



UNIVERSITY OF
LIMERICK
OLLSCOIL LUIMNIGH

Limerick Digital
Cancer Research
Centre



Assigning treatment regimens to Irish patients in head and neck squamous cell carcinoma with large language models

Meghana Kshirsagar

Motivation

- **Personalized Treatment:** To understand the highly variable outcomes of treatments for a personalized treatment approach.
- **Patient-Specific Prognostication:** Analyzing large cohort data can lead to more accurate patient-specific prognostication.
- **Targeted Therapy:** The shift towards targeted therapies in cancer treatment, due to the significant side effects of conventional methods.

Barriers in adoption of AI driven clinical decision making

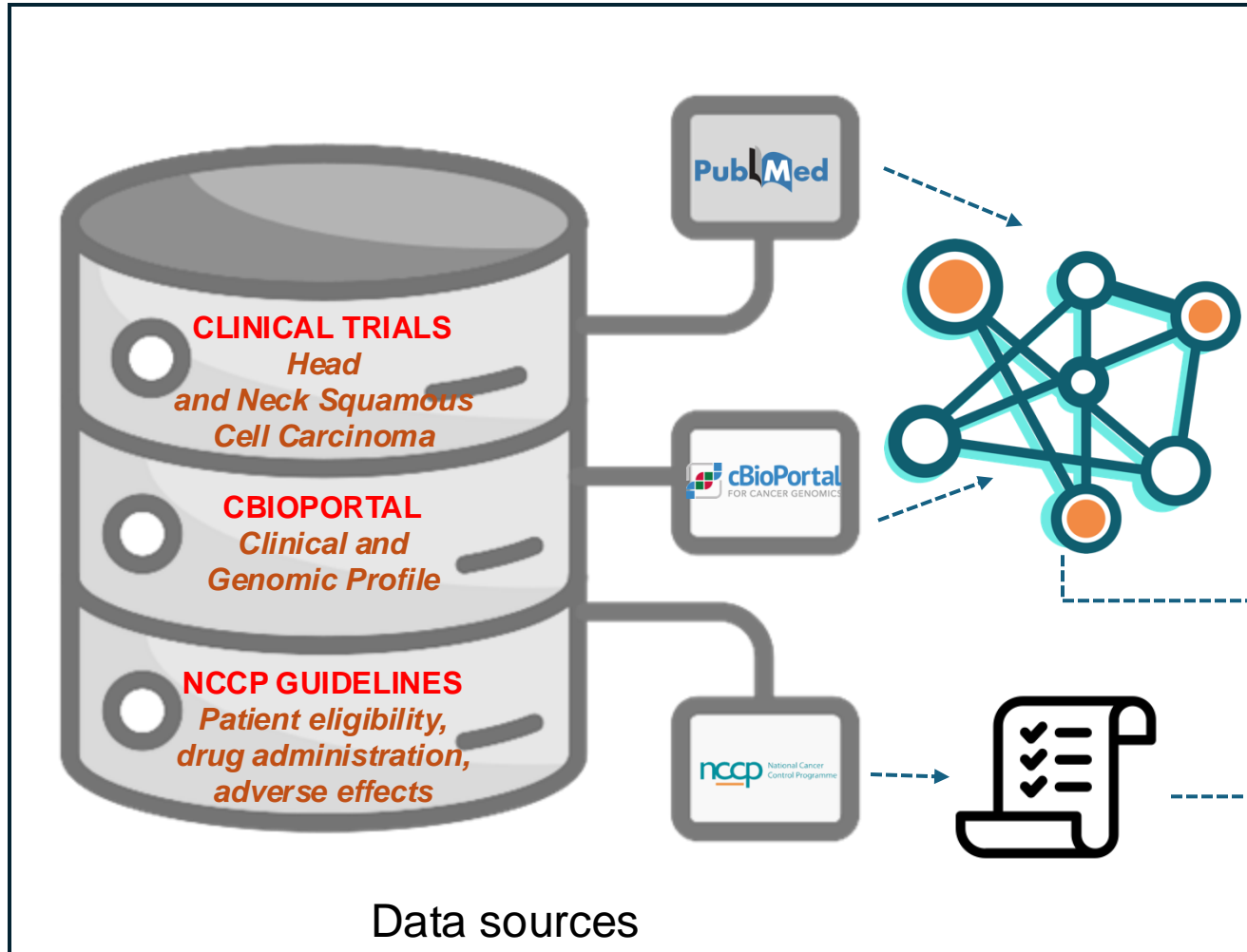


Towards a Trustworthy AI driven Clinical decision making

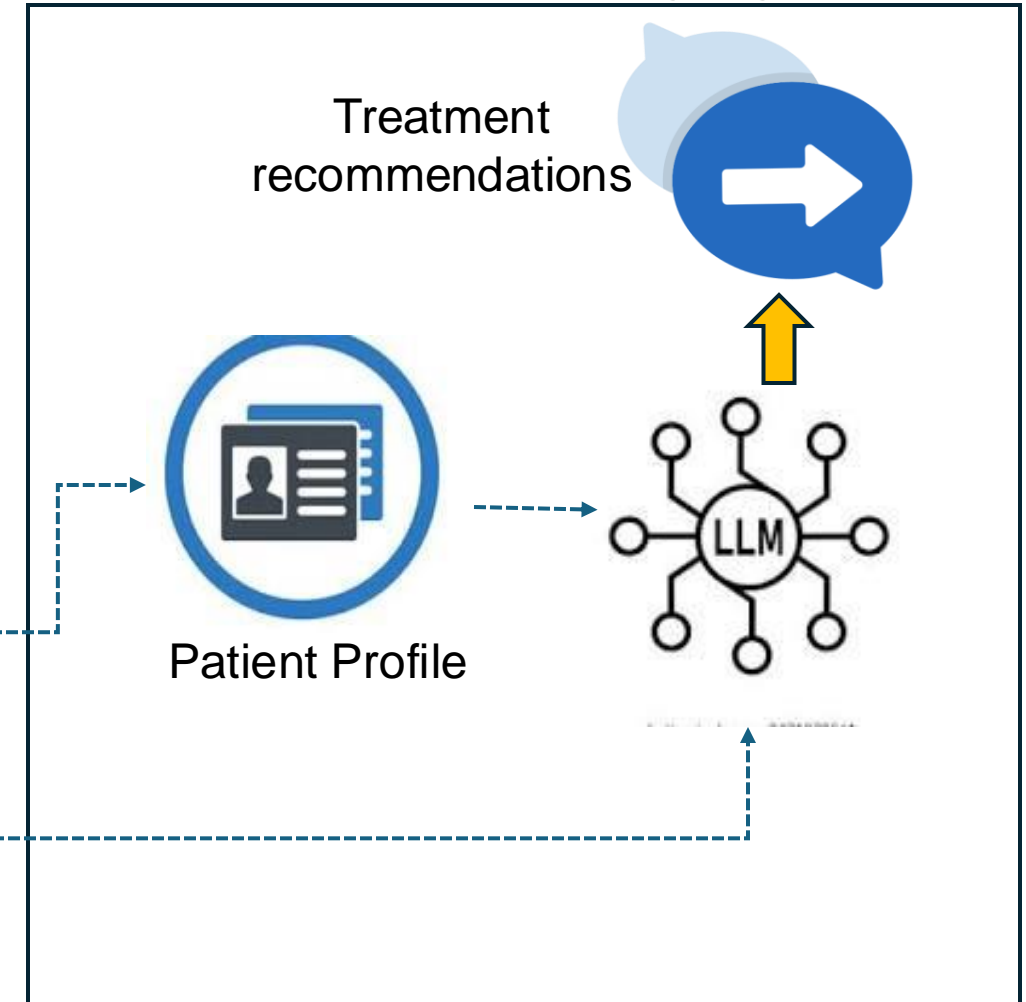


Methodology

Stage I: Build curated dataset



Stage II: Personalized Treatment Recommendations (ongoing)



