GPT4V hierarchical data extraction

By team NoLimits

Team NO LIMITS

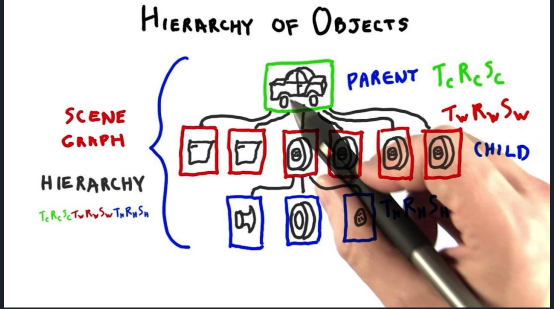
Jean-Pierre Bianchi
MS in Machine Learning
ML engineer (Omdena)
Ex founder & CTO
LinkedIn

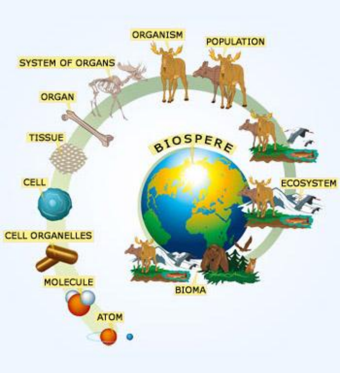
Information is hierarchical

Reality is made of objects, which are made of objects. Atoms make cells, which make organs, which make bodies, which make families & groups etc

The concept of object, as clearly defined by dictionaries, is used by everyone in every sentence, but it is quite hard to make algorithms understand and see what humans do so easily.

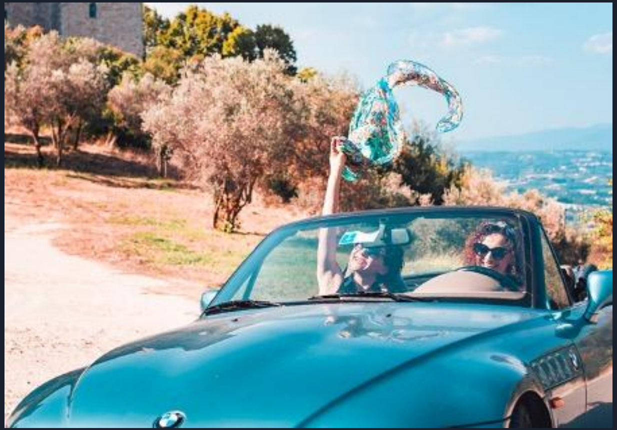
This project aims to extract information from images with GPT4V, and visualize it with Neo4J in a hierarchical knowledge graph.





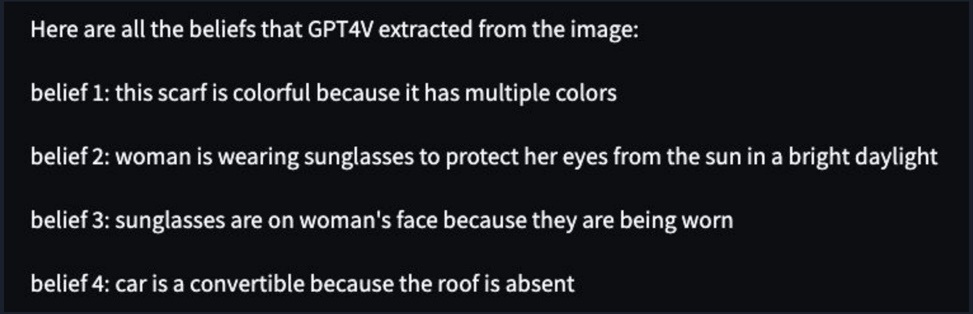
The power of GPT4V

In this image, anyone can immediately recognize a convertible car, trees, a scarf, women with sunglasses enjoying a sunny day, and even an old building.



GPT4V is extremely good at interpreting a scene, and produce not only the objects in it but also explanations (‘beliefs’) about them.

