Active-Inference-Based Evidence Collection Method for Disinformation Verification

Nichika Maeno[1], Shin'ichi Arakawa[1], Masayuki Murata[1], and Daichi Kominami[1]

***Abstract*— Social media platforms have emerged as vital channels for situational awareness during disasters, facilitating the rapid sharing of real-time updates. However, these platforms can also facilitate the spread of disinformation. For disaster-related claims that can be corroborated in the physical world, estimating on-site conditions using observations from IoT sensors can provide strong supporting evidence for disinformation verification. Nevertheless, collecting data from all available information sources is undesirable in terms of system load; therefore, minimizing access to information sources is important. In this paper, we propose an active-inference-based information collection method and implement it as an evidence collection service within a disinformation countermeasure platform. Active inference provides a framework in which an agent estimates real-world states from observations and selects actions to reduce uncertainty. The agent selectively acquires observations and updates its state estimates, and this iterative process enables efficient evidence collection for disinformation verification. Simulation results under a real-data-driven river-flooding scenario show that the proposed method generally obtains evidence useful for disinformation verification with fewer source accesses than the baseline methods.**

Index Terms – Disinformation, Active inference, Free energy principle, Information gain, Posterior distribution, River flooding

I. INTRODUCTION

In recent years, the widespread adoption of services such as social media has made the Internet an increasingly important means of information gathering [1]. In particular, during natural disasters such as earthquakes and floods, social media can facilitate the rapid spread of crucial information, helping people grasp rapidly changing conditions. However, these platforms can also spread false and misleading content, which may increase public confusion and impede rescue operations [2].

Among disaster-related disinformation, unverified claims about real-world conditions can be assessed by identifying discrepancies between online assertions and the actual situation. In such cases, estimating on-site conditions from observations obtained by IoT sensors, such as cameras and water-level gauges, can provide strong supporting evidence. Nevertheless, there are a large number of fixed cameras and IoT sensors, and exhaustive data collection from these sensors may impose a significant load on data-providing services. For this reason, selective acquisition of the necessary information is required to reduce access frequency.

[1]Nichika Maeno, Shin'ichi Arakawa, Masayuki Murata, and Daichi Kominami are with Graduate School of Information Science and Technology, The University of Osaka, Osaka, Japan, e-mail: {n-maeno, arakawa, murata, d-kominami}@ist.osaka-u.ac.jp

Human information collection for disinformation verification typically does not rely on exhaustive access to all available information sources. In real-world disaster situations, human decision makers often possess prior knowledge, such as hazard maps and other publicly available contextual information. Using this knowledge and the information obtained so far, they select information sources expected to be effective for disinformation verification. They then access the selected sources and estimate the actual situation from the acquired information. If substantial uncertainty remains, they select additional sources expected to reduce uncertainty and repeat this process.

Active inference provides a unified framework for modeling perception and action of humans and living organisms under uncertainty about the external environment [3]. In active inference, humans and other living organisms are modeled as agents, and an agent combines prior beliefs with sensory inputs to infer a posterior distribution over possible states of the external environment. Based on the inferred results, the agent selects and executes actions so as to reduce uncertainty about the external state and to obtain preferred sensory inputs. By iteratively performing inference and action, the agent adapts to the surrounding environment.

Information collection for disinformation verification can be viewed as a sequential decision-making problem under uncertainty. When suitable prior beliefs are constructed from hazard maps and other contextual knowledge, and sensor and camera data are treated as observations, information source selection can be formulated within the active inference framework. This framework is particularly suitable for this task because it naturally integrates prior knowledge from geospatial data into the prior distribution and models the human-like cognitive process of reducing uncertainty through a single principle of free energy minimization.

This paper aims to model human processes of state estimation and selective information acquisition to systematically implement an evidence collection service for disinformation verification using active inference. We cast the flood-verification setting into the active inference framework by treating information sources that report flooding conditions as observations and constructing prior beliefs from multiple geospatial layers. We then conduct simulations under synthetic flood scenarios, and the results suggest that the proposed approach can efficiently collect supporting evidence for disinformation verification through selective access to information sources.

II. Related Work

A. Current State of Disinformation

Disinformation has emerged as a serious social issue in recent years. The Global Risks Report 2025 [4] identifies "Misinformation and Disinformation" as one of the most critical risks in the near term.