Retrieving PubMed data

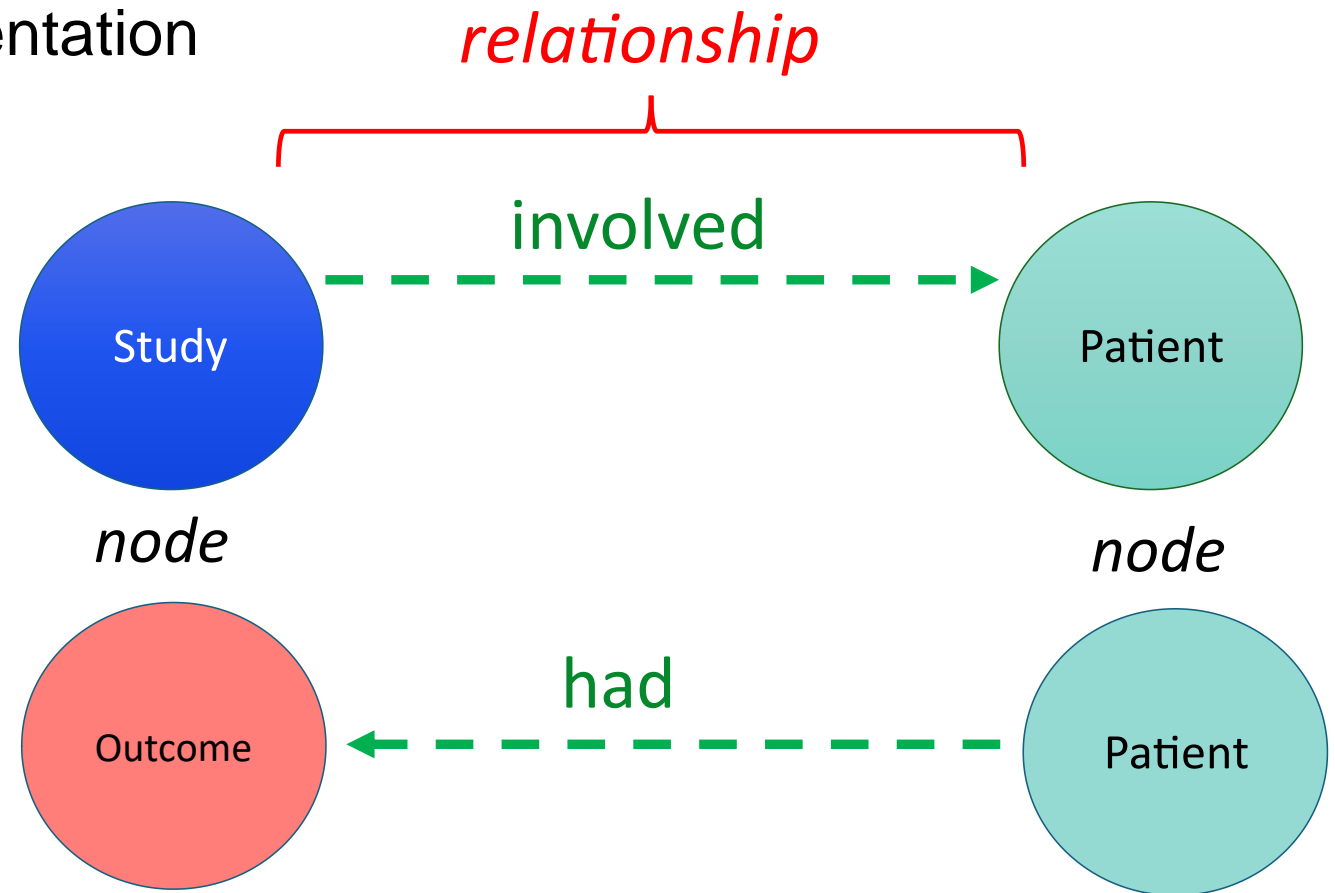
```
1 import os
2 import scrapy
3 import nltk
4 from scrapy.http import Request
5 nltk.download('punkt')
6
7 class PubMedSpider(scrapy.Spider):
8     name = "pubmed_spider"
9
10    custom_settings = {
11        'FEED_FORMAT': 'json',
12        'FEED_URI': '../scraped_data/overview.json',
13        'LOG_LEVEL': 'DEBUG',
14        'RETRY_ENABLED': True,
15        'RETRY_TIMES': 3,
16        'RETRY_HTTP_CODES': [500, 502, 503, 504, 429],
17        'DOWNLOAD_DELAY': 2,
18        'ROBOTSTXT_OBEY': False,
19    }
```

```
1 [
2     {
3         "clinical_trial_id": "31679945",
4         "metadata": {
5             "url": "https://pubmed.ncbi.nlm.nih.gov/31679945/",
6             "title_of_study": "Pembrolizumab alone or with chemotherapy versus... (cut off for demo.)",
7             "stored_file_name_summary": "31679945_summary.txt",
8             "stored_file_name_full": null,
9             "figures": {
10                 "figures_folder": null,
11                 "figures_files": []
12             },
13             "type": "Clinical Trial",
14             "extension": ".txt",
15             "pmcid": null,
16             "doi": "10.1016/S0140-6736(19)32591-7",
17             "drugs_mentioned_in_title": [
18                 "Cetuximab",
19                 "Pembrolizumab"
20             ],
21             "tokens": {
22                 "title_tokens": 33,
23                 "summary_content_tokens": 692,
24                 "title_and_summary_total_tokens": 725,
25                 "title_and_full_article_total_tokens": null
26             }
27         },
28     },
29 ]
```

Automatic knowledge graph creation with prompt engineering

- Structured Knowledge Representation
- Contextual Understanding

```
Study 1: {  
  "id": "36219809",  
  "title": "Pembrolizumab With or Without Chemotherapy in Recurrent or Metastatic Head and Neck Squamous Cell Carcinoma: Updated Results of the Phase III KEYNOTE-048 Study",  
