Introduction

If almost 60 percent of our members say that they spend three hours or more of every working day on administrative tasks, they cannot be there for their patients during this time. I simply think it’s a scandal how much manpower and working time is wasted on data collection and documentation.

- Chairman of the „Marburger Bund“, Dr. Susanne Johna (2022)

[Ma22] Marburger-Bund: Zu wenig Personal, zu viel Bürokratie, unzulängliche Digitalisierung (quotation translated from German)

Evaluating Adapter-based Knowledge-enhanced Language Models in the Biomedical Domain

Alexander Fichtl

16.10.2023, Master Thesis Final Presentation

Chair of Software Engineering for Business Information Systems (sebis)
Department of Computer Science
School of Computation, Information and Technology (CIT)
Technical University of Munich (TUM)
wwwmatthes.in.tum.de

Outline

1. **Motivation**

2. **Background**
   - Knowledge-enhanced Language Models (KELMs)
   - Adapters
   - Evaluation Tasks

3. **Research Questions**

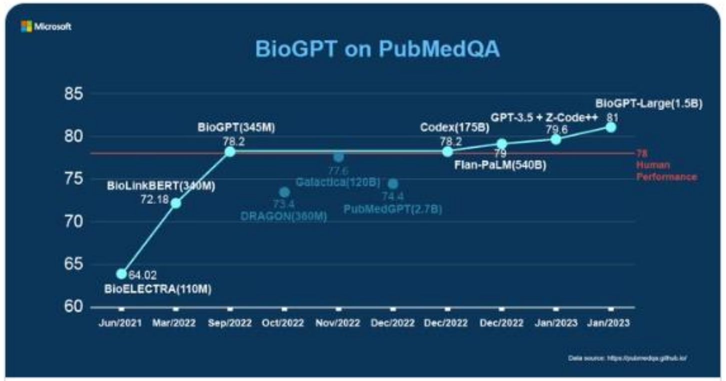
4. **Methodology and Results**
   - Literature Review
   - Model Experiments
   - Research Survey

5. **Shortcomings and Future Research**

Motivation

Microsoft Research
@MSFTResearch

BioGPT, a domain-specific generative model pre-trained on large-scale biomedical literature, has achieved human parity, outperformed other general and scientific LLMs, and could empower biologists in various scenarios of scientific discovery. Learn more: msft.it/6014eAnLq



oup.com
BioGPT: pre-trained transformer for biomed text generation & mining

Original Investigation

April 28, 2023

Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum

John W. Ayers, PhD, MA[1,2]; Adam Poliak, PhD[3]; Mark Dredze, PhD[4]; et al

Key Points

Question Can an artificial intelligence chatbot assistant, provide responses to patient questions that are of comparable quality and empathy to those written by physicians?

Findings In this cross-sectional study of 195 randomly drawn patient questions from a social media forum, a team of licensed health care professionals compared physician's and chatbot's responses to patient's questions asked publicly on a public social media forum. The chatbot responses were preferred over physician responses and rated significantly higher for both quality and empathy.

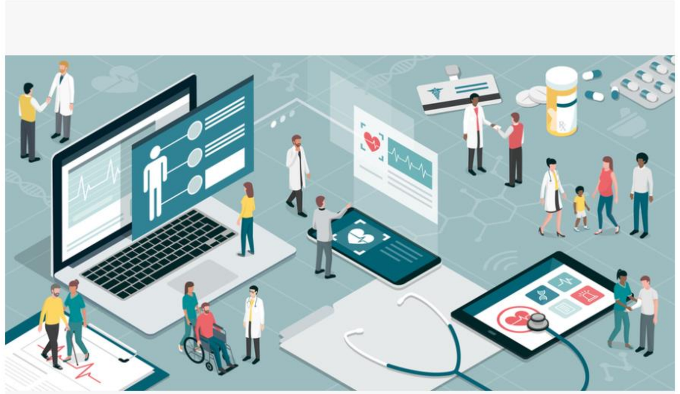
Meaning These results suggest that artificial intelligence assistants may be able to aid in drafting responses to patient questions.

MIT News
ON CAMPUS AND AROUND THE WORLD

Large language models help decipher clinical notes

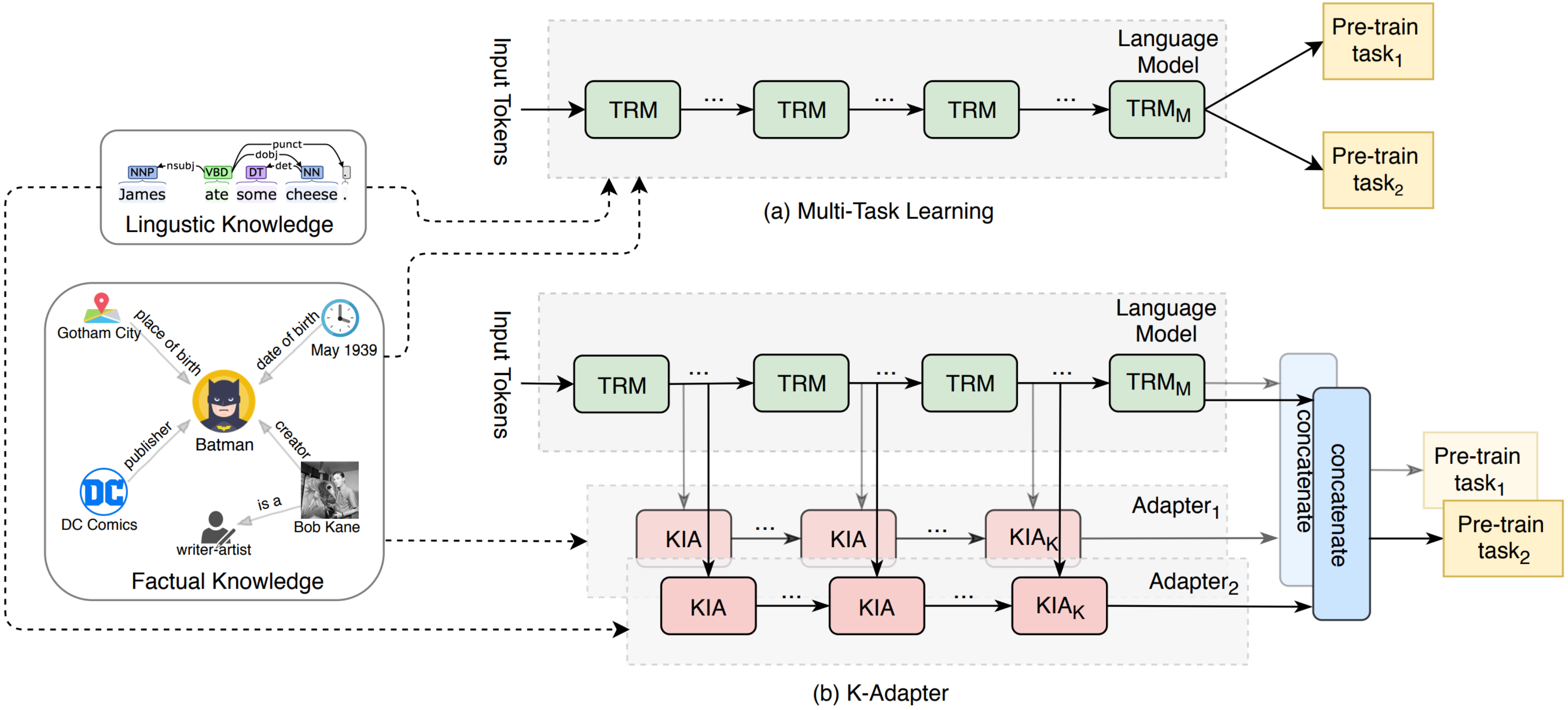
Researchers used a powerful deep-learning model to extract important data from electronic health records that could assist with personalized medicine.

Rachel Gordon | MIT CSAIL
December 1, 2022



[Mi23] Microsoft: *BioGPT: generative pre-trained transformer for biomedical text generation and mining*
[Go22] Gordon, R.: *Large language models help decipher clinical notes*
[Ay23] *Ayers*, J., Poliak, A., Dredze, M., et al.: *Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum*

Background: What are KELMs?

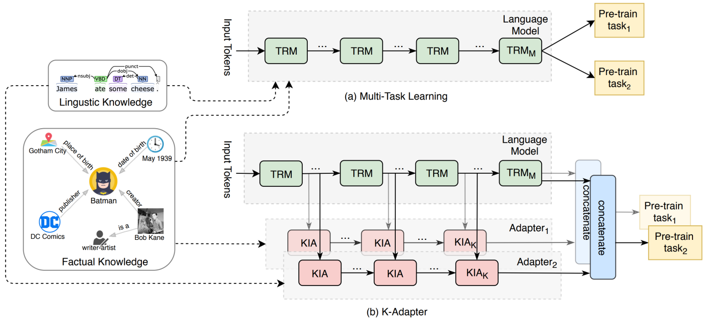


[Wa21a] Wang, R., Tang, D., Duan, N., Wei, Z., Huang, X., Ji, J., Cao, G., Jiang, D., Zhou, M.: K-ADAPTER: Infusing Knowledge into Pre-Trained Models with Adapters

Background: What are KELMs?

Medical natural language inference task (NLI):

- **Patient Premise**: No history of blood clots or DVTs, has never had chest pain prior to one week ago
- **Hypothesis**: Patient has angina

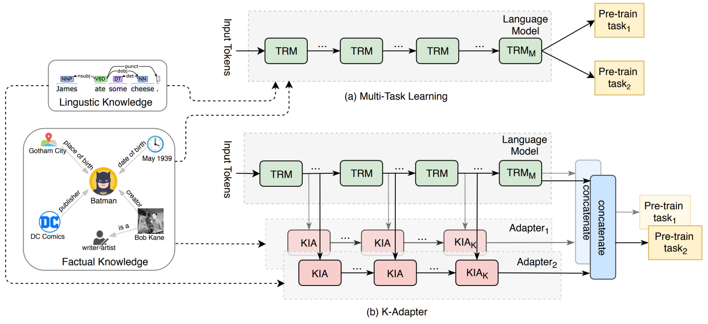


[Lu21] Lu, Q., Dou, D., Nguyen, T.H.: Parameter-Efficient Domain Knowledge Integration from Multiple Sources for Biomedical Pre-trained Language Models

Background: What are KELMs?

Medical natural language inference task (NLI):

- **Patient Premise**: No history of blood clots or DVTs, has never had chest pain prior to one week ago
- **Hypothesis**: Patient has angina
- **Correct Classification:** Entailment



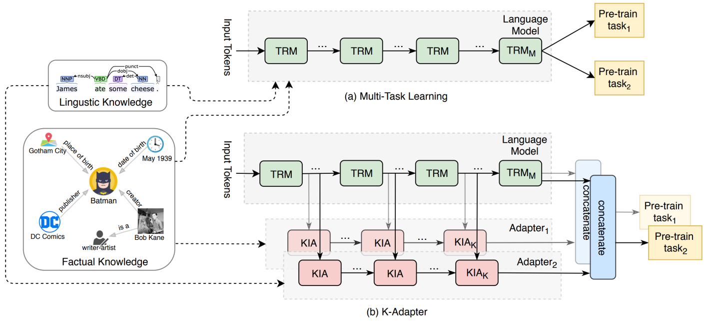
[[Lu21]] Lu, Q., Dou, D., Nguyen, T.H.: *Parameter-Efficient Domain Knowledge Integration from Multiple Sources for Biomedical Pre-trained Language Models*

Background: What are KELMs?

Medical natural language inference task (NLI):

- **Patient Premise**: No history of blood clots or DVTs, has never had chest pain prior to one week ago
- **Hypothesis**: Patient has angina
- **Correct Classification:** Entailment

➡ Correct inference more likely if model specifically learned synonyms and relations



[Lu21] *Lu, Q., Dou, D., Nguyen, T.H.: Parameter-Efficient Domain Knowledge Integration from Multiple Sources for Biomedical Pre-trained Language Models*

Background: What are KELMs?