GPT4V could describe the scene clearly, and extract a list of beliefs, which are spot on.



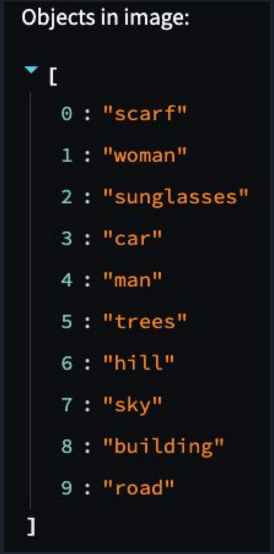
Here are all the beliefs that GPT4V extracted from the image:

belief 1: this scarf is colorful because it has multiple colors

belief 2: woman is wearing sunglasses to protect her eyes from the sun in a bright daylight

belief 3: sunglasses are on woman's face because they are being worn

belief 4: car is a convertible because the roof is absent



Objects in image:

```
▾ [
    0 : "scarf"
    1 : "woman"
    2 : "sunglasses"
    3 : "car"
    4 : "man"
    5 : "trees"
    6 : "hill"
    7 : "sky"
    8 : "building"
    9 : "road"
]
```

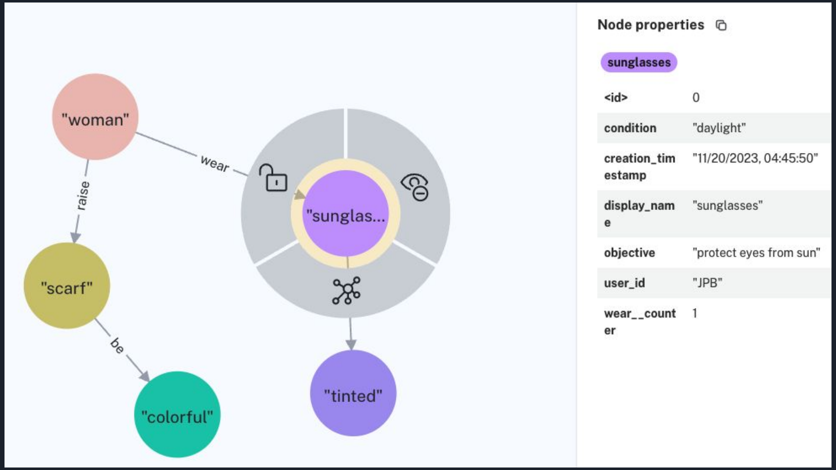
GPT4 to go even further

We went much further by prompting GPT4 to produce the relationships ('actions') between objects.

And also, the 'objective' and the 'conditions' in which such actions happens as you can see.

GPT4 was spot on and could extract detailed explanations of why women wear sunglasses and in which conditions.

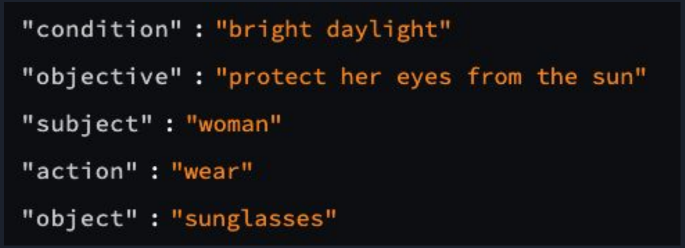
So, each object was represented by 5 fields, which is ideal to visualize them in a Neo4J graph.