  "patients and methods": "Patients were randomly assigned (1:1) to pembrolizumab, pembrolizumab-chemotherapy, or cetuximab-chemotherapy. Efficacy was evaluated in programmed death ligand 1 (PD-L1) combined positive score (CPS)  $\geq 20$ , CPS  $\geq 1$ , and total populations, with no multiplicity or alpha adjustment.",  
  "findings": "The median study follow-up was 45.0 months (interquartile range, 41.0-49.2; n = 882). At data cutoff (February 18, 2020), overall survival improved with pembrolizumab in the PD-L1 CPS  $\geq 20$  (hazard ratio [HR], 0.61; 95% CI, 0.46 to 0.81) and CPS  $\geq 1$  populations (HR, 0.74; 95% CI, 0.61 to 0.89) and was noninferior in the total population (HR, 0.81; 95% CI, 0.68 to 0.97). Overall survival improved with pembrolizumab-chemotherapy in the PD-L1 CPS  $\geq 20$  (HR, 0.62; 95% CI, 0.46 to 0.84), CPS  $\geq 1$  (HR, 0.64; 95% CI, 0.53 to 0.78), and total (HR, 0.71; 95% CI, 0.59 to 0.85) populations. The objective response rate on second-course pembrolizumab was 27.3% (3 of 11). PFS2 improved with pembrolizumab in the PD-L1 CPS  $\geq 20$  (HR, 0.64; 95% CI, 0.48 to 0.84) and CPS  $\geq 1$  (HR, 0.79; 95% CI, 0.66 to 0.95) populations and with pembrolizumab-chemotherapy in the PD-L1 CPS  $\geq 20$  (HR, 0.64; 95% CI, 0.48 to 0.86), CPS  $\geq 1$  (HR, 0.66; 95% CI, 0.55 to 0.81), and total (HR, 0.73; 95% CI, 0.61 to 0.88) populations. PFS2 was similar after pembrolizumab and longer after pembrolizumab-chemotherapy on next-line taxanes and shorter after pembrolizumab and similar after pembrolizumab-chemotherapy on next-line nontaxanes.",  
  "conclusion": "With a 4-year follow-up, first-line pembrolizumab and pembrolizumab-chemotherapy continued to demonstrate survival benefit versus cetuximab-chemotherapy in recurrent/metastatic head and neck squamous cell carcinoma. Patients responded well to subsequent treatment after pembrolizumab-based therapy."  
},
```