Prior research points out that disinformation is driven mainly by two motivations: pecuniary and ideological [5]. Pecuniary incentives arise under the attention economy, in which publishers obtain advertising revenue in proportion to traffic. Ideological motives lead actors to advance preferred candidates or political agendas. As the attention economy continues to grow and social polarization intensifies, disinformation is expected to remain a serious societal challenge.

In addition, the widespread availability of advanced AI technologies raises concerns that AI-generated disinformation may be posted at scale by bots, and that deepfake videos and images may circulate widely. To address the increasing sophistication of AI-driven manipulation, deepfake detection techniques have been actively studied [6]. For example, Xu et al. [7] proposed a new framework called a set convolutional neural network, which treats faces across multiple video frames as a set to detect face manipulation. As another example, Blue et al. [8] proposed a method for detecting audio deepfakes by estimating human vocal-tract characteristics from speech signals.

While these techniques are effective for detecting manipulated or AI-generated media, they are not sufficient on their own to determine whether a claim is true or false. For claims that can be corroborated by real-world conditions, one approach is to estimate the actual situation from reliable data and compare it with the reported information. Therefore, using physical-world information such as sensors and camera data can provide useful evidence.

B. Disinformation Countermeasure Platform

To address disinformation, Fujitsu Limited and nine industry/academia organizations have developed and are operating an integrated disinformation countermeasure platform [9]. The platform provides an integrated framework for disinformation detection, evidence collection, analysis, and impact assessment. In particular, the platform collects and manages related information about unverified content as supporting evidence and supports disinformation verification and impact assessment by analyzing consistency and contradictions across evidence items and estimating potential social impact. Based on this approach, the platform promotes research and development of four technical components: disinformation detection, evidence collection and integrated management, comprehensive analysis, and social impact assessment.

In this paper, we study methods for collecting IoT sensor data as supporting evidence within the evidence collection component.

III. ACTIVE-INFERENCE-BASED EVIDENCE COLLECTION METHOD

A. Overview of Active Inference

This subsection provides an overview of the active inference framework [3], [10]. In the active inference framework, an agent interacts with its surrounding environment. Hidden states of the environment generate observations received by the agent. The agent infers environmental states from these observations and influences the environment through its actions. In this paper, the hidden state of the external environment is denoted by s', observations by o, and the agent's actions by a. The environmental state inferred by the agent is denoted by s and is distinguished from the hidden state s'.

The generative model represents the agent's beliefs about how observations and states are generated. For a time horizon T, the generative model is defined as

$$P(o_{0:T}, s_{0:T} \mid a_{0:T-1}) = P(s_0) \prod_{\tau=0}^{T-1} P(s_{\tau+1} \mid s_\tau, a_\tau) \prod_{\tau=0}^{T} P(o_\tau \mid s_\tau), \tag{1}$$

where $P(o_\tau \mid s_\tau)$ denotes the observation likelihood, which is the conditional distribution of observation o_τ given state s_τ. The transition model $P(s_{\tau+1} \mid s_\tau, a_\tau)$ specifies how the state evolves under action a_τ. The term $P(s_0)$ denotes the prior belief over the initial state.

Under this generative model, the agent infers the state s_τ from observation o_τ. This inference corresponds to computing the posterior distribution $P(s_\tau \mid o_\tau)$. In active inference, an approximate posterior $Q(s_\tau)$ is introduced, and state inference is formulated as the problem of finding $Q(s_\tau)$ that minimizes the variational free energy:

$$F_\tau = \mathbb{E}_{Q(s_\tau)}\left[\ln \frac{Q(s_\tau)}{P(o_\tau, s_\tau)}\right]. \tag{2}$$

Action selection is also formulated in terms of the generative model and expected free energy minimization. Preferences over observations are encoded as a prior distribution $P(o_\tau)$. Given a policy π, which denotes a sequence of actions, the expected free energy $G_\tau(\pi)$ is bounded below as

$$\begin{aligned} G_\tau(\pi) \geq & -\underbrace{\mathbb{E}_{Q(o_\tau|\pi)}[D_{KL}[Q(s_\tau \mid o_\tau, \pi) \parallel Q(s_\tau \mid \pi)]]}_{\text{information gain}} \\ & -\underbrace{\mathbb{E}_{Q(o_\tau|\pi)}[\log P(o_\tau)]}_{\text{prior preference}}. \end{aligned} \tag{3}$$

The first term is referred to as information gain and represents the expected divergence between the posterior over states conditioned on an observation and the predictive distribution over states prior to observation. The second term quantifies the degree to which observations predicted under a policy align with the agent's prior preferences. Therefore, because both terms appear with negative signs in $G_\tau(\pi)$, minimizing expected free energy corresponds to both reducing uncertainty about hidden states and generating observations consistent with the agent's prior preferences.

