New terms/technology used

- Hindsight relabeling
- Visual navigation model (VNM: ViNG)

Goal

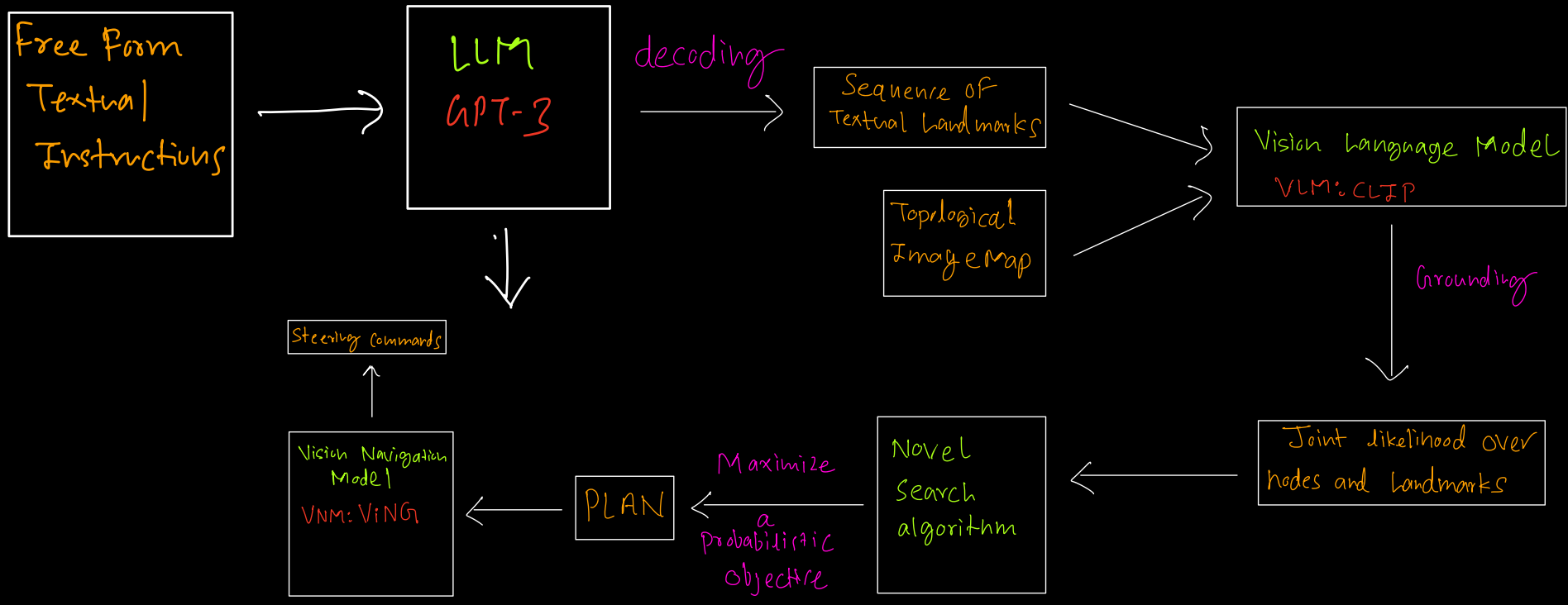
Enabling a self supervised system for robotic navigation to execute natural language instructions by leveraging the capabilities of pre-trained models without any user annotated navigational Data.

