



Using LLMs for data- and knowledge management in radiation oncology

Fabio Dennstädt

Projects of the ISRO



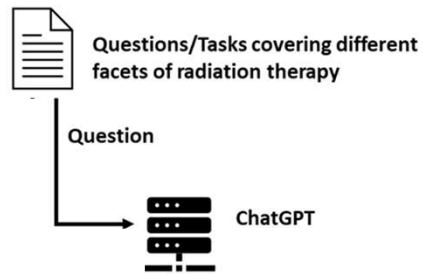
> [Adv Radiat Oncol](#). 2023 Nov 4;9(3):101400. doi: 10.1016/j.adro.2023.101400. eCollection 2024 Mar.

Exploring Capabilities of Large Language Models such as ChatGPT in Radiation Oncology

Fabio Dennstädt¹, Janna Hastings^{2 3}, Paul Martin Putora^{1 4}, Erwin Vu¹, Galina F Fischer¹, Krisztian Süveg¹, Markus Glatzer¹, Elena Riggerbach⁴, Hông-Linh Hà⁴, Nikola Cihoric⁴

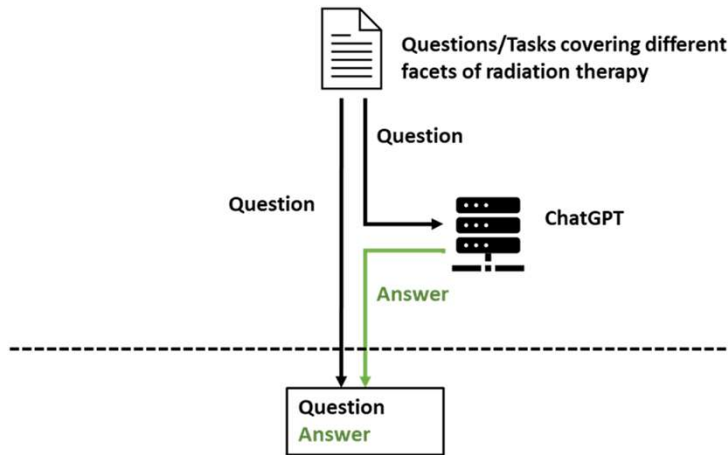
Exploring Capabilities of Large Language Models such as ChatGPT in Radiation Oncology

Fabio Dennstädt¹, Janna Hastings^{2 3}, Paul Martin Putora^{1 4}, Erwin Vu¹, Galina F Fischer¹, Krisztian Süveg¹, Markus Glatzer¹, Elena Riggenschach⁴, Hông-Linh Hà⁴, Nikola Cihoric⁴



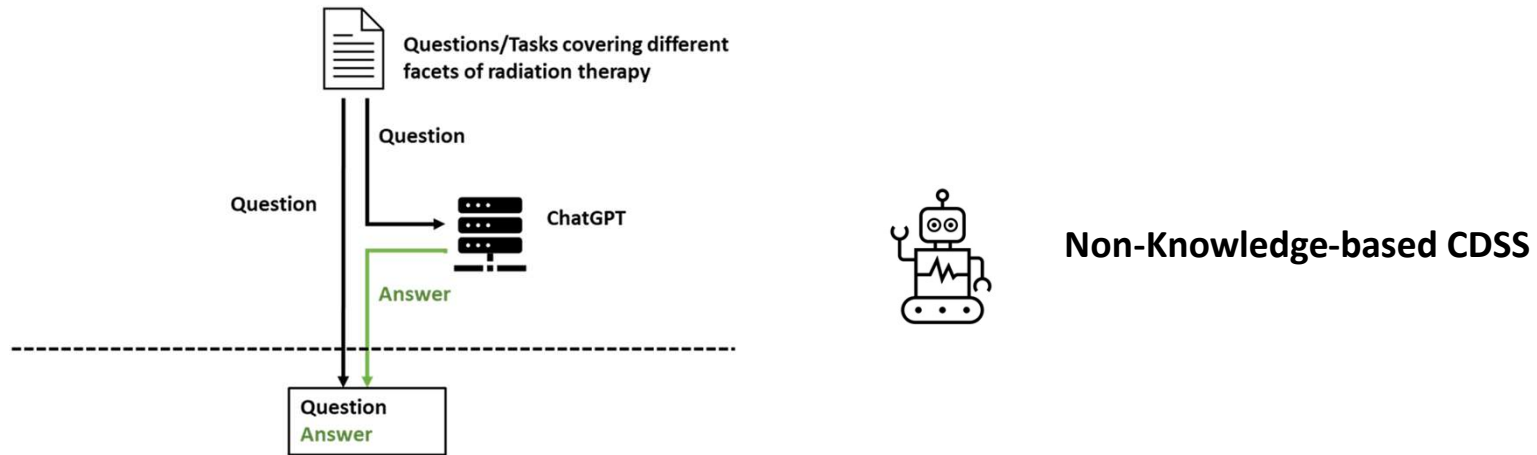
Exploring Capabilities of Large Language Models such as ChatGPT in Radiation Oncology

Fabio Dennstädt¹, Janna Hastings^{2,3}, Paul Martin Putora^{1,4}, Erwin Vu¹, Galina F Fischer¹, Krisztian Süveg¹, Markus Glatzer¹, Elena Riggenschach⁴, Hông-Linh Hà⁴, Nikola Cihoric⁴



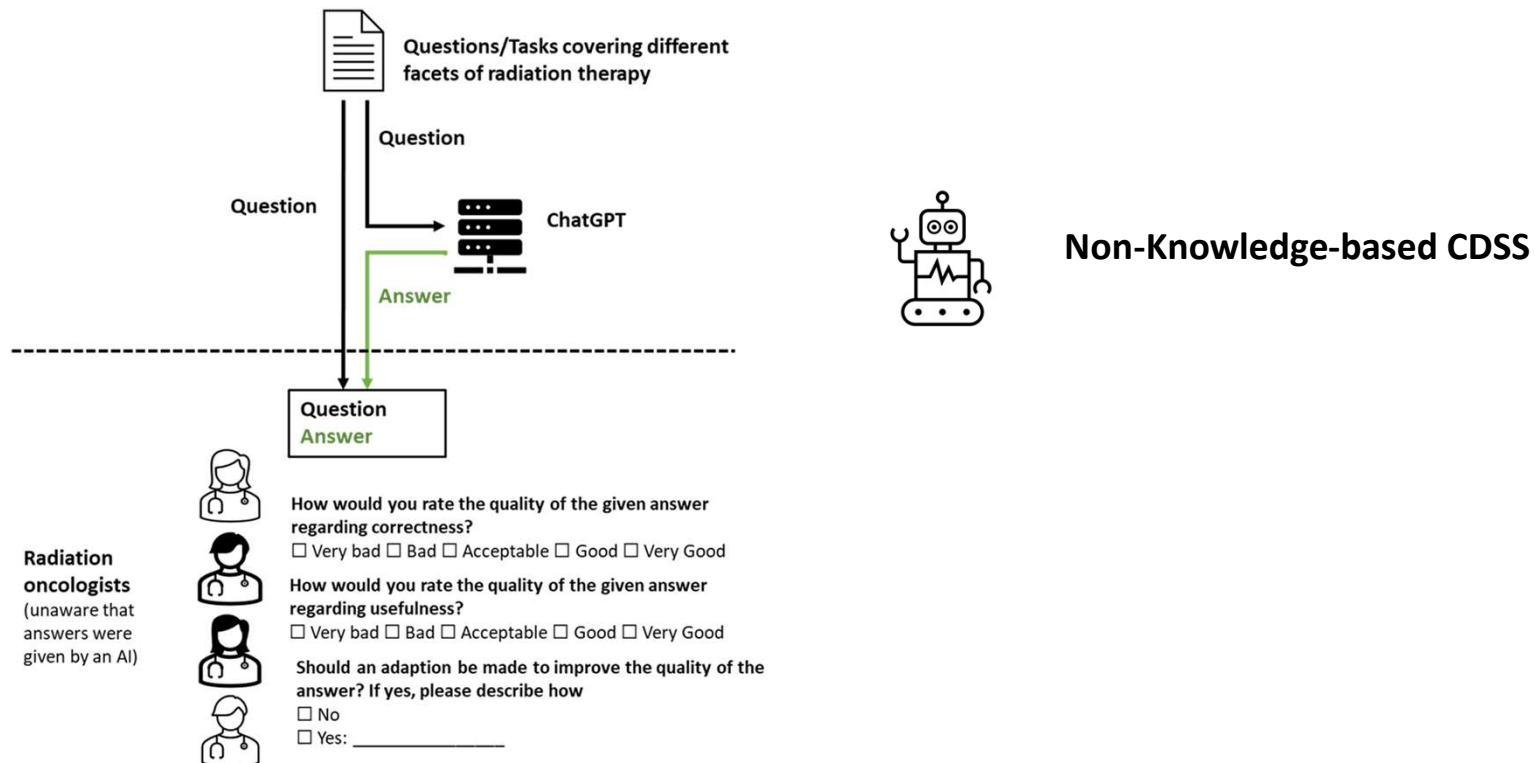
Exploring Capabilities of Large Language Models such as ChatGPT in Radiation Oncology

Fabio Dennstädt¹, Janna Hastings^{2,3}, Paul Martin Putora^{1,4}, Erwin Vu¹, Galina F Fischer¹, Krisztian Süveg¹, Markus Glatzer¹, Elena Riggenschach⁴, Hông-Linh Hà⁴, Nikola Cihoric⁴



Exploring Capabilities of Large Language Models such as ChatGPT in Radiation Oncology

Fabio Dennstädt¹, Janna Hastings^{2 3}, Paul Martin Putora^{1 4}, Erwin Vu¹, Galina F Fischer¹, Krisztian Süveg¹, Markus Glatzer¹, Elena Riggenschach⁴, Hông-Linh Hà⁴, Nikola Cihoric⁴



Example of „good answer“

“A patient with limited disease small-cell lung cancer had chemotherapy with concurrent radiotherapy for 6 weeks. After 4 weeks of the treatment, he is beginning to develop symptoms of pain during swallowing (VAS 6/10), which he describes as burning. He also suffers from impaired swallowing, which is why he reduced eating. What is the most likely reason for the symptoms and how could you help the patient?”.