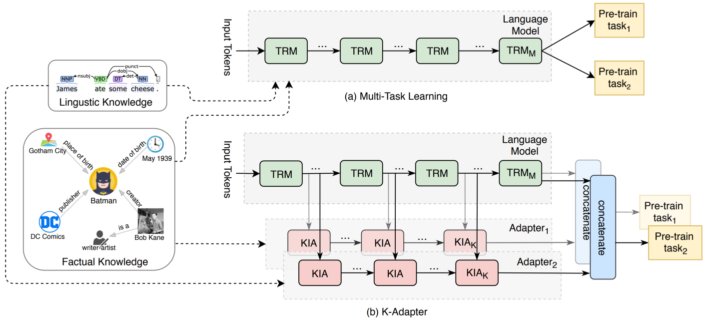
Relevance

- Active research area
- Can address hallucinations
- Superior performance over vanilla LMs

Shortcomings

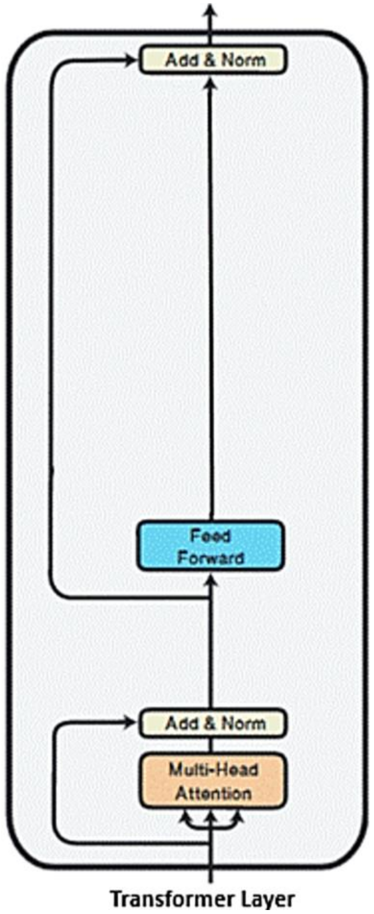
- Long Training Periods
- Catastrophic forgetting

➡ Lightweight “Adapters”



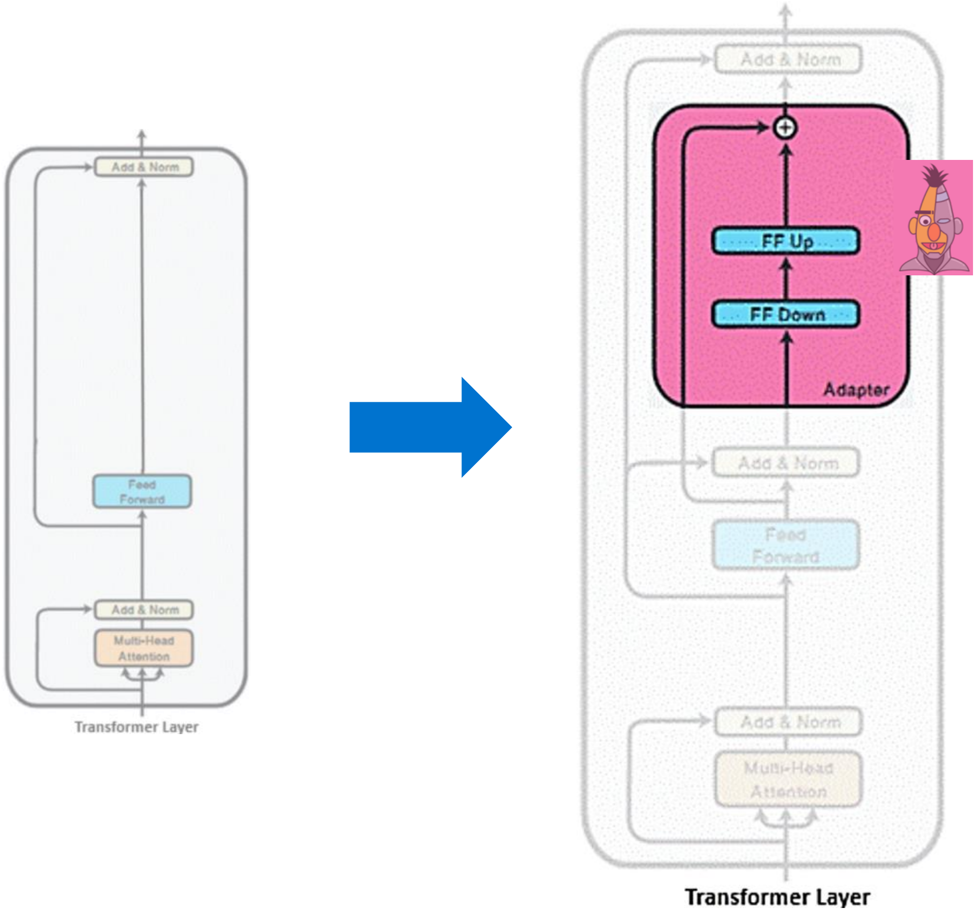
[We21a] Wei, X., Wang, S., Zhang, D., Bhatia, P., Arnold A.: Knowledge Enhanced Pretrained Language Models: A Compreshensive Survey
[Wa21a] Wang, R., Tang, D., Duan, N., Wei, Z., Huang, X., Ji, J., Cao, G., Jiang, D., Zhou, M.: K-ADAPTER: Infusing Knowledge into Pre-Trained Models with Adapters

Background: What is an Adapter?



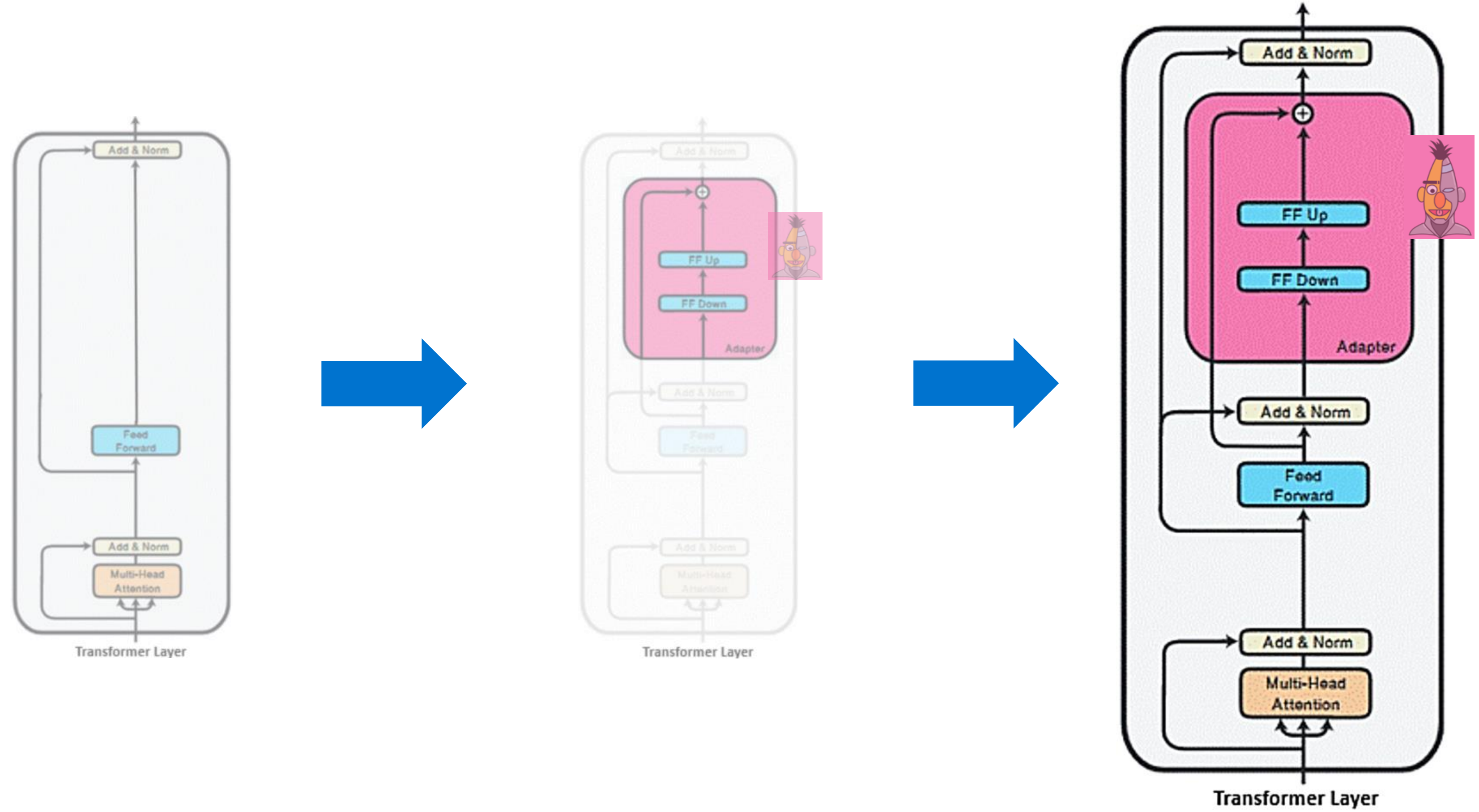
[Va17] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., et al.: *Attention Is All You Need*
[Pf20a] Pfeiffer, J., Rücklé, A., Poth, C., Kamath, A., Vulić, I., Ruder, S., Cho, K., Gurevych, I.: *AdapterHub: A Framework for Adapting Transformers*

Background: What is an Adapter?



[Va17] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., et al.: Attention Is All You Need
[Pf20a] Pfeiffer, J., Rücklé, A., Poth, C., Kamath, A., Vulić, I., Ruder, S., Cho, K., Gurevych, I.: AdapterHub: A Framework for Adapting Transformers

Background: What is an Adapter?

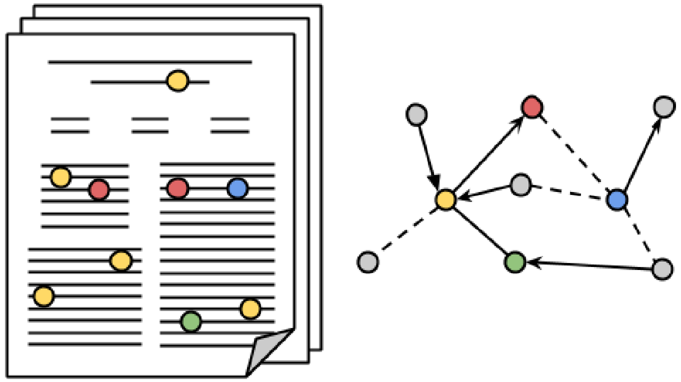


[Va17] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., et al.: Attention Is All You Need

[Pf20a] Pfeiffer, J., Rücklé, A., Poth, C., Kamath, A., Vulić, I., Ruder, S., Cho, K., Gurevych, I.: AdapterHub: A Framework for Adapting Transformers

Background: Evaluation Tasks

Dataset	Task	Train	Dev	Test	EvaluationMetrics
HoC	DocumentClassification	1295	186	371	MicroF1
PubMedQA	QuestionAnswering	450	50	500	Accuracy
BioASQ	QuestionAnswering	670	75	140	Accuracy
MedNLI	Natural Language Inference	11232	1395	1,422	Accuracy



BLURB

Biomedical **L**anguage **U**nderstanding and **R**easoning **B**enchmark

[Gu20] Gu, Y.,Tinn, R., Cheng, H., et al.: Domain-Specific Language Model Pretraining for Biomedical Natural Language Processing

Research Questions

RQ1: What adapter-based approaches to knowledge-enhancement exist, and how do they compare to each other?

- Systematic Literature Review (SLR)

RQ2: Can we improve existing approaches with new methods and data from a private ontology?

- SLR
- Thesis experiments with OntoChem KGs

RQ3: Is the research on biomedical KELMs relevant to medical professionals, and what factors hinder or support the deployment of the technology in practice?

- Research survey

Systematic Literature Review

RQ1: What adapter-based approaches to knowledge-enhancement exist, and how do they compare to each other?

- Systematic Literature Review (SLR)

RQ2: Can we improve existing approaches with new methods and data from a private ontology?

- SLR
- Thesis experiments

RQ3: Is the research on biomedical KELMs relevant to medical professionals, and what factors hinder or support the deployment of the technology in practice?

