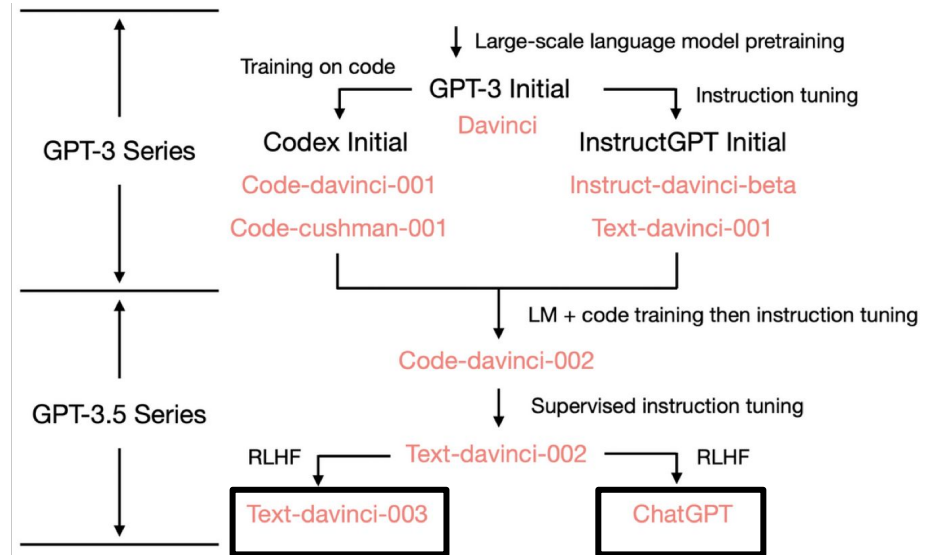
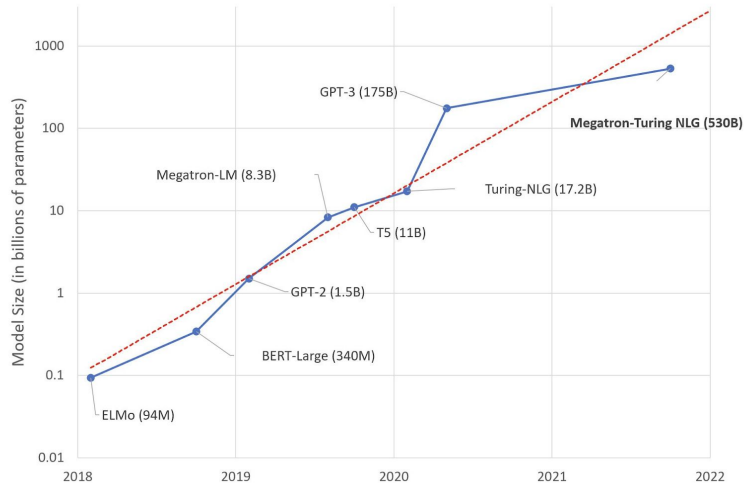


Harnessing GPT-3.5 For Generation of Analogies Between Paragraphs Describing Scientific Processes



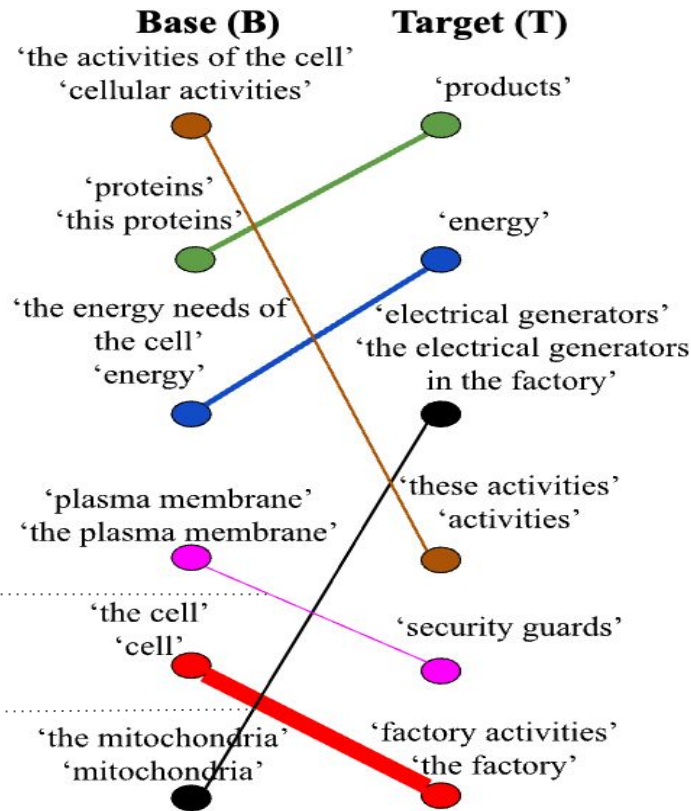
Our work: we tackle a more **realistic** setting – **analogies** between **natural language procedural texts** describing **situations or processes**