Example of „good answer“

“A patient with limited disease **small-cell lung cancer** had **chemotherapy with concurrent radiotherapy** for 6 weeks. After 4 weeks of the treatment, he is beginning to develop **symptoms of pain** during swallowing (VAS 6/10), which he describes as burning. He also suffers from **impaired swallowing**, which is why he reduced eating. What is the most likely reason for the symptoms and how could you help the patient?”.

“The most likely reason for the patient's symptoms of pain during swallowing and impaired swallowing is radiation-induced esophagitis, which is a common side effect of concurrent chemoradiotherapy for lung cancer. The symptoms can be due to inflammation and injury of the esophageal mucosa caused by the radiation.

To help the patient, the following measures can be taken:

- 1. Pain management: Over-the-counter pain relievers or prescription pain medication can be used to manage the pain.***
- 2. Diet modification: Soft, bland foods and frequent small meals can help alleviate symptoms and improve swallowing. Avoiding spicy and acidic foods can also help.***
- 3. Mouth care: Good oral hygiene, such as regular mouth rinses and gargles, can help to soothe the throat and prevent infections.***
- 4. Lubricating agents: Using agents such as honey, glycerin, or saliva substitutes can help to lubricate the throat and improve swallowing.***

It is important to consult a medical professional for proper evaluation and management of the patient's symptoms, as the above measures may not be sufficient to alleviate symptoms in all cases. Additionally, if the symptoms persist or worsen, further investigations such as endoscopy may be necessary to rule out other causes.”

“The most likely reason for the patient's symptoms of pain during swallowing and impaired swallowing is radiation-induced esophagitis which is a common side effect of concurrent chemoradiotherapy for lung cancer. The symptoms can be due to inflammation and injury of the esophageal mucosa caused by the radiation.

To help the patient, the following measures can be taken:

- 1. Pain management: Over-the-counter pain relievers or prescription pain medication can be used to manage the pain.***
- 2. Diet modification: Soft, bland foods and frequent small meals can help alleviate symptoms and improve swallowing. Avoiding spicy and acidic foods can also help.***
- 3. Mouth care: Good oral hygiene, such as regular mouth rinses and gargles, can help to soothe the throat and prevent infections.***
- 4. Lubricating agents: Using agents such as honey, glycerin, or saliva substitutes can help to lubricate the throat and improve swallowing.***

It is important to consult a medical professional for proper evaluation and management of the patient's symptoms, as the above measures may not be sufficient to alleviate symptoms in all cases. Additionally, if the symptoms persist or worsen, further investigations such as endoscopy may be necessary to rule out other causes.”

***“The most likely reason for the patient's symptoms of pain during swallowing and impaired swallowing is radiation-induced esophagitis which is a common side effect of concurrent chemoradiotherapy for lung cancer. The symptoms can be due to inflammation and injury of the esophageal mucosa caused by the radiation.*”**



To help the patient, the following measures can be taken:

1. Pain management: Over-the-counter pain relievers or prescription pain medication can be used to manage the pain.



2. Diet modification: Soft, bland foods and frequent small meals can help alleviate symptoms and improve swallowing. Avoiding spicy and acidic foods can also help.



3. Mouth care: Good oral hygiene, such as regular mouth rinses and gargles, can help to soothe the throat and prevent infections.



4. Lubricating agents: Using agents such as honey, glycerin, or saliva substitutes can help to lubricate the throat and improve swallowing.



It is important to consult a medical professional for proper evaluation and management of the patient's symptoms, as the above measures may not be sufficient to alleviate symptoms in all cases. Additionally, if the symptoms persist or worsen, further investigations such as endoscopy may be necessary to rule out other causes.”

“The most likely reason for the patient's symptoms of pain during swallowing and impaired swallowing is radiation-induced esophagitis which is a common side effect of concurrent chemoradiotherapy for lung cancer. The symptoms can be due to inflammation and injury of the esophageal mucosa caused by the radiation.



To help the patient, the following measures can be taken:

1. Pain management: Over-the-counter pain relievers or prescription pain medication can be used to manage the pain.



2. Diet modification: Soft, bland foods and frequent small meals can help alleviate symptoms and improve swallowing. Avoiding spicy and acidic foods can also help.



3. Mouth care: Good oral hygiene, such as regular mouth rinses and gargles, can help to soothe the throat and prevent infections.



4. Lubricating agents: Using agents such as honey, glycerin, or saliva substitutes can help to lubricate the throat and improve swallowing.



It is important to consult a medical professional for proper evaluation and management of the patient's symptoms, as the above measures may not be sufficient to alleviate symptoms in all cases. Additionally, if the symptoms persist or worsen, further investigations such as endoscopy may be necessary to rule out other causes.”

Score of 5.00 regarding both correctness and usefulness

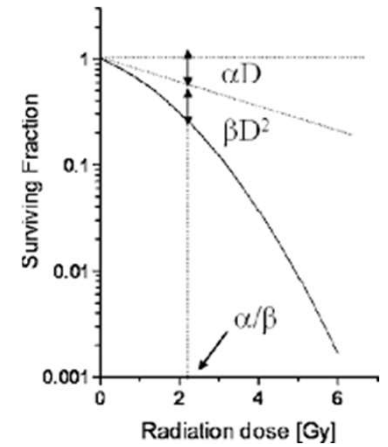
Example of „bad answer“

- “What is 20x3Gy in EQD2 based on the linear-quadratic model using an $\alpha/\beta=2$?”.

Example of „bad answer“

- “What is 20x3Gy in EQD2 based on the linear-quadratic model using an $\alpha/\beta=2$?”.

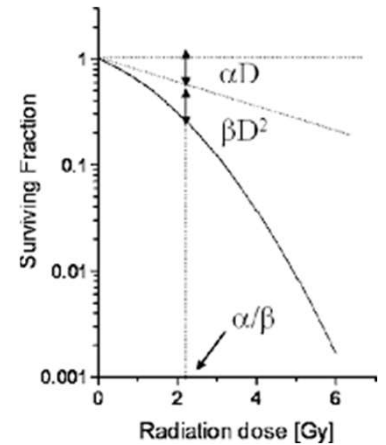
In this case the correct answer would be EQD2 = 75 Gy.



Example of „bad answer“

- “What is 20x3Gy in EQD2 based on the linear-quadratic model using an $\alpha/\beta=2$?”.

In this case the correct answer would be EQD2 = 75 Gy.



“The equivalent dose in 2 Gy fractions (EQD2) can be calculated using the linear-quadratic model and an α/β of 2 as follows:

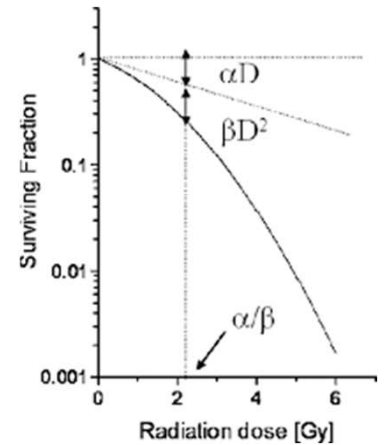
$$EQD2 = 20 \times 3Gy * (2 + (2/\alpha/\beta)) = 20 \times 3Gy * (2 + (2/2)) = 20 \times 3Gy * 4 = 240 Gy$$

So, the EQD2 for 20 fractions of 3 Gy each with an α/β of 2 is 240 Gy.”

Example of „bad answer“

- “What is 20x3Gy in EQD2 based on the linear-quadratic model using an $\alpha/\beta=2$?”.

In this case the correct answer would be EQD2 = 75 Gy.



“The equivalent dose in 2 Gy fractions (EQD2) can be calculated using the linear-quadratic model and an α/β of 2 as follows:

$$EQD2 = 20 \times 3Gy * (2 + (2/\alpha/\beta)) = 20 \times 3Gy * (2 + (2/2)) = 20 \times 3Gy * 4 = 240 Gy$$

So, the EQD2 for 20 fractions of 3 Gy each with an α/β of 2 is 240 Gy.”