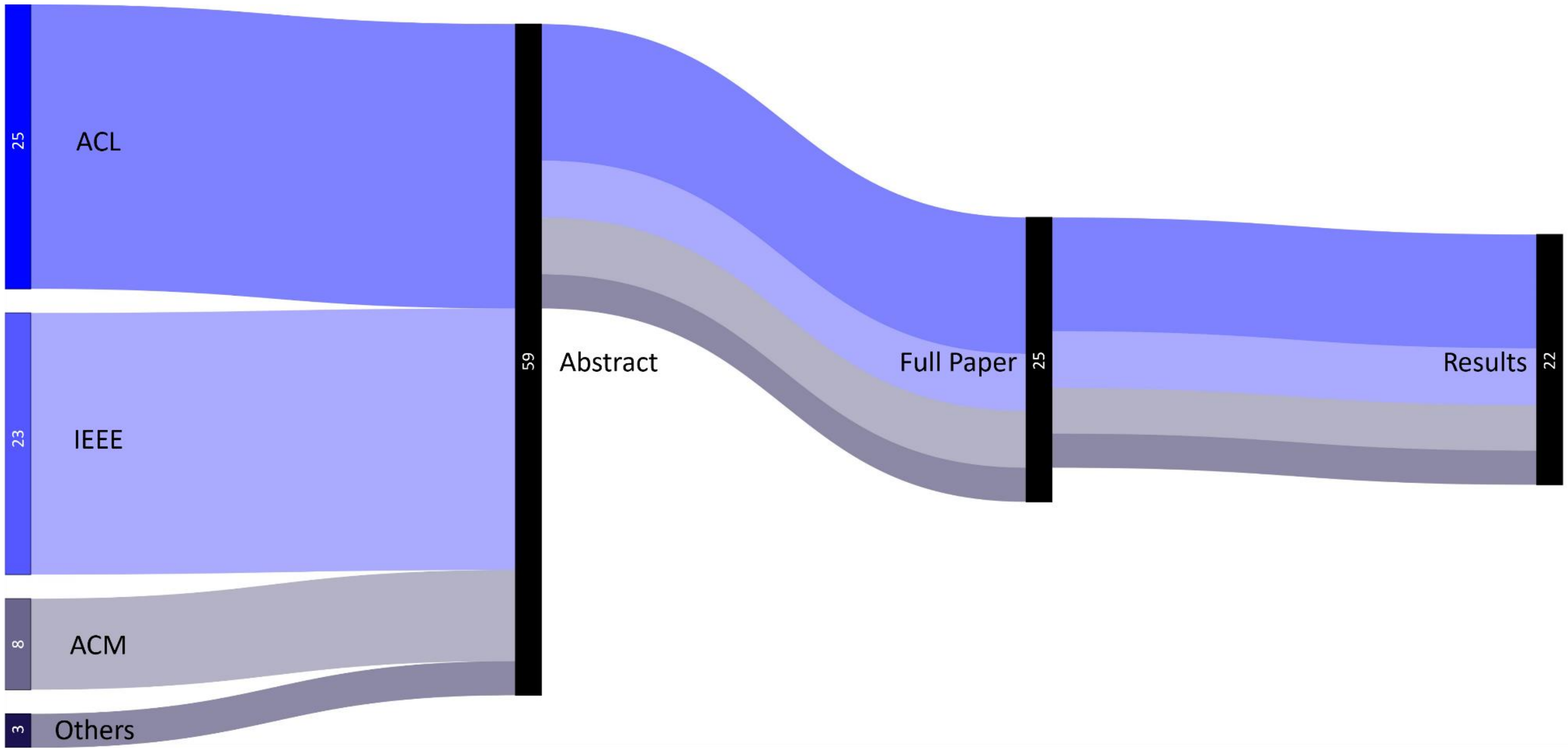
- Research survey

SLR: Methodology

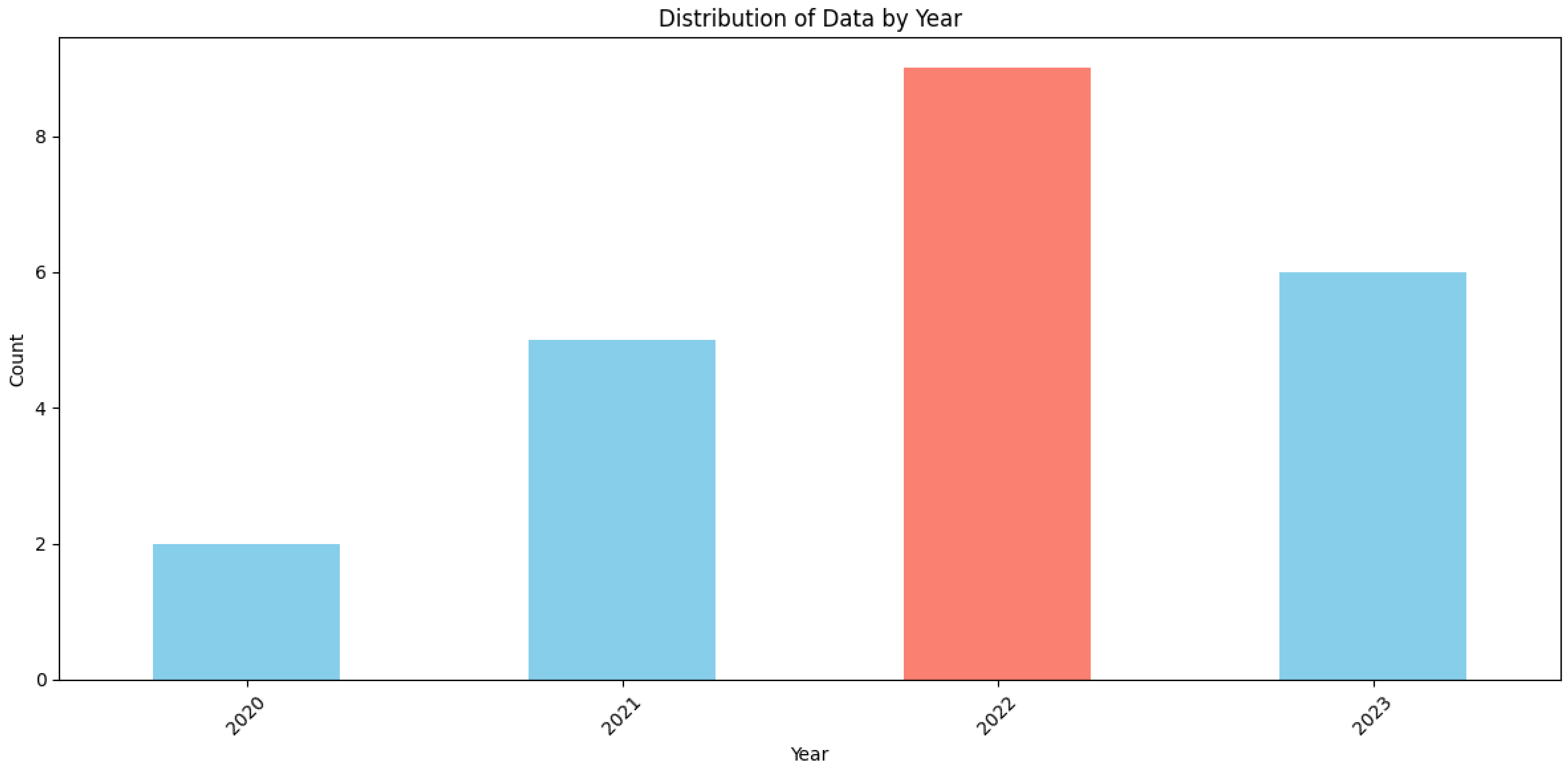
Criteria	Inclusion Criteria	Exclusion Criteria
Platforms	Papers published on ACM, ACL , or IEEE Xplore	Papers from other platforms (if not essential)
Search String	Papers matching the search string ("adapter" OR "adapter-based") AND ("language model" OR "nlp" OR "natural language processing") AND ("injection" OR "knowledge")	Papers that do not match the search string
Thematic Relation	Papers which address the topic of adapter-based knowledge-enhanced language models	Papers where Adapters were used for NLP, but for use-cases other than knowledge-enhancement
Language	Papers written in English or in German are included.	Papers not written English or German
Publication Date	Papers published after February 2, 2019 (publication of the Houlsby Adapter, the first LLM adapter)	Papers published before February 2, 2019
Duplicates	Publications not yet part of the selection process	Duplicate versions of the same Papers

[Ki09] Kitchenham, B., Brereton, P., Budgen, D., et al.: Systematic literature reviews in software engineering – A systematic literature review

SLR: Results

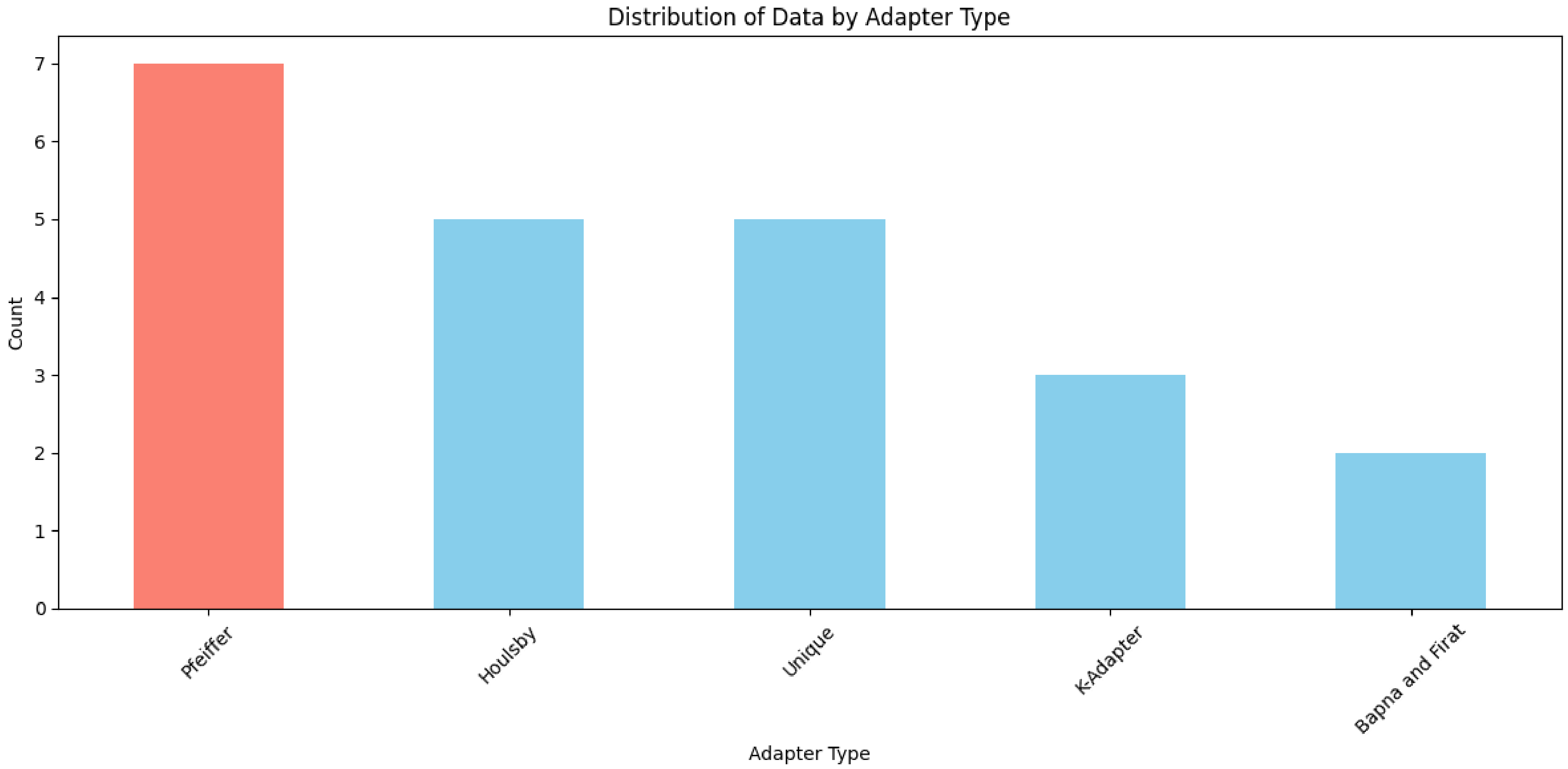

25
ACL
23
IEEE
8
ACM
3
Others
59
Abstract
Full Paper
25
Results
22

SLR: Results

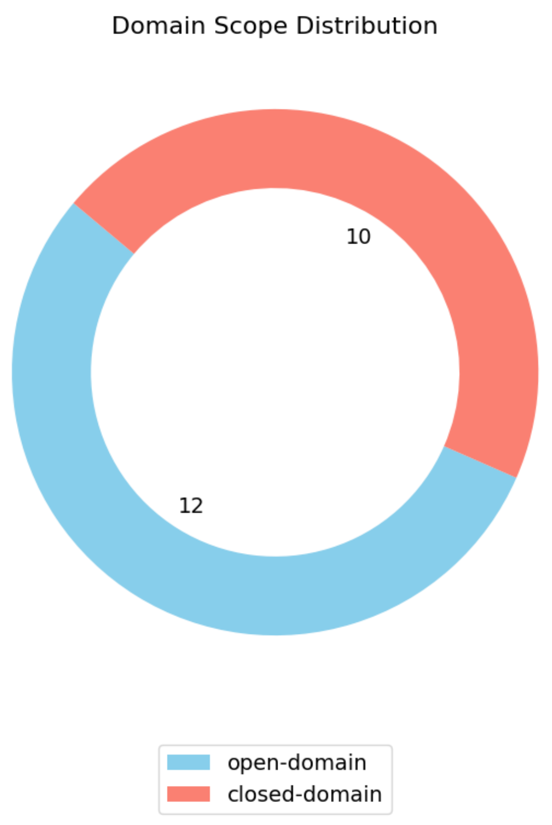


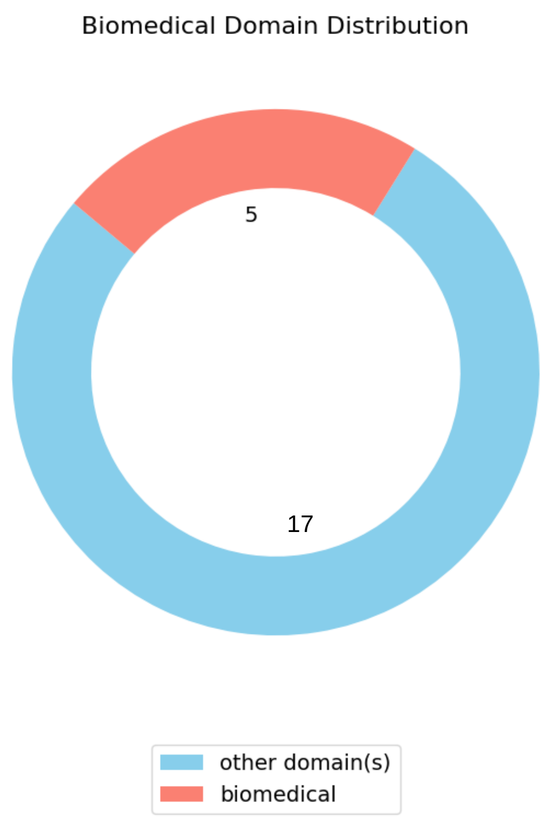
Distribution of Data by Year
Count
8
6
4
2
0
2020
2021
2022
2023
Year