Build Knowledge Graphs

Stage 1:
Supervised

Stage II :
Unsupervised

Stage 1: *Supervised*

```
system_prompt_for_nodes = """
Extract the following nodes from the provided medical study text. Extract in the given schema:

{"study": "extract The summary of the study title", "patients": "Who were the patients",
"treatment": "What was the treatment given to them",
"stage": "what stage of the cancer did the pateints have", "drug": "what drugs where given to the patients",
"adverse_effects": "What effects did the drug or the treatment had",
"age": "What was the patients age", "gender": "were the patients male, female or both",
"response": "How did the patients responded to the given treatment",
"outcomes": "summary of the study outcomes", "biomarkers": "What were the biomakers involved in the study",
"period": "study period", "survival": "survival rate of the patients", "deaths": "death rate of the patients"
}

Do not include any prefix or suffix to the answer. Leave an empty string if a node doesn't exist.

"""
```

Stage 1: *Supervised*

```
system_prompt_for_nodes = """
Extract the following nodes from the provided medical study text. Extract in the given schema:

{"study": "extract The summary of the study title", "patients": "Who were the patients",
"treatment": "What was the treatment given to them",
"stage": "what stage of the cancer did the pateints have", "drug": "what drugs where given to the patients",
"adverse_effects": "What effects did the drug or the treatment had",
"age": "What was the patients age", "gender": "were the patients male, female or both",
"response": "How did the patients responded to the given treatment",
"outcomes": "summary of the study outcomes", "biomarkers": "What were the biomakers involved in the study",
"period": "study period","survival": "survival rate of the patients", "deaths": "death rate of the patients"
}

Do not include any prefix or suffix to the answer. Leave an empty string if a node doesn't exist.

"""
```

Nodes:

Study
Patients
Treatment
Drug
Adverse effects
Age
Gender
Response
Outcomes
Biomarkers
Period
Survival
Deaths

Stage 1: *Supervised*

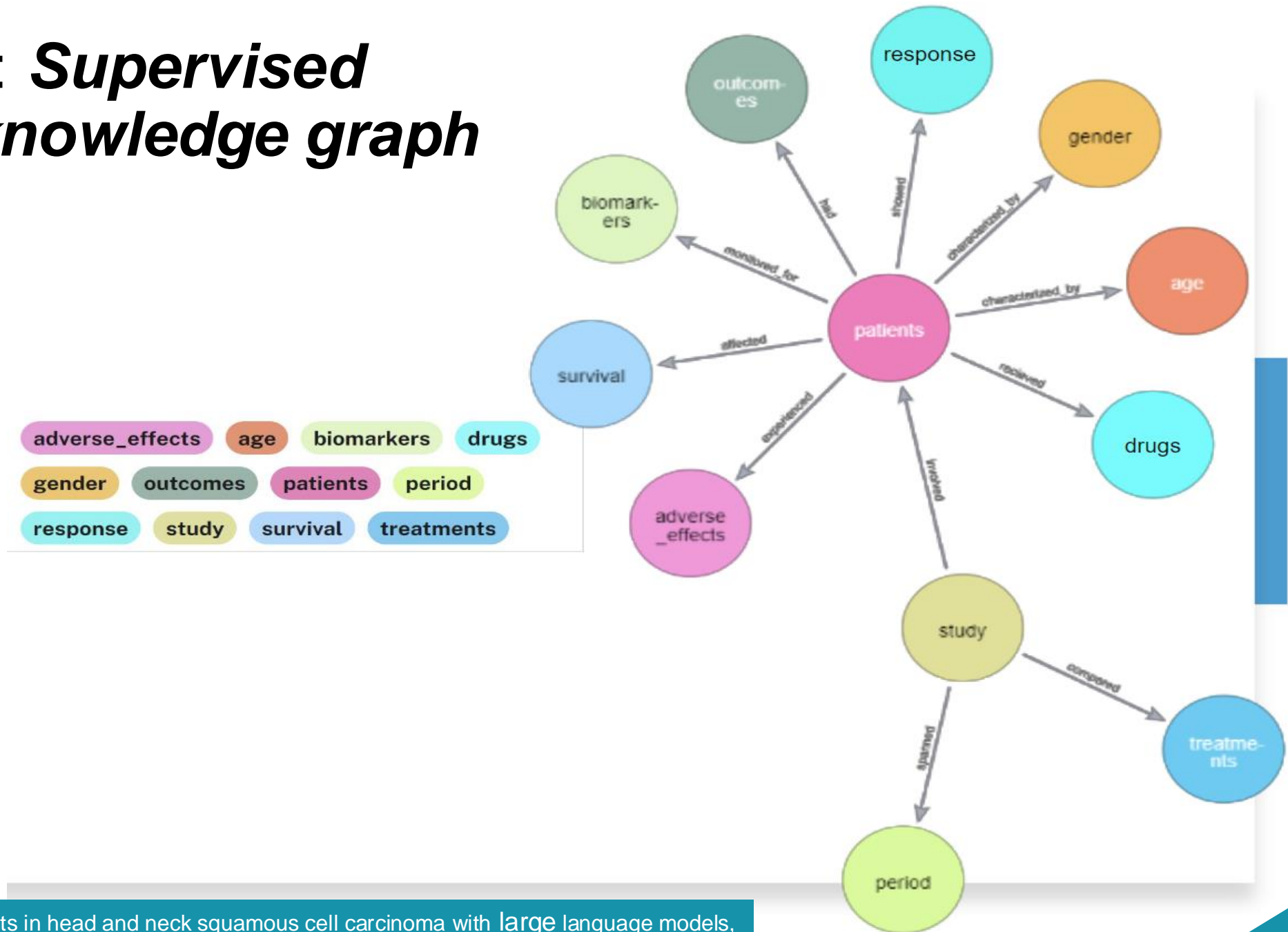
```
✓ system_prompt_for_basic_relationship = """  
Based on the extracted nodes and information from the provided medical study text, create relationships triplets  
describing the connections between the nodes in a knowledge graph format.  
Each triplet should consist of a head node, a relationship, and a tail node.  
The expected relationship format is : node|relationship|node. Do not include any prefix or suffix  
|
```