➔ Overall mixed result with good performance on some of the questions.

→ Overall mixed result with good performance on some of the questions.

Multicentric follow-up study with medically optimized LLM in collaboration with task group „digitization and AI“ of DEGRO ongoing.

LLM-based data-extraction from clinical text


LLM-based data-extraction from clinical text



Diagnosis lists
(free text)

„Breast cancer **left side** at 12.00 NST **cT1c** cN0 G2 R0 ER
100%
- Screening detected
- 20.09.22 **segmentectomy** left breast
- Histology: **15 mm** breast cancer NST, pN0 (sn) (0/2)
...“

LLM-based data-extraction from clinical text



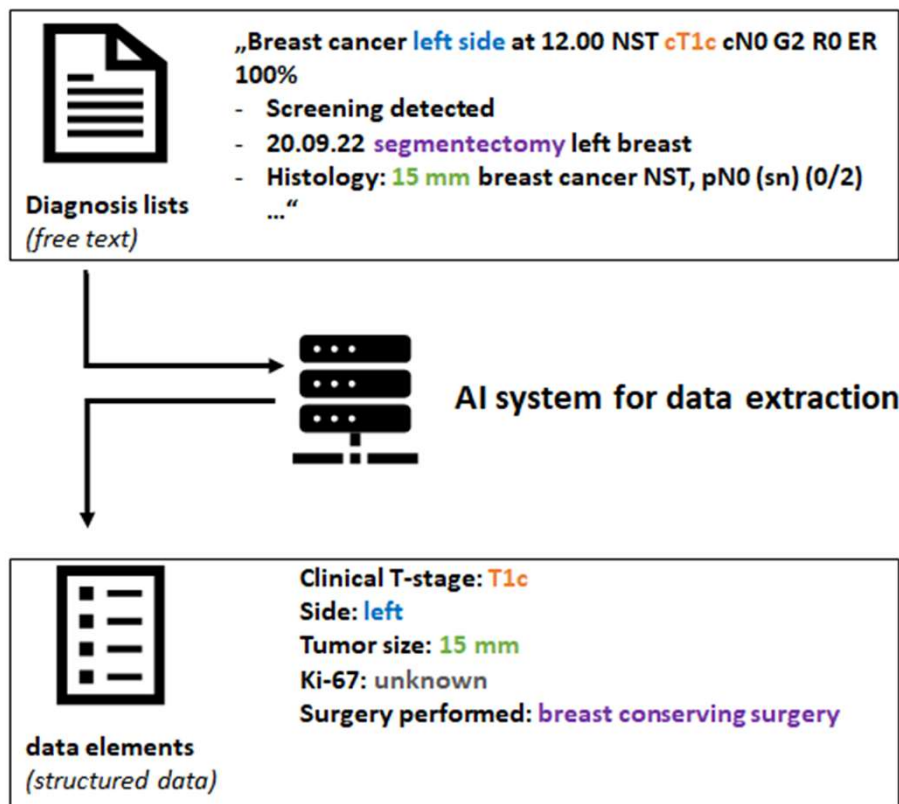
„Breast cancer **left side** at 12.00 NST **cT1c** cN0 G2 R0 ER
100%
- Screening detected
- 20.09.22 **segmentectomy** left breast
- Histology: **15 mm** breast cancer NST, pN0 (sn) (0/2)
...“

Diagnosis lists
(free text)



AI system for data extraction

LLM-based data-extraction from clinical text



Ongoing study, testing on
>200 diagnosis lists in
radiotherapy

Ongoing study, testing on
>200 diagnosis lists in
radiotherapy

CDE	Correct values given by LLM
Age	100%
Laterality	96%
Menopausal Status	99%
molecular Type	99%
Laterality	96%
L-Status	96%
progesteron receptor status	95%
estrogen receptor status	95%
R-Status	94%
...	

Ongoing study, testing on >200 diagnosis lists in radiotherapy

- 31 different CDEs

CDE	Correct values given by LLM
Age	100%
Laterality	96%
Menopausal Status	99%
molecular Type	99%
Laterality	96%
L-Status	96%
progesteron receptor status	95%
estrogen receptor status	95%
R-Status	94%
...	

Ongoing study, testing on >200 diagnosis lists in radiotherapy

- 31 different CDEs
- 7 \geq 95% performance

CDE	Correct values given by LLM
Age	100%
Laterality	96%
Menopausal Status	99%
molecular Type	99%
Laterality	96%
L-Status	96%
progesteron receptor status	95%
estrogen receptor status	95%
R-Status	94%
...	

Ongoing study, testing on >200 diagnosis lists in radiotherapy

- 31 different CDEs
- 7 \geq 95% performance
- 13 \geq 90% performance

CDE	Correct values given by LLM
Age	100%
Laterality	96%
Menopausal Status	99%
molecular Type	99%
Laterality	96%
L-Status	96%
progesteron receptor status	95%
estrogen receptor status	95%
R-Status	94%
...	

SMARAGD project

SMARAGD project

- **Smart Radiology Goes Digital**



SMARAGD project

- Smart Radiology Goes Digital



Development of a software system to automate extraction and transfer of data along the care path of breast cancer patients, including the domains of radiology, pathology and radiation therapy

SMARAGD project



- Smart Radiology Goes Digital

Development of a software system to automate extraction and transfer of data along the care path of breast cancer patients, including the domains of radiology, pathology and radiation therapy