SLR: Results


Distribution of Data by Adapter Type
Count
0
1
2
3
4
5
6
7
Pfeiffer
Houlsby
Unique
K-Adapter
Bapna and Firat
Adapter Type

SLR: Results


Domain Scope Distribution
10
12
open-domain
closed-domain


Biomedical Domain Distribution
5
17
other domain(s)
biomedical

SLR: Results

paper	adapter type	scope	coverage	biomed	task	nickname
Zou et al. (2022)	K-Adapter	open	/	no	RC	K-MBAN
Moon et al. (2021)	Houlsby	open	/	no	MT	/
Yu et al. (2023)	Unique	open	/	no	SL	CSBERT
Quian et al. (2022)	Unique	open	/	no	SR	/
Li et al. (2023)	Houlsby	closed	multi	no	SF	/
Li et al. (2023)	Unique	open	/	no	SC	CKGA
Nguyen et al.(2023)	Pfeiffer	open	/	no	SA	/
Lai et al. (2023)	Pfeiffer	closed	single	yes	QA, NLI, EL	KEBLM
Guo et al. (2022)	Unique	open	/	no	NER	/
Chronopoulou et al. (2023)	Bapna and Firat	closed	both	no	LM	Adapter-Soup
Wold et al. (2022)	Houlsby	open	/	no	LAMA	/
Chronopoulou et al. (2022)	Unique	closed	multi	no	LM	/
Hung et al. (2022)	Pfeiffer	closed	multi	no	TOD	DS-TOD
Emelin et al. (2022)	Houlsby	closed	multi	no	TOD	/
Xu et al. (2022)	Bapna and Firat	open	/	no	KGD	KnowBERT
Kær et al. (2021)	Pfeiffer	closed	multi	yes	NER, STC	mDAPT
Lu et al. (2021)	K-Adapter	closed	single	yes	NLI	DAKI
Majewska et al. (2021)	Pfeiffer	open	/	no	EE	/
Lauscher et al. (2020)	Houlsby	open	/	no	GLUE	/
Meng et al. (2021)	Pfeiffer	closed	single	yes	BLURB	MoP
Wang et al. (2020)	K-Adapter	open	/	no	RCL, ET, QA	K-Adapter
Xie et al. (2022)	Pfeiffer	closed	single	yes	ES	KeBioSum

SLR: Results

paper	adapter type	scope	coverage	biomed	task	nickname
Zou et al. (2022)	K-Adapter	open	/	no	RC	K-MBAN
Moon et al. (2021)	Houlsby	open	/	no	MT	/
Yu et al. (2023)	Unique	open	/	no	SL	CSBERT
Quian et al. (2022)	Unique	open	/	no	SR	/
Li et al. (2023)	Houlsby	closed	multi	no	SF	/
Li et al. (2023)	Unique	open	/	no	SC	CKGA
Nguyen et al.(2023)	Pfeiffer	open	/	no	SA	/
Lai et al. (2023)	Pfeiffer	closed	single	yes	QA, NLI, EL	KEBLM
Guo et al. (2022)	Unique	open	/	no	NER	/
Chronopoulou et al. (2023)	Bapna and Firat	closed	both	no	LM	Adapter-Soup
Wold et al. (2022)	Houlsby	open	/	no	LAMA	/
Chronopoulou et al. (2022)	Unique	closed	multi	no	LM	/
Hung et al. (2022)	Pfeiffer	closed	multi	no	TOD	DS-TOD
Emelin et al. (2022)	Houlsby	closed	multi	no	TOD	/
Xu et al. (2022)	Bapna and Firat	open	/	no	KGD	KnowBERT
Kær et al. (2021)	Pfeiffer	closed	multi	yes	NER, STC	mDAPT
Lu et al. (2021)	K-Adapter	closed	single	yes	NLI	DAKI
Majewska et al. (2021)	Pfeiffer	open	/	no	EE	/
Lauscher et al. (2020)	Houlsby	open	/	no	GLUE	/
Meng et al. (2021)	Pfeiffer	closed	single	yes	BLURB	MoP
Wang et al. (2020)	K-Adapter	open	/	no	RCL, ET, QA	K-Adapter
Xie et al. (2022)	Pfeiffer	closed	single	yes	ES	KeBioSum

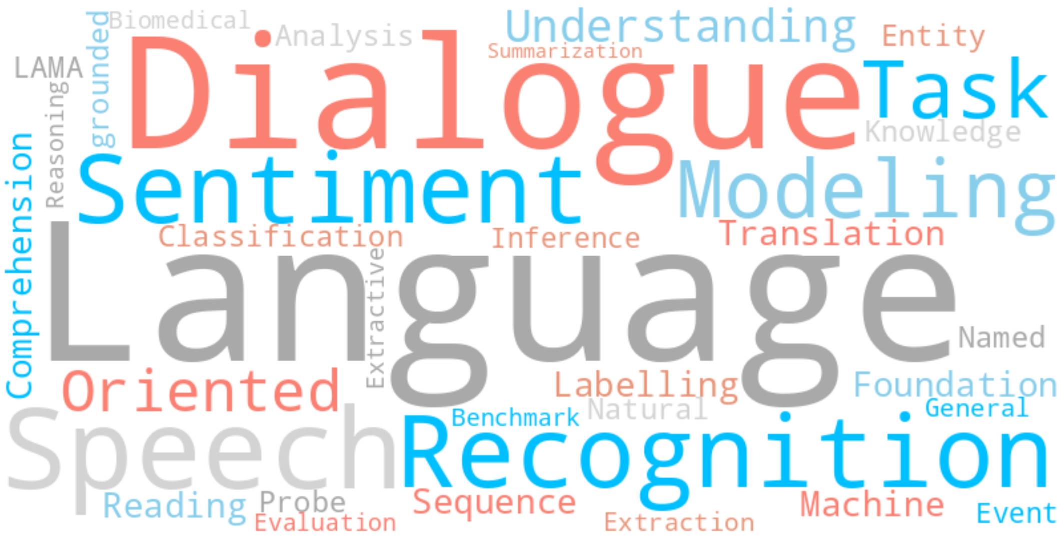
SLR: Summary

RQ1: What adapter-based approaches to knowledge-enhancement exist, and how do they compare to each other?

SLR: Summary

RQ1: What adapter-based approaches to knowledge-enhancement exist, and how do they compare to each other?

- The SLR portrayed, analysed, and compared existing approaches
- We found the best setting for our model experiments
- We provide a novel and extensive resource for other researchers in the field



Model Experiments

RQ1: What adapter-based approaches to knowledge-enhancement exist, and how do they compare to each other?

- Systematic Literature Review (SLR)

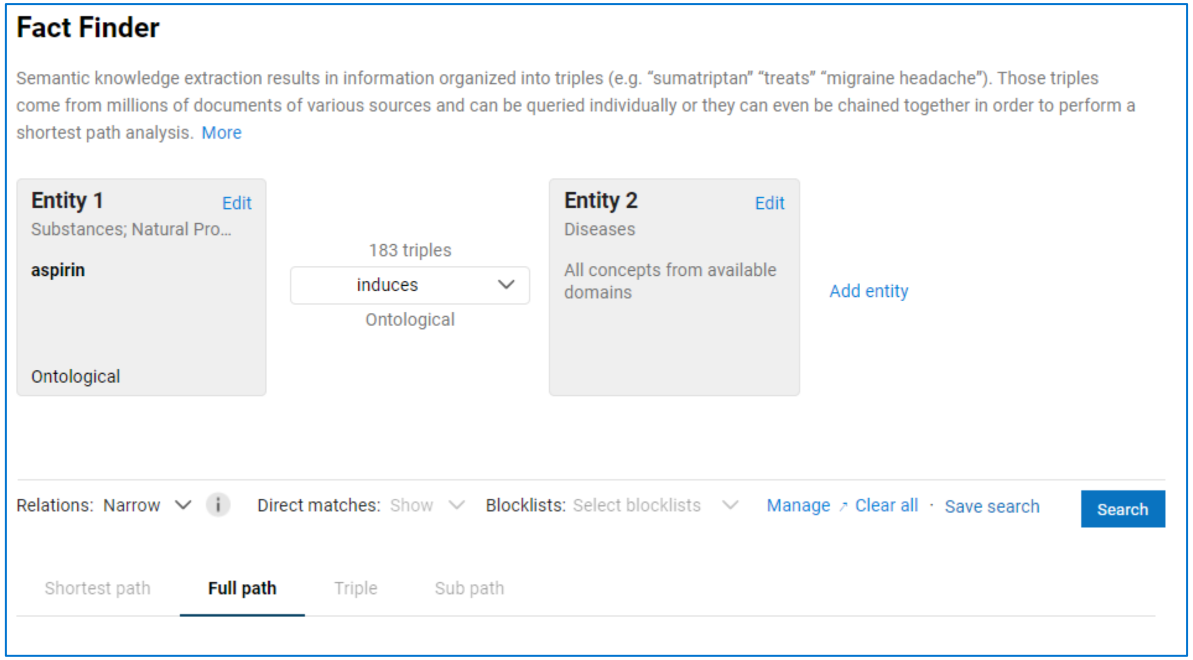
RQ2: Can we improve existing approaches with new methods and data from a private ontology?

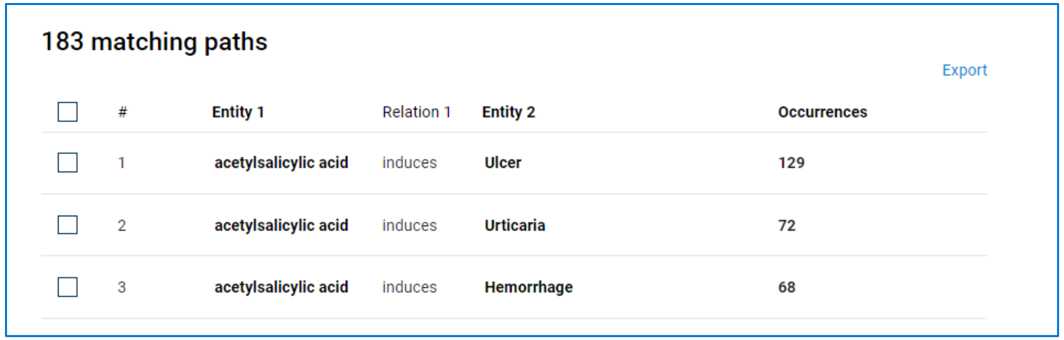
- SLR
- Thesis experiments with OntoChem KGs

RQ3: Is the research on biomedical KELMs relevant to medical professionals, and what factors hinder or support the deployment of the technology in practice?