Node properties

sunglasses

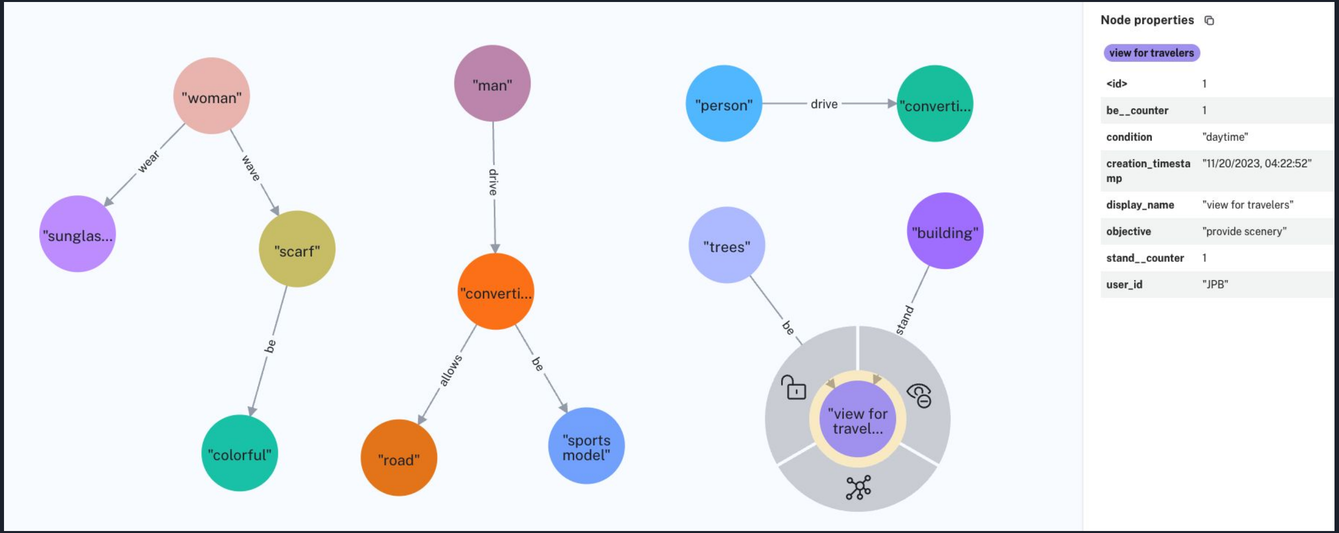
<id>	0
condition	"daylight"
creation_timestamp	"11/20/2023, 04:45:50"
display_name	"sunglasses"
objective	"protect eyes from sun"
user_id	"JPB"
wear__counter	1



```
"condition" : "bright daylight"
"objective" : "protect her eyes from the sun"
"subject" : "woman"
"action" : "wear"
"object" : "sunglasses"
```

Solution

- Prompt GPT4V to extract not only objects in a picture, but also the relationship (‘beliefs’) with other objects, the objective and conditions in which this relationship applies
- However, GPT4V was not as good as GPT4 to extract the 5 fields of information from the text it produced, so we used GPT4 for the final post-processing.
- We developed an API for Neo4J to easily create a hierarchical knowledge graph in one line of code.



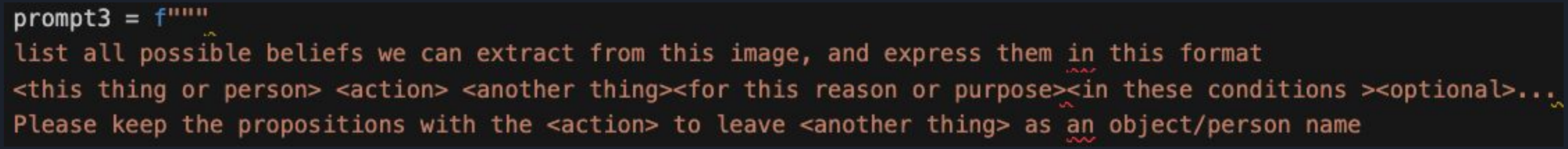
Node properties

view for travelers

<id>	1
be__counter	1
condition	"daytime"
creation_timestamp	"11/20/2023, 04:22:52"
display_name	"view for travelers"
objective	"provide scenery"
stand__counter	1
user_id	"JPB"