- Collaboration between academic and industry partners, funded by Innosuisse

 INSELSPITAL

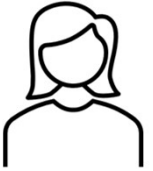

Berner
Fachhochschule

Swiss Institute for Translational
and Entrepreneurial Medicine


 Wemedoo
Clinical Information Specialists



ID Information und
Dokumentation im
Gesundheitswesen 



Breast cancer
patient



Radiology reports



Tumor board



Breast cancer patient



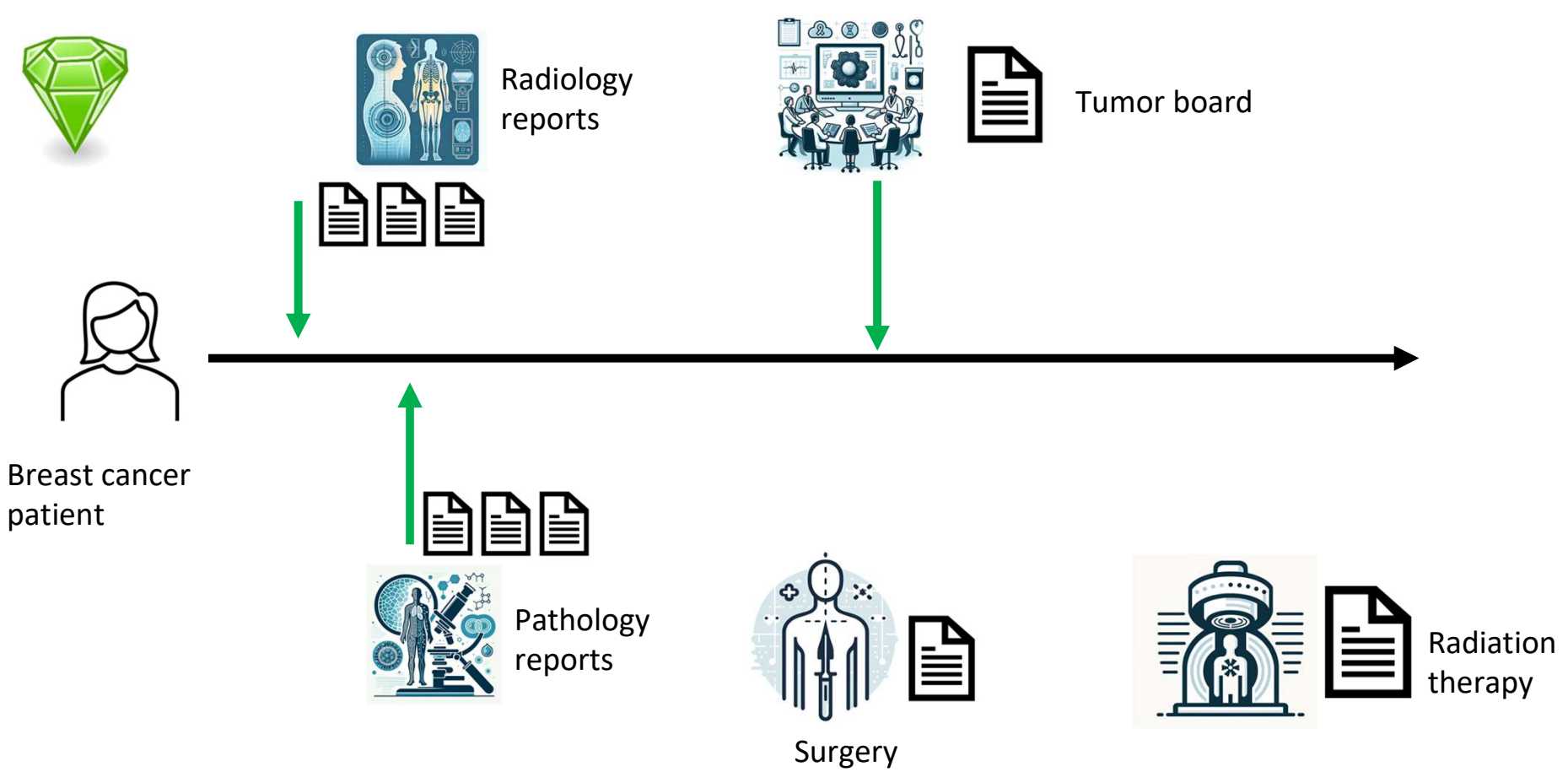
Pathology reports

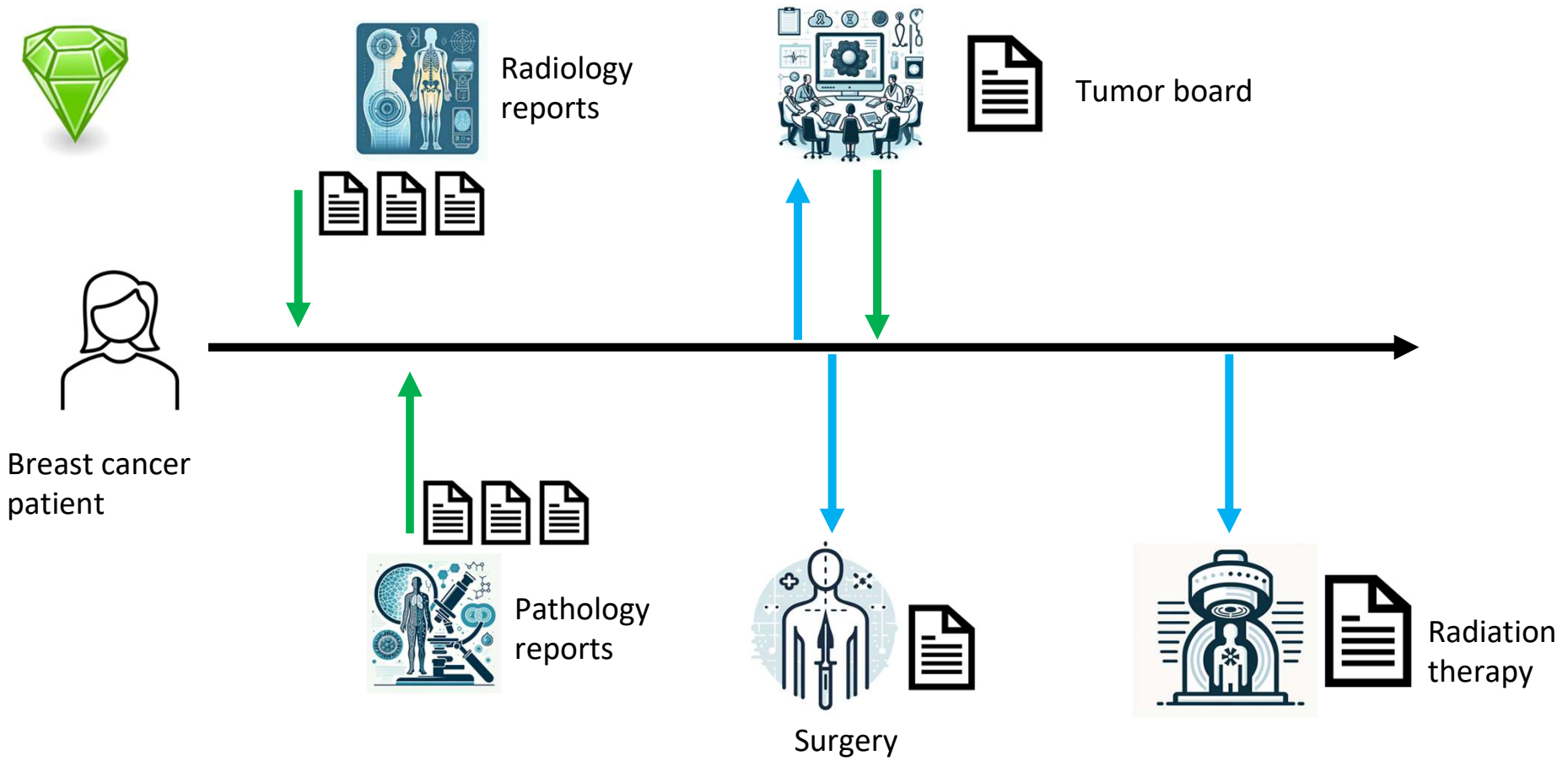


Surgery

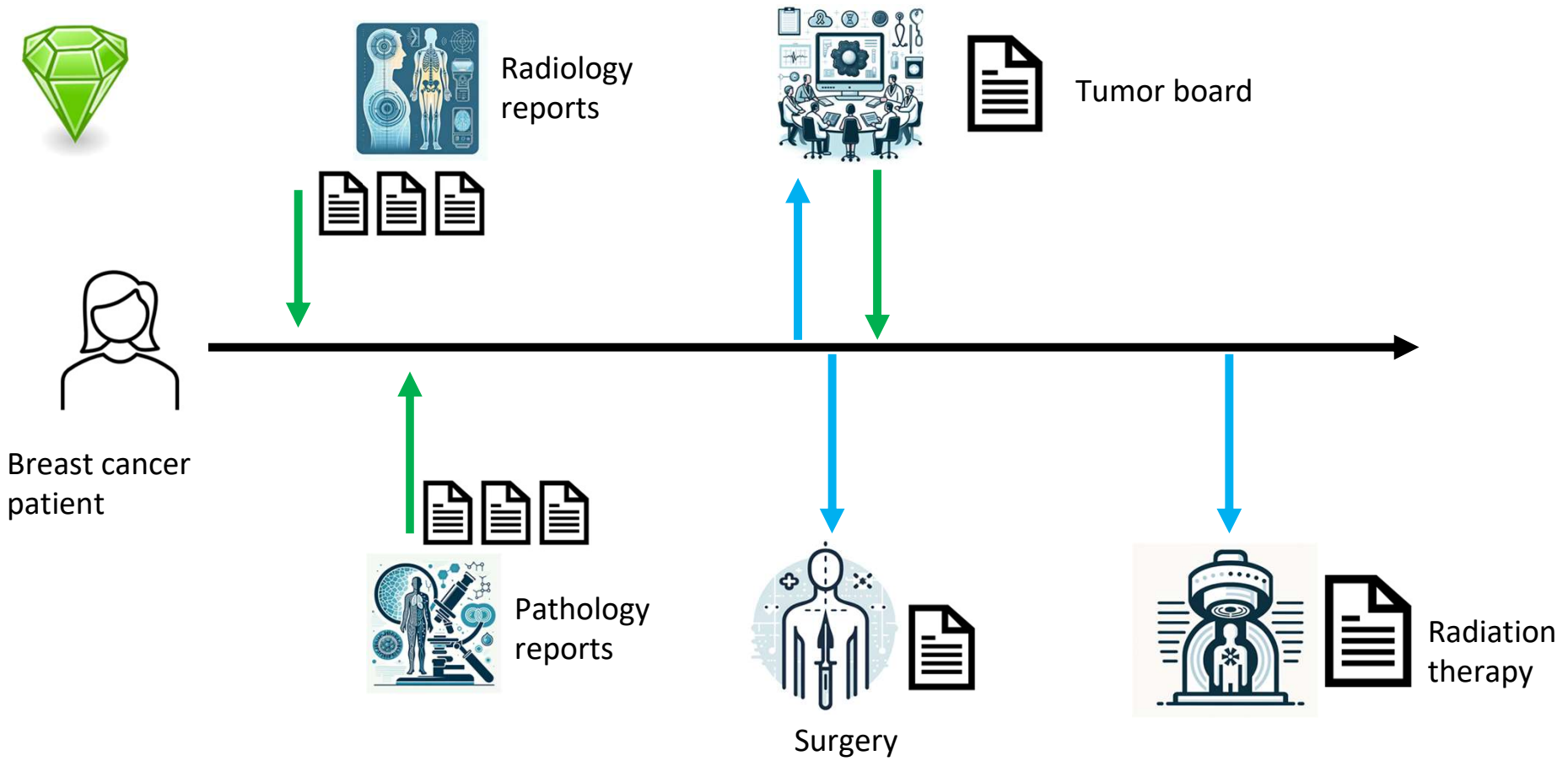


Radiation therapy







➔ Extraction of CDE values using LLMs
