The Legibility Methods in Agent Systems

Siyuan Li*

School of Computer Science & Technology, Huazhong University of Science and Technology, Wuhan, 10487, China

Abstract. To tackle problems in the domains of human-machine collaboration and multiagent cooperation (e.g., multiagent sequential decision-making, path planning, and navigation), "legibility" — the agent's ability to convey its intentions through its behavior — can provides a reliable foundation for efficient and seamless collaboration. This paper systematically reviews the most recent research on legibility in agent-based scenarios, concentrating on its theoretical foundations, core methodologies, evaluation metrics, and future directions. The analysis demonstrates that existing methods (such as reward shaping, inverse reinforcement learning, and network flow optimization) have increased the efficiency of collaboration as well as the speed and success rate of intention inference. These methods hold both theoretical significance and practical value in real-world scenarios like warehouse management. But there are still issues with scalability in dynamic target scenarios and adaptability to partially observable environments. Future research could look at adaptive legibility techniques combined with inverse reinforcement learning as well as legibility coordination mechanisms in multiagent dynamic interactions.

1 Introduction

In human-agent collaboration scenarios like navigation and warehouse management, as well as in multiagent systems like autonomous driving and urban traffic scheduling, other agents or humans can only optimize their decisions in a targeted way if they quickly comprehend an agent's intent. However, traditional agents frequently ignore the transmission efficiency and success rate of their own intents in favor of concentrating only on maximizing individual efficiency [1]. Collaborative failure can result from mistakes in determining an agent's intent when their behavior is ambiguous.

The concept of legibility originates from the field of robot motion planning, referring to an agent's ability to enable observers (humans or other agents) to quickly and accurately infer its goals through behavioral design [2]. Early studies on legibility in the domain of human-robot interaction mainly concentrated on how devices, like robot arms, make legible movements under various viewpoints and different interference conditions [3, 4]. Collaboration can go more smoothly when legibility-related techniques are used to help humans or other agents better understand the agent's intent.

* Corresponding author's email: bikuhiku@hust.edu.cn

Legibility research has moved from single-agent motion to multiagent decision-making scenarios in recent years. In the Lead-Follow Maze (LFM) task, for instance, the leading agent created legible trajectories that allowed followers to identify the target 50% of steps ahead of time and reduced task completion time by 40% [1]; in warehouse robot collaboration, the use of legible paths reduced the time it took for human supervisors to deduce agents' intentions by 14% [5].

The current trajectories of legibility are extensive. However, review articles on the subject seldom address recent progress, particularly in the domain of multiagent systems. This paper will systematically review the research status of legibility in agent systems from four aspects: theoretical foundations, core methods, evaluation metrics, and future directions, aiming to provide a comprehensive overview for researchers in this field.

2 The Theoretical Foundations of Legibility

2.1 The definition of legibility

The essence of legibility lies in minimizing an observer's uncertainty about an agent's objectives through behavioral design. Dragan et al. first defined legibility mathematically [6]: For a trajectory segment $\xi_{S\to Q}$ (from initial state S to state Q at time t), legibility manifests as the probability that the observer infers the true objective G^* from $\xi_{S\to Q}$, i.e.,

$$L(\xi_{S\to Q}) = P(G^*|\xi_{S\to Q}) \tag{1}$$

Capelli et al. further extended this definition in the context of multi-robot systems, proposing that a multi-robot system is legible if and only if users can understand its "spatial goal" and/or "coordination goal" solely by observing its motion patterns through implicit communication [7].

The spatial aim refers to the general movement direction of the multi-robot system, or essentially, “where the robot team is going”. It is a macro-level description of the system's movement endpoint or trend, not the local trajectory of individual robots, and must be reflected in the constant trend of collective movement. In contrast, the coordination aim relates to the relative mobility modes and aggregation patterns among robots, or essentially, “what the robot team is doing”. It is determined by multi-robot systems' distributed coordination control techniques (e.g., establishing fixed formations, converging to a common point, or covering a certain area) and represents the system's collective behavioral logic used to complete tasks.

2.2 Quantification methods for legibility

In the process of further applying the formal definition of legibility to practical implementation frameworks, it is necessary to adopt methods to quantify the impact of legibility on systems. Existing core models can be categorized into three main types.