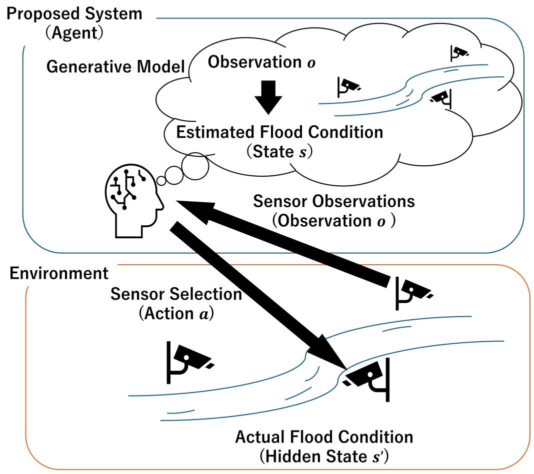

Fig. 1. Overview of the proposed evidence collection system

Under the above model, the agent repeatedly executes the following cycle: (1) acquiring observations generated by the external environment, (2) inferring the current state by minimizing variational free energy, (3) selecting a policy by minimizing expected free energy, and (4) executing actions prescribed by the selected policy to modify the external environment.

B. Overview of Proposed Method

In this paper, we assume that the disinformation countermeasure platform described in Section II-B handles a disinformation-verification task related to river flooding. Specifically, we consider a scenario in which a social media post claims that flooding has occurred at a specific location near a river. Based on this claim, an information collection request is issued through the platform. In response to this request, as shown in Fig. 1, the proposed method collects information from nearby IoT sensors and estimates the occurrence of flooding.

We formulate the proposed method within the active inference framework. We treat information from sources as observations, the flooding state and the currently accessed source as the state, and the source selection as the action. Based on this formulation, we model flooding state estimation and source selection as free energy minimization and thereby enable efficient information collection.

C. Active Inference Formulation of the Proposed Method

In this scenario, we define the state variable s_τ in the active inference model as a pair $(s_\tau^{\text{flood}}, s_\tau^{\text{src}})$, where s_τ^{flood} represents the flooding state around the river and s_τ^{src} denotes the information source accessed by the system at time τ. For simplicity, we represent the flooding state around the river as a discrete $n \times n$ grid map in which each cell takes one of two values: "flooded" or "not flooded". We index each cell by (i, j), where i and j denote the horizontal and vertical indices, respectively $(i, j \in \{0, 1, \ldots, n-1\})$. Let $s_{\tau,(i,j)}^{\text{flood}}$ denote the flooding state of cell (i, j) at time τ; then s_τ^{flood} is defined as

$$s_\tau^{\text{flood}} = \left(s_{\tau,(0,0)}^{\text{flood}}, \ldots, s_{\tau,(0,n-1)}^{\text{flood}}, s_{\tau,(1,0)}^{\text{flood}}, \ldots, s_{\tau,(n-1,n-1)}^{\text{flood}}\right). \quad (4)$$

In our formulation, the observation variable o_τ in the active inference model corresponds to the information obtained when the system accesses an information source at time τ. The observation o_τ indicates the flooding state of the grid cell containing the accessed source; thus,

$$o_\tau \in \{\text{“flooded”}, \text{“not flooded”}\}. \quad (5)$$

The act of changing the information source observed by the agent is represented as an action a_τ. Accordingly, the number of available actions equals the number of possible values of s_τ^{src}.

We define one time step as the time required for the agent to select an information source and obtain an observation from it. Here, the transition model for s_τ^{src} represents source switching. When the agent executes action a_τ, it changes the currently accessed source, and the next-step source state is deterministically given by $s_{\tau+1}^{\text{src}} = a_\tau$. For the flooding state factor s_τ^{flood}, the transition model represents how the flooding state may evolve during the time required for the agent to select a source and obtain an observation. For the experiments reported in Section IV, we assume a static flooding state, so the transition model for s_τ^{flood} assumes no change in the flooding state during each agent step.