➔ Synoptic Report

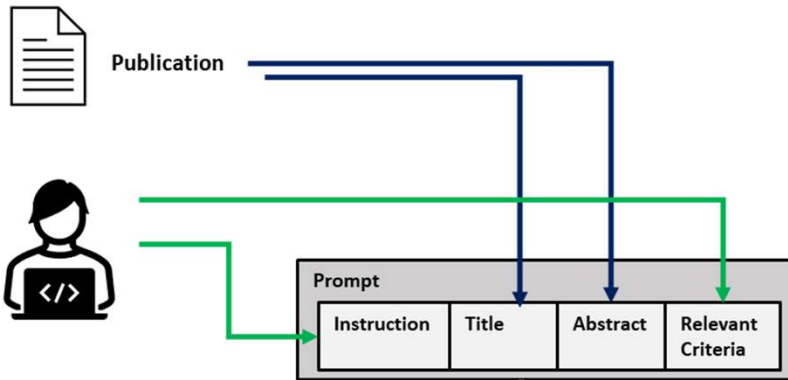


... results will be published towards end of 2024

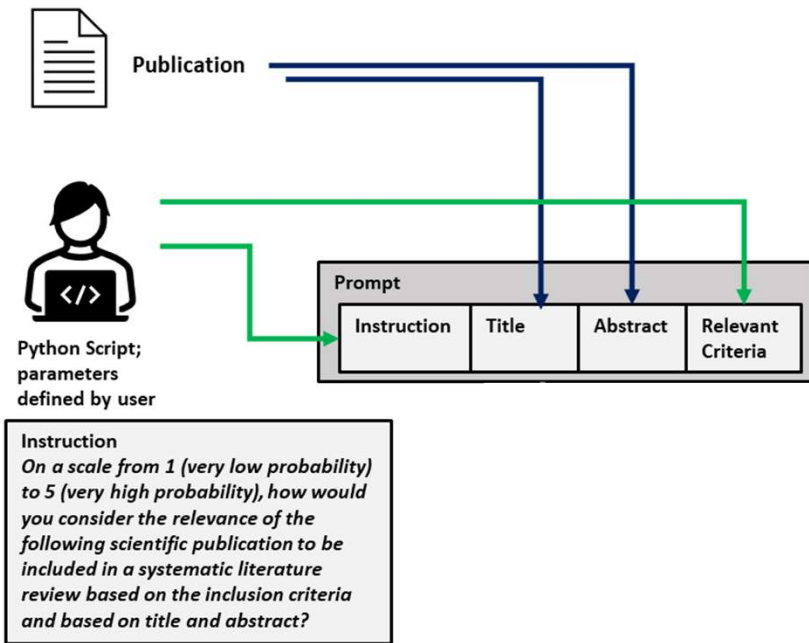
-  Extraction of CDE values using LLMs
-  Synoptic Report

LLM-based citation screening

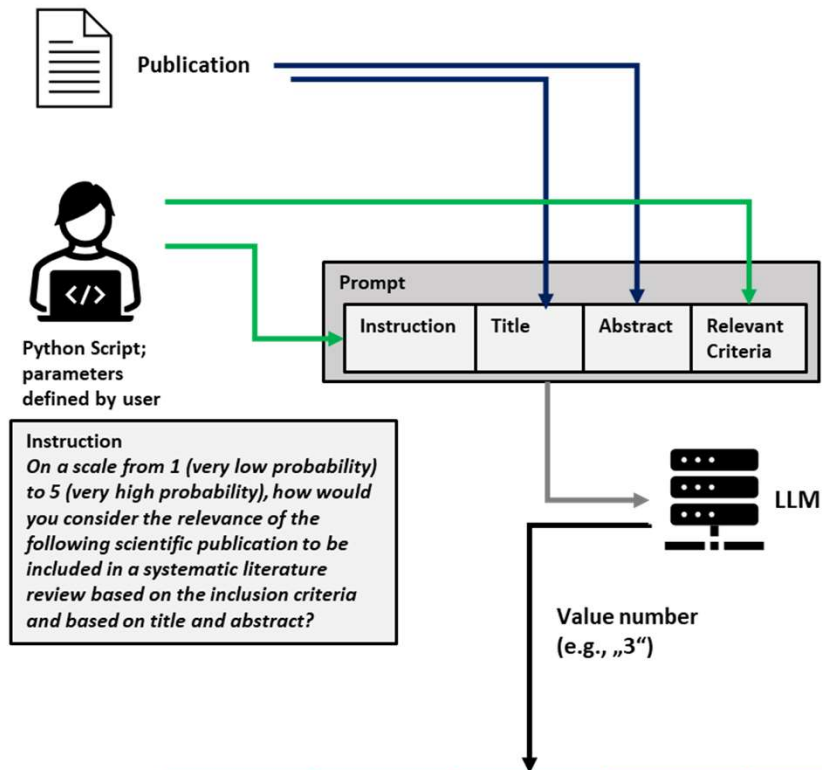
LLM-based citation screening



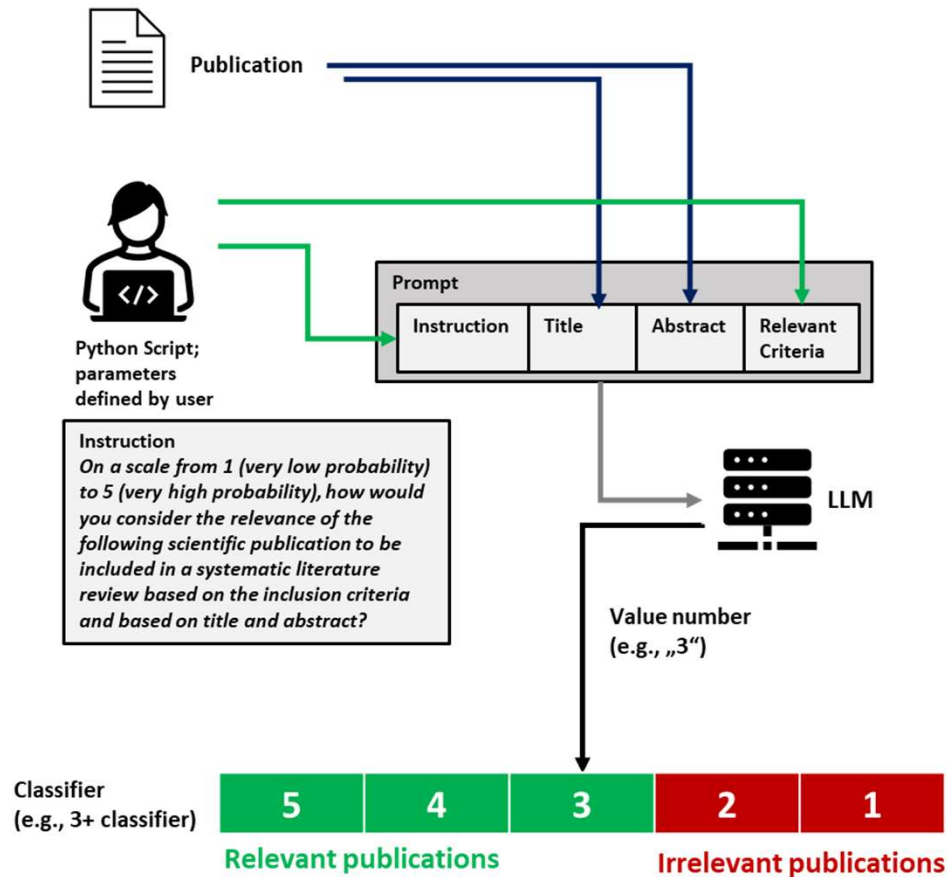
LLM-based citation screening



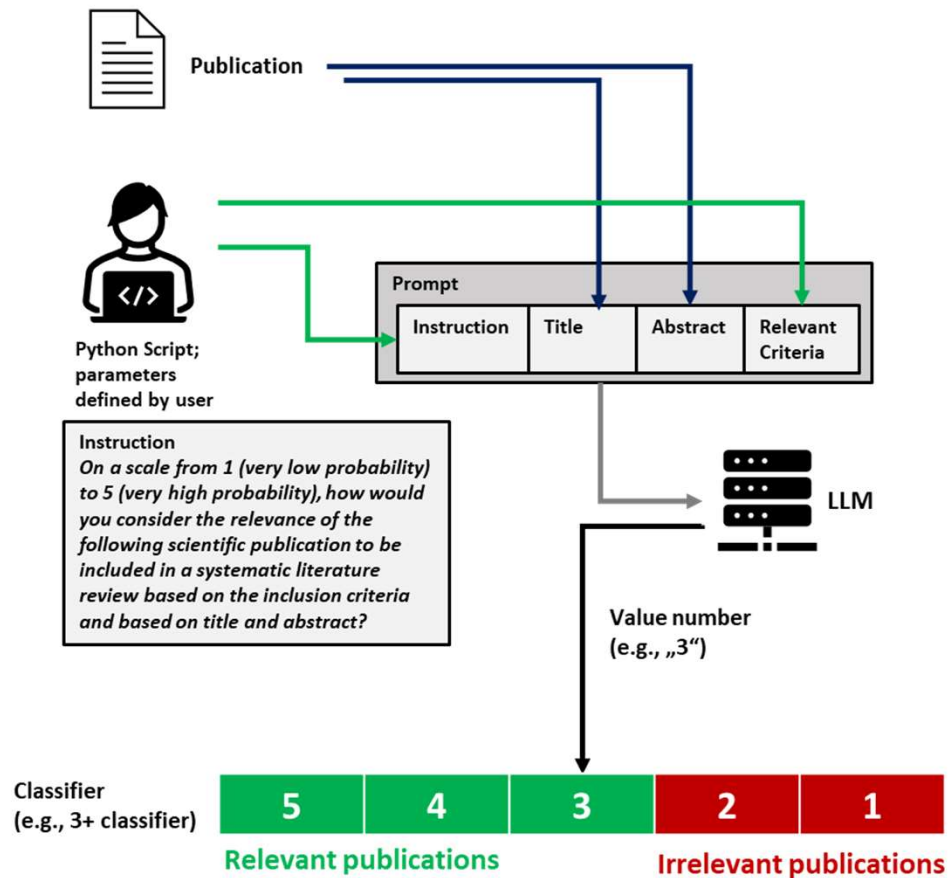
LLM-based citation screening



LLM-based citation screening



LLM-based citation screening



Sensitivity (=Recall) / Specificity in citation screening on 10 different biomedical topics

