

# Doctor.ai: Navigate a Medical Knowledge Graph Using English

Sixing Huang, Derek Ding, Emil Pastor, Irwan Butar Butar, Shiny Zhu  
Supported by Maruthi Prithivirajan, Joshua Yu and Daniel Ng


# Why a chatbot? Democratize medical KG for all


More and more knowledge graphs in healthcare, but only experts can use them.


- For people who can neither read nor write, let alone program (in Cypher)
- For the visually impaired
  
- Voice is 2-3 times faster than typing
- Hands-free interactions are required in some scenarios (surgical rooms and dust-free labs)
- Between patient-doctor there is an information asymmetry, patients need to know more
  
- Simplify complex concepts
- Connect the personal, specific medical histories to the public, general & authoritative knowledge

# Interface


**Doctor.ai**

Which pathogen causes HIV infection 

 Human immunodeficiency virus 1 (HIV-1), Human immunodeficiency virus 2 (HIV-2)


Which drug treats HIV infection? 


Dolutegravir sodium, Etravirine, Delavirdine mesylate, Lamivudine, nevirapine and zidovudine, Dolutegravir, abacavir and lamivudine, Efavirenz, emtricitabine and tenofovir disoproxil, Lopinavir and ritonavir, Doravirine, Raltegravir potassium, Zidovudine and lamivudine, Atazanavir sulfate, Ritonavir,


Type the message ... 


**Doctor.ai**

lamivudine, Tipranavir, Abacavir sulfate, Fosamprenavir calcium, Emtricitabine and tenofovir alafenamide, Efavirenz, Emtricitabine and tenofovir disoproxil, Elvitegravir, Abacavir, lamivudine and zidovudine, Indinavir sulfate, Darunavir and cobicistat, Enfuvirtide, Rilpivirine hydrochloride, tenofovir disoproxil fumarate and emtricitabine, Bictegravir, emtricitabine and tenofovir alafenamide, Cabotegravir and rilpivirine