- Research survey

Model Experiments: Methodology




183 matching paths

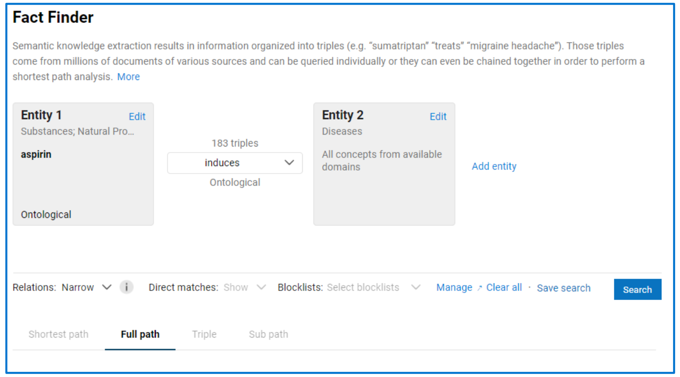
Export

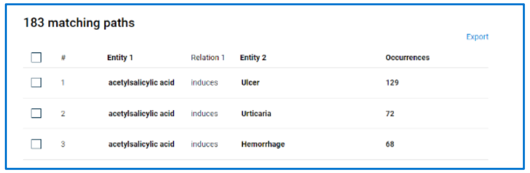
#	Entity 1	Relation 1	Entity 2	Occurrences
1	acetylsalicylic acid	induces	Ulcer	129
2	acetylsalicylic acid	induces	Urticaria	72
3	acetylsalicylic acid	induces	Hemorrhage	68

[On23] Ontochem: SciWalker platform

Model Experiments: OntoChem Knowledge Graphs

Onto20Rel	#Triplets
relates to	708,076
induces	502,512
modulates	326,534
treats	225,279
inhibits	219,720
is analyzed by	195,291
produces	173,979
increases activity of	148,673
contains	133,241
increases	110,803
detects	93,373
decreases activity of	85,425
prevents	82,574
increases expression of	80,771
expresses	62,142
attenuates	54,865
decreases expression of	51,152
binds to	49,206
is a	47,435
affects expression of	37,399
Total	**3,388,450**

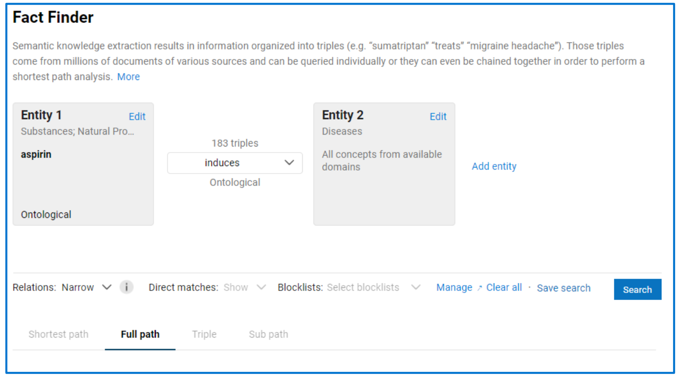


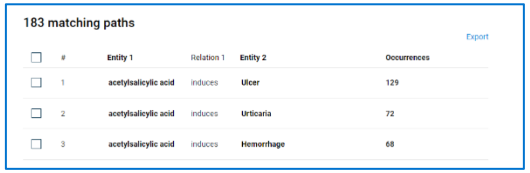


Model Experiments: OntoChem Knowledge Graphs

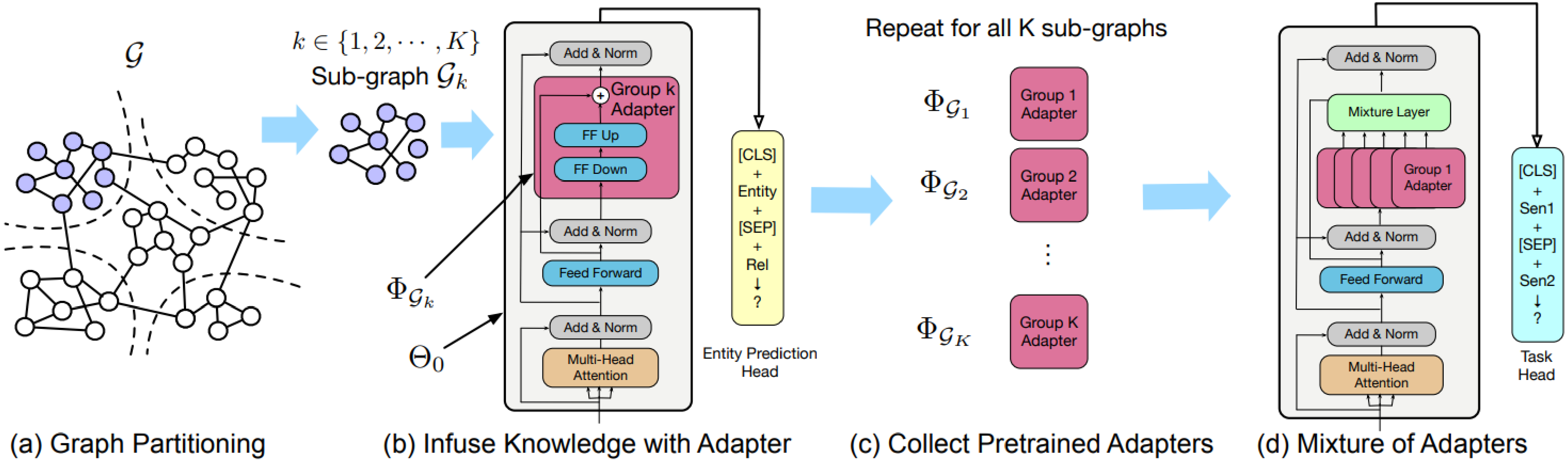
Onto20Rel	#Triplets
relates to	708,076
induces	502,512
modulates	326,534
treats	225,279
inhibits	219,720
is analyzed by	195,291
produces	173,979
increases activity of	148,673
contains	133,241
increases	110,803
detects	93,373
decreases activity of	85,425
prevents	82,574
increases expression of	80,771
expresses	62,142
attenuates	54,865
decreases expression of	51,152
binds to	49,206
is a	47,435
affects expression of	37,399
Total	**3,388,450**

OntoType20Rel	#Triplets
[protein] relates to [disease]	295,841
[substance] induces [physiology]	282,721
[food] contains [compound]	269,211
[substance] treats [disease]	247,348
[biomarker] of [disease]	205,604
[substance] is analyzed by [method]	130,275
[plant] produces [compound]	102,270
[protein] induces [physiology]	85,411
[compound] increases activity of [protein]	85,196
[compound] decreases activity of [protein]	72,311
[substance] inhibits [physiology]	68,728
[protein] is a [biomarker]	65,558
[anatomy] produces [protein]	64,206
[substance] prevents [disease]	60,260
[protein] induces [disease]	59,577
[substance] modulates [protein]	54,533
[protein] is analyzed by [method]	54,250
[method] treats [disease]	35,768
[method] detects [physiology]	33,504
[protein] modulates [physiology]	24,332
	2,296,904





Model Experiments: Methodology



[Me21] Meng, Z., Liu, F., Clark, T., et al.: Mixture-of-Partitions: Infusing Large Biomedical Knowledge Graphs into BERT
[Pf21] Pfeiffer, J., Kamath, A., Rücklé, A., et al.: AdapterFusion: Non-Destructive Task Composition for Transfer Learning

Model Experiments: Results

↓ model | dataset →	HoC	PubMedQA	BioASQ7b	MedNLI
SciBERT-base	$80.52_{\pm 0.60}$	$57.38_{\pm 4.22}$	$75.93_{\pm 4.20}$	$81.19_{\pm 0.54}$
+ MoP	$81.79^{\dagger}_{\pm 0.66}$ ↑	$54.66_{\pm 3.10}$	$78.50^{\dagger}_{\pm 4.06}$ ↑	$81.20_{\pm 0.37}$
+ KEBLM	/	59.0	/	82.14
BioBERT-base	$81.41_{\pm 0.59}$	$60.24_{\pm 2.32}$	$77.50_{\pm 2.92}$	$82.42_{\pm 0.59}$
+ MoP	$82.53^{\dagger}_{\pm 1.08}$ ↑	$61.04_{\pm 4.81}$ ↑	$80.79^{\dagger}_{\pm 4.40}$ ↑	$82.93_{\pm 0.55}$ ↑
+ KEBLM	/	**68.00**	/	84.24
+ DAKI	/	/	/	83.41
PubMedBERT-base	$82.25_{\pm 0.46}$	$55.84_{\pm 1.78}$	$87.71_{\pm 4.25}$	$84.18_{\pm 0.19}$
+ MoP	$\mathbf{83.26}^{\dagger}_{\pm 0.32}$ ↑	$62.84^{\dagger}_{\pm 2.71}$ ↑	$90.64^{\dagger}_{\pm 2.43}$ ↑	$\mathbf{84.70}_{\pm 0.19}$ ↑
+ OntoType20Rel (ours)	$82.17_{\pm 0.62}$	$55.40_{\pm 5.57}$	$86.36_{\pm 3.07}$	$83.94_{\pm 0.63}$
+ Onto20Rel (ours)	$82.39_{\pm 0.65}$ ↑	$56.12_{\pm 2.91}$ ↑	$84.36_{\pm 4.73}$	$83.97_{\pm 0.59}$
BioLinkBERT-base	$82.21_{\pm 0.87}$	$56.76_{\pm 3.00}$	$91.29_{\pm 3.18}$	$84.1_{\pm 0.03}$
+MoP (ours)	$82.36_{\pm 0.57}$ ↑	$63.62^{\dagger}_{\pm 5.31}$ ↑	$91.50_{\pm 2.25}$ ↑	$83.78_{\pm 0.09}$
+OntoType20Rel (ours)	$82.37_{\pm 0.42}$ ↑	$60.46_{\pm 5.81}$ ↑	$\mathbf{92.14}_{\pm 2.30}$ ↑	$82.84_{\pm 0.34}$
+Onto20Rel (ours)	$82.24_{\pm 1.25}$ ↑	$63.28^{\dagger}_{\pm 4.46}$ ↑	$90.57_{\pm 3.14}$	$83.69_{\pm 0.55}$

[Me21] Meng, Z., Liu, F., Clark, T., et al.: Mixture-of-Partitions: Infusing Large Biomedical Knowledge Graphs into BERT

[Gu20] Gu, Y.,Tinn, R., Cheng, H., et al.: Domain-Specific Language Model Pretraining for Biomedical Natural Language Processing

Model Experiments: Results

↓ model | dataset →	HoC	PubMedQA	BioASQ7b	MedNLI
SciBERT-base	$80.52_{\pm 0.60}$	$57.38_{\pm 4.22}$	$75.93_{\pm 4.20}$	$81.19_{\pm 0.54}$
+ MoP	$81.79^{\dagger}_{\pm 0.66} \uparrow$	$54.66_{\pm 3.10}$	$78.50^{\dagger}_{\pm 4.06} \uparrow$	$81.20_{\pm 0.37}$
+ KEBLM	/	59.0	/	82.14
BioBERT-base	$81.41_{\pm 0.59}$	$60.24_{\pm 2.32}$	$77.50_{\pm 2.92}$	$82.42_{\pm 0.59}$
+ MoP	$82.53^{\dagger}_{\pm 1.08} \uparrow$	$61.04_{\pm 4.81} \uparrow$	$80.79^{\dagger}_{\pm 4.40} \uparrow$	$82.93_{\pm 0.55} \uparrow$
+ KEBLM	/	**68.00**	/	84.24
+ DAKI	/	/	/	83.41
PubMedBERT-base	$82.25_{\pm 0.46}$	$55.84_{\pm 1.78}$	$87.71_{\pm 4.25}$	$84.18_{\pm 0.19}$
+ MoP	$\mathbf{83.26}^{\dagger}_{\pm 0.32} \uparrow$	$62.84^{\dagger}_{\pm 2.71} \uparrow$	$90.64^{\dagger}_{\pm 2.43} \uparrow$	$\mathbf{84.70}_{\pm 0.19} \uparrow$
+ OntoType20Rel (ours)	$82.17_{\pm 0.62}$	$55.40_{\pm 5.57}$	$86.36_{\pm 3.07}$	$83.94_{\pm 0.63}$
+ Onto20Rel (ours)	$82.39_{\pm 0.65} \uparrow$	$56.12_{\pm 2.91} \uparrow$	$84.36_{\pm 4.73}$	$83.97_{\pm 0.59}$
BioLinkBERT-base	$82.21_{\pm 0.87}$	$56.76_{\pm 3.00}$	$91.29_{\pm 3.18}$	$84.1_{\pm 0.03}$
+MoP (ours)	$82.36_{\pm 0.57} \uparrow$	$63.62^{\dagger}_{\pm 5.31} \uparrow$	$91.50_{\pm 2.25} \uparrow$	$83.78_{\pm 0.09}$
+OntoType20Rel (ours)	$82.37_{\pm 0.42} \uparrow$	$60.46_{\pm 5.81} \uparrow$	$\mathbf{92.14}_{\pm 2.30} \uparrow$	$82.84_{\pm 0.34}$
+Onto20Rel (ours)	$82.24_{\pm 1.25} \uparrow$	$63.28^{\dagger}_{\pm 4.46} \uparrow$	$90.57_{\pm 3.14}$	$83.69_{\pm 0.55}$

[Me21] Meng, Z., Liu, F., Clark, T., et al.: *Mixture-of-Partitions: Infusing Large Biomedical Knowledge Graphs into BERT*
[Gu20] Gu, Y.,Tinn, R., Cheng, H., et al.: *Domain-Specific Language Model Pretraining for Biomedical Natural Language Processing*

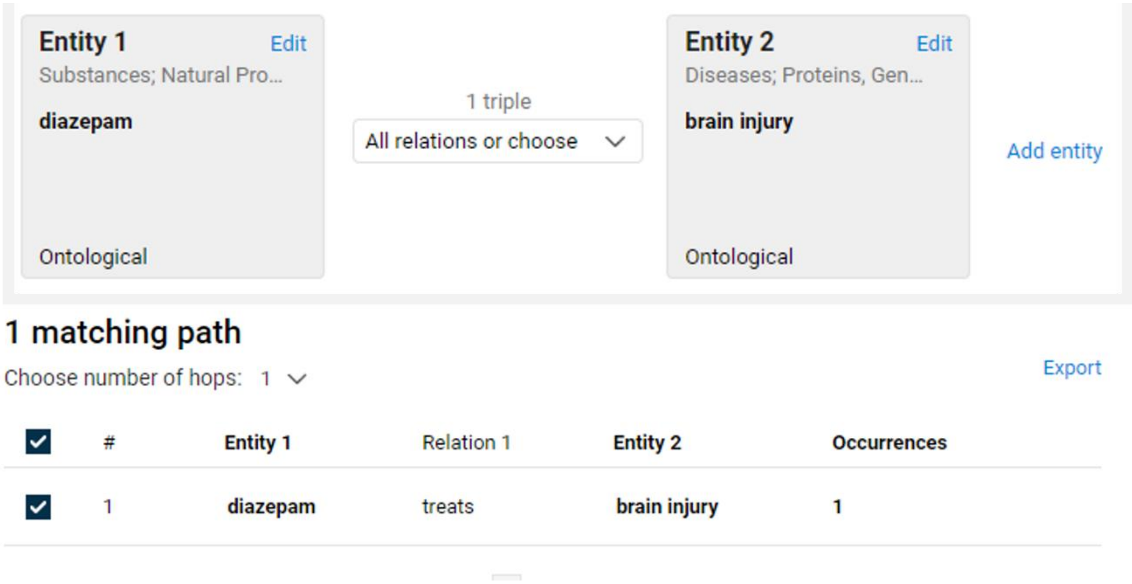
Model Experiments: Results (Qualitative Probing)

Question: Can Diazepam be beneficial in the treatment of traumatic brain injury?