- 94.48%/31.78% for FlanT5 model
- 97.58%/19.12% for OpenHermes-NeuralChat model
- 81.93%/75.19% for Mixtral model
- 97.58%/38.34% for the Platypus 2 model

Performance on a new topic: „CDSS in radiation oncology“

Performance on a new topic: „CDSS in radiation oncology“



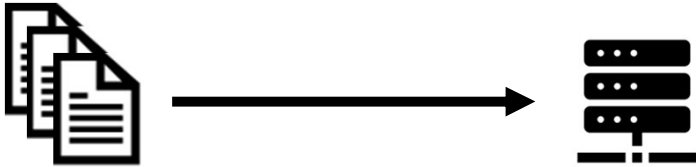
521 publications

36 relevant

485 irrelevant

(as determined by
humans)

Performance on a new topic: „CDSS in radiation oncology“



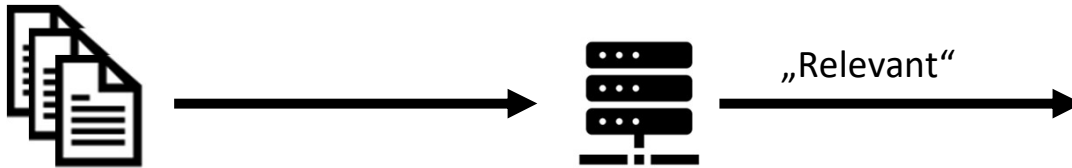
521 publications

36 relevant

485 irrelevant

(as determined by
humans)

Performance on a new topic: „CDSS in radiation oncology“



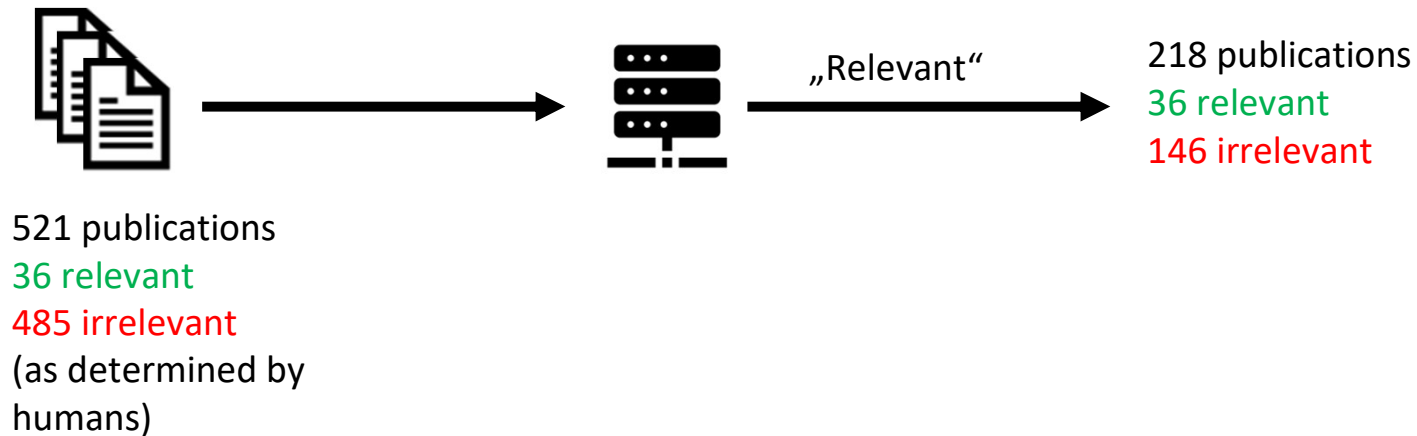
521 publications

36 relevant

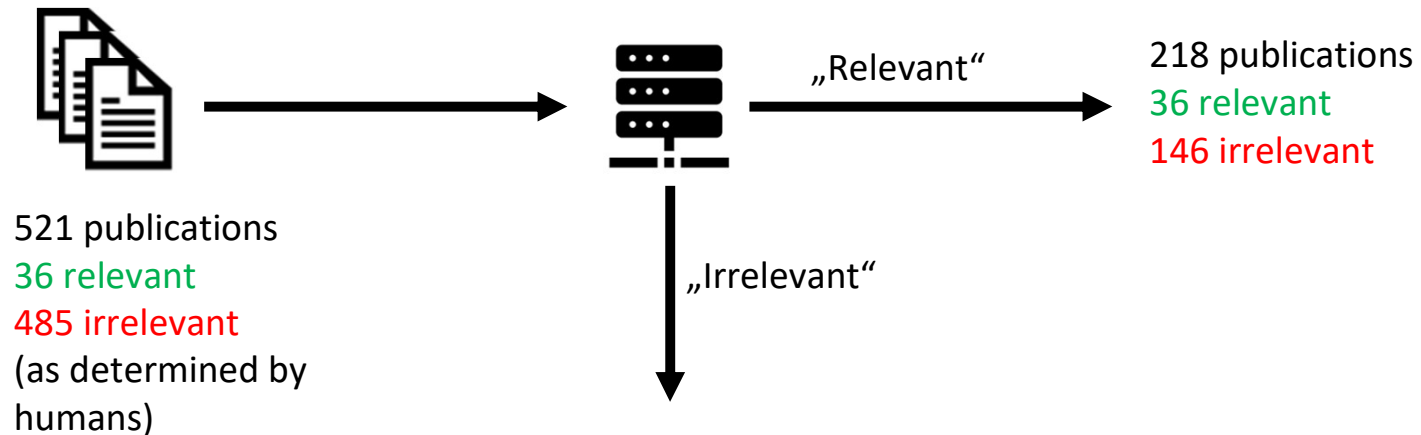
485 irrelevant

(as determined by
humans)

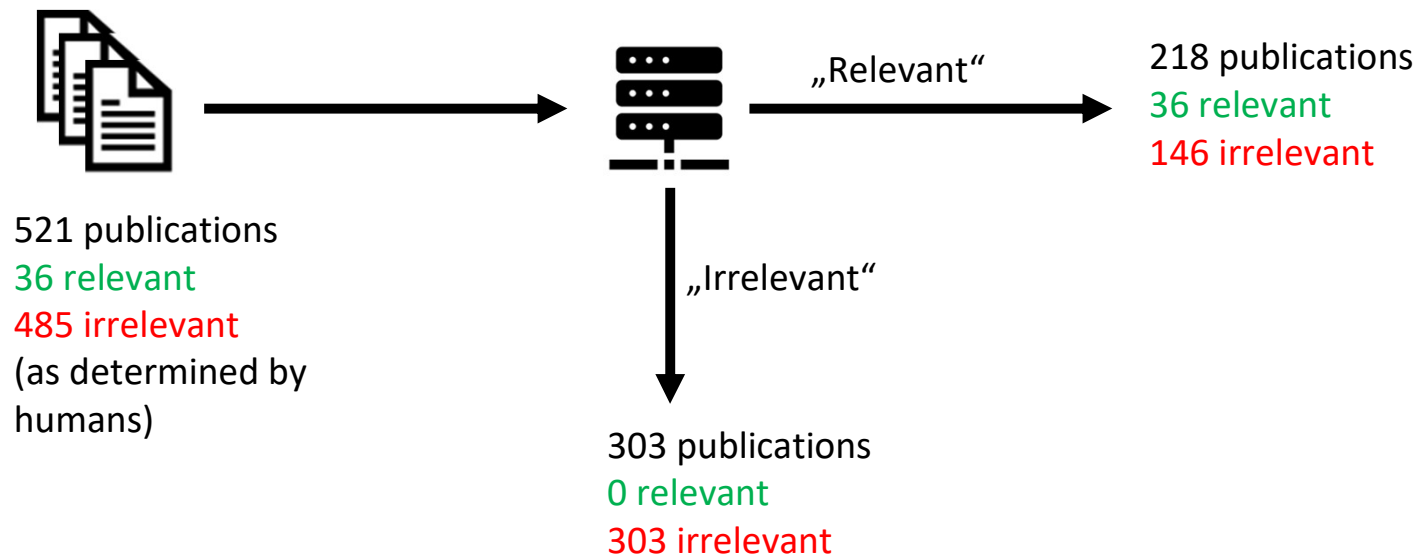
Performance on a new topic: „CDSS in radiation oncology“