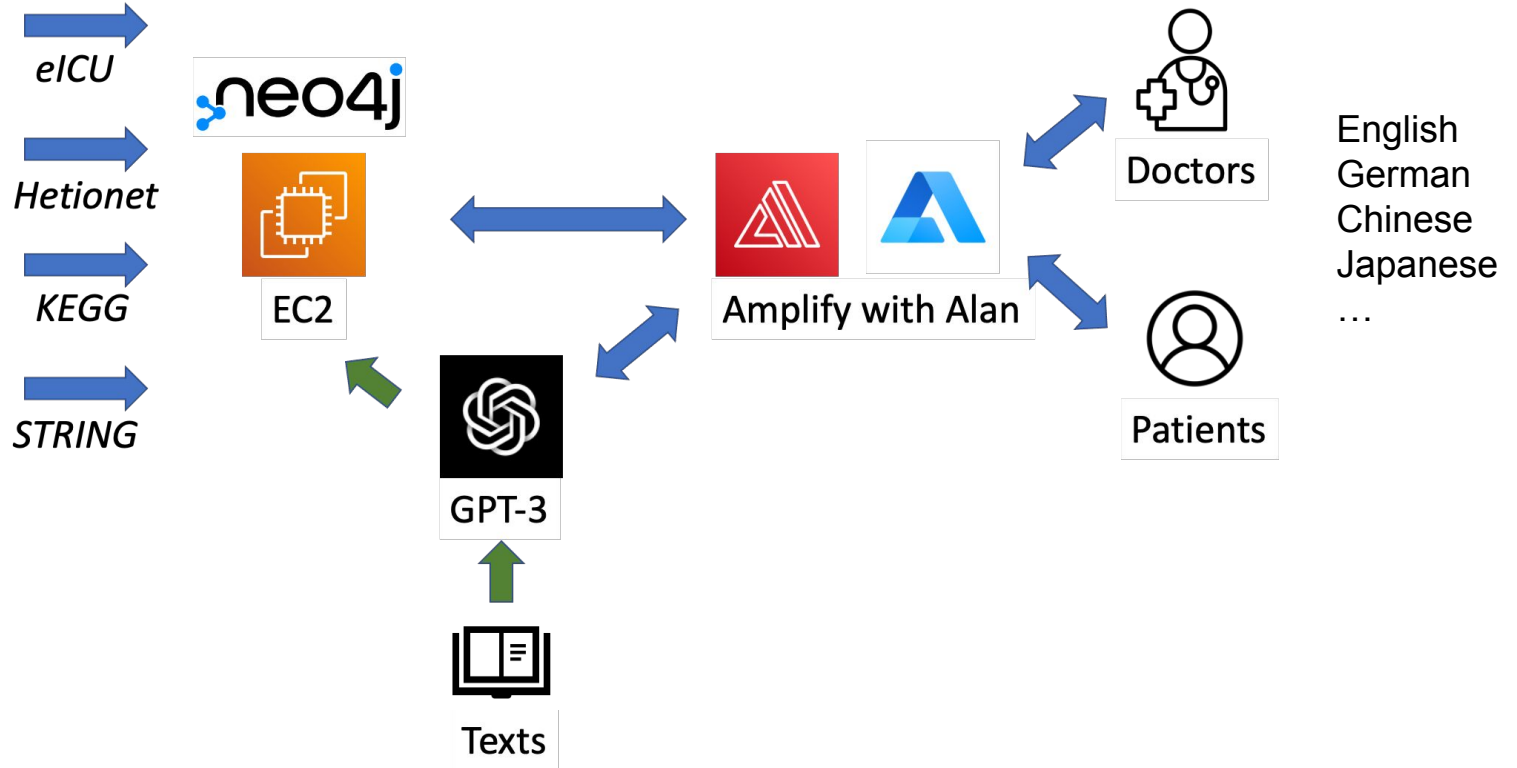


Which gene causes Spondylocarpotarsal synostosis syndrome 

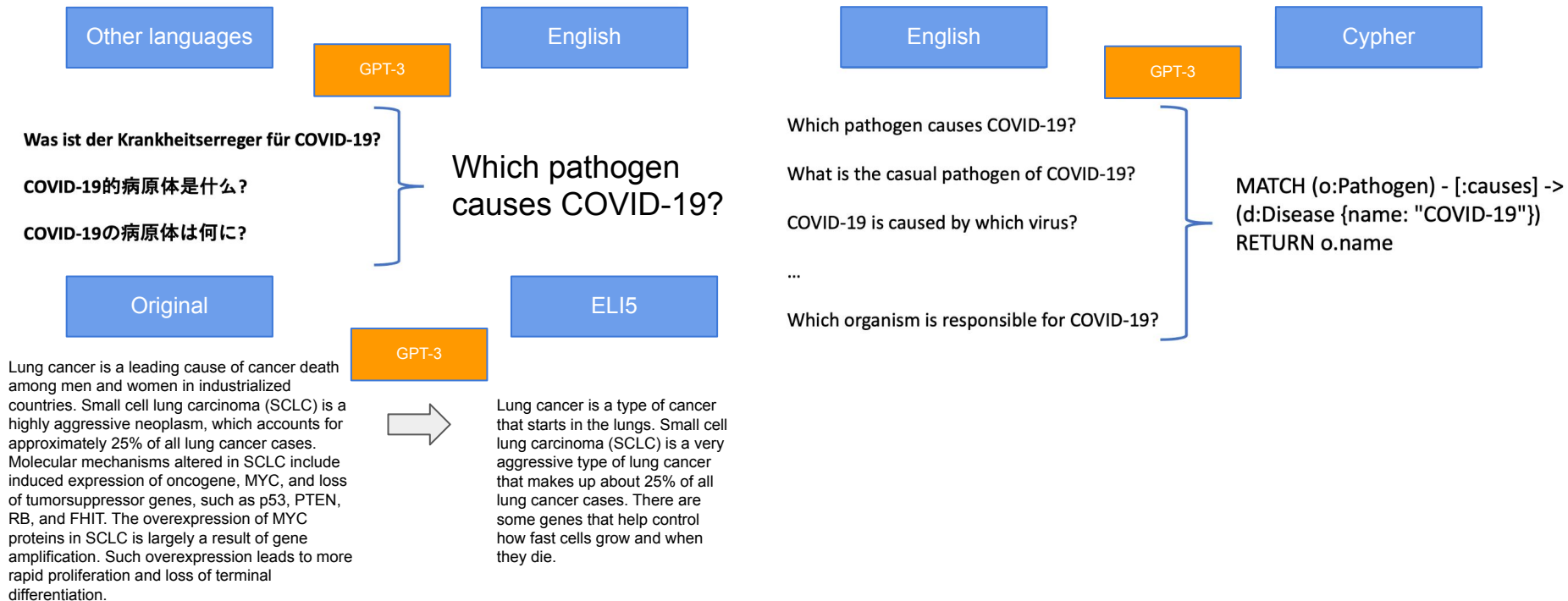
 FLNB

Type the message ... 