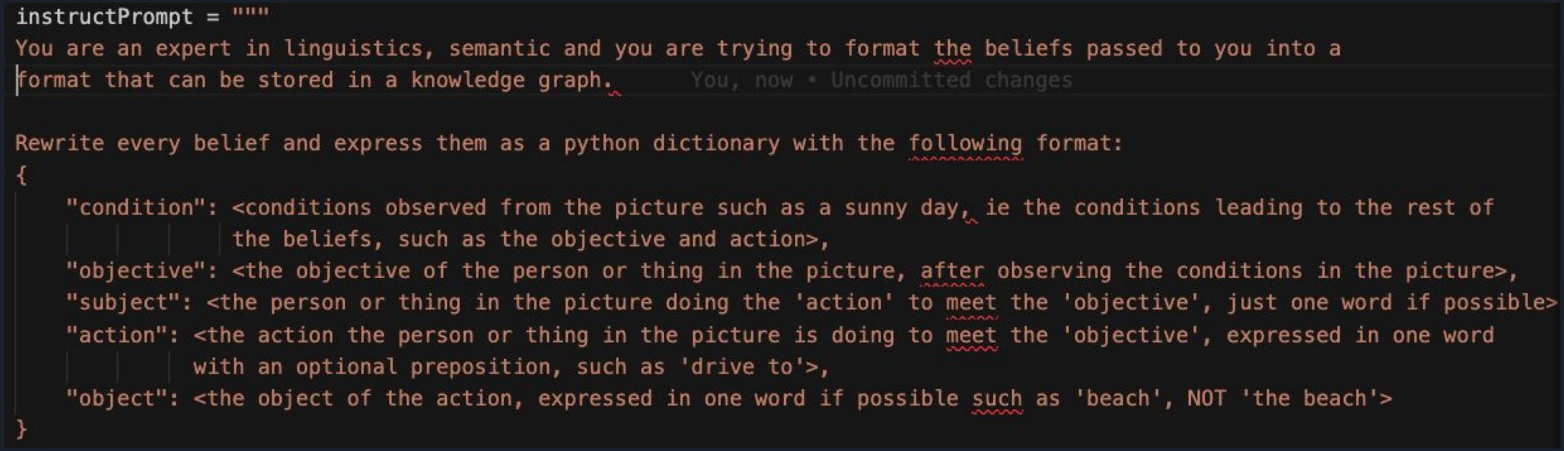
Prompts

GPT4V prompt to force it to produce a ‘belief’ about every object



```
prompt3 = f"""
list all possible beliefs we can extract from this image, and express them in this format
<this thing or person> <action> <another thing><for this reason or purpose><in these conditions ><optional>...
Please keep the propositions with the <action> to leave <another thing> as an object/person name
```

GPT4 prompt to extract 5 fields from every belief.



```
instructPrompt = """
You are an expert in linguistics, semantic and you are trying to format the beliefs passed to you into a
format that can be stored in a knowledge graph.      You, now • Uncommitted changes

Rewrite every belief and express them as a python dictionary with the following format:
{
    "condition": <conditions observed from the picture such as a sunny day, ie the conditions leading to the rest of
                 the beliefs, such as the objective and action>,
    "objective": <the objective of the person or thing in the picture, after observing the conditions in the picture>,
    "subject": <the person or thing in the picture doing the 'action' to meet the 'objective', just one word if possible>
    "action": <the action the person or thing in the picture is doing to meet the 'objective', expressed in one word
               with an optional preposition, such as 'drive to'>,
    "object": <the object of the action, expressed in one word if possible such as 'beach', NOT 'the beach'>
}
```

Advantages

Representing data as objects in a hierarchical structure has many advantages:

- Complexity vs compression: the notion of object in itself is a powerful compression technique, ie complex objects are represented by one word and simply linked to their parts in a graph.
- Efficiency: once the knowledge graph is built, it represents the knowledge of humanity, which evolves but doesn't change drastically.
  - It is not necessary to re-learn it all at every training, which translates into much shorter training times*
  - Because of the compression advantage, this should also translate into smaller LLM's with new architectures incorporating the knowledge graph
  - Inference is much faster (human-like)
- Explainability: being able to classify objects like humans also allows to explain which beliefs were decisive in the process
- Because the objects are those humans use, the graph can be inspected and edited by humans (unlike the data features hidden in gigantic matrices)

*Providing a knowledge-base memory to LLM's is a very exciting prospect and could be a game changer!!

