**GOAL-DRIVEN REASONING**

What do we need for goal-driven navigation?

**VISION-AND-LANGUAGE NAVIGATION (VLN)**

- **Challenges:**
  - Association between language and images
  - No map or GPS

**INSTRUCTION GROUNDING**

- We expect textual attention to match with agent’s actions — ground instruction to actions.
- Incorrect grounding from existing work

**SELF-MONITORING AGENT**

- **Visual grounding:**
  - Progress monitoring
  - Grounding qualitative analysis
  - Experiments & analysis

**END-TO-END LEARNED BACKTRACKING AGENT**

- **Backtracking framework**
  - Progress marker: which way to go?
  - Rollback: forward or rollback?

**EXPERIMENTS & ANALYSIS**

- Navigation Error (NE): mean of the shortest path distance between the agent’s final position and the goal location.
- Success Rate (SR): the percentage of final positions less than 2m away from the goal location.
- Success rate weighted by Path Length (SPL): SPL weighted with trajectory lengths.

**STATE-THIRD PARTY USE**

- Information about the use of Georgia Tech’s facilities for non-Georgia Tech purposes.
- Use of images and graphics from various sources.
- Collaboration and contributions from multiple organizations and individuals.

**Acknowledgment:**

- Images and content from various sources and contributors.
- Support from multiple institutions and partners.

---

**Self-Monitoring Navigation Agent via Auxiliary Progress Estimation**

Chih-Yao Ma, Jiase Lu, Zuxuan Wu, Ghasan AlRegib, Zuolet Kira, Richard Socher, and Caiming Xiong

1Georgia Institute of Technology, University of Maryland, College Park, 2Salesforce Research

[GitHub Link 1](https://github.com/chihyaoma/selfmonitoring-agent)

[GitHub Link 2](https://github.com/chihyaoma/regretful-agent)