Context: The present experiment examined the effects of diazepam, a positive modulator at the GABA(A) receptor, on survival and cognitive performance in traumatically brain-injured animals.

Predictions:

BioLinkBERT:	no
BioLinkBERT +OntoType20Rel:	yes
True Label:	yes

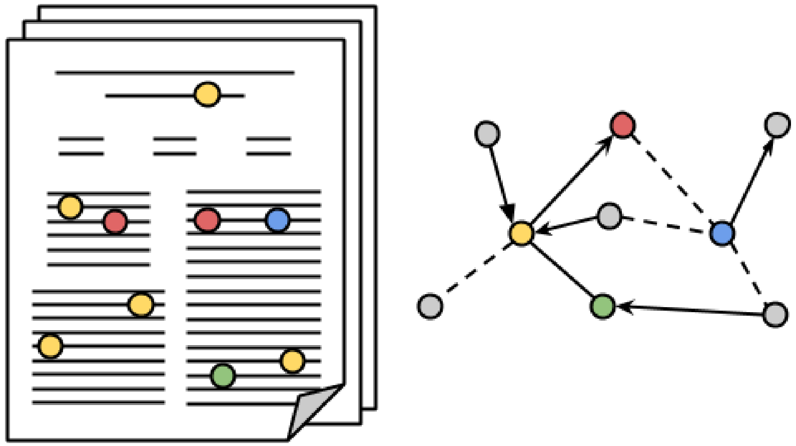


[On23] *Ontochem: SciWalker platform*

Model Experiments: Summary

RQ2: Can we improve existing approaches with new methods and data from a private ontology?

- We were able to improve KELMs with OntoChem KGs
- We proposed Onto20Rel and OntoType20Rel
- Diversification of KG Sources



Research Survey

RQ1: What adapter-based approaches to knowledge-enhancement exist, and how do they compare to each other?

- Systematic Literature Review (SLR)

RQ2: Can we improve existing approaches with new methods and data from a private ontology?

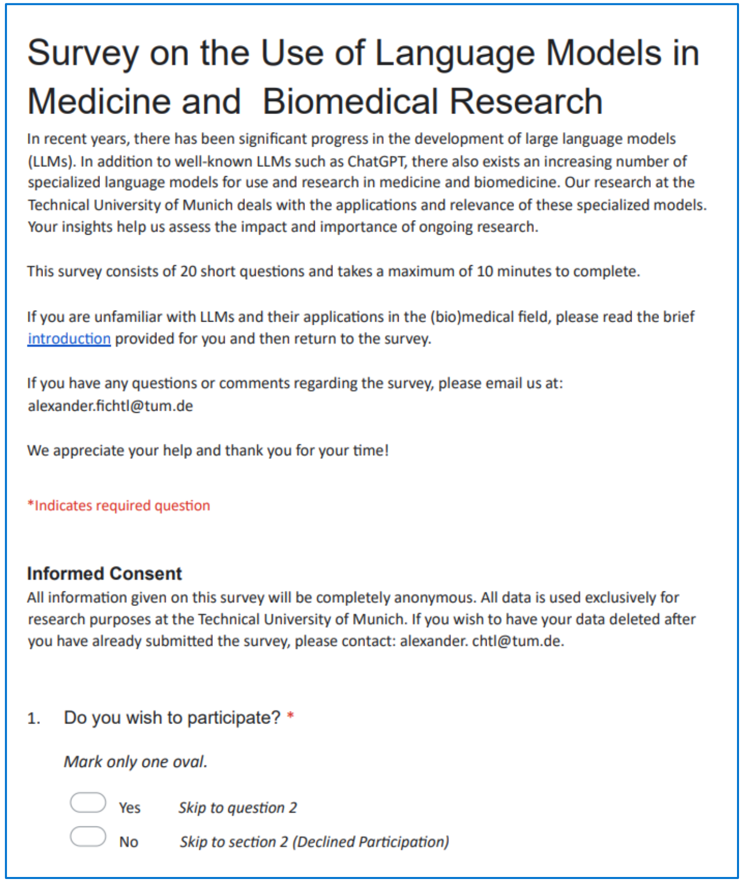
- SLR
- Thesis experiments with OntoChem KGs

RQ3: Is the research on biomedical KELMs relevant to medical professionals, and what factors hinder or support the deployment of the technology in practice?

- Research survey

Research Survey: Methodology

- **Extent:** Twenty questions, including a brief introduction to biomedical NLP
- **Targeting:** Medical centres and hospitals in Munich and the surrounding countryside, student groups at TUM and LMU, personal contacts of the authors and advisor
- **Related work:** Survey focus on NLP in biomedicine not yet represented in literature. Comparison with survey on AI in general



Survey on the Use of Language Models in Medicine and Biomedical Research

In recent years, there has been significant progress in the development of large language models (LLMs). In addition to well-known LLMs such as ChatGPT, there also exists an increasing number of specialized language models for use and research in medicine and biomedicine. Our research at the Technical University of Munich deals with the applications and relevance of these specialized models. Your insights help us assess the impact and importance of ongoing research.

This survey consists of 20 short questions and takes a maximum of 10 minutes to complete.

If you are unfamiliar with LLMs and their applications in the (bio)medical field, please read the brief introduction provided for you and then return to the survey.

If you have any questions or comments regarding the survey, please email us at: alexander.fichtl@tum.de

We appreciate your help and thank you for your time!

*Indicates required question

Informed Consent

All information given on this survey will be completely anonymous. All data is used exclusively for research purposes at the Technical University of Munich. If you wish to have your data deleted after you have already submitted the survey, please contact: alexander. chtl@tum.de.

1. Do you wish to participate? *

Mark only one oval.

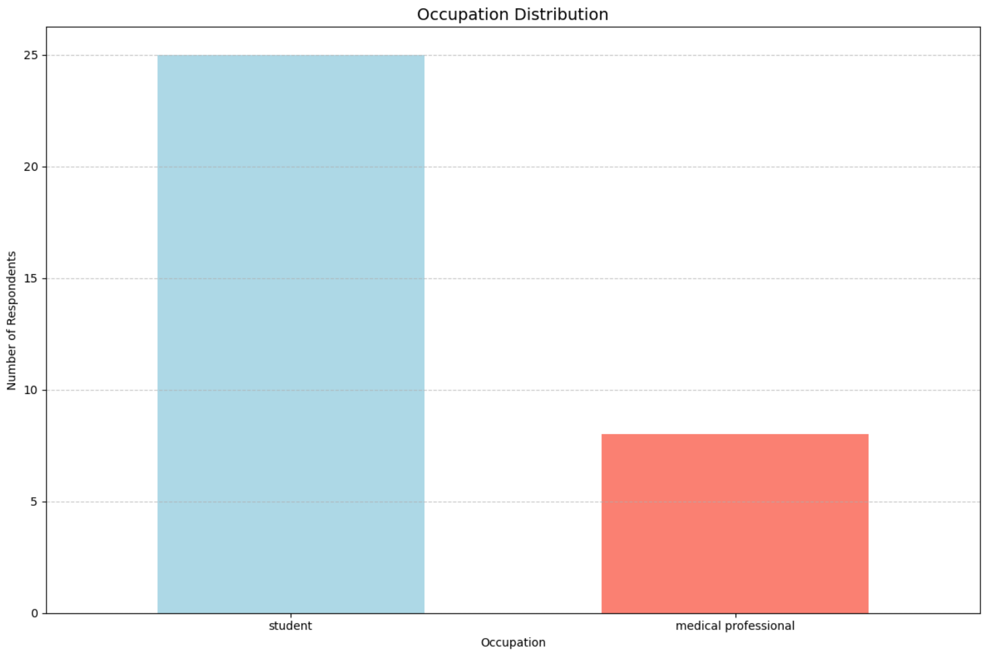
- Yes — *Skip to question 2*
- No — *Skip to section 2 (Declined Participation)*

[Sc21] Scheetz, S., Rothschild, P., McGuinness, M., et al.: A survey of clinicians on the use of artificial intelligence in ophthalmology, dermatology, radiology and Radiation Oncology

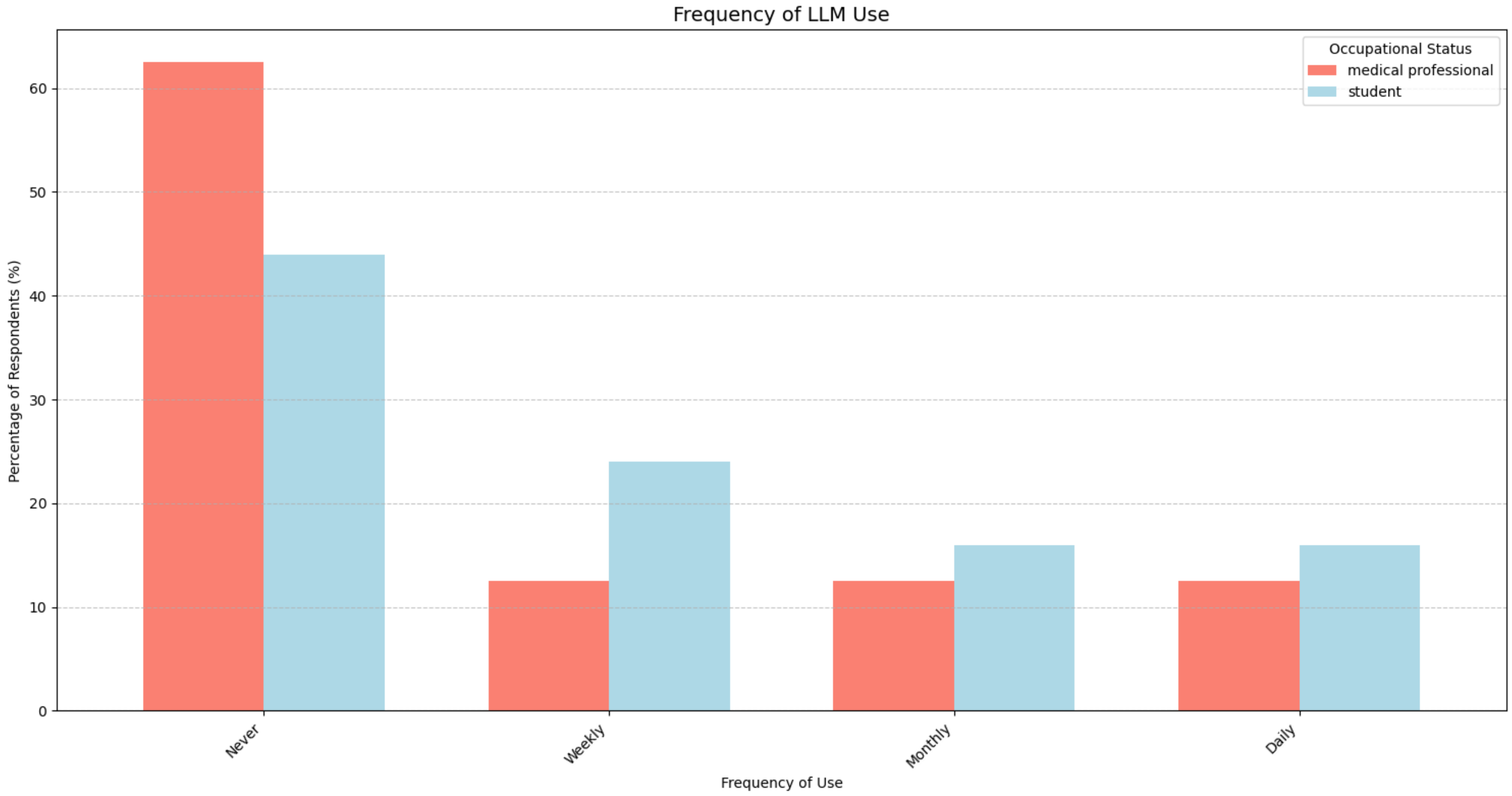
Research Survey: Results

- **Participants:** 34 participants, with 33 usable questionnaires
- **Distribution:** 25 students, 8 medical professionals
  - Experience levels from <5 to >30 years
  - Specialties vary strongly across medical disciplines