Markets & revenue stream

A hierarchical knowledge based approach can greatly impact several markets:

- Behavior modeling (psychology, personal assistant / advisor, predictions)
- CV: full explainability, better accuracy, less/no errors, no hallucinations
- Robotics, AV, guided systems improve when one can localize & recognize objects
- Healthcare: better diagnostics, and therefore better treatments
- ML: algorithms improvement, dimensionality reduction
- LLM: training time and size reductions

It is extremely difficult to evaluate the potential revenues from so many big markets at this stage. This would require a full blown business plan. But one can already see how providing true explainability and improved accuracy can be a game changer!

Revenues are certainly in the billions to whoever takes the lead.

Next steps / backlog

This project is a ‘proof of concept’, not a commercializable app, simply because there are so many ways to use it to improve ML & AI algorithms, even LLM’s, which in turn will impact many important markets.

GPT4V gave outstanding results, but they could be improved for instance by recognizing synonyms to avoid generating several nodes for the same objects.

Although correct, the results were not deterministic, so the code should be improved to increment the graph with new beliefs.

With funding, we would address the critical growth areas of any startup:

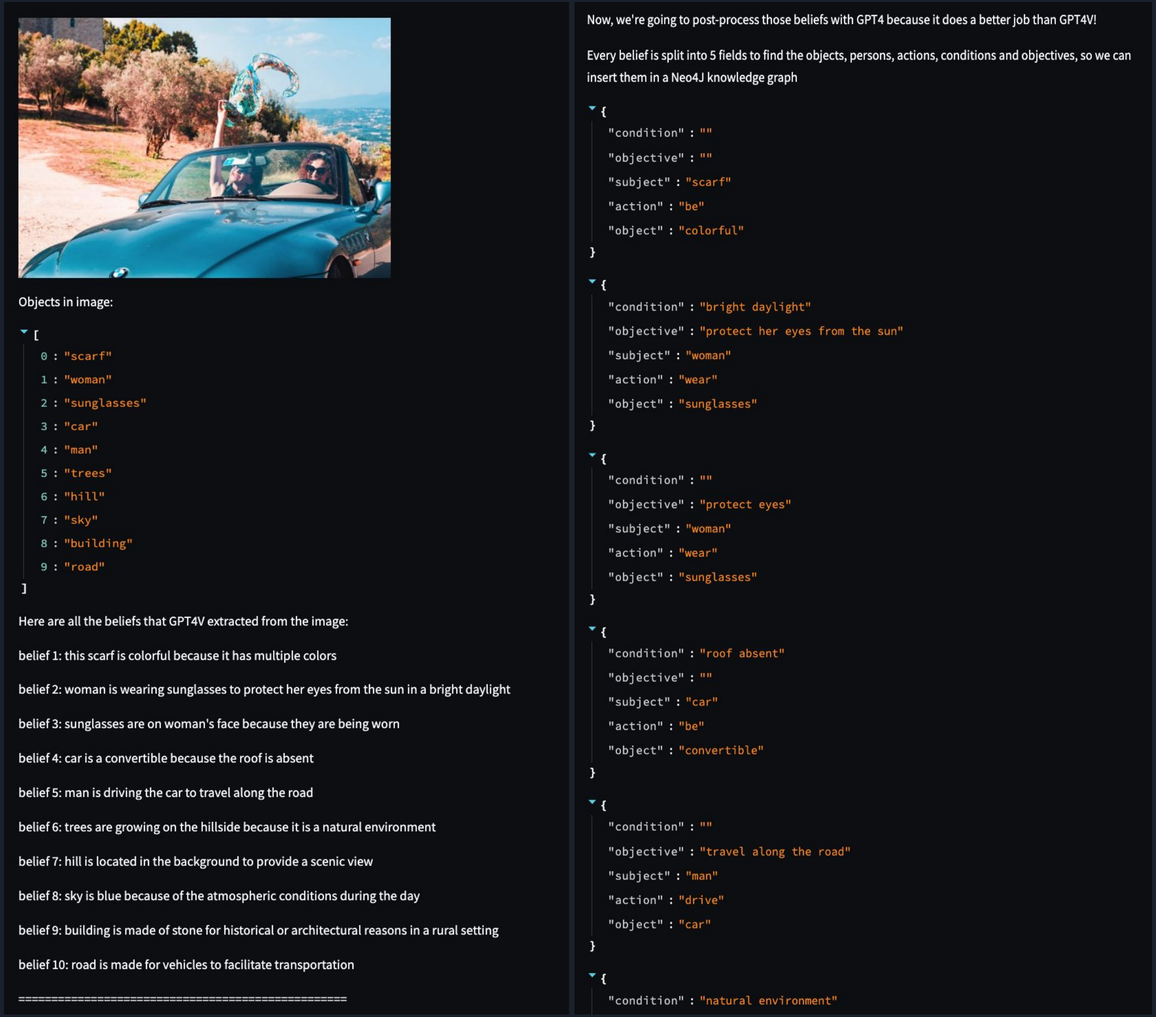
- Improve the current solution to create a deeper and totally reliable hierarchical knowledge base
  - Requires more programmers, access to expensive hardware
  - Improve all aspects, test, extensive and reliable CI/CD
  - Then turn it into a professional product, tested, deployed, and scalable
- Identify the first ‘easy’ applications in terms of impact on potential markets
  - object recognition in general
  - Then specializing it for robotics, AV, diagnostics etc
- Find key partners to try our technology
  - Requires staff with industry knowledge and connections
  - Requires marketing experts
- Become leaders in such technology
  - Requires hiring experts in targeted fields and showcasing our potentials by publishing articles and papers
  - Put a solid management structure in place

