



Robot probes city grid

Description

Sentences are of course sequences of tokens (or words), and one of the major tasks of large language models (LLMs) is to predict the plausible continuation of sentences from some start condition, such as a prompt in a chatbot dialogue box, taking account of the previous flow of tokens within a given context window.

Several LLMs such as ChatGPT4 now combine their linguistic acumen with image feature detection to exhibit a kind of spatial reasoning, a precursor to true multi-modal AI.

I loaded a photograph to ChatGPT and told it, "I want to swap the positions of the jug and the miniature teapot. I only have one hand free."



It correctly instructed me in this simple task.

Given the constraints, here's a methodical approach:

1. With your free hand, pick up the jug and place it on the table next to the miniature teapot.
2. Now, pick up the miniature teapot and place it where the jug was.
3. Finally, move the jug to where the teapot originally was.

I took another picture and typed: "Thanks. Now how do I progress to this arrangement?"

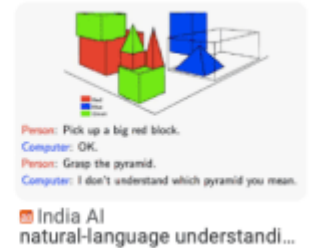
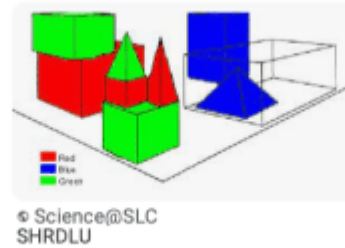


Ah, the classic teapot-inside-a-jug maneuver! Here's a step-by-step for the one-handed academic:

1. Gently lift the miniature teapot out of the jug and place it on the table.
2. Now, pick up the jug and set it onto the coaster or mat where the teapot was.
3. Lastly, lift the miniature teapot and nestle it inside the jug.

Perhaps my initial instruction to ChatGPT was ambiguous. The sequence isn't quite right, but the system commended my table-top exercise: "Your artistic arrangement evokes the idea of containment and whimsy - perfect for a midday mental breather. Enjoy the view!"