Stage 1: *Supervised*

```
system_prompt_for_basic_relationship = ""  
Based on the extracted nodes and information from the provided medical study text, create relationships triplets  
describing the connections between the nodes in a knowledge graph format.  
Each triplet should consist of a head node, a relationship, and a tail node.  
The expected relationship format is : node|relationship|node. Do not include any prefix or suffix
```

```
['Study|involved|Patients',  
 'Study|compared|Treatment',  
 'Patients|received|Drug',  
 'Patients|experienced|Adverse effects',  
 'Patients|characterized by|Age',  
 'Patients|characterized by|Gender',  
 'Patients|showed|Response',  
 'Patients|had|Outcomes',  
 'Patients|monitored for|Biomarkers',  
 'Study|spanned|Period',  
 'Patients|affected|Survival',  
 'Patients|experienced|Deaths']
```

Stage 1: *Supervised* Resulting knowledge graph



Stage 2: *Unsupervised*

```
system_prompt_for_advanced_relationships = """
```

```
Based on the extracted nodes and information from the provided medical study text,  
create relationships triplets describing the connections between the nodes in a knowledge graph format.  
Each triplet should consist of a head node, a relationship, and a tail node.
```

```
The expected relationship format is : node|relationship|node. Do not include any prefix or suffix
```

```
Create relationships between the nodes based on the extracted information and the provided medical study text
```

```
Nodes:
```

```
"""
```


Stage 2: *Unsupervised*

```
system_prompt_for_advanced_relationships = """
Based on the extracted nodes and information from the provided medical study text,
create relationships triplets describing the connections between the nodes in a knowledge graph format.
Each triplet should consist of a head node, a relationship, and a tail node.
The expected relationship format is : node|relationship|node. Do not include any prefix or suffix

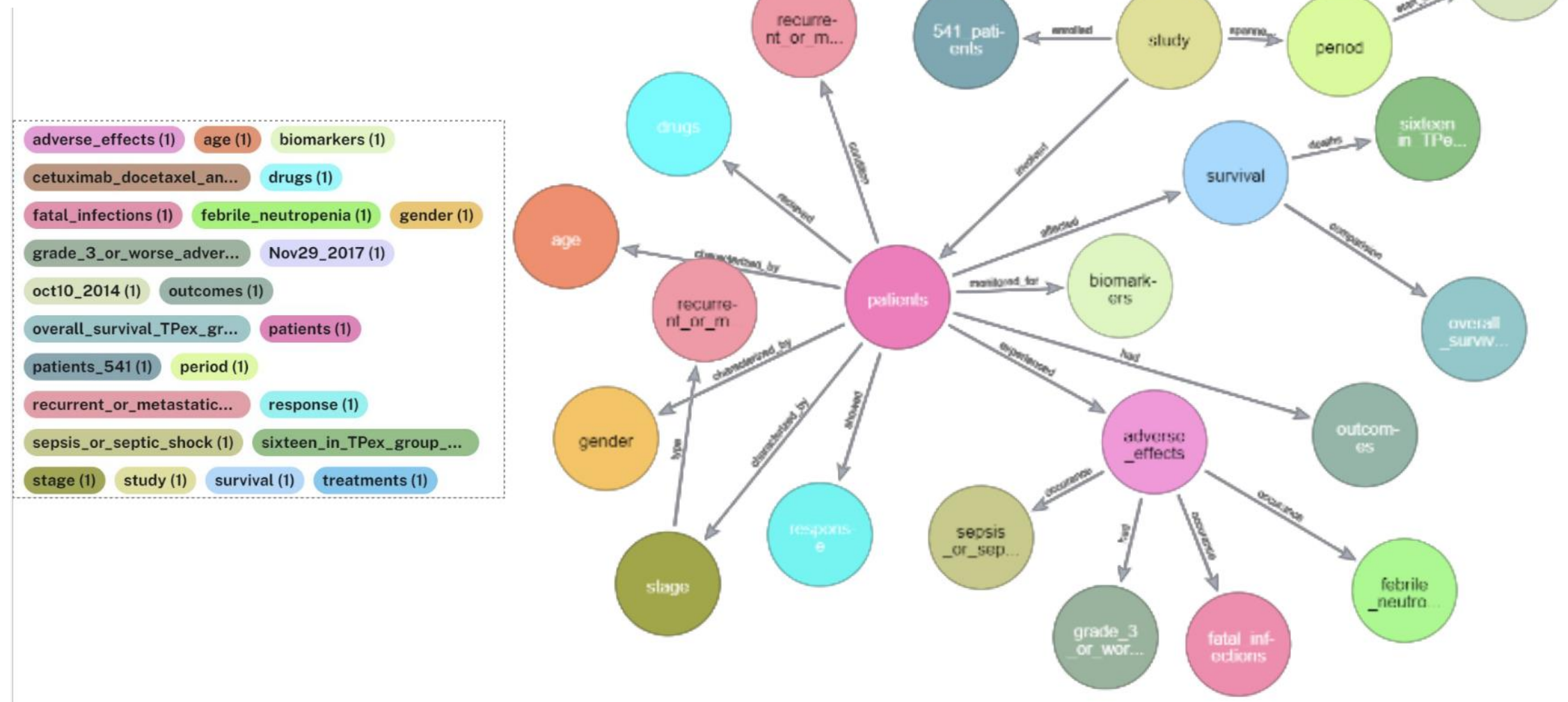
Create relationships between the nodes based on the extracted information and the provided medical study text

Nodes:
"""
```

