

Sander Puts

CTO support for tumour registration

Maastrro Science Day 2024

 **Maastrro**



**Maastricht
University**







Tumour registries



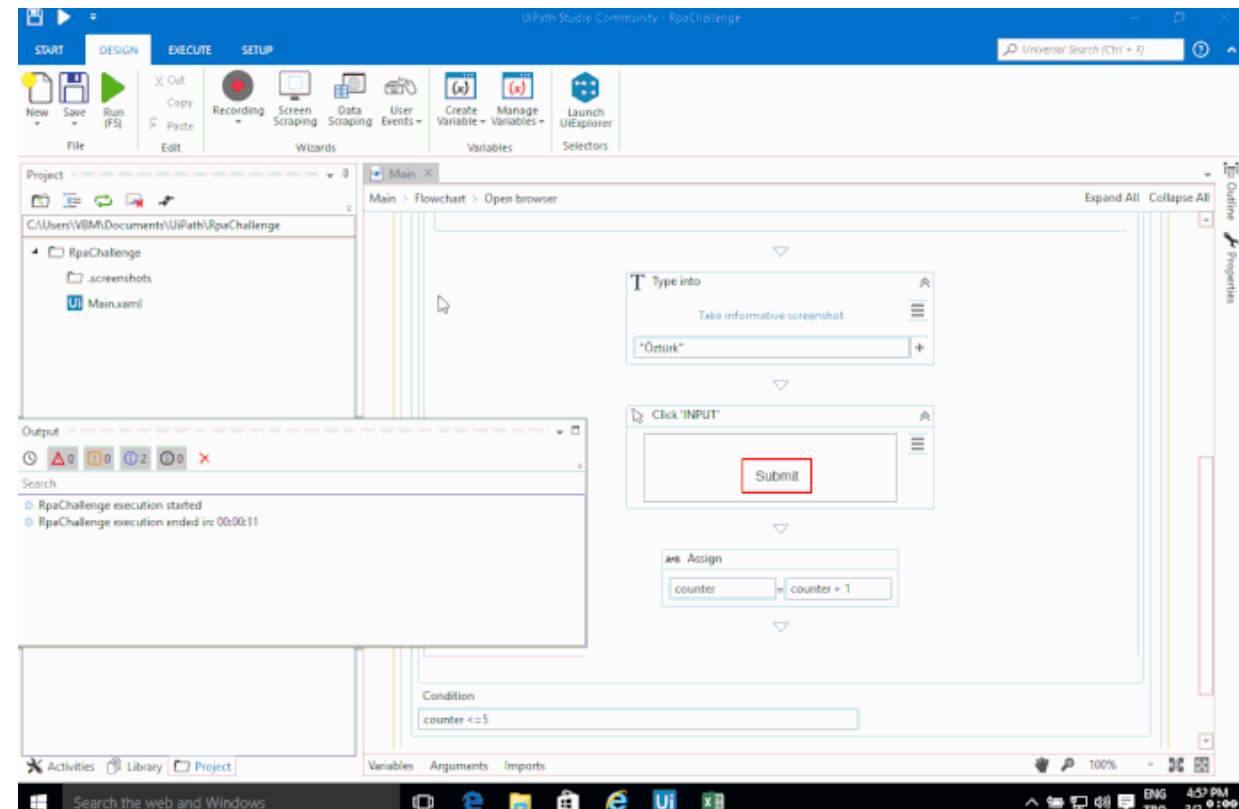
Epic



Mix of source systems

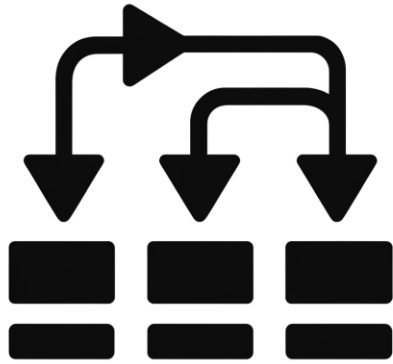
Possible solution

RPA: Robotic Process Automation

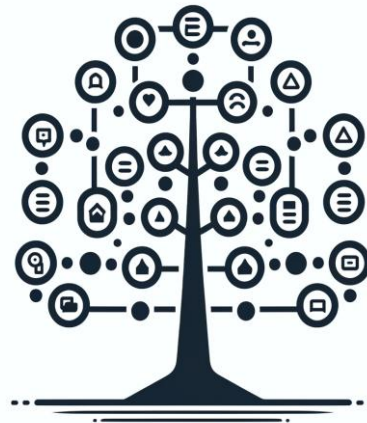




CTO Challenges



Data transformations



Classifications (CTCAE)