# The architecture



# Natural Language Understanding (NLU): GPT-3



# GPT-3

tion Examples Playground Upgrade Help MGI

Playground Summarize for a 2nd grader Save View code Share ...

```
1 #How many times did patient id_1 visit the ICU?
2 MATCH (p:Patient {patient_id: "id_1"}-[:HAS_STAY]->(v:PatientUnitStay) RETURN COUNT(v)
3
4 #When did patient id_1 visit the ICU?
5 MATCH (p:Patient {patient_id: "id_1"}-[:HAS_STAY]->(v:PatientUnitStay) RETURN v.hospitaldischargeyear
6
7 #Which drug treats COVID-19?
8 MATCH (d:Compound)-[:treats]->(c:Disease {name: "COVID-19"}) RETURN d.name
9
10
11 #Which pathogen causes Kyasanur Forest disease?
12 MATCH (o:Pathogen)-[:causes]->(d:Disease {name: "Kyasanur Forest disease"}) RETURN o.name
13
14 #Which pathogen causes COVID-19?
15 MATCH (o:Pathogen)-[:causes]->(d:Disease {name: "COVID-19"}) RETURN o.name
16
17 #Which gene causes Christianson syndrome?
18 MATCH (g:Gene)-[:r1:associates]->(d:Disease {name: "Christianson syndrome"}) RETURN g.name
19
20 #Tell me something about the disease named "Christianson syndrome"
21 MATCH (d:Disease {name: "Christianson syndrome"}) RETURN d.description
22
23 #I have Dyspepsia, Hiccup and Edema. What can be the cause of this?
24 MATCH (s1:Symptom {name: "Dyspepsia"}) <-[:presents]- (d:Disease) MATCH (s2:Symptom {name: "Hiccup"}) <-[:presents]- (d:Disease)
25 MATCH (s3:Symptom {name: "Edema"}) <-[:presents]- (d:Disease) RETURN d.name
26
27 #Tell me something about the disease called COVID-19?
28 MATCH (d:Disease {name: "COVID-19"}) RETURN d.description
29
```

1. Choose this  
code-davinci-001

2. Paste your #English-Cypher examples first

3. Your question

4. Click this  
Generate

5. The answer is correct!

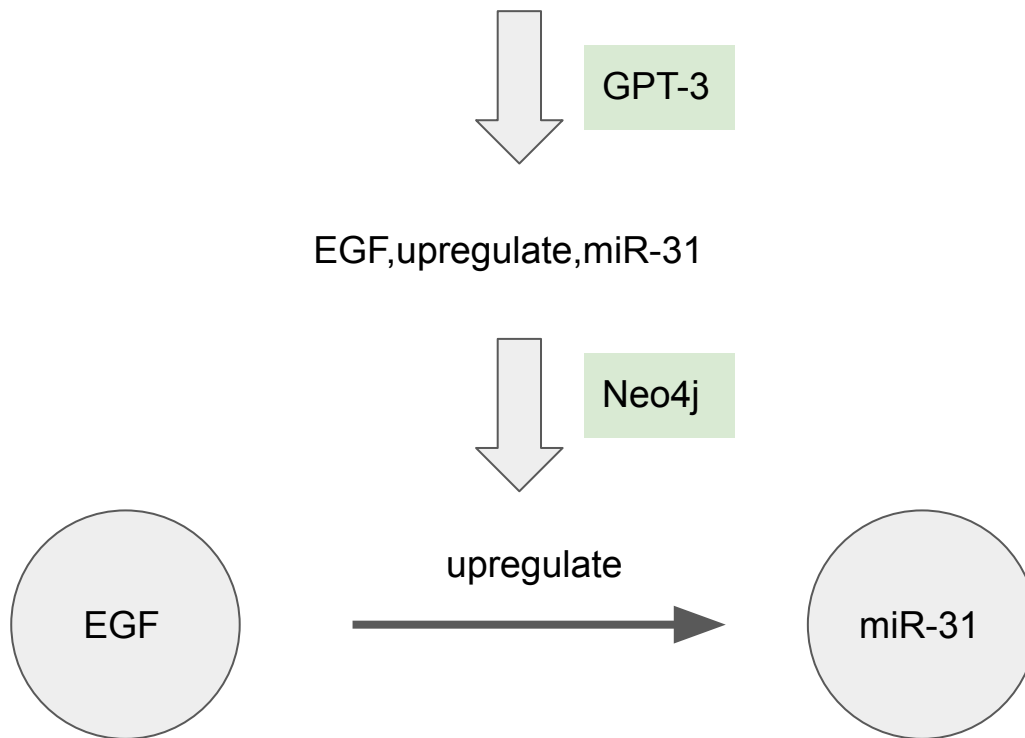
Temperature 0  
Response length 50  
Stop sequences  
Enter sequence and press Tab  
Top P 1  
Frequency penalty 0  
Presence penalty 0  
Best of 1  
Inject start text  
Inject restart text  
Show model settings