The first type leverages two special cases of f-divergence: Kullback-Leibler (KL) divergence and total variation divergence. In the Legible MDP proposed by Miura et al. [8], the idea of using negative KL divergence to shape the legibility reward function was first envisioned. Specifically, for an observer’s estimated belief b_t and the belief b^* representing the true target, their KL divergence distance is defined as:

$$dist(b_t, b^*) = D_{KL}(b_t||b^*) = \sum_{n=1}^{N} b_t(r_n)\, log\, \frac{b_t(r_n)}{b^*(r_n)} \tag{2}$$

Where rn denotes the reward function for the n-th target.

In the PoLMDP model proposed by Faria et al. [3], total variation divergence is used to replace KL divergence. The distance is calculated as:

$$dist(b_t, b^*) = \frac{1}{2} D_{TV}(b_t, b^*) = \frac{1}{2} \sum_{\theta \in \Theta} |b_t(\theta) - b^*(\theta)| \tag{3}$$

Where Θ denotes the set of all possible objectives of the agent.

The second category is based on information theory, which uses metrics such as information entropy and cross-entropy to represent the gap between the estimated target and the actual target. In the IGL (Information Gain based Legibility) model proposed by Liu et al. [9], the legibility reward function is designed as:

$$L = \mathcal{H}(\tau_t) - \mathcal{H}(\tau_t, a) \tag{4}$$

Where τ_t denotes the agent's action trajectory at time t, and a represents the action selected by the agent at time t. The definitions of $\mathcal{H}(\tau_t)$(information entropy of the trajectory τ_t) and $\mathcal{H}(\tau_t, a)$(information entropy after action a is executed following τ_t) are as follows:

$$\mathcal{H}(\tau_t) = -\sum_{g \in \mathcal{G}} b(g|\tau_t)\, log\, b(g|\tau_t) \tag{5}$$

$$\mathcal{H}(\tau_t, a) = \sum_{s_{t+1} \in S} T^o(s_{t+1} s_t, a) \mathcal{H}(\tau_t \circ s_{t+1}) \tag{6}$$

Here, $b(g|\tau_t)$ is the conditional probability that the observer predicts the agent's goal g based on the trajectory τ_t, $\mathcal{G}$ is the set of all possible goals of the agent, and T^o represents the state transition probability in the observer's cognition.

Persiani et al. [10], on the other hand, attempted to construct the agent's true policy distribution $P_R(\Pi)$ and the observer's inferred policy distribution $P_R^H(\Pi|s, a)$, and used their cross-entropy to represent the gap between the two distributions:

$$D\big(P_R(\Pi), P_R^H(\Pi|s, a)\big) = -\, log\, P_R^H(\pi_R|a, s) \tag{7}$$

The third category leverages flow networks. For pathfinding scenarios, Bernardini et al. (2024) [11] transformed legibility into a "minimum path overlap" problem. In partially observable environments, they construct an s-Legibility graph $G_{po}^{(s)}$, converting legibility verification into a maximum flow problem. They conclude that a system satisfies s-legibility (i.e., an observer needs at most s observations to infer the target) if and only if there exist D edge-disjoint paths in $G_{po}^{(s)}$.

3 Methods and Models of Legibility

3.1 Methods integrating reinforcement learning

Methods integrating reinforcement learning primarily achieve legibility by shaping legibility reward functions, and they offer strong overall scalability. For example, the MAAL framework proposed by Liu et al. [1] can be well integrated into existing multi-agent reinforcement learning algorithms. Currently, most research on legibility focuses on this aspect.

The PoLMDP model is quite similar to the Legible MDP model. However, Faria et al. [5] proposed a new legibility reward function and adopted total variation distance instead of Kullback-Leibler (KL) divergence. This avoids reliance on the agent's interaction history, allowing the model to be solved using traditional MDP methods. As a result, it achieves better time complexity in multi-agent sequential decision-making problems (reducing from the PSPACE-complete complexity of Legible MDPs to the PSPACE complexity of PoLMDPs).

In online guessing games, compared with methods that solely pursue optimal policies, PoLMDP increased the success rate of human subjects in guessing the agent's goal from 70% to 85%, and reduced the average inference time from 18.22 seconds to 15.67 seconds [5].

The MAAL framework is built upon the Legible Interactive POMDP (LI-POMDP)—an extension of the Interactive POMDP (I-POMDP). The framework first conducts belief modeling: each agent A^i infers two types of beliefs via Bayesian updates: Fisrt is the peer goal belief $\widehat{b^{-i}}$, which represents A^i's prediction of other agents' goals; second is the peer's belief about its own goal $\widehat{b^{-i}}$(derived through recursive reasoning, i.e., A^i infers how other agents perceive its own goals).