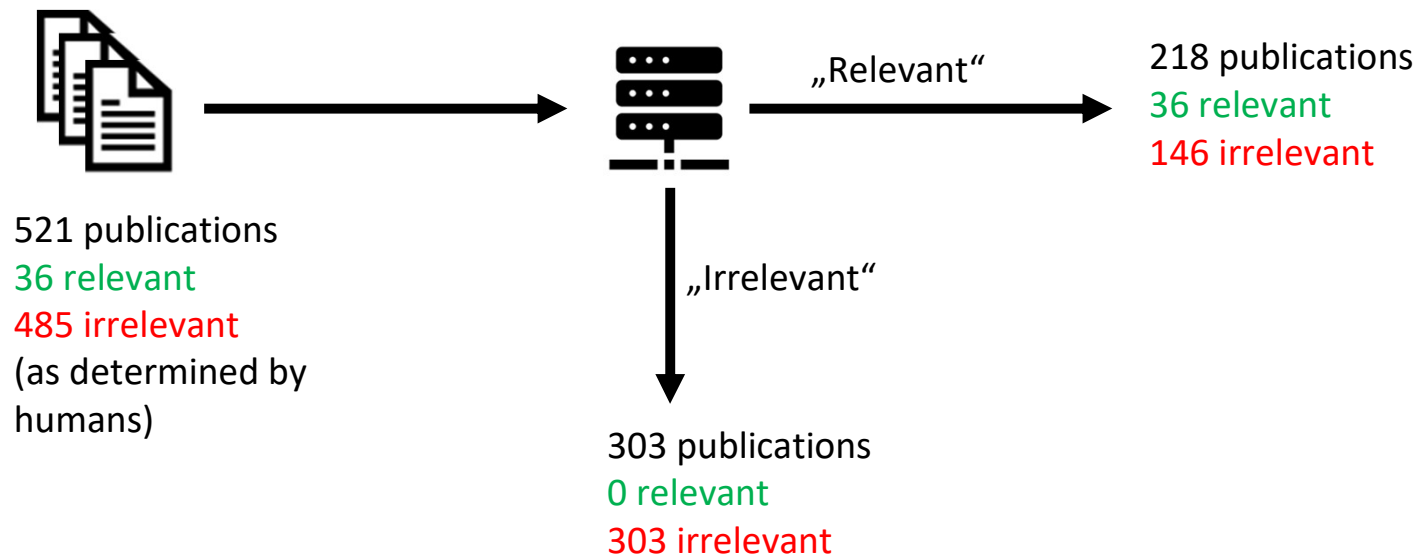
Performance on a new topic: „CDSS in radiation oncology“



Performance on a new topic: „CDSS in radiation oncology“

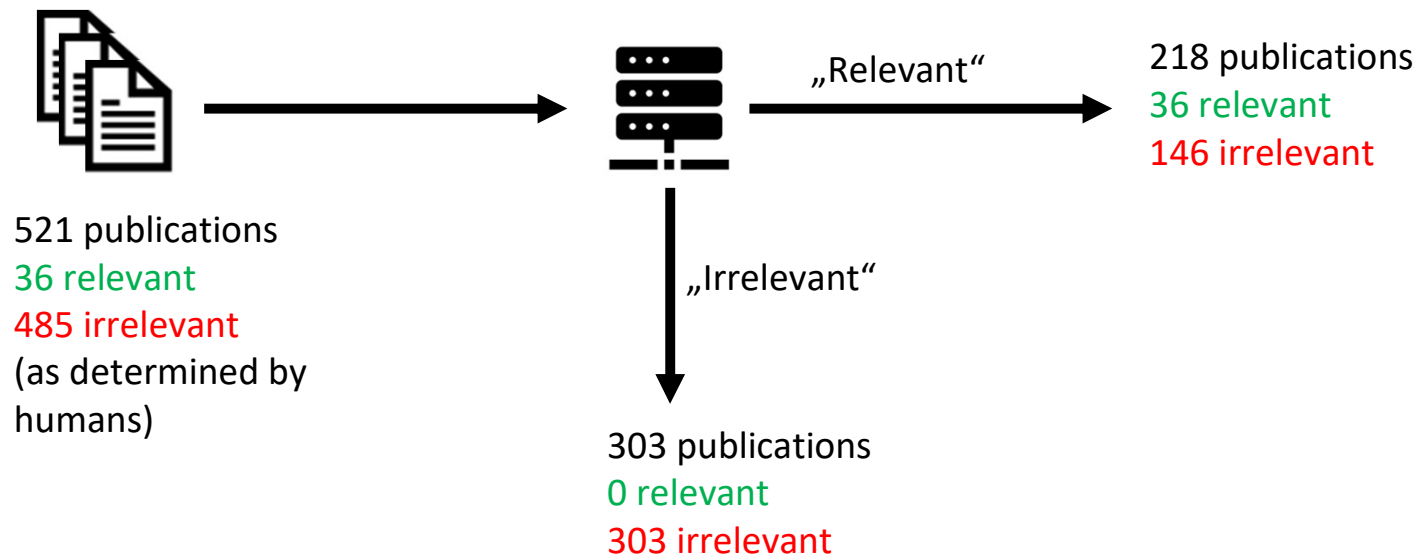


Performance on a new topic: „CDSS in radiation oncology“



→ Exclusion of 58 % irrelevant publications without missing a relevant one

Performance on a new topic: „CDSS in radiation oncology“



→ Exclusion of 58 % irrelevant publications without missing a relevant one

Manuscript under review

What are other people doing?

Med-PaLM

Med-PaLM

- LLM developed by Google, aligned to the medical domain

Med-PaLM

- LLM developed by Google, aligned to the medical domain
- First LLM to surpass the mark on USMLE style questions, published in December 2022

Med-PaLM

- LLM developed by Google, aligned to the medical domain
- First LLM to surpass the mark on USMLE style questions, published in December 2022
- 60% pass mark, Med-PaLM achieved 67%

Med-PaLM

- LLM developed by Google, aligned to the medical domain
- First LLM to surpass the mark on USMLE style questions, published in December 2022
- 60% pass mark, Med-PaLM achieved 67%



Lay assessment of helpfulness of answers

Med-PaLM

- LLM developed by Google, aligned to the medical domain
- First LLM to surpass the mark on USMLE style questions, published in December 2022
- 60% pass mark, Med-PaLM achieved 67%



Lay assessment of helpfulness of answers

Med-PaLM2

- successor of Med-PaLM, results published in March 2023

Med-PaLM2

- successor of Med-PaLM, results published in March 2023
- achieved 85% on USMLE questions (+18% in 4 months)

Med-PaLM2

- successor of Med-PaLM, results published in March 2023
- achieved 85% on USMLE questions (+18% in 4 months)
- reported to answer questions at an „expert doctor level“

The Open Medical-LLM Leaderboard: Benchmarking Large Language Models in Healthcare

Published April 19, 2024



Hugging Face

The Open Medical-LLM Leaderboard: Benchmarking Large Language Models in Healthcare



Published April 19, 2024

T ▲	Model ▲	Average 1 ▲	MedMCQA ▲	MedQA ▲	MMLU Anatomy ▲	MMLU Clinical Knowledge ▲	MMLU College Biology
◆	bongbongs/NewMes-v6	87.13	70.91	76.36	94.81	94.72	96.53
◆	aaditya/OpenBioLLMlama-70B	86.06	74.01	78.16	83.9	92.93	93.83
◆	Jayant9928/orpo_med_v7	84.42	58.95	69.36	86.67	95.47	94.44
◆	johnsnowlabs/JSL-MedLlama-3-8B-v4.0	84.12	58.74	61.74	89.63	94.72	96.53
●	Med-PaLM 2 (5 Shots)	84.09	71.3	79.7	77.8	88.3	94.4
◆	Jayant9928/orpo_med_v6	83.46	59.02	62.92	88.15	94.34	93.75
◆	ProbeMedicalYonseiMAILab/medllama3-v16.2	83.1	73.06	79.18	80	84.91	91.67
◆	bongbongs/NewMes-v5.2	83.1	73.06	79.18	80	84.91	91.67
●	GPT-4	82.97	69.5	78.8	80	86.4	95.1
◆	Jayant9928/orpo_med_v5	82.83	59.77	63.32	88.89	93.58	91.67
◆	ProbeMedicalYonseiMAILab/medllama3-v16.1	82.12	70.81	78.32	79.26	86.79	90.28

The Open Medical-LLM Leaderboard: Benchmarking Large Language Models in Healthcare