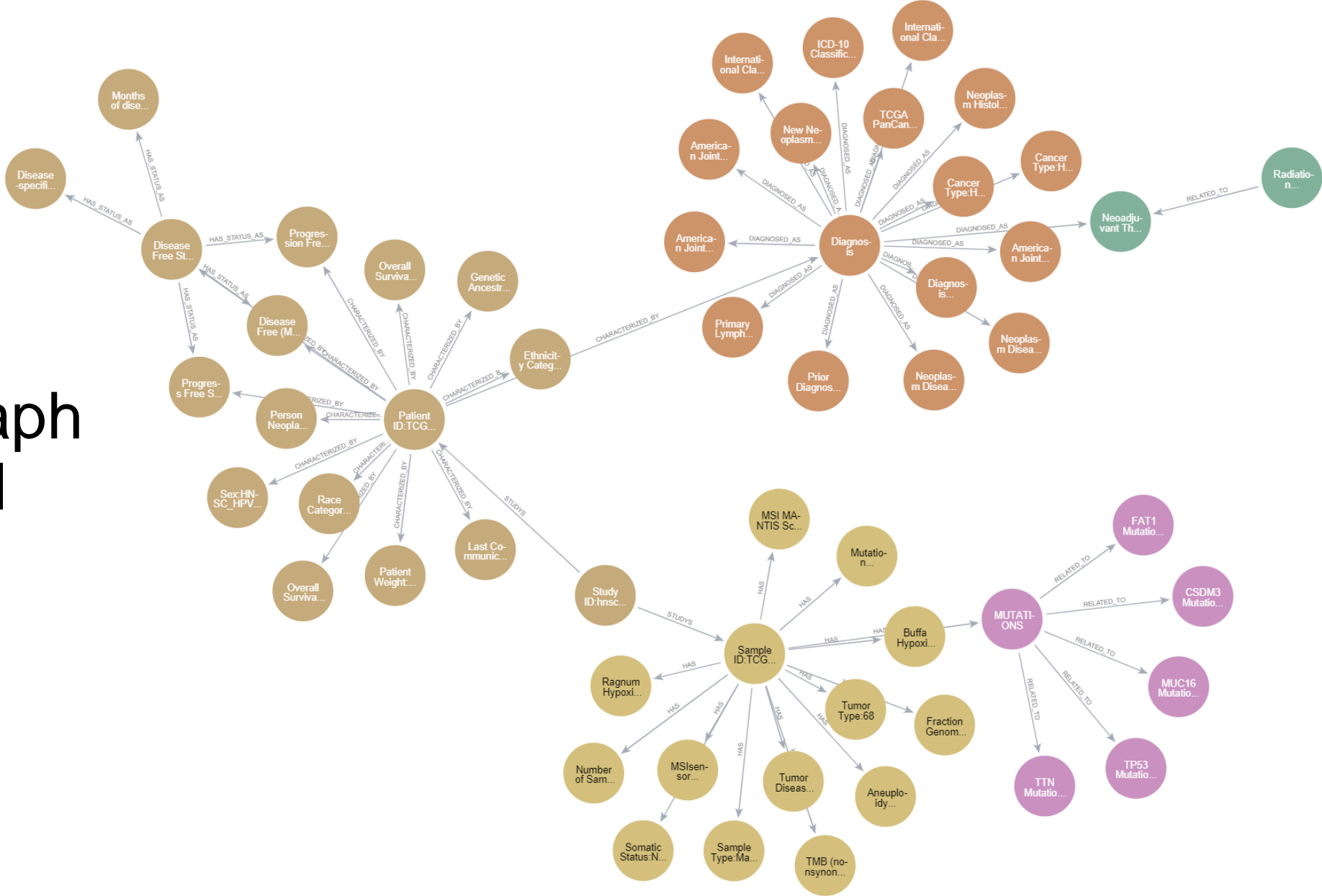
```
['Study|involved|Patients',
 'Study|compared|Treatment',
 'Patients|received|Drug',
 'Patients|experienced|Adverse effects',
 'Patients|characterized by|Age',
 'Patients|characterized by|Gender',
 'Patients|showed|Response',
 'Patients|had|Outcomes',
 'Patients|monitored for|Biomarkers',
 'Study|spanned|Period',
 'Patients|affected|Survival',
 'Patients|experienced|Deaths']
```

```
{'33684370': ['study|enrollment|541 patients',
 'patients|condition|recurrent or metastatic head and neck squamous-cell carcinoma',
 'treatment|comparison|Cetuximab, docetaxel, and cisplatin versus platinum, fluorouracil, and cetuximab',
 'stage|type|recurrent or metastatic head and neck squamous-cell carcinoma',
 'period|start date|Oct 10, 2014',
 'period|end date|Nov 29, 2017',
 'survival|comparison|Overall survival TPEX group (14.5 months) vs. EXTREME group (13.4 months)',
 'outcomes|comparison|Overall survival did not differ significantly between the groups',
 'patients|adverse events|grade 3 or worse adverse events during chemotherapy',
 'adverse_effects|occurrence|fatal infections',
 'adverse_effects|occurrence|febrile neutropenia',
 'adverse_effects|occurrence|sepsis or septic shock',
 'deaths|association|16 patients in TPEX group and 21 in EXTREME group'],
```