Subsequently, an intrinsic reward is designed based on the KL divergence gain (ΔD_{KL}), which balances task rewards (e.g., rewards for reaching the target) and legibility rewards. Finally, concatenated features—including the environmental state, the agent's own goal, and peer beliefs—are fed into the policy network, and the policy is updated using algorithms such as Q-Learning or DQN. Compared to the PoLMDP framework, MAAL is applicable to more complex scenarios and has obtained a formal proof of completeness.

For example, in the Lead-Follow Maze scenario, the MAAL+QL framework (with $\beta = 0.01$) achieved an 89% increase in average reward compared to traditional Independent Q-Learning (IQL), and its Policy Correctness Ratio (PCR, i.e., target inference accuracy) reached 98%—significantly outperforming methods like VDN and QMIX. In the SMAClite scenario, after 10,000 training episodes, the win rate fluctuation of MAAL+DQN was less than 0.15, which is superior to VDN (with a fluctuation greater than 0.5) [1].

Additionally, based on inverse reinforcement learning (IRL), Zeng et al. proposed an IRL framework that integrates non-negative matrix factorization (NMF) [12]. This framework attempts to simultaneously learn the reward function and the legibility reward function in a reverse manner using a set of "expert trajectories."The framework first performs single-objective IRL initialization: it learns the initial reward function R_{init} from the rational trajectory set D_r and the initial legibility function L_{init} from the legible trajectory set D_l, respectively. Subsequently, using an initial weight ω, it constructs a joint objective function $\mathcal{R} = \omega(R) + (1-\omega)L$. Finally, multiple groups of joint functions are combined to form matrix $\mathcal{V}$, which is decomposed via NMF into a basis matrix $\mathcal{W}$ and a weight matrix $\mathcal{H}$: The basis matrix outputs the ultimately optimized reward function and legibility function; The weight matrix outputs the dynamically adjusted weight ω^*.

3.2 Methods integrating flow networks

To address the legibility problem in multi-agent pathfinding, Bernardini et al. propose a network flow-based solution for the Partially Observable-Legibility Problem (PO-LP) [11], with the following key steps:

First, it performs graph transformation: converting the original environment graph into an s-Legibility graph $G_{po}^{(s)}$. Redundancy is reduced by merging observably equivalent subpaths—for example, subpaths that are indistinguishable to the observer (such as paths composed of different unobservable edges but generating identical observation sequences).

Next, it verifies s-legibility via a maximum flow algorithm: checking whether there exist $|D|$) edge-disjoint paths in $G_{po}^{(s)}$(where $|D|$ denotes the number of potential destinations for the agents). If such paths exist, the system is proven to satisfy s-legibility—meaning the observer only needs at most s observations to uniquely infer an agent's destination.

Finally, under budget constraints, it optimizes the trade-off between legibility and efficiency: using a minimum cost flow algorithm to minimize the total cost of the paths (e.g., path length or energy consumption). This step ensures that while the paths remain sufficiently legible, their practical efficiency (in terms of resource usage) is also optimized, achieving a balanced "legibility-efficiency" performance.

4 Typical Experimental Scenarios and Evaluation Metrics

Currently, typical experimental scenarios in this field are mainly divided into human-machine collaboration scenarios and multi-agent collaboration scenarios (Table 1).

Table 1. Typical Scenarios for Multi-Agent Collaboration

Environment Type	Discrete Space	Continuous Space
Discrete Actions	Lead-Follow Maze Grid World Deep Q-Network Tunnel Experiment	StarCraft Multi-Agent Challenge Autonomous Driving at Pedestrian Crosswalks
Discrete Actions		Particle Simple Navigation

For multi-agent collaboration, the main evaluation metrics fall into two categories: one is metrics for overall efficiency, such as task success rate, average time consumption, win rate stability, convergence speed, and path cost; the other is metrics related to legibility, such as prediction correctness ratio (PCR), prediction time ratio (PTR), and legibility delay. Metrics like cumulative reward are a synthesis of these two categories.