➢ Preliminary survey in preparation for survey at TUM Klinikum rechts der Isar (MRI)

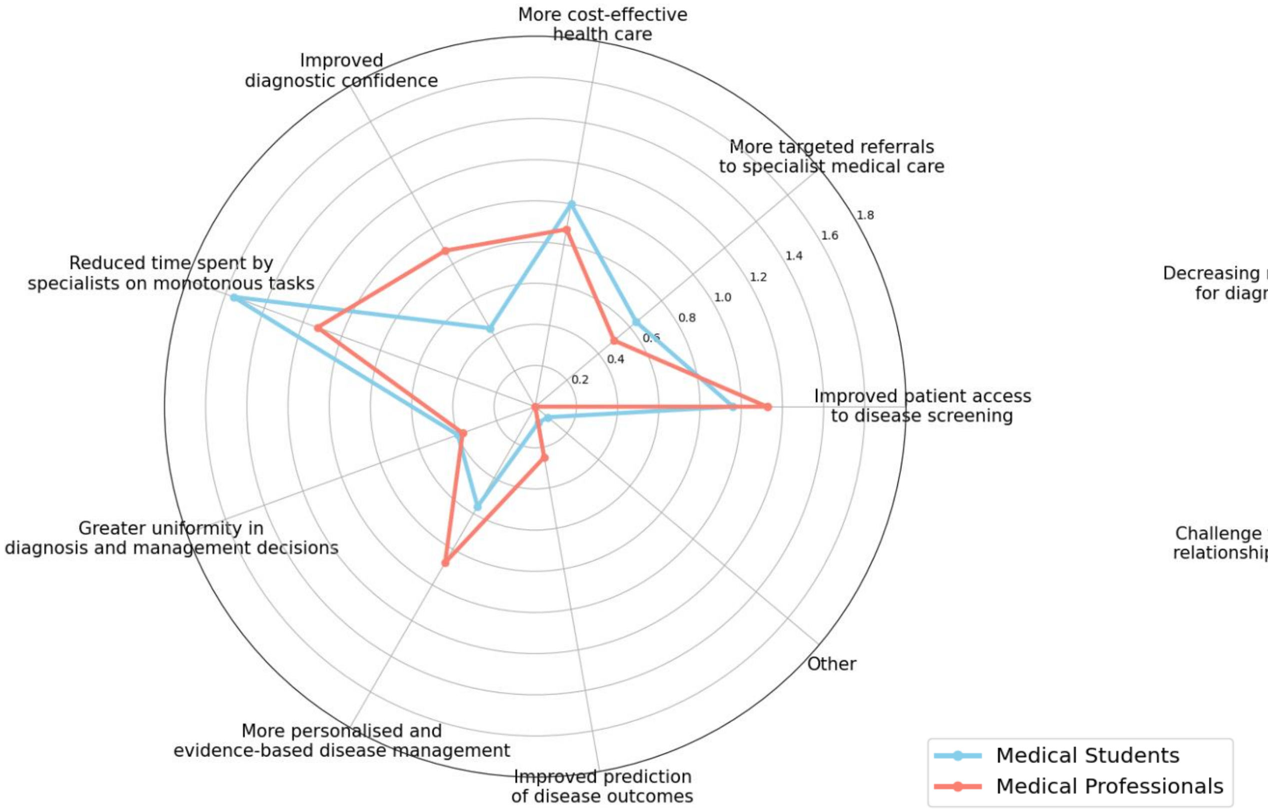


Research Survey: Results

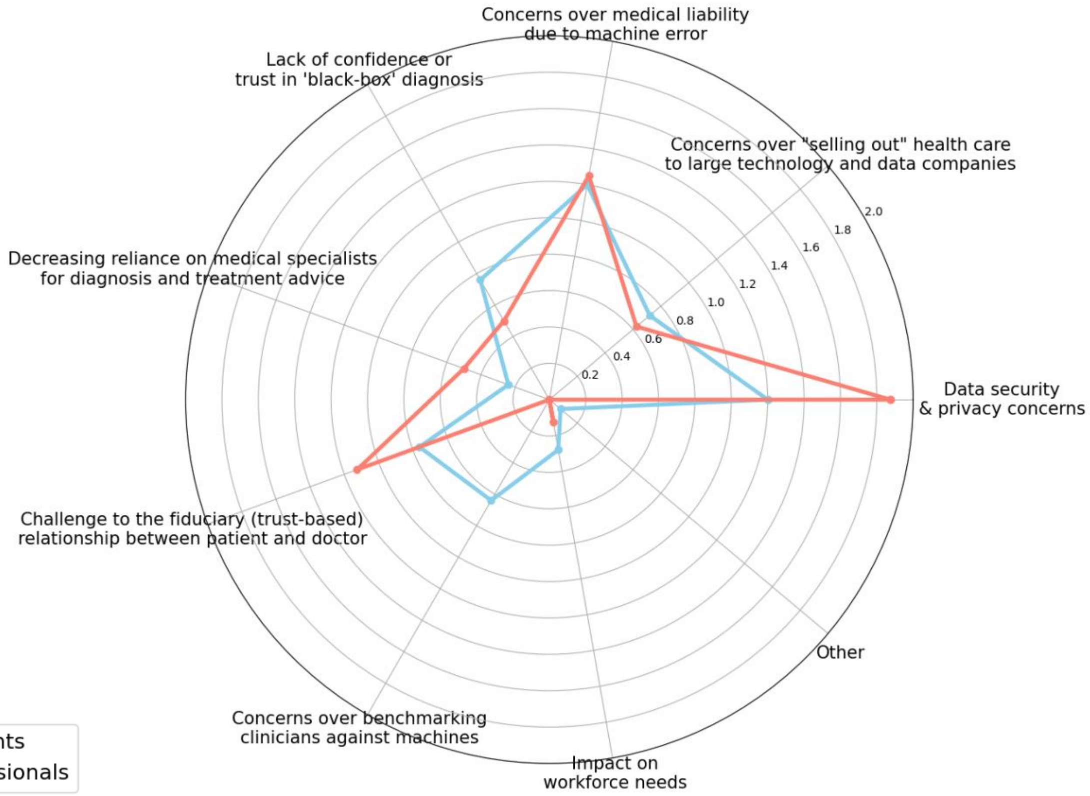


Research Survey: Results

Chances



Concerns



Research Survey: Results

“For me, teaching and training needs to be more rigorous so that those using the technology can detect potential issues. I do think that a free chatbot replacing an expensive doctor’s visit to tell you that you need bed rest and plenty of water is viable, but I think that it will take a decade to become mainstream.”

- Medical student

“More information about possible use cases should be provided in training. Also, experts should be hired who, in collaboration with doctors, can devise customized strategies or concepts tailored to individual needs."

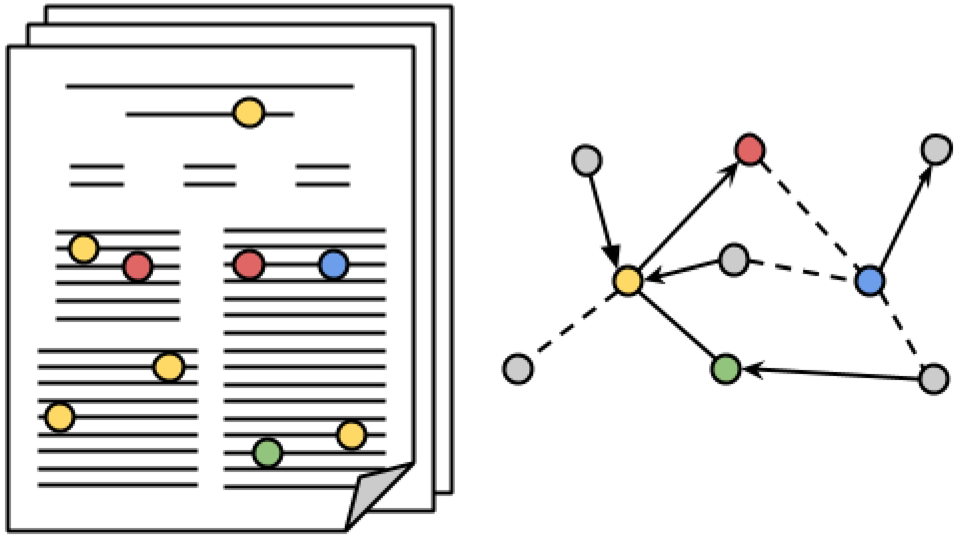
- Pathologist, 10-20 years of experience, translated from German

Responses edited for readability and privacy

Research Survey: Results

RQ3: Is the research on biomedical KELMs relevant to medical professionals, and what factors hinder or support the deployment of the technology in practice?

- The medical community is gaining familiarity with NLP and LLMs
- Participants recognize and anticipate chances and advantages
- There are prevalent concerns, like data security and the impact on the doctor-patient trust relationship, but notably not regarding model performance



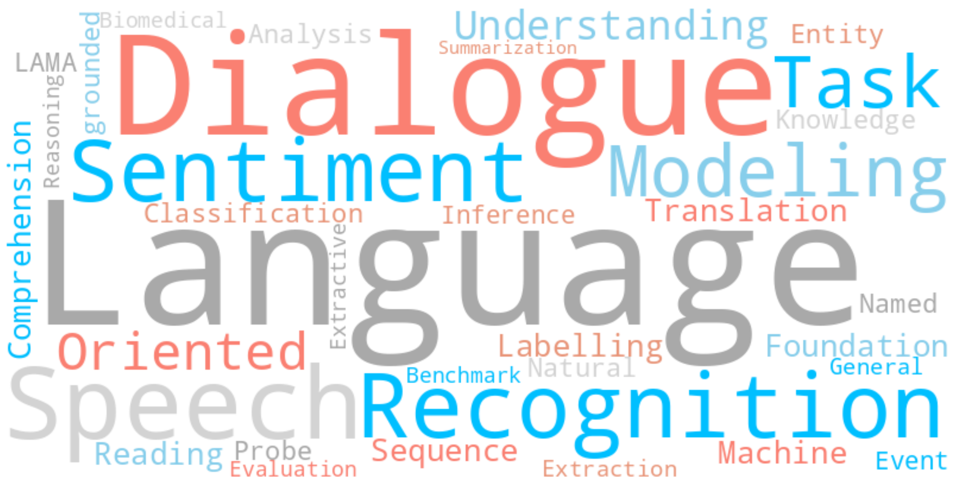
Shortcomings and Future Research

Main Shortcomings:

- Incomplete KGs
- Survey Response Rates

Future Research Opportunities:

- OntoChem Potential
- KG Merging
- …



Thanks!

Uhrenturm der TUM

B.Sc.