Stage 2: *Unsupervised* Resulting knowledge graph



Knowledge Graph from cBioPortal



NCCP Guidelines

```
Treatment regimen = ""  
<User-query>
```

Here are the eligibility and exclusion guidelines for the following targeted therapies

Cetuximab - <NCCP-Cetuximab>

Pembrolizumab - <NCCP-pembrolizumab>

Nivolumab - <NCCP-Nivolumab>

Here are the indication for administering the treatments

Cetuximab - <NCCP-Cetuximab-indication>

Pembrolizumab - <NCCP-pembrolizumab-indication>

Nivolumab - <NCCP-Nivolumab-indication>

Here are the guidelines for drug administration

Cetuximab - <NCCP-Cetuximab-administration>

Pembrolizumab - <NCCP-pembrolizumab-administration>

Nivolumab - <NCCP-Nivolumab-administration>

Example: Clinical Predictor Tool

prompt

I'll give you 10 clinical trial summaries for cetuximab treatments in Head and Neck Squamous Cell Carcinomas (HNSCC). Each summary provides detailed information from the study, including patient demographics, methods, results, and conclusions. Based on this information, learn to identify patterns and factors that led to the recommendation of the treatment for treating HNSCC. Understand and learn from the study's findings, methods, patient information, and conclusions to assist in making informed treatment recommendations for HNSCC patients.

patient profile

Create a treatment plan for the patient profile given below. The patient has HNSCC.

Patient Profile:

Name: John Doe

Age: 40 years

Gender: Male

Ethnicity: Caucasian

Medical History:

Diagnosed with high-grade mucoepidermoid carcinoma (MEC) of the right submandibular salivary gland.

Underwent non-radical resection of the tumor in January 2006 at another institution.

Experienced local disease progression and cervical lymph node involvement three months post-surgery.

Received three cycles of paclitaxel, cisplatin, fluorouracil (TPF) regimen, but disease progressed systemically with diffuse subcutaneous neoplastic infiltrates.

Underwent a second-line chemotherapy with carboplatin and vinorelbine, but further progression occurred after two cycles.

Developed metastatic lesions in mediastinal lymph nodes, pleura, and bones as evidenced by total body [18F]fluorodeoxyglucose (FDG) positron emission tomography (PET) with CT-scan imaging.

Immunohistochemical analysis of the primary tumor specimen showed intense and diffuse staining for EGFR. Cytogenetic fluorescence in situ hybridization (FISH) analysis demonstrated polysomy of chromosome 7p12 (CEP7).

Current Symptoms:

Rapidly progressive disease with metastatic involvement of mediastinal lymph nodes, pleura, bones, and subsequently brain metastases leading to recurrent convulsive seizures.

