

Knowledge Synthesis in Dynamic Human-Swarm Interactions using LLMs

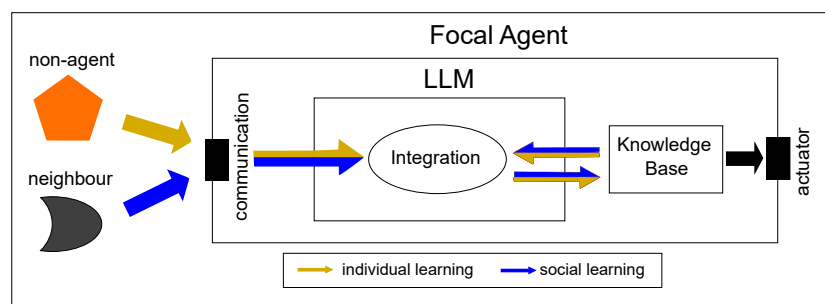
Boubacar Ballo¹, Absera Yihunie¹, Lilly Schwarzenbach²[0009-0009-1143-4369], Hanan Salam¹[0000-0001-6971-5264], and Eliseo Ferrante^{1,2}[0000-0002-2213-8356]

¹ Department of Computer Science, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates

² Department of Computer Science, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

Collective exploration and understanding of an unknown environment remains an ongoing challenge in swarm robotics. Constructing rich descriptions of the environment is often made difficult by the models' reliance on simple, non-semantic signals and their cognitive simplicity [1]. This problem is further aggravated when humans interact with swarms in the environment: human sources represent dynamic pieces of information appearing and disappearing as time passes, as humans may dynamically leave and join the scene or generate novel pieces of information dynamically [4].

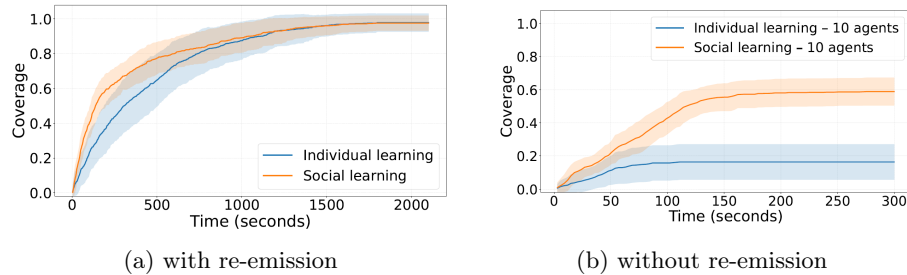
To overcome these issues, we propose to integrate Large Language Models (LLMs) in the control loop of the agents. Agents move independently through the environment. They start out without any prior knowledge, i.e. their summary or Knowledge Base is empty. As an agent encounters either another agent or a third party actor, e.g. a human interlocutor, it receives the information that interlocutor possesses. The LLM integrates the received information into the existing knowledge base while respecting a maximum size of the Knowledge Base (250 tokens) [3]. Once the Knowledge Base is updated, it shares the contents with other agents in its vicinity.



To measure the success of our framework, we use Coverage Score as our metric, where σ represents the agent summary, C the ground truth description of the entire environment, n the number of claims in C , and $\hat{\ell}_{j,i}$ is the entailment decision per claim.

$$\text{Cov}(\sigma, C) = \frac{1}{n} \sum_{j=1}^n \mathbf{1} \left[\bigvee_{i=1}^m [\hat{\ell}_{j,i} = 1] \right] \in [0, 1]$$

Entailment can be defined as one sentence implying the other [2]. In our case, this means that at least one sentence in the Knowledge Base constitutes a premise that unambiguously entails a piece of information in the ground truth.



In our experiments, 10 agents perform a random walk in a 2D bounded arena without any prior knowledge. Any knowledge they gain about the environment has to come either from an informant or another agent. We consider two scenarios: 1) individual learning only, 2) individual and social learning. In the first case, the agents don't share information with other members of the swarm, only with the informants. In the second case, they communicate with both the third party actors and other swarm agents.

Our preliminary results show that both individual agents and social learners are able to capture the entire information in the environment given enough time as long as the information is presented repeatedly. When there is no re-emission, i.e. information is presented once and when it disappears, it is lost forever, neither condition achieves full coverage. However, social learner perform much better as chances of encountering the information in time are much higher. Those who have missed it can learn about it from others. These results demonstrate the potential of our framework for information aggregation in dynamic environments, making this an interesting direction for future work.

References

1. Hamann, H.: Swarm robotics: A formal approach, vol. 221. Springer (2018)
2. Korman, D.Z., Mack, E., Jett, J., Renear, A.H.: Defining textual entailment. *Journal of the Association for Information Science and Technology* **69**(6), 763–772 (2018)
3. Wan, F., Huang, X., Cai, D., Quan, X., Bi, W., Shi, S.: Knowledge Fusion of Large Language Models (Jan 2024)
4. Zhen, Z., Chen, Y., Wen, L., Han, B.: An intelligent cooperative mission planning scheme of uav swarm in uncertain dynamic environment. *Aerospace Science and Technology* **100**, 105826 (2020)