Large Model Navigation
LM-NAV

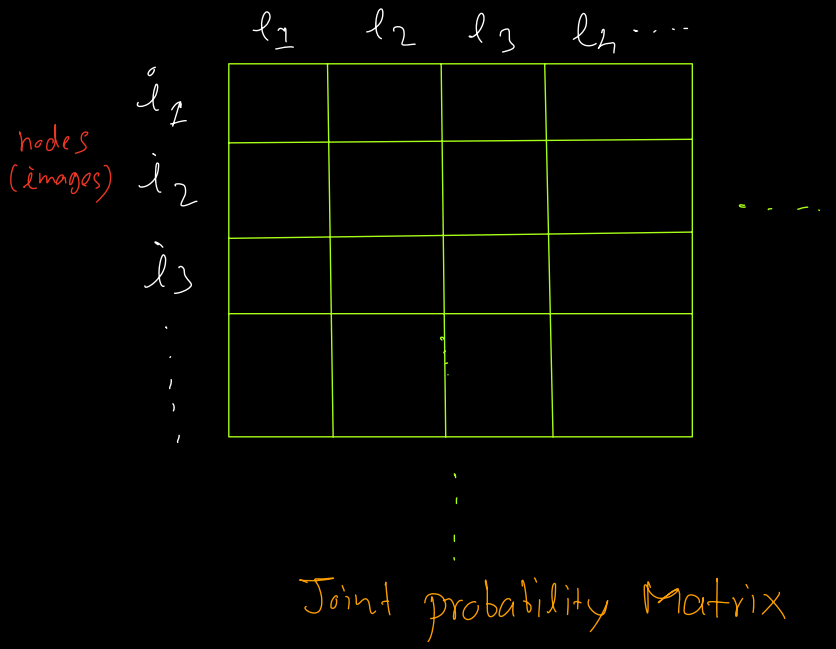
LLM VLM VNM

Issues with previous approaches

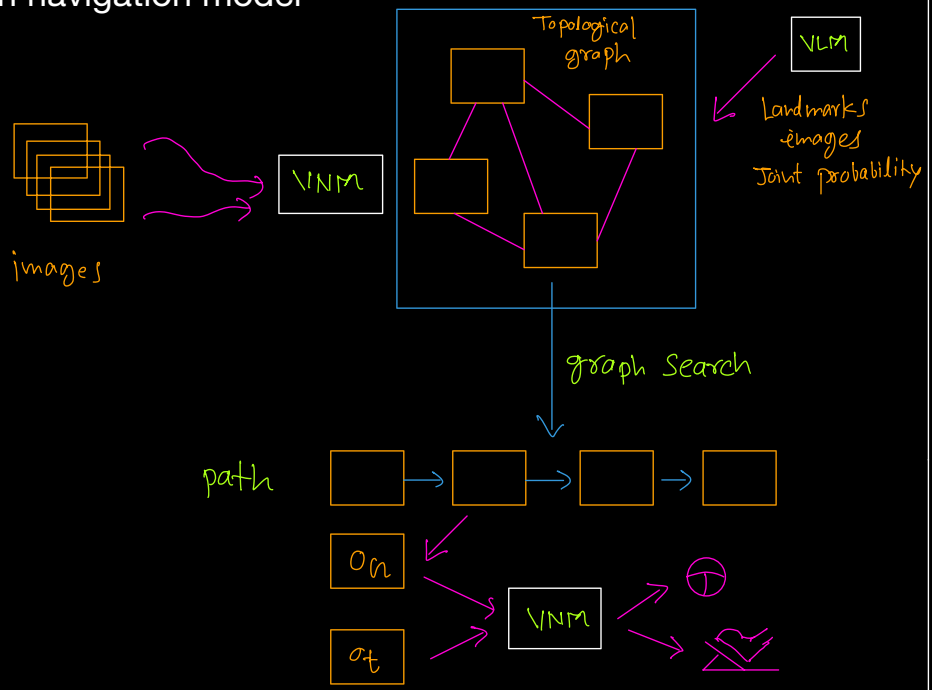
- Learning from trajectories annotated with language instructions (1-5)
- Clunky goal specification using images or locations.
- Tend to operate on structured data (15 - 19).
- Requires photo realistic simulators (22 - 30), (1,3 - 5). (46-49)
- Language annotation for robot experience; access to low level skills that can follow textual commands (31 - 44) (13)
- Similar work
  - Self supervised methods for navigation(6 - II) (51)
  - Automatically generate labels using on board sensors and hindsight relabeling.
  - Still unable to parse high level instructions like free form text.



Vision language model (clip)



Vision navigation model



Landmark generation

- prompt engineering is used to generate landmarks
- In context learning examples are given so that the model can reason about more complex examples
- Prompt engineering is also used for VLM