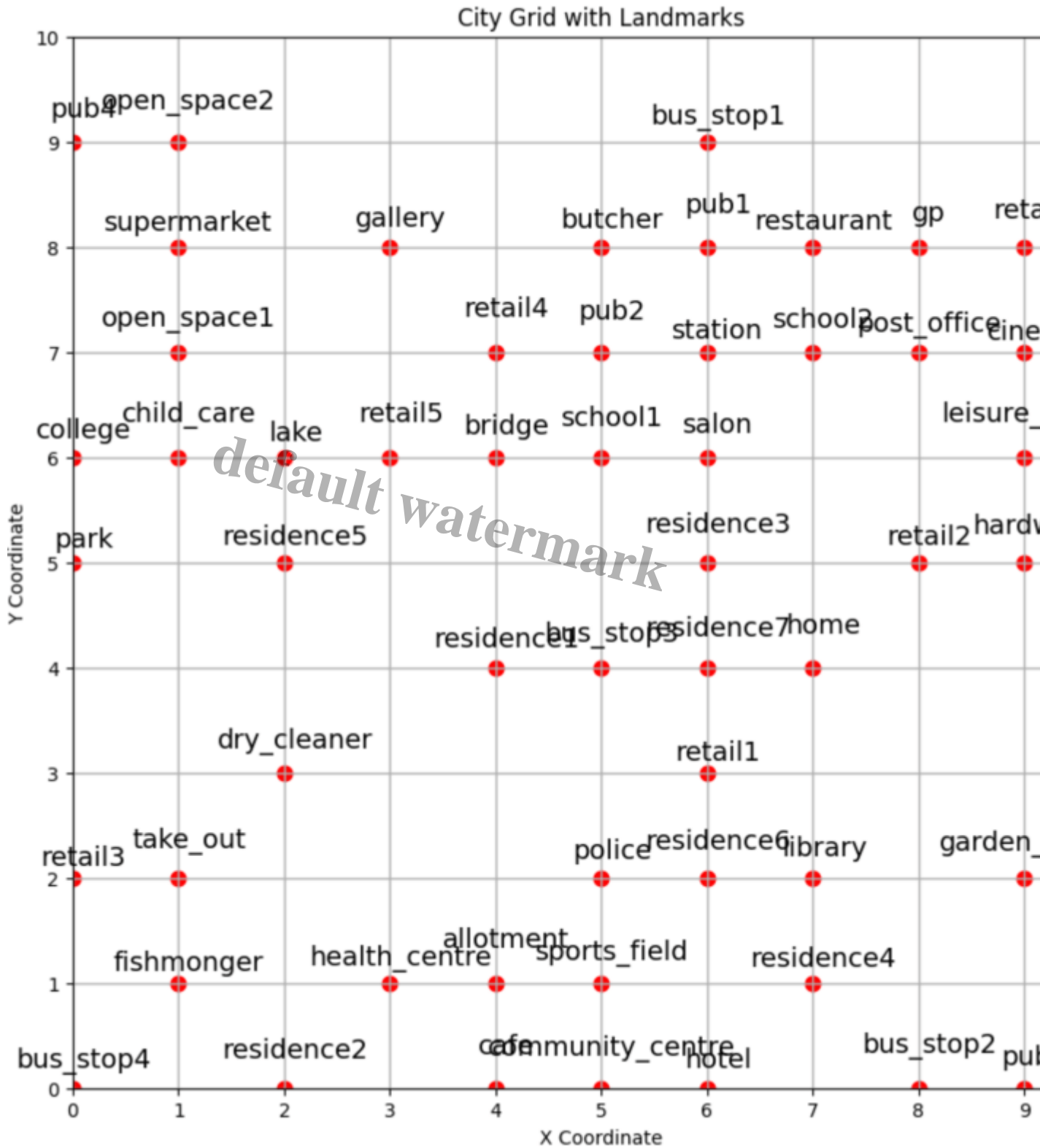
This exercise and related attempts at moving objects in photographic space reminded me of the seminal AI language work by Terry Winograd. He invented a program called SHRDLU for planning the movement of blocks as if operated by a robot arm. As described in his 1971 PhD thesis, the program responded to natural language commands and answered questions about the arrangement of blocks. His approach drew on rule-based and logic-based programming, which is a very different approach to AI than neural network-based LLMs. Here's a Google search on some of the images from that work.



Planning

Sequencing is a major aspect of language. It's also key in any domain of planning and even *time series forecasting*. The latter usually involves statistical methods to extrapolate something like market trends. On the other hand, AI *planning* devises a series of steps to accomplish some goal, relevant to moving objects around a construction site, planning a production process, controlling a self-drive car or planning a journey.

To extricate ChatGPT from its rich literary tendencies I gave it an image of a banal urban grid plan generated by algorithm, but uploaded to ChatGPT as a pixel image (i.e. without vector information).



I asked ChatGPT to describe the image. It projected much more into the image than went into the random disposition of feature words on a grid.

This map represents a grid-like layout of a city or urban space, delineated by X and Y coordinates, spanning from 0 to 10 on both axes. The map is titled "City Grid with Landmarks".

Landmarks and facilities are represented as red points with accompanying labels. Here's a categorized description:

Recreation & Open Spaces:

- Open_space1 and open_space2 appear to be recreational or green areas.
- There's a lake located near the center.
- A sports_field is located towards the bottom.

Transportation:

- There are four bus stops labeled: bus_stop1, bus_stop2, bus_stop3, and bus_stop4, scattered throughout the grid.

Residential Areas:

- There are residences labeled from residence1 to residence7. Additionally, there's a point labeled "home".

Retail & Commercial:

- Several retail points are spread across the map, labeled from retail1 to retail6.
- There are pubs labeled as pub1, pub2, and pub3.
- Other commercial points include a supermarket, gallery, butcher, restaurant, salon, dry_cleaner, fishmonger, and take_out.