Implicit knowledge

Tumour follow up → New malignancy: yes



Type



Code ICD-O-3



Date



Language Models

Trained to predict the next word in a sentence:

The cat is chasing the _____



dog 5%

mouse 70%

squirrel 20%

boy 5%

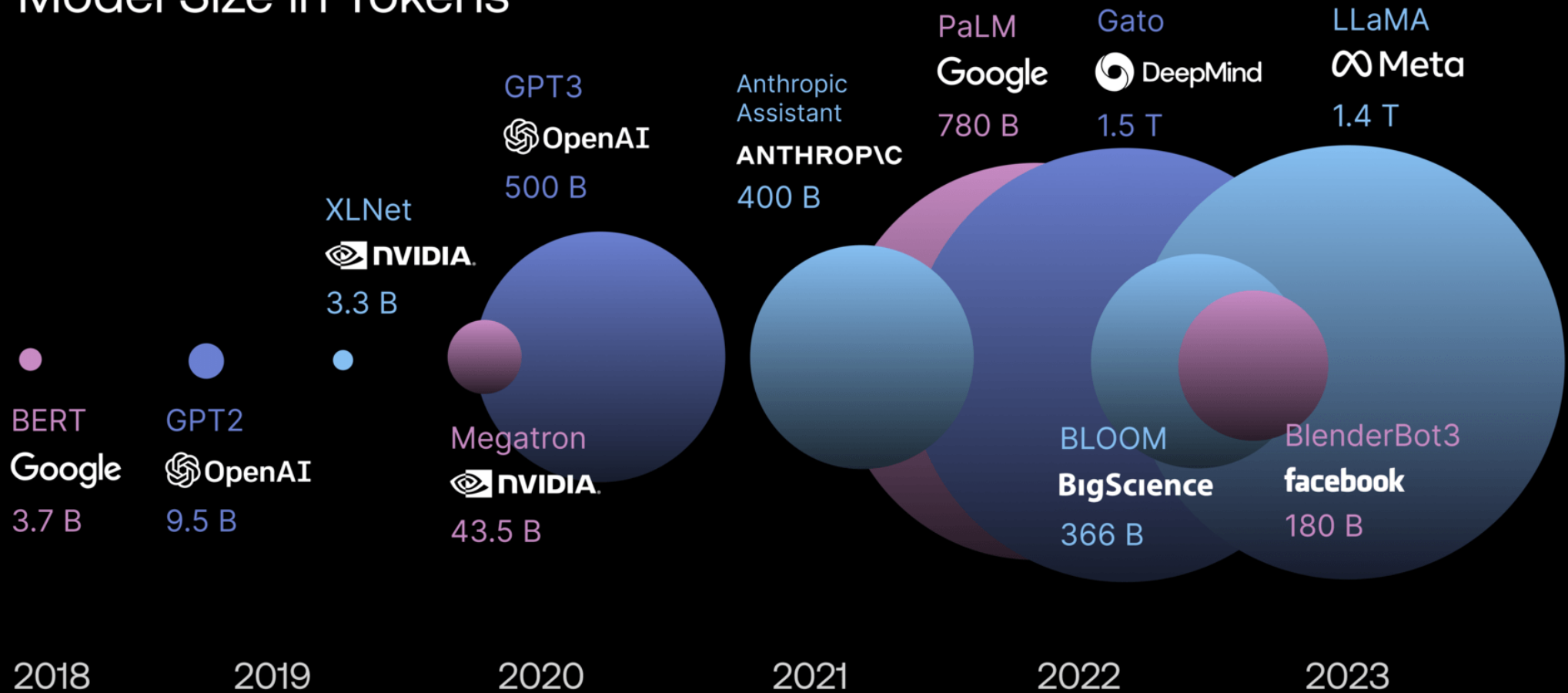
house 0%



ME WANTS MORE DATA

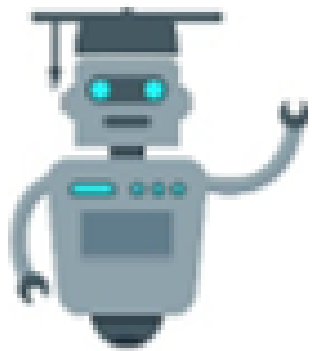


Model Size in Tokens





Who was the first person to land on Moon?