491 Plain text

Demo

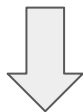
# Use GPT-3 to extract relationships from raw texts

“EGF up-regulates miR-31 through the C/EBP $\beta$  signal cascade in oral carcinoma”



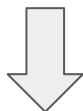


“In chicken, adiposity is influenced by hepatic stearoyl-CoA desaturase (SCD) 1. This gene is up-regulated by low-fat high-carbohydrate diet and down-regulated by addition of polyunsaturated fatty acids (PUFA).. ...”



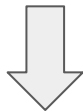
Crosslingual Coreference resolves pronouns and acronyms

In chicken, ... hepatic stearoyl-CoA desaturase (SCD) 1. hepatic stearoyl-CoA desaturase (SCD) 1 is up-regulated by ...



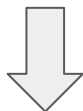
spaCy splits the sentences

[“In chicken ...”,  
“hepatic stearoyl-CoA desaturase (SCD) 1 is up-regulated by ...”]



Hugging face filters the list with NER

“hepatic stearoyl-CoA desaturase (SCD) 1 is up-regulated by ...”



GPT-3 extracts relationships

low-fat high-carbohydrate diet,upregulate,SCD1  
addition of PUFA,downregulate,SCD1

# Import the relationships

**Data Importer** Version 0.1.1-beta [Send feedback](#) [Show Results](#) [Run Import](#) ... | ?

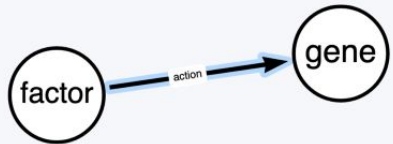
**Files** [Browse](#)

**gene\_regulation.csv**

- factor heat shock
- action upregulate
- gene cplB

**Graph Model**

Add node

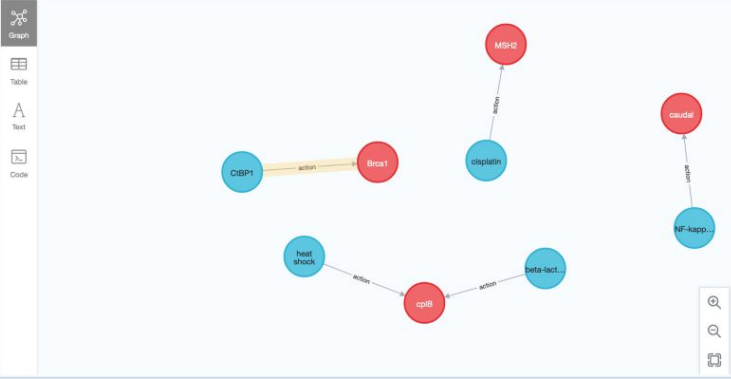


```
graph LR; factor((factor)) -- action --> gene((gene))
```