Educational & Childcare:

- There's a school labeled as "school1" and another just as "school".
- A college is located on the far left.
- Child care is located near the college.

Health & Wellness:

- A general practitioner (gp) and a health_centre are present.

Civic & Public Services:

- Points labeled as police, post_office, and library represent essential civic services.
- There's a community_centre towards the bottom.

Leisure & Entertainment:

- A cinema and leisure_centre are located towards the right.

- A garden_centre is towards the bottom-right.

Other Landmarks:

- The grid also contains other landmarks like a bridge, hardware store, hotel, and an allotment.

Though I didn't ask, it even proffered an evaluation: "The layout seems to provide a mix of residential, commercial, recreational, and essential services, which gives an impression of a well-balanced urban space. It is quite a dynamic representation of city elements!"

I asked the system about grid coordinates: "What is at (6, 9)?" It answered correctly: "at the coordinate (6,9) on the map, there is a bus_stop1 located."

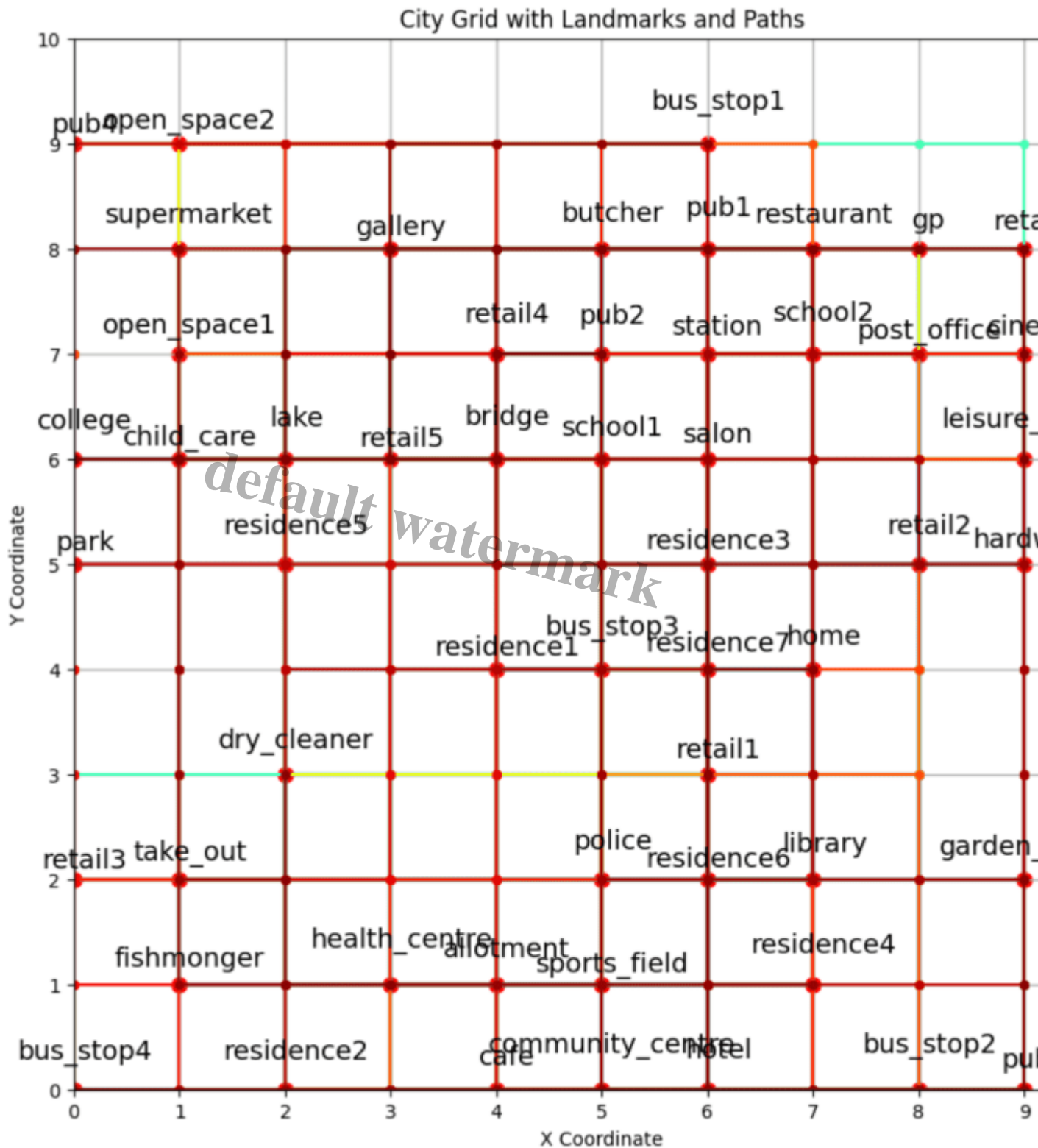
I asked if following just the grid lines I can get from there to the lake without passing through other landmarks? It delivered a series of steps that looked plausible, but didn't precisely match the landmark positions on the map.