The observation likelihood $P(o_\tau \mid s_\tau)$ represents the probability of observing "flooded" or "not flooded" at the location associated with the currently accessed source. While $P(o_\tau \mid s_\tau)$ can be extended to incorporate sensor noise, in Section IV, we assume noise-free observations. In particular, the likelihood assigns probability 1 to the observation consistent with the corresponding flooding state and 0 otherwise.

We assume independence between the initial flooding state and the initial source state. Accordingly, the prior belief over the initial state is factorized as

$$P(s_0) = P(s_0^{\text{flood}})\,P(s_0^{\text{src}}). \quad (6)$$

For $P(s_0^{\text{src}})$, we assume that the system is not accessing any information source at the initial time and denote this initialization condition by $s^{\text{src}} = \varnothing$. Accordingly, $P(s_0^{\text{src}})$ assigns probability 1 to $s^{\text{src}} = \varnothing$. Here, $\varnothing$ is an initialization marker only and is not included in the source state space or the action set.

$P(s_0^{\text{flood}})$ represents the occurrence probabilities of flooding patterns inferred from hazard maps and geospatial information. As an initial heuristic construction of the geospatial prior, we use four geospatial layers: the GSI pale color map, a landform classification map for flood control, an elevation map [11], and the planned-scale flood inundation assumption map provided by the MLIT Hazard Map Portal Site [12]. We introduce a resistance value representing how difficult it is for flooding to spread and an evidence value representing the hazard level indicated by the hazard map, and construct $P(s_0^{\text{flood}})$ as follows:

TABLE I
LANDFORM-TO-RESISTANCE SCORES USED TO CONSTRUCT $P(s_0^{\mathrm{FLOOD}})$

Landform class	Resistance score
River	0.001
Former river channel	10
Floodplain	30
Other	50
Natural levee	120
Plateau	200

1) Compute a resistance value on a fine 32×32 grid from the landform classification map using fixed landform-to-resistance scores (Table I), and adjust it using elevation so that higher cells have larger resistance. For the $n \times n$ flooding state grid, the minimum fine-grid resistance value in each coarse cell is used as the representative resistance value.
2) For each coarse cell k, compute an evidence value E_k from the planned-scale flood inundation map based on the colored-area ratio: $E_k = 100.0$ (ratio $\geq 5\%$), $E_k = 10.0$ (ratio $\geq 1\%$), and $E_k = 0.0$ otherwise.
3) Identify cells containing river segments from the pale color map as candidate flood origins, and generate a flood-expansion order by repeatedly adding the adjacent cell with the smallest resistance.
4) For each step in the flood-expansion order, compute a weight for each intermediate flooding pattern based on the evidence value of the newly added cell.
5) Repeat Steps 3–4 for all origins, sum weights for identical patterns, and normalize the aggregated weights to obtain $P(s_0^{\mathrm{flood}})$.

With the above definitions, the generative model over a finite horizon T is written in the same form as Eq. (1) in Section III-A. After obtaining an observation from the accessed information source at time τ, the agent updates its belief by computing an approximate posterior $Q(s_\tau)$ that minimizes the variational free energy in Eq. (2). In our implementation, $Q(s_\tau)$ is computed using a fixed-point iteration method to minimize variational free energy. Since the state is defined as $s_\tau = (s_\tau^{\mathrm{flood}}, s_\tau^{\mathrm{src}})$, the flooding state is estimated from the marginal posterior over the flooding state factor, $Q(s_\tau^{\mathrm{flood}})$.

For action selection, we do not seek any particular observation, so we set the prior preference $P(o_\tau)$ to be uniform over the observation space. With this setting, the prior-preference term in Eq. (3) becomes a policy-independent constant, and minimizing the expected free energy is equivalent to maximizing the expected information gain. In this study, the agent compares the expected free energy of all candidate actions and selects the one with the smallest value. As a result, the agent preferentially selects information sources that are expected to be more informative for estimating the flooding state.

In this way, we apply the active inference framework to flood-related disinformation verification by formulating flooding state estimation as variational free energy minimization and information source selection as expected free energy minimization.