Base: Animal Cell

The plasma membrane encloses the animal cell. It controls the movement of materials into and out of the cell. The Nucleus controls the activities of the cell. These cellular activities require energy. The Mitochondria extract energy from food molecules to provide the energy needs of the cell. Animal cells must also synthesize a variety of proteins and other organic molecules necessary for growth and repair. Ribosomes produce these proteins. The cell may use these proteins or move them out of the cell for use in other cells. To move organic molecules, the cell contains a complex system of membranes that create channels within the cell. This system of membranes is called the endoplasmic reticulum.

Target: Factory

Security guards monitor the doors of the factory. They control the movement of people into and out of the factory. Factory activities may be coordinated by a control center. These activities require energy. The electrical generators in the factory provide energy. The factory synthesizes products from raw materials using machines. The factory has hallways to move products through it.



Find Mappings by Questions (FMQ)

Mining Analogies From ProPara – Examples

Examples for analogies mined by our method (FMQ):

B1 Prompt: Describe how oxygen reaches cells in the body



Self analogy

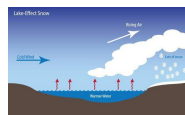


T1 Prompt: What do lungs do?

B2 Prompt: How does rain form?

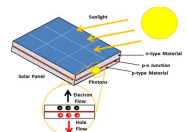


Close analogy

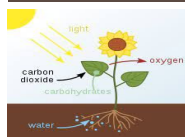


T2 Prompt: How does snow form?

B3 Prompt: How does a solar panel work?



Far analogy



T3 Prompt: What happens during photosynthesis?

Method	Not	Sub	Self	Close	Far
SBERT	0	0	89	11	0
FMV	28	15	26	20	11
FMQ	21	16	29	18	16

Top-100 of the ranking. Analogies prevalence in data: ~3%

Evaluating the Mappings – The ProPara dataset

PROMPT: Describe the process by which hurricanes form

Warm water floats up from the ocean.

A pocket of low pressure air is created near the surface of the ocean.

Warm air from outside areas pushes into the low pressure area.

The warm air keeps rising and forms clouds. The wind is getting fed moisture that evaporates from the ocean. This moisture causes the swirling air and clouds to grow bigger.

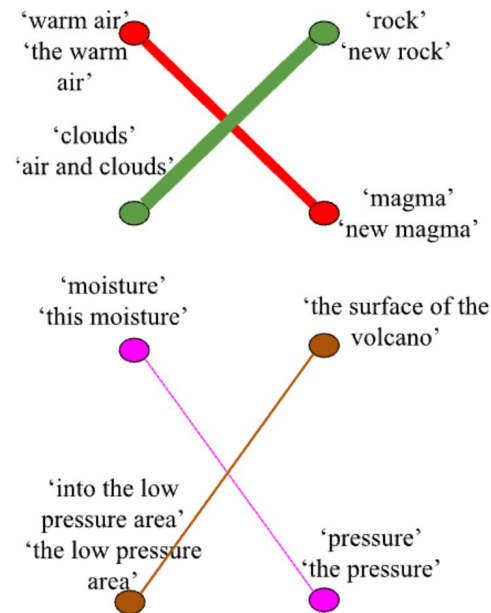
As it gets bigger the wind rotates faster.

PROMPT: What causes a volcano to erupt?