Answer: **Yuri Gagarin** was the first person to land on the Moon. **X**



Correct Answer: **Neil Armstrong** was the first person to land on the Moon. **✓**

LLM Challenges



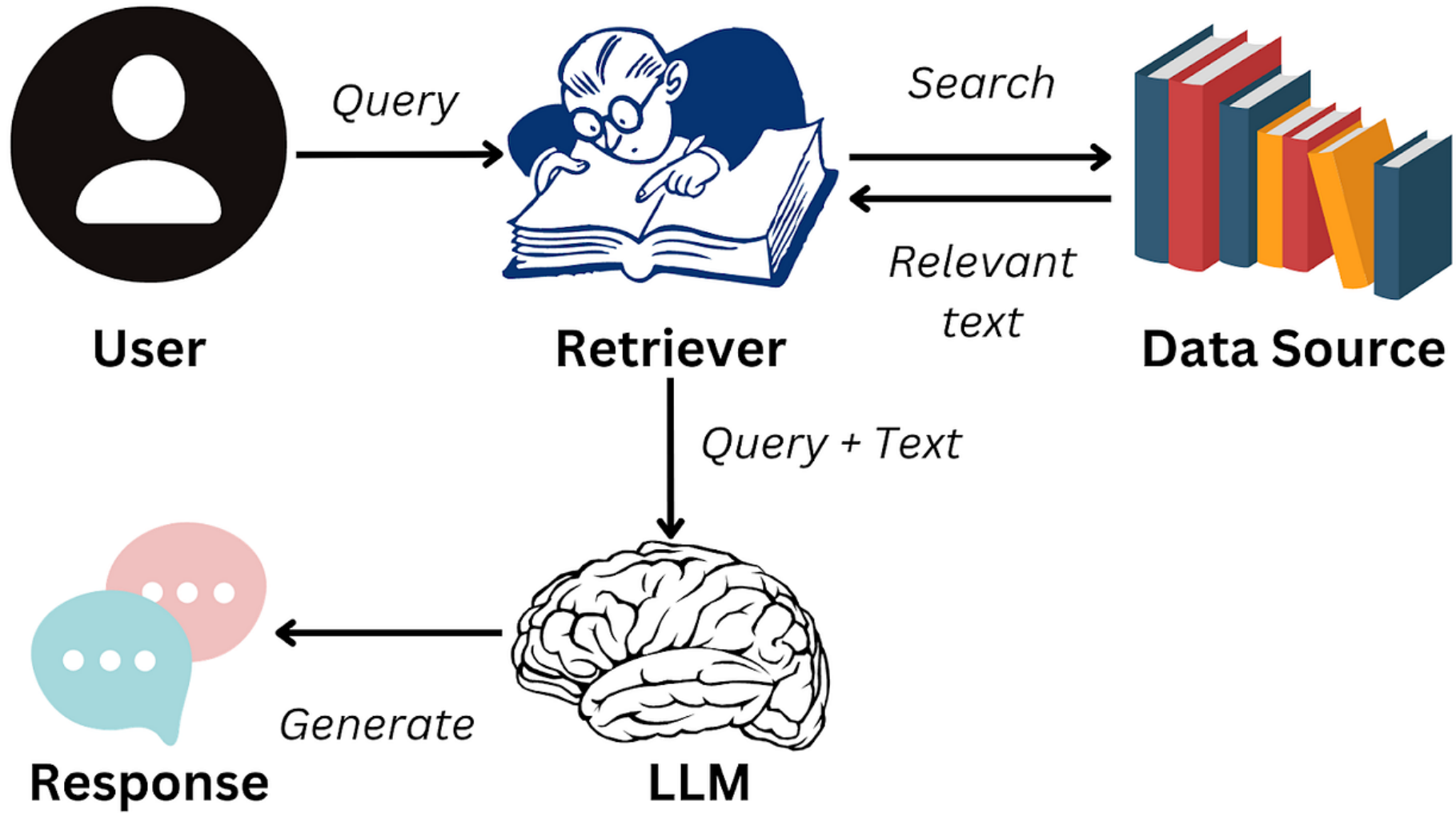
Privacy



Hardware



Trust & reliability



Retrieval Augmented Generation (RAG)