Working solution

A proof of concept has been deployed at gpt4v-demo.streamlit.app

The code is at https://github.com/jpbianchi/

See our video for more details.



Objects in image:

```
▾ [
    0 : "scarf"
    1 : "woman"
    2 : "sunglasses"
    3 : "car"
    4 : "man"
    5 : "trees"
    6 : "hill"
    7 : "sky"
    8 : "building"
    9 : "road"
  ]
```

Here are all the beliefs that GPT4V extracted from the image:

belief 1: this scarf is colorful because it has multiple colors

belief 2: woman is wearing sunglasses to protect her eyes from the sun in a bright daylight

belief 3: sunglasses are on woman's face because they are being worn

belief 4: car is a convertible because the roof is absent

belief 5: man is driving the car to travel along the road

belief 6: trees are growing on the hillside because it is a natural environment

belief 7: hill is located in the background to provide a scenic view

belief 8: sky is blue because of the atmospheric conditions during the day

belief 9: building is made of stone for historical or architectural reasons in a rural setting

belief 10: road is made for vehicles to facilitate transportation

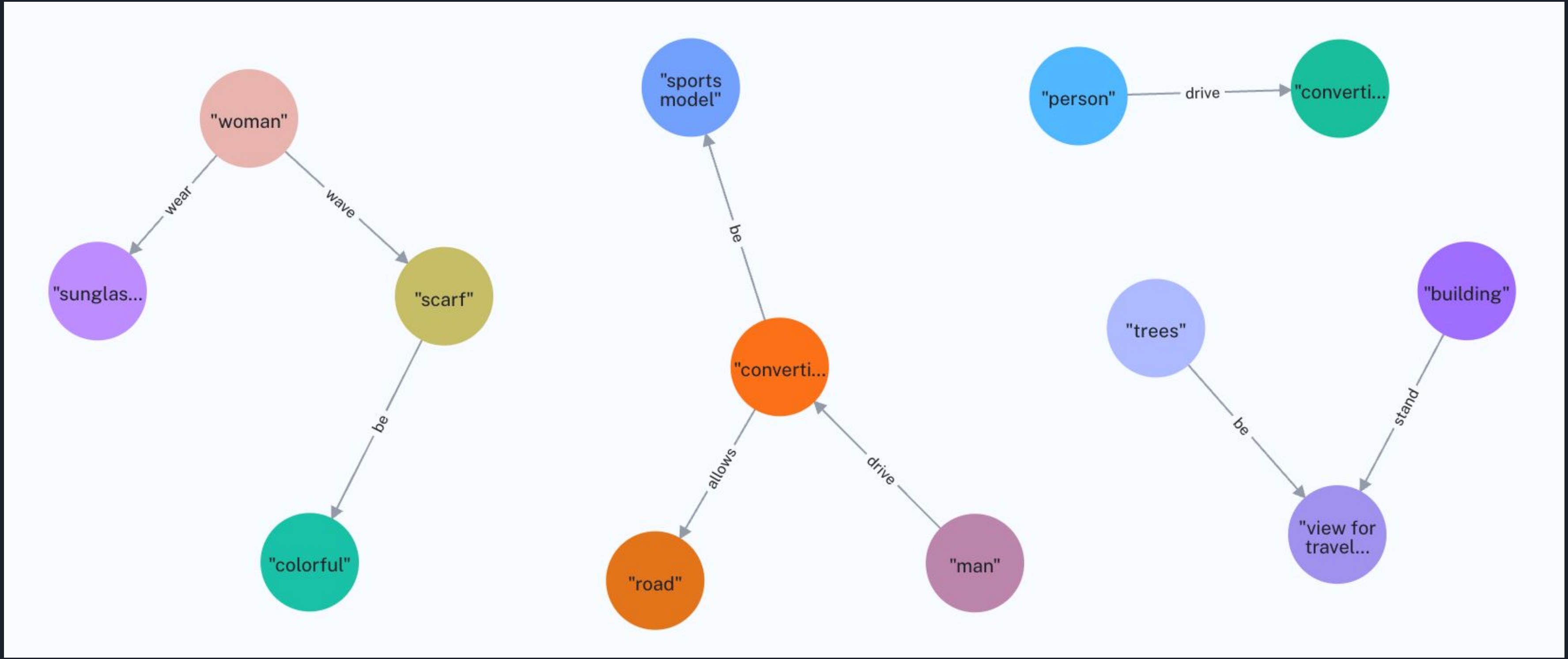
==