Magma rises from deep in the earth. The magma goes into volcanos. The volcanos pressure the magma upwards.

The pressure causes the magma to push through the surface of the volcano. The lava cools. The lava forms new rock. New magma is pressured to the surface of the volcano.

The volcano bursts through the rock that formed after the last eruption.



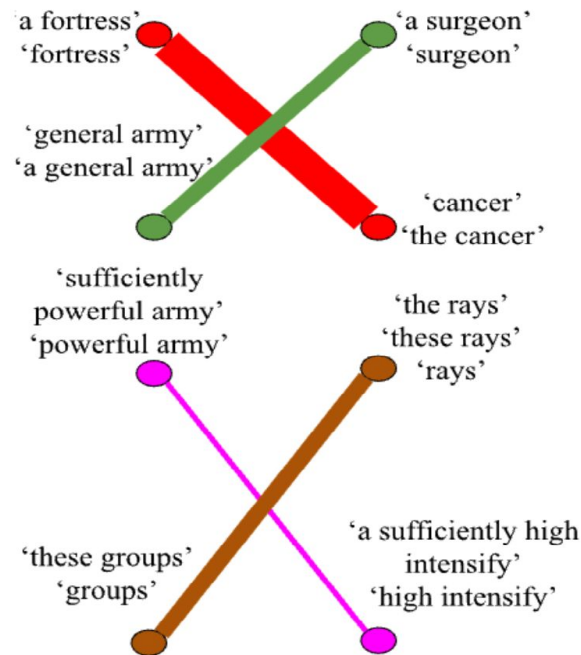
Evaluating the Mappings – Analogous Stories

Base: The general

A general was trying to destroy a fortress which was situated at the center of a country with roads leading to it, by using his army. He needed to use his army as a complete group in order to destroy the fortress. However, he could not march his army down a road to the fortress because the roads were mined to explode when large groups of men passed over them. After considerable thought, he knew just what to do. He divided his army up into small groups of men, and by sending these groups, simultaneously, from a number of different directions, they converged on the fortress, making up a sufficiently powerful army to destroy it.

Target: The surgeon

A surgeon was trying to destroy a cancer which was situated in the central region of a patient's brain, by using a type of ray. He needed to use these rays at a high intensity in order to destroy the cancerous tissue. However, at such an intensity the healthy brain tissue will also be destroyed. After considerable thought, he knew just what to do. He divided the rays up into batches of low-intensity rays, and then by sending them, simultaneously, from a number of different directions, they converged on the cancer, making up a sufficiently high intensity to destroy it.



ChatGPT – In-Context Learning Phase

I'm going to write two paragraphs that are analogous.

The Source paragraph is about the animal cell.

The Target paragraph is about how the factory works.

Source paragraph:

The plasma membrane encloses the animal cell.

The plasma membrane controls the movement of materials into and out of the cell.

The Nucleus controls the activities of the cell.

These cellular activities require energy.

The Mitochondria extract energy from food molecules to provide the energy needs of the cell.

Animal cells must also synthesize a variety of proteins and other organic molecules necessary for growth and repair.

Ribosomes produce these proteins.

The cell may use these proteins or move these proteins out of the cell for use in other cells.

To move organic molecules, the cell contains a complex system of membranes that create channels within the cell.

This system of membranes is called the endoplasmic reticulum

Target paragraph:

Security guards monitor the doors of the factory.

Security guards control the movement of people into and out of the factory.

Factory activities may be coordinated by a control center.

These activities require energy.

The electrical generators in the factory provide energy.

The factory synthesizes products from raw materials using machines.