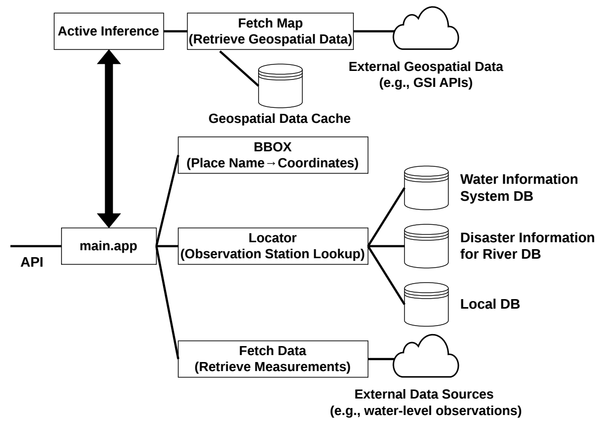

Fig. 2. Implemented architecture of the evidence collection service within the disinformation countermeasure platform.

D. System Architecture and Platform Integration

Fig. 2 shows the architecture of the evidence collection service implemented within the disinformation countermeasure platform. The active-inference-based source selection and belief update method described in Section III-C forms the core decision module of this service.

The process starts when the service receives an API request that specifies the target region for verification, such as a place name or a geographic area. If a place name is given, a BBOX module converts it into geographic coordinates and defines the analysis region. Given this region, a Fetch Map module retrieves the geospatial layers used in Section III-C, which may be cached to reduce repeated access to external services. The system then constructs the prior belief over flooding states from these layers. The system also uses a Locator module to identify candidate information sources in the region by querying local and public databases. The sources found in this step define the candidate values of the source state factor s_τ^{src}.

After the prior belief and candidate sources are prepared, the system repeats the following process. At each time step, the system first selects the next source by minimizing the expected free energy, as described in Section III-C. It then obtains an observation through the Fetch Data module and updates the belief over flooding states.

This process is repeated until a stopping condition is satisfied (e.g., the uncertainty becomes sufficiently small). The system then returns flooding-related indicators via the API.

IV. EVALUATION

A. Simulation Settings

To evaluate the core decision logic within the operational platform, we extracted the evidence collection module and conducted a simulation under a synthetic flooding scenario. This scenario is inspired by a past case of flood-related disinformation about Musashi-Kosugi (Kawasaki, Japan).

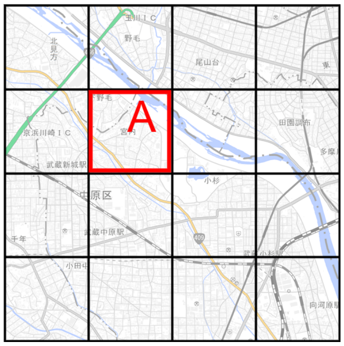

Fig. 3. Grid map of the Musashi-Kosugi region used in the simulation. Created by editing the GSI Tiles [11]

In the incident, the disinformation post claimed flooding by reusing images captured during an actual flood in the same area a few years earlier. To model this situation in our simulation, we assume that a social media post reports flooding at a particular place around Musashi-Kosugi. In this scenario, the disinformation countermeasure platform localizes the reported location using the attached metadata and visual cues in the post and issues an evidence collection request to assess whether the location is flooded. We further assume that no trustworthy information source, such as a public camera or an IoT sensor, is available at the reported location. Therefore, the system must infer the flooding state of the reported location from observations obtained at nearby reliable sources.

Based on this setting, we make the following assumptions:

- Following the grid-based representation described in Sec. III-C, we set $n = 4$ and discretize the target region into a 4×4 grid (Fig. 3). We refer to the cell containing the reported location as the target cell.
- We abstract real-world devices such as water-level gauges and cameras as information sources.
- One information source is available in each cell except for the target cell, where no source is available.
- When the agent accesses an information source, it obtains an accurate observation of the flooding state of the cell in which the source is located.
- The flooding state is static over the sequence of source accesses.

B. Evaluation Method

Under the scenario described in Sec. IV-A, we evaluate whether the proposed method can correctly determine the flooding state of the target cell with fewer accesses to information sources. In the following experiments, we assume that the reported location falls within cell A in Fig. 3. We designate this cell as the target cell and denote its coordinates by (X,Y). The agent's belief that the target cell is flooded at time τ is denoted by b_τ and is computed as follows:

$$b_\tau = \sum_{s_\tau \in S} \mathbb{I}\left(s^{\text{flood}}_{\tau,(X,Y)} = \text{flooded}\right) Q(s_\tau), \tag{7}$$

where $Q(s_\tau)$ denotes the approximate posterior at step τ, and S is the set of possible states. The indicator function $\mathbb{I}(\cdot)$ is defined as

$$\mathbb{I}\left(s^{\text{flood}}_{\tau,(X,Y)} = \text{flooded}\right) = \begin{cases} 1, & s^{\text{flood}}_{\tau,(X,Y)} = \text{flooded}, \\ 0, & s^{\text{flood}}_{\tau,(X,Y)} = \text{not flooded}. \end{cases} \tag{8}$$