Now, we're going to post-process those beliefs with GPT4 because it does a better job than GPT4V!

Every belief is split into 5 fields to find the objects, persons, actions, conditions and objectives, so we can insert them in a Neo4J knowledge graph

```
▾ {
    "condition" : ""
    "objective" : ""
    "subject" : "scarf"
    "action" : "be"
    "object" : "colorful"
  }
▾ {
    "condition" : "bright daylight"
    "objective" : "protect her eyes from the sun"
    "subject" : "woman"
    "action" : "wear"
    "object" : "sunglasses"
  }
▾ {
    "condition" : ""
    "objective" : "protect eyes"
    "subject" : "woman"
    "action" : "wear"
    "object" : "sunglasses"
  }
▾ {
    "condition" : "roof absent"
    "objective" : ""
    "subject" : "car"
    "action" : "be"
    "object" : "convertible"
  }
▾ {
    "condition" : ""
    "objective" : "travel along the road"
    "subject" : "man"
    "action" : "drive"
    "object" : "car"
  }
▾ {
    "condition" : "natural environment"
```

Typical Neo4J graph produced by the app


"woman"
wear
wave
"sunglas...
"scarf"
be
"colorful"
"sports model"
be
"converti...
allows
drive
"road"
"man"
"person"
drive
"converti...
"trees"
"building"
be
stand
"view for travel...