Predicting plausible paths

In my previous post ([AI learns ABC](#)) I investigated GPT-style language models as a means of predicting sequences in a more abstract domain, as in the case of deriving the order of the entire alphabet from fragments of the alphabet.

With help from ChatGPT4 I since devised a program in Python to generate shortest paths between any two randomly selected landmarks on my map here, and to generate the sequence of landmarks that a path would encounter. So that's a series of landmark lists (500) of varying lengths, a bit like the alphabet fragments from my previous post. There was no neural network programming in this aspect of the exercise. Here's a sample of the paths.

- supermarket lake residence5 dry_cleaner residence2
- health_centre allotment sports_field residence6 retail1
- school1 bus_stop3 police sports_field community_centre cafe residence2 bus_stop4
- post_office gp restaurant
- supermarket retail4 bridge residence1 allotment cafe
- garden_centre library residence6 police take_out retail3
- cafe residence2 fishmonger take_out park college open_space1
- cinema post_office school2 station salon residence3 residence7 retail1 residence6
- lake retail5 bridge residence1 police
- school2 home library residence6



The paths of course overlap in the map. Trained on those paths, my AI model was then able to make up plausible paths from any landmark. I show some of the output below, starting with the gp.

gp restaurant school2 station pub2 retail4 residence3 bridge bus_stop3 pub1 butcher pub1
butcher open_spacel1 residence6 hotel police residence6 library residence4 sports_field

I had hoped for a single continuous path longer than any of the training examples, a bit like the derivation of the entire alphabet in the previous post. Considering the discontinuities, there's no guarantee that the model has synthesised new path sequences that are not in the training corpus without testing for that.

References

- Coyne, R. (1985). Knowledge-based planning systems and design: a review. • *Architectural Science Review* **28**(4): pp. 95-103.
- Peixeiro, M. (2023). TimeGPT: The First Foundation Model for Time Series Forecasting. • Towards Data Science 25 October. Retrieved 27 October 2023, from <https://towardsdatascience.com/timegpt-the-first-foundation-model-for-time-series-forecasting-bf0a75e63b3a>.
- Schank, R. C. and R. P. Abelson (1977). *Scripts, Plans, Goals and Understanding: An Inquiry into Human Knowledge Structures*. Hillsdale, N.J.: Erlbaum.
- Suchman, L. A. (1987). *Plans and situated actions: the problem of human-machine communication*. Cambridge: Cambridge University Press.
- Winograd, T. (1971). *PhD thesis: Procedures as a representation for data in a computer program for understanding natural language*. MA: MIT.

Note

- Featured image from Bing Dall-e prompted with "Apocalyptic teapot robot probes city grid photorealistic."

Category

1. Artificial Intelligence

Tags

1. Planning
2. sequence

Date Created

October 28, 2023

Author

rcoyne99