Human-machine collaboration scenarios lack relatively unified experimental settings and are often conducted among people of different age groups without relevant technical backgrounds based on needs—for example, testing subjects' prediction of the movement goals of a robotic arm, or testing subjects' prediction of the spatial goals and coordination goals of an agent group. The main evaluation metrics of these experiments include the time and accuracy of subjects' inference of the agents' intentions, as well as subjects' confidence scores in the agents.

5 Challenges and Future Directions

Through the review and analysis of methods related to agent behavior legibility above, it is found that although current research has extended from transparent environments to partially observable ones, there are still difficulties in handling large-scale or relatively complex tasks,

as well as scenarios with a high proportion of unobservable environments. Additionally, there is a lack of adaptation to scenarios with dynamic obstacles and dynamic targets. This creates a gap between legibility methods and practical applications, making them unable to cope with relatively complex and time-sensitive tasks in the real world.

In future research, efforts can be made to explore the integration of frameworks such as MAAL, IRL, and PoLMDP with task decomposition methods to expand the application boundaries of legibility. Meanwhile, to address more complex problems, federated learning can be used to achieve distributed legibility strategy optimization, avoiding resource competition caused by multiple agents pursuing legibility simultaneously. For edge devices, knowledge distillation can be applied. Moreover, combining variational inference with reinforcement learning can help design adaptive legibility strategies in dynamic target scenarios, such as Transformer-based belief prediction prediction networks.

6 Conclusion

This paper focuses on "agent behavior legibility" and provides a systematic review of this domain. The definition of legibility is explained theoretically as "reducing observers' uncertainty in inferring goals through behavioral design," its evolution from single-agent to multi-agent scenarios is traced, and three quantitative approaches based on f-divergence, information theory, and flow networks are categorized, clearly presenting the theoretical logic and evaluation dimensions.

In terms of methodology, it focuses on analyzing mainstream approaches that combine legibility with reinforcement learning or network flow: the former optimizes path legibility through graph transformation and flow algorithms, supplementing the use of inverse reinforcement learning in legibility function learning to enhance the methodological system; the latter balances legibility and task efficiency through reward shaping, with typical frameworks adapted to multi-agent reinforcement learning algorithms.

Reference

1. Y. Liu, Y. Pan, Y. Zeng, B. Ma, P. Doshi, Active legibility in multiagent reinforcement learning. Artif. Intell. 346, 104357 (2025)
2. A.D. Dragan, K.C.T. Lee, S.S. Srinivasa, Legibility and predictability of robot motion. in Proceedings of the 8th ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI 2013), Tokyo, Japan, 301–308 (2013)
3. S. Nikolaidis, A. Dragan, S. Srinivasa, Viewpoint-based legibility optimization. in Proceedings of the 11th ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI 2016), Christchurch, New Zealand, 271–278 (2016)
4. B. Busch, J. Grizou, M. Lopes, F. Stulp, Learning legible motion from human–robot interactions. Int. J. Soc. Robot. 9(3-4), 517–530 (2017)
5. M. Faria, F.S. Melo, A. Paiva, “Guess what I'm doing”: Extending legibility to sequential decision tasks. Artif. Intell. 330, 104107 (2024)
6. A.D. Dragan, S. Bauman, J. Forlizzi, S.S. Srinivasa, Effects of robot motion on human-robot collaboration. in Proceedings of the 10th ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI 2015), 51–58 (2015)
7. B. Capelli, M. Santos, L. Sabattini, Towards the legibility of multi-robot systems. J. Hum.-Robot Interact. 13(2), Article 21 (2024)
8. S. Miura, A.L. Cohen, S. Zilberstein, Maximizing legibility in stochastic environments. in 2021 30th IEEE Int. Conf. on Robot & Human Interactive Communication (RO-

MAN), 1053–1059 (2021)

9. Y. Liu, Y. Zeng, B. Ma, Y. Pan, H. Gao, X. Huang, Improvement and evaluation of the policy legibility in reinforcement learning. in Proceedings of the 22nd Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2023), 3044–3046 (2023)
10. M. Persiani, T. Hellström, Policy regularization for legible behavior. Neural Comput. Appl. 35, 16781–16790 (2023)
11. S. Bernardini, F. Fagnani, A. Neacsu, S. Franco, Optimizing pathfinding for goal legibility and recognition in cooperative partially observable environments. Artif. Intell. 333, 104148 (2024)
12. B. Zeng, Y. Pan, J. Tang, Y. Zeng, Automate legibility through inverse reinforcement learning. ACM Trans. Autonom. Adapt. Syst. (2025)