Published April 19, 2024

T ▲	Model ▲	Average 1 ▲	MedMCQA ▲	MedQA ▲	MMLU Anatomy ▲	MMLU Clinical Knowledge ▲	MMLU College Biology
◆	bongbongs/NewMes-v6	87.13	70.91	76.36	94.81	94.72	96.53
◆	aaditya/OpenBioLLMlama-70B	86.06	74.01	78.16	83.9	92.93	93.83
◆	Jayant9928/orpo_med_v7	84.42	58.95	69.36	86.67	95.47	94.44
◆	johnsnowlabs/JSL-MedLlama-3-8B-v4.0	84.12	58.74	61.74	89.63	94.72	96.53
●	Med-PaLM 2 (5 Shots)	84.09	71.3	79.7	77.8	88.3	94.4
◆	Jayant9928/orpo_med_v6	83.46	59.02	62.92	88.15	94.34	93.75
◆	ProbeMedicalYonseiMAILab/medllama3-v16.2	83.1	73.06	79.18	80	84.91	91.67
◆	bongbongs/NewMes-v5.2	83.1	73.06	79.18	80	84.91	91.67
●	GPT-4	82.97	69.5	78.8	80	86.4	95.1
◆	Jayant9928/orpo_med_v5	82.83	59.77	63.32	88.89	93.58	91.67
◆	ProbeMedicalYonseiMAILab/medllama3-v16.1	82.12	70.81	78.32	79.26	86.79	90.28

The Open Medical-LLM Leaderboard: Benchmarking Large Language Models in Healthcare



Published April 19, 2024

Open Source

T ▲	Model ▲	Average 1 ▲	MedMCQA ▲	MedQA ▲	MMLU Anatomy ▲	MMLU Clinical Knowledge ▲	MMLU College Biology
◆	bongbongs/NewMes-v6	87.13	70.91	76.36	94.81	94.72	96.53
◆	aaditya/OpenBioLLMlama-70B	86.06	74.01	78.16	83.9	92.93	93.83
◆	Jayant9928/orpo_med_v7	84.42	58.95	69.36	86.67	95.47	94.44
◆	johnsnowlabs/JSL-MedLlama-3-8B-v4.0	84.12	58.74	61.74	89.63	94.72	96.53
●	Med-PaLM 2 (5 Shots)	84.09	71.3	79.7	77.8	88.3	94.4
◆	Jayant9928/orpo_med_v6	83.46	59.02	62.92	88.15	94.34	93.75
◆	ProbeMedicalYonseiMAILab/medllama3-v16.2	83.1	73.06	79.18	80	84.91	91.67
◆	bongbongs/NewMes-v5.2	83.1	73.06	79.18	80	84.91	91.67
●	GPT-4	82.97	69.5	78.8	80	86.4	95.1
◆	Jayant9928/orpo_med_v5	82.83	59.77	63.32	88.89	93.58	91.67
◆	ProbeMedicalYonseiMAILab/medllama3-v16.1	82.12	70.81	78.32	79.26	86.79	90.28

What about media other than text?

LLaVA-Med

Visual input example, Biomedical image:



LLaVA-Med

Visual input example, Biomedical image:



User

What is this image about?

LLaVA-Med

Visual input example, Biomedical image:



User

What is this image about?

LLaVA-Med

The image is a coronal view of a computed tomography (CT) scan of the abdomen and pelvis. It provides a detailed view of the internal organs and structures within the abdominal and pelvic cavities.

What about predictions regarding LLMs?

What the future of 2022 was imagined regarding LLMs in 2021



On June 30, 2022, what will be the state-of-the-art few-shot or transfer accuracy on the Massive Multitask Language Understanding dataset?

Forecast: 57.1 %

Crowd Forecast

Forecast History

More Info

News

71 forecasters
Start: 2021-07-24
End: 2021-09-29

From MMLU_college_medicine:

Acute Myeloid Leukemia is a malignant cancer in which myeloid progenitor cells become rapidly dividing and retain their semi-stem cell state. Under normal circumstances, which of the following cell types would arise from this stem cell?

From MMLU_college_medicine:

Acute Myeloid Leukemia is a malignant cancer in which myeloid progenitor cells become rapidly dividing and retain their semi-stem cell state. Under normal circumstances, which of the following cell types would arise from this stem cell?

A: NK cell

B: Platelet

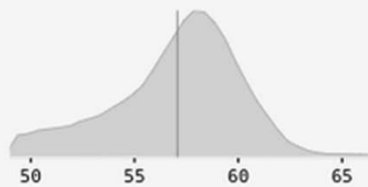
C: T cell

D: Lymphoid progenitor

What the future of 2022 was imagined regarding LLMs



On June 30, 2022, what will be the state-of-the-art few-shot or transfer accuracy on the Massive Multitask Language Understanding dataset?



Forecast: 57.1 %

Crowd Forecast

Forecast History

More Info

News

71 forecasters
Start: 2021-07-24
End: 2021-09-29

What the future of 2022 was imagined regarding LLMs



On June 30, 2022, what will be the state-of-the-art few-shot or transfer accuracy on the Massive Multitask Language Understanding dataset?

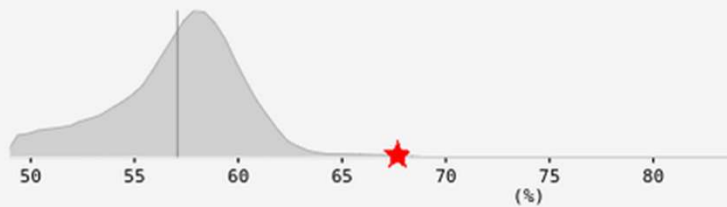
Crowd Forecast

Forecast History

More Info

News

Forecast: 57.1 %



71 forecasters
Start: 2021-07-24
End: 2021-09-29

What the future of 2022 was imagined regarding LLMs



On June 30, 2022, what will be the state-of-the-art few-shot or transfer accuracy on the Massive Multitask Language Understanding dataset?

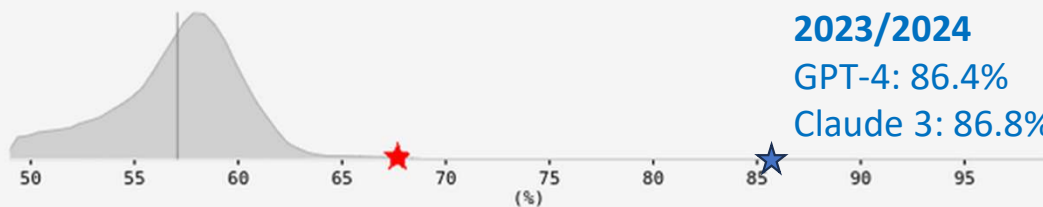
Crowd Forecast

Forecast History

More Info

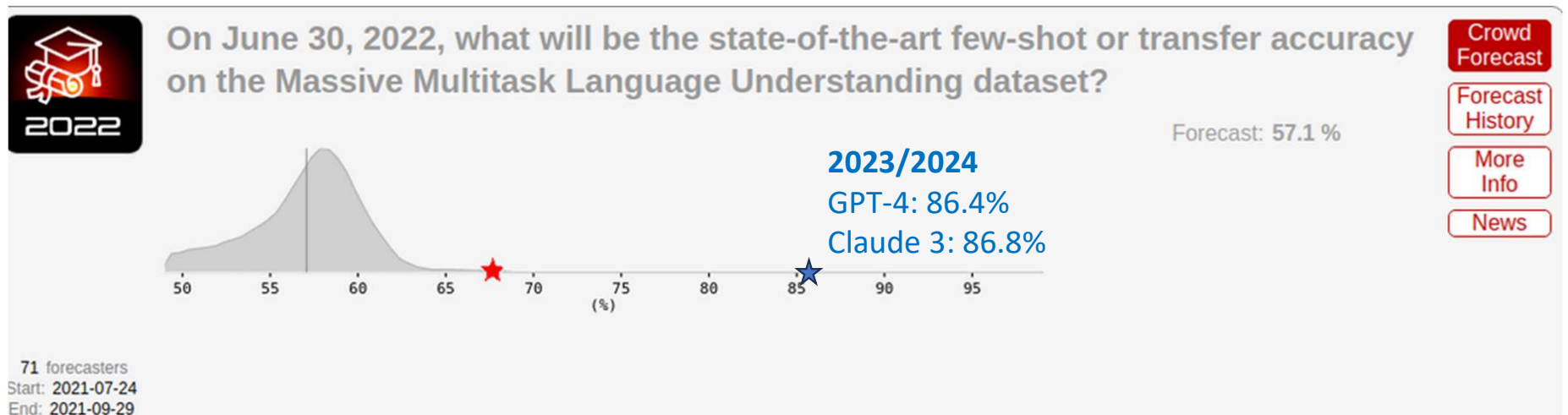
News

Forecast: 57.1 %



71 forecasters
Start: 2021-07-24
End: 2021-09-29

What the future of 2022 was imagined regarding LLMs



→ Underprediction

Conclusion

Conclusion

- LLMs are very powerful for different tasks regarding data- and knowledge management in radiation therapy, as well as general medicine

Conclusion

- LLMs are very powerful for different tasks regarding data- and knowledge management in radiation therapy, as well as general medicine
- Many groups are working on many different use cases

Conclusion

- LLMs are very powerful for different tasks regarding data- and knowledge management in radiation therapy, as well as general medicine
- Many groups are working on many different use cases
- **The progress in the field is very rapid**

Thank you for your attention.