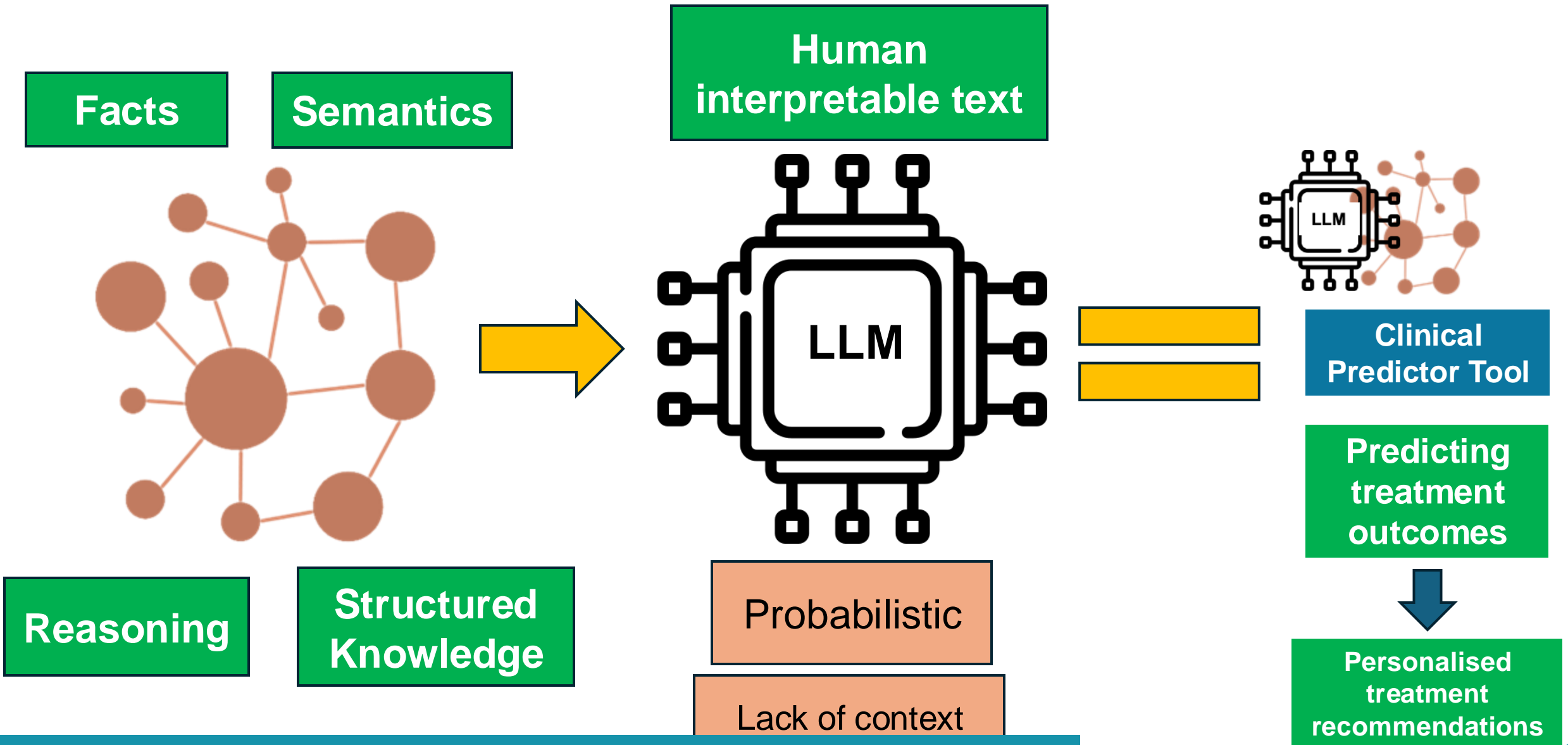
Neurologic examination showed diffuse slowing of background activity in EEG and multiple parenchymal brain metastases on CT-scan.

response from the clinical predictor tool

Given John's medical history and treatment response, the most appropriate treatment plan would involve:

- **Continuation of Cetuximab:** Despite the development of brain metastases, John showed a partial response to cetuximab in combination with cisplatin, with a reduction in extracranial metastatic lesions. Therefore, continuing cetuximab as part of his treatment regimen is warranted.
- **Consideration of Brain Metastases Management:** Since John developed brain metastases, it's essential to continue managing his neurologic symptoms and monitor for any further progression or recurrence. Depending on his neurological status and response to palliative whole-brain irradiation, additional interventions such as stereotactic radiosurgery or systemic therapies targeting brain metastases may be considered.
- **Supportive Care:** Given the rapidly progressive nature of John's disease and his worsening neoplastic pleural effusion, palliative and supportive care measures should be prioritized to manage symptoms, improve quality of life, and provide comfort.

Clinical Predictor Tool



Acknowledgements

- Culhane Lab
- Leah Wyne
- Ilias Kourousis

References

- Albarqawi, A. (2023) 'Language models in the biomedical and clinical tasks', *Medium*, available: <https://albarqawi.medium.com/language-models-in-the-biomedical-and-clinical-tasks-b0fa4eefc210> [accessed 28 Nov 2023].
- Chung, C.H., Li, J., Steuer, C.E., Bhateja, P., Johnson, M., Masannat, J., Poole, M.I., Song, F., Hernandez-Prera, J.C., Molina, H., Wenig, B.M., Kumar, S., Kuperwasser, C., Stephens, P.J., Farinhas, J.M., Shin, D.M., Kish, J.A., Muzaffar, J., Kirtane, K., Rocco, J.W., Schell, M.J., Saba, N.F., and Bonomi, M. (2022) 'Phase II Multi-institutional Clinical Trial Result of Concurrent Cetuximab and Nivolumab in Recurrent and/or Metastatic Head and Neck Squamous Cell Carcinoma', *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, 28(11), 2329–2338, available: <https://doi.org/10.1158/1078-0432.CCR-21-3849>.
- Dennis, M.J., Sacco, A.G., Qi, Y., Bykowski, J., Pittman, E., Chen, R., Messer, K., Cohen, E.E.W., and Gold, K.A. (2022) 'A phase I study of avelumab, palbociclib, and cetuximab in patients with recurrent or metastatic head and neck squamous cell carcinoma', *Oral Oncology*, 135, 106219, available: <https://doi.org/10.1016/j.oraloncology.2022.106219>.

Thank you!

Q & A

Contact:

meghana.kshirsagar@ul.ie