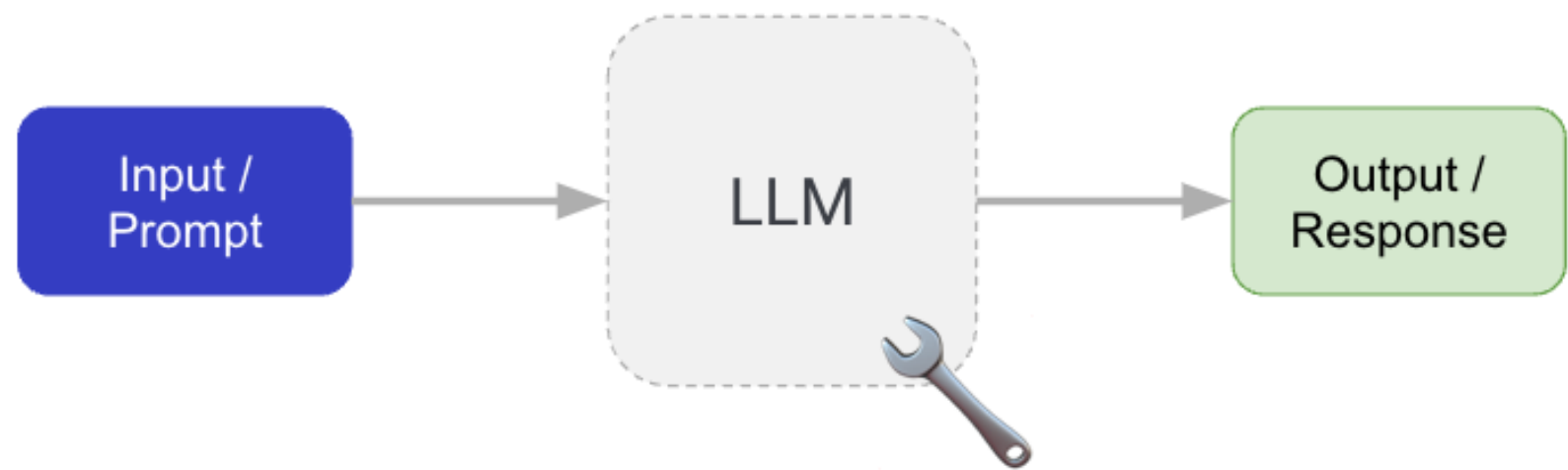
	Meta Llama 3 8B	Gemma 7B - It Measured	Mistral 7B Instruct Measured
MMLU 5-shot	68.4	53.3	58.4
GPQA 0-shot	34.2	21.4	26.3
HumanEval 0-shot	62.2	30.5	36.6
GSM-8K 8-shot, CoT	79.6	30.6	39.9
MATH 4-shot, CoT	30.0	12.2	11.0

	Meta Llama 3 70B	Gemini Pro 1.5 Published	Claude 3 Sonnet Published
MMLU 5-shot	82.0	81.9	79.0
GPQA 0-shot	39.5	41.5 CoT	38.5 CoT
HumanEval 0-shot	81.7	71.9	73.0
GSM-8K 8-shot, CoT	93.0	91.7 11-shot	92.3 0-shot
MATH 4-shot, CoT	50.4	58.5 Minerva prompt	40.5