**Mapping Details**

Type `neo4j$`

```
$ MATCH p=(*)->(*) RETURN p LIMIT 25
```

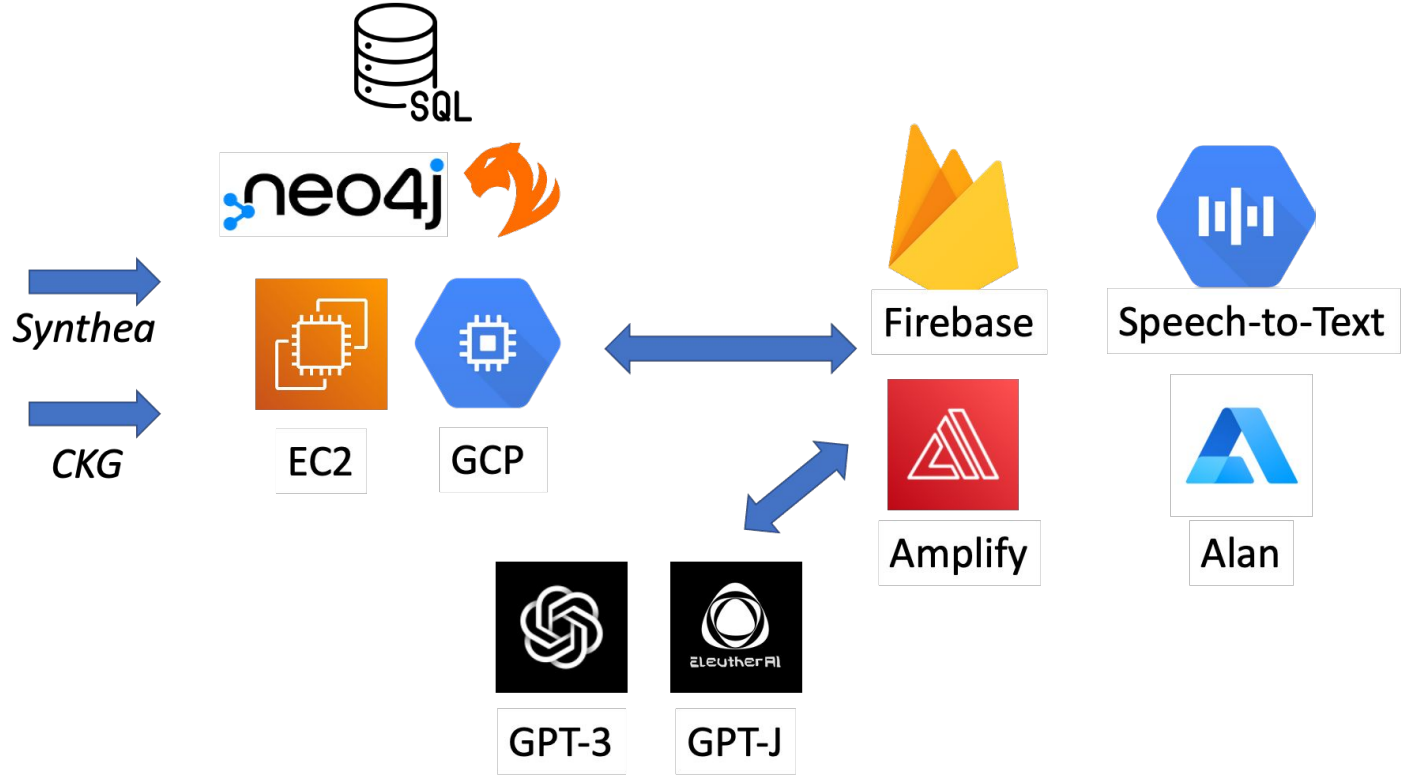


**Relationship Properties**

Property	Value
action	
<id>	4
action	downregulate

`:$play start`

# Interchangeable components for different purposes



# Publications and repos

[Doctor.ai, an AI-Powered Virtual Voice Assistant for Health Care](https://medium.com/p/8c09af65aabb) (https://medium.com/p/8c09af65aabb)

[Transfer Knowledge Graphs to Doctor.ai](https://towardsdatascience.com/transfer-knowledge-graphs-to-doctor-ai-cc21765fa8a6)

(https://towardsdatascience.com/transfer-knowledge-graphs-to-doctor-ai-cc21765fa8a6)

[Use Crosslingual Coreference, spaCy, Hugging face and GPT-3 to Extract Relationships from Long Texts](https://medium.com/geekculture/use-crosslingual-coreference-spacy-hugging-face-and-gpt-3-to-extract-relationships-from-long-17a9f7f48b9a)

(https://medium.com/geekculture/use-crosslingual-coreference-spacy-hugging-face-and-gpt-3-to-extract-relationships-from-long-17a9f7f48b9a)

[Relationship Extraction with GPT-3](https://medium.com/p/bb019dcf41e5) (https://medium.com/p/bb019dcf41e5)

[ELI5 Medical Texts with GPT-3](https://medium.com/p/a5c2c4580977) (https://medium.com/p/a5c2c4580977)

[https://github.com/dgg32/doctorai\\_eli5](https://github.com/dgg32/doctorai_eli5)

[https://github.com/dgg32/huggingface\\_gpt3](https://github.com/dgg32/huggingface_gpt3)

# Thank you

My collaborators, Neo4j and the Graph community

MGI

Basecamp Research