software & Network Security** and **Mental Health | Suicide Prevention**

Welcome to visit us in South Bend!

http://www.meng-jiang.com/lab.html

https://github.com/DM2-ND

mjiang2@nd.edu

UNIVERSITY OF NOTRE DAME

College of Engineering