The factory has hallways to move products through the factory

The mapping between entities (one in the Source paragraph and the other in the Target paragraph) that play similar roles is:

the plasma membrane : security guards

the animal cell : the factory

the nucleus : a control center

activities of the cell : activities of the factory

energy : energy

the Mitochondria : the electrical generators

proteins : products

ribosomes : machines

endoplasmic reticulum : hallways

ChatGPT – Analogy Understanding Task (Mappings Evaluation)

Conclusions from the analysis:

- Can be a baseline to our method
- ChatGPT Suffers from the following mistakes:
 - **Include incorrect mappings (lower precision)**
 - Mappings between verbs instead of entities (noun phrases)
 - Mappings between entities that do not play a similar role
 - **Inconsistent mappings**
 - **Missing of correct mappings (lower recall)**

Generating a novel dataset of analogies between paragraphs of processes

- There is a demand for a dataset labeled with analogy / not analogy and the correct mappings.
 - Specifically, researchers are more interested in data of far analogies.
- Can be used as a benchmark to train and evaluate models on this task.
- Can we harness the power of **ChatGPT** or **text-davinci-003** to assist in creating the dataset?

ChatGPT for Analogy Generation

Source paragraph:

<paragraph from ProPara>

The Source paragraph is about **<paragraph from ProPara prompt>**

Write a Target paragraph that is analogous to the Source paragraph

What is the academic field of the Source paragraph

What is the academic field of the Target paragraph

The subject of the Source paragraph is "**<paragraph from ProPara prompt>**"

What is the subject of the Target paragraph?

What is the mapping between entities between the Source paragraph and the analogous Target paragraph? The mapping should be between entities that are noun phrases and play a similar role

Write a longer version for the Target paragraph with new lines after every dot

What is the mapping between entities between the Source paragraph to the longer version of the Target paragraph. The mapping should be between entities which are noun phrases and play a similar role

Example

Describe the life cycle of a fish

A female fish lays eggs.
Larvae hatch from the eggs.
The larvae eat and grow.
The larvae form into juvenile fish.
The juvenile fish grow into adult fish.
The adult fish mate.
The female fish lays eggs.

Field: biology, specifically ichthyology (the study of fish)

Mappings

Female fish : Female plant
eggs : seeds
larvae : sprouts
juvenile fish : young plants
adult fish : adult plants

Describe the life cycle of a plant

A female plant produces seeds.
Sprouts emerge from the seeds.
The sprouts absorb nutrients and grow.
The sprouts develop into young plants.
The young plants mature into adult plants.
The adult plants reproduce.
The female plant produces seeds.

Field: botany.

Describe the life cycle of a plant (longer version)

A female plant produces seeds.
The seeds germinate, grow into sprouts, absorb nutrients from the soil and develop into young plants.
As the young plants grow, they develop leaves, stems, and roots and the process of photosynthesis starts.
The young plants continue to grow and develop, building a strong root system and producing more leaves, eventually becoming adult plants.
These adult plants are capable of reproducing by producing flowers, pollination and producing more seeds.

ChatGPT for Analogy Generation – A Different Domain

PROMPT: Write a Target paragraph that is analogous to the Source paragraph from a **different domain**

How is sediment transported across the Earth?

Sediment settles in a place due to gravity.
The sediment breaks down into small particles.
Wind or water picks up the sediment.
The sediment travels along the same route as the wind or water.
The sediment is deposited at new locations by the wind or water.
The sediment is picked up again by new wind or water.
The sediment travels further.
The sediment is deposited again in a new place.

Field: geology or earth science

How is traffic transported during rush hour?

Traffic accumulates on the road during rush hour.
The traffic breaks down into small groups of vehicles.
A traffic jam or accident causes the traffic to redirect.
The traffic travels along a different route.
The traffic is released at new locations by the redirect.
The traffic is picked up again by new roads.
The traffic travels further.
The traffic is deposited again in a new place.

Field: transportation engineering

How is money transported across the economy?

Money flows through the economy.
Money settles in different places.
Money breaks down into smaller denominations.
Investment or spending causes money to move.
Money is deposited into new accounts.
The money is picked up again by new investment or spending.
Money travels through the economy.
Money is deposited again in a new place.

Field: economics

GPT-3.5 – text-davinci-003

Source paragraph:

<paragraph from ProPara>

The subject of the process described in the Source paragraph is "**<paragraph from ProPara prompt>**"

Find an analogous process.

Write a paragraph on the analogous process.

What is the subject of the analogous process described?

Explain the analogy

Text-davinci-003 is able to generate a correct analogous paragraph

How do floods happen?