We evaluate the belief b_τ specifically for the target cell after each information source access because our primary objective is the verification of a particular claim rather than estimating the entire region's state. In our simulation setting, once observations from all 15 available sources have been acquired, the belief b_τ is no longer updated. To assess the system's convergence properties in the best and worst-case scenarios, we consider two extreme cases for the ground-truth flooding condition: (i) all cells are not flooded and (ii) all cells are flooded.

- **Proposed (Active Inference)**: sources are selected by minimizing the expected free energy, as described in Sec. III-C.
- **Random**: sources are selected uniformly at random without replacement. Results are averaged over 100 runs.
- **High-to-low**: sources are accessed in descending order of the prior flooding probability of each cell under $P(s^{\text{flood}}_0)$.
- **Low-to-high**: sources are accessed in ascending order of the prior flooding probability of each cell under $P(s^{\text{flood}}_0)$.

C. Results

Fig. 4 and Fig. 5 show how b_τ changes under two true flooding conditions. The vertical axis represents the belief that the target cell is flooded; values closer to 1 indicate stronger belief that flooding has occurred at the target cell. In Fig. 4, the vertical axis uses a logarithmic scale, whereas in Fig. 5, it uses logarithmic spacing in terms of the distance from 1 to visualize convergence toward 1. In this simulation, the computed belief does not take the exact values 0 or 1 because of numerical precision in the Python implementation.

When all cells are not flooded, the belief that the target cell is flooded decreases and converges to a value close to zero. This indicates that the system can output evidence supporting the conclusion that flooding has not occurred at the target cell. Conversely, in the all-flooded case, the belief increases toward one, supporting the conclusion that the target cell is flooded. These results suggest that the proposed method can provide evidence useful for verifying flood-related disinformation.

We next compare the proposed method with baselines. In both the all-not-flooded case and the all-flooded case, the proposed method approaches values close to 0 or 1 with fewer source accesses than the Random baseline. For example, using $b_\tau \leq 10^{-3}$ as the convergence threshold in Fig. 4, the proposed method reaches the threshold after 3 accesses, whereas the Random baseline requires 13 accesses on average. This indicates that action selection based on expected free energy minimization is more effective for

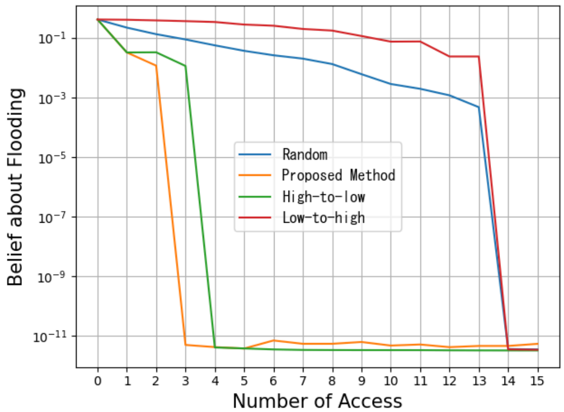

Fig. 4. Value of b_τ when all grid cells in Fig. 3 are not flooded.

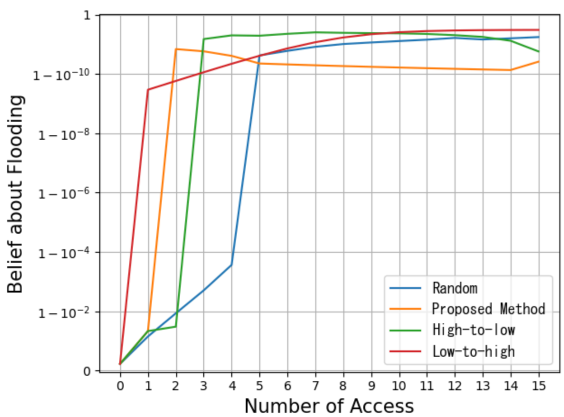

Fig. 5. Value of b_τ when all grid cells in Fig. 3 are flooded.

reducing uncertainty about the target cell. In the all-not-flooded case, the High-to-low baseline approaches