Alexander Fichtl

Technical University of Munich (TUM)
TUM School of CIT
Department of Computer Science (CS)
Chair of Software Engineering for Business Information Systems (sebis)

Boltzmannstraße 3
85748 Garching bei München

alexander.fichtl@tum.de
+49 160 97578593
wwwmatthes.in.tum.de

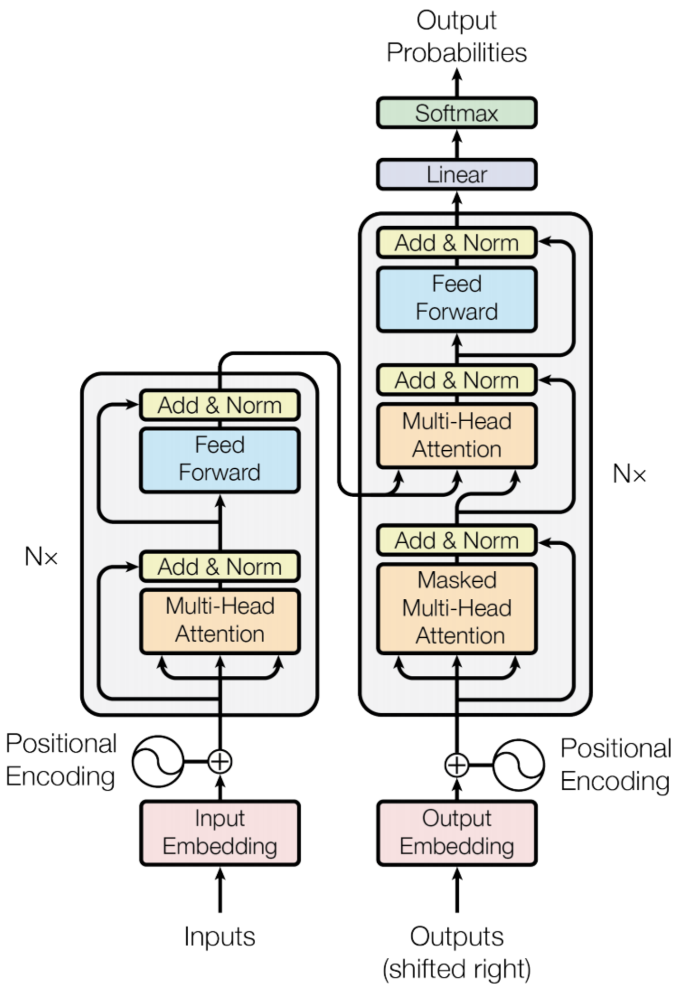
Appendix A: PubMedQA

Question:
Do preoperative statins reduce atrial fibrillation after coronary artery bypass grafting?
Context:
(Objective) Recent studies have demonstrated that statins have pleiotropic effects, including anti-inflammatory effects and atrial fibrillation (AF) preventive effects [...]
(Methods) 221 patients underwent CABG in our hospital from 2004 to 2007. 14 patients with preoperative AF and 4 patients with concomitant valve surgery [...]
(Results) The overall incidence of postoperative AF was 26%. ***Postoperative AF was significantly lower in the Statin group compared with the Non-statin group (16% versus 33%, p=0.005).*** Multivariate analysis demonstrated that independent predictors of AF [...]
Long Answer:
(Conclusion) Our study indicated that preoperative statin therapy seems to reduce AF development after CABG.
Answer: yes

Figure 1: An instance (Sakamoto et al., 2011) of PubMedQA dataset: Question is the original question title; Context includes the structured abstract except its conclusive part, which serves as the Long Answer; Human experts annotated the Answer yes. Supporting fact for the answer is ***highlighted***.

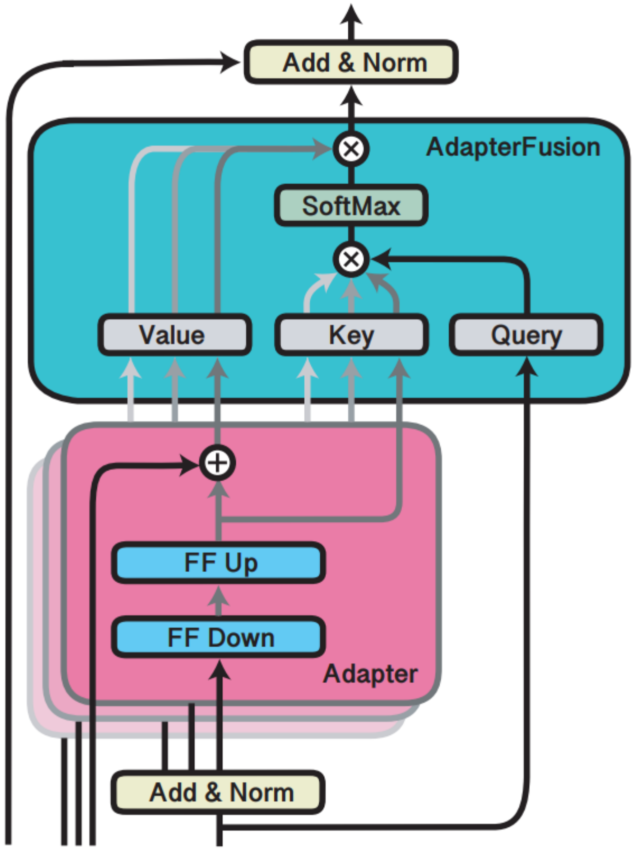
[Ji19] Jin,Q., Dhingra, B., Liu, Z., Cohen, W., et al.: PubMedQA: A Dataset for Biomedical Research Question Answering

Appendix B: Transformers



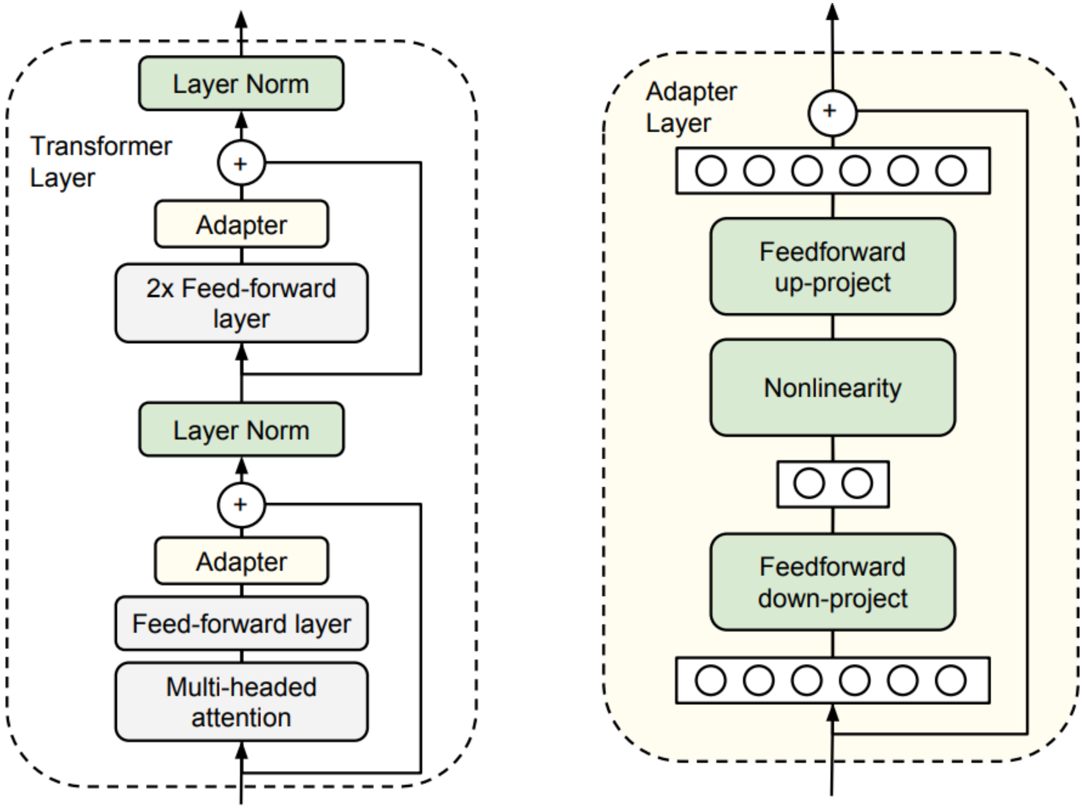
[Va17] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., et al.: Attention Is All You Need

Appendix C: Adapter Fusion



[Pf21] *Pfeiffer, J., Kamath, A., Rücklé, A., et al.:* AdapterFusion: Non-Destructive Task Composition for Transfer Learning

Appendix D: Houlsby Adapter



[Pf21] *Houlsby, N., Giurgiu, A., Jastrzebski, S., et al.: Parameter-Efficient Transfer Learning for NLP:*

Appendix E: Initial Research Questions

Can adapter-based approaches outperform other knowledge injection methods in downstream tasks (Blurb, claim verification)?

- Literature review
- Thesis experiments

How does the performance of KELMs in closed domains compare to open domain performance?

- Literature review
- Thesis experiments

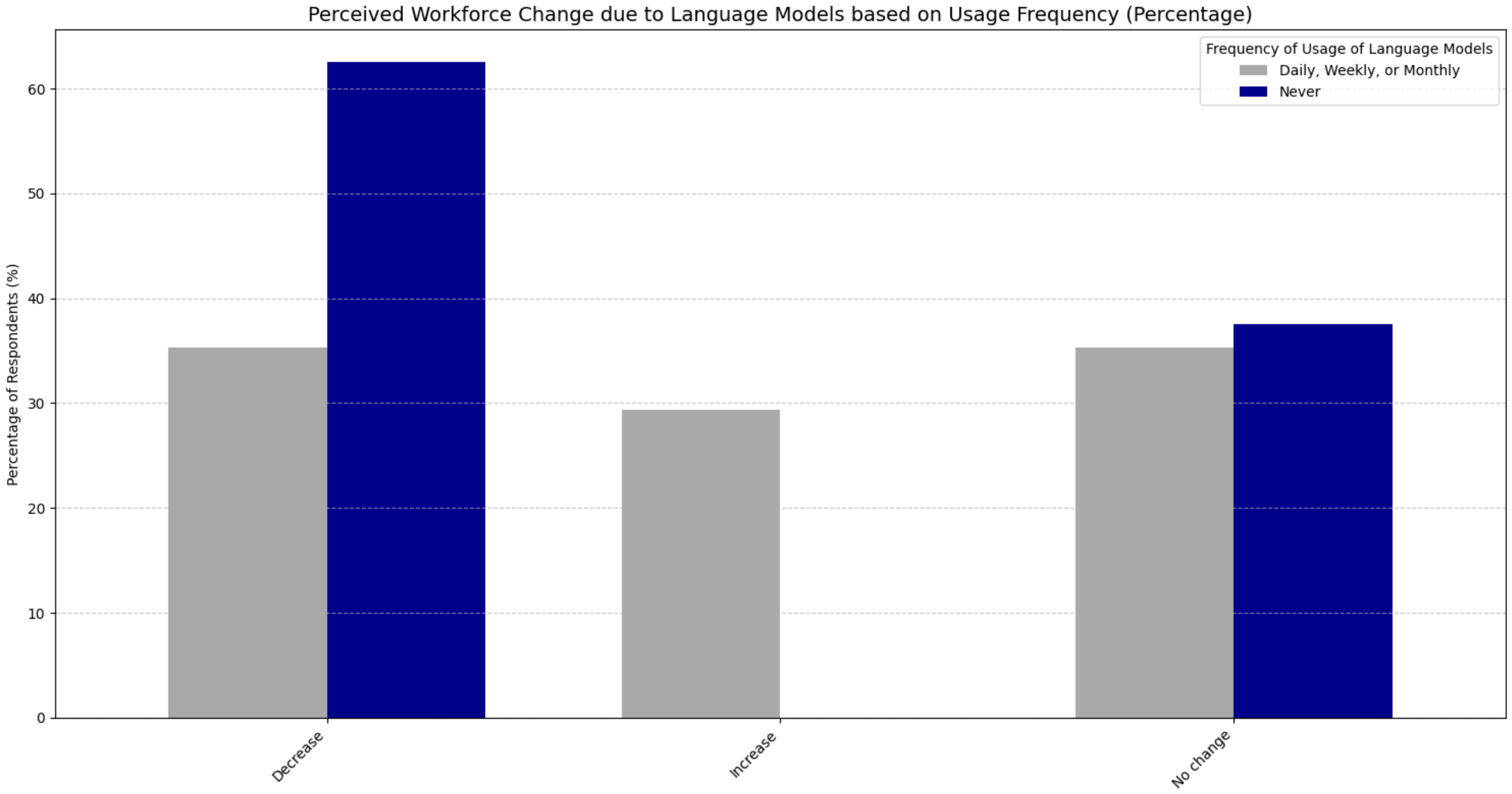
How do models trained on a private ontology (e.g., OntoChem) compare to models trained on public ontologies (e.g., UMLS)?

- Thesis experiments
- SciWalker

Is there interest in the results of this thesis amongst medical professionals and can they make use of biomedical KELMs?

- Survey and Interviews
- Mini-workshops on knowledge enhancement with adapters

Appendix F: Perceived Implication



Appendix G: Survey Participant Background

1. What is your (aspired) specialty?	2. Years of practice in your field of expertise?
Psychiatry	Currently in training/Student
Genetics	Currently in training/Student
Microbiology	Currently in training/Student
Data Scientist (Medical)	<5 years
August Staimann	Currently in training/Student
biostatistics	<5 years
Optometry	Currently in training/Student
training/Student	Currently in training/Student
medical technology	Currently in training/Student
Undecided	Currently in training/Student
	5-10 years
Plastic surgery	Currently in training/Student
designing	<5 years
Data science	5-10 years
Consultant medical doctor (obstetrics and gynaecology)	Currently in training/Student
Brain computer interfaces	Currently in training/Student
Innere	Currently in training/Student
Psychiatrie	Currently in training/Student
Chirurgie	Currently in training/Student
Experimentelle unfallchirurgische Forschung	<5 years
Innere Medizin (hausärztlich)	20-30 years
Psychiatrie und Psychotherapie	Currently in training/Student
Innere Medizin	Currently in training/Student
Allgemeinmedizin	Currently in training/Student
Innere	Currently in training/Student
Pädiatrie	Currently in training/Student
Physiologie Forschung	>30 years
Öffentliche Apotheke	<5 years
Notfallpflege	>30 years
Orthopädie	Currently in training/Student
Chirurgin	Currently in training/Student
TBD	Currently in training/Student
Pathologie	10-20 years