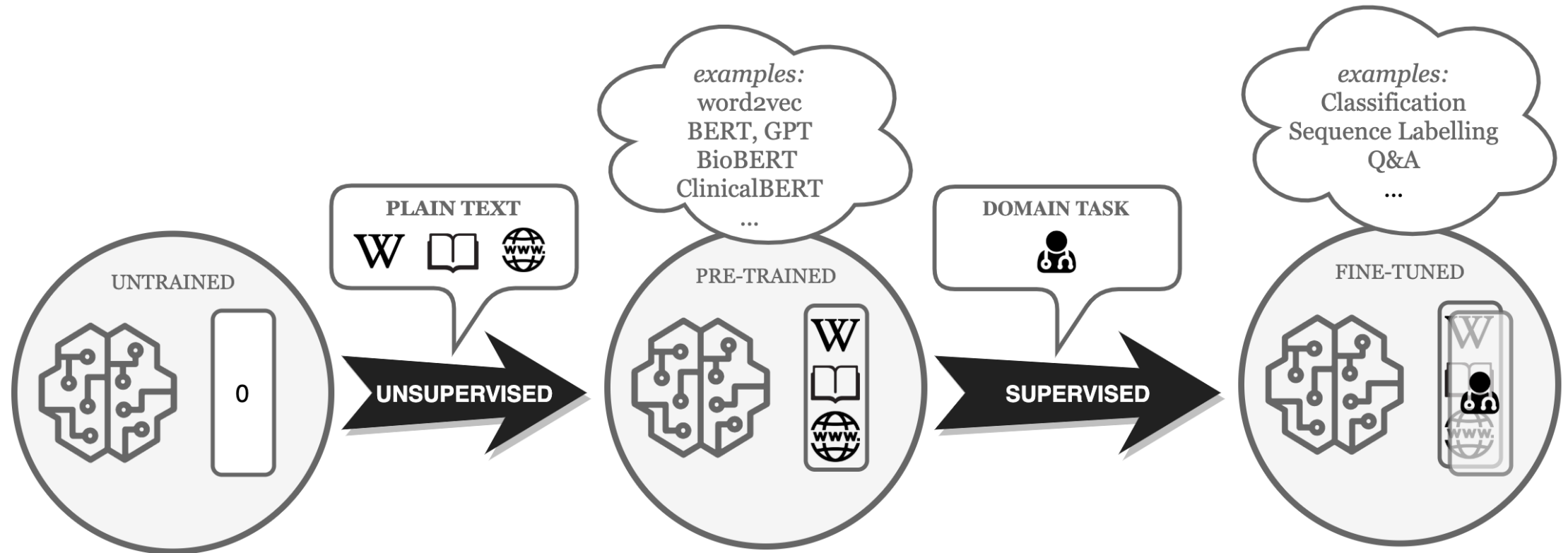
Prompt Engineering



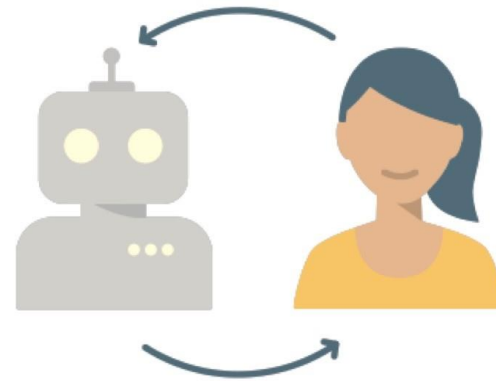
Fine-tuning



Fine-tuning



Several challenges: Keep humans-in-the-loop



MEDSTRUCT-NLP
Assistance, Classification and Information Extraction for Medical Free-Text Reporting

View Annotated: Autocheck:

Annotated Report

Thoraxgrote massa zichtbaar in de linker bovenkwab met maximale diameter op 8.46 van 4.7 x 3.0 cm. **Mogelijke** ingroei in het mediastrium. Satelliet laesies zichtbaar op 8.41 met een grootte van 1.3 cm. Lymfeklier zichtbaar op station 7 met een lente van circa 5.2 cm. **Geen** lymfeklieren aan de contralaterale zijde zichtbaar. **Geen** consolidatie in de middenkwab. **Geen** atelectase. Conclusie tumor met satelliet laesies in de linker bovenkwab.

- Present
- Involved
- Context Modifier
- Context Target

Example text from Article-1 (NL)

TNM-8 Lung

Nederlands

T3 N2 M

Primary Tumor

4.7 cm Left Side

Present Involved

Satellite Nodule (T3) Ipsilateral Tumor (T4)

Lymph Nodes

7 Subcarinal x