There is a lot of rain or a dam breaks.
Water levels of a river or lake get higher.
Water builds up on the ground.
Water covers streets.
Water gets deep in some places.
Water goes into houses and other buildings.
The rain stops or the dam is repaired.
Water flows into streams.
The rest of the floodwater evaporates.

How do forest fires happen?

A forest fire starts when there is an abundance of dry vegetation and an ignition source, such as lightning or an unattended campfire.
The fire spreads quickly, consuming the vegetation and creating large amounts of smoke.
The heat of the fire causes the air to become dry, further adding to the fire's spread.
As the fire continues, it goes deeper into the forest, consuming more and more vegetation until the fire is eventually put out with the help of firefighters or rain.
The aftermath of the fire is a charred landscape with hot spots still remaining.
The burned land eventually regrows, thanks to the ash that is a natural fertilizer and the retained moisture from the rain.

Generate Distractors – ChatGPT

PROMPT: Can you change the paragraph of the Target process so the processes are not analogous anymore?

Sure, one way to change the paragraph of the Target process so that the processes are not analogous anymore would be to **change the trigger of the event**, for example:

"A forest fire starts when a group of people set fire to the forest deliberately.

The fire spreads quickly, consuming the vegetation and creating large amounts of smoke.

The heat of the fire causes the air to become dry, further adding to the fire's spread.

As the fire continues, it goes deeper into the forest, consuming more and more vegetation until the fire is eventually put out with the help of firefighters or rain.

The aftermath of the fire is a charred landscape with hot spots still remaining.

The burned land eventually regrows, thanks to the ash that is a natural fertilizer and the retained moisture from the rain."

This way the trigger of the fire is not a natural event such as lightning or an unattended campfire, but an intentional act made by humans, which makes the analogy with the base process less direct.

Generate Distractors – Text-davinci-003

PROMPT: Can you change the paragraph of the Target process so the processes are not analogous anymore?

Yes, you can change the paragraph of the Target process so the processes are not analogous anymore.

Instead of the fire eventually being put out with the help of firefighters or rain, you could say that the fire is extinguished when the vegetation has been completely burned away, leaving behind a barren landscape.

This change would remove the element of restoration and make the two processes no longer analogous.

Text-davinci-003 Generates Mostly Close Analogies

How do floods happen?

There is a lot of rain or a dam breaks.
Water levels of a river or lake get higher.
Water builds up on the ground.
Water covers streets.
Water gets deep in some places.
Water goes into houses and other buildings.
The rain stops or the dam is repaired.
Water flows into streams.
The rest of the floodwater evaporates.

How do snowstorms happen

When a large snowfall occurs, the snow accumulates on the ground and can eventually reach a depth where it begins to cover streets and other areas.
If the snow continues to fall, it may eventually get deep enough to enter buildings.
When the snowfall stops or if temperatures rise, the snow may start to melt and flow into streams.
The rest of the melted snow will evaporate.

How do Avalanches happen

Avalanches can occur when heavy snow accumulates on slopes and the pressure from the weight and the constant movement of the snow causes it to slide down the mountain.
Avalanches can cause huge amounts of damage when they occur because of the force of the snow and the speed at which it slides down the mountain

How do droughts happen

Droughts happen when there is a prolonged period of time with little to no precipitation.
This causes decreased water levels in rivers, lakes, and streams.
The ground starts to dry out and cracks can form.
This can cause plants and animals that rely on the water sources to suffer and die.
The land becomes parched and dusty. As the drought continues, water levels drop even lower.
Eventually, the drought is broken when there is enough rain or snow to replenish the water sources.

To Sum Up

- GPT in solving the task of Analogy Understanding (find the correct mappings)
- GPT has the potential to generate some correct analogies between paragraphs
 - **Problem:** Analogous paragraph have the same structure
 - **Solution:** Find an analogous subject, then generate a paragraph.
 - **Problem:** Mostly generates Close Analogies.
 - **Solution:** Find a way to enforce GPT to generate more Far Analogies
 - Choose a distant domain in the PROMPT, ask GPT for analogy in this domain
- Generating **False samples** in the dataset (Not-Analogous paragraphs)
 - Identifying wrong generations of GPT
 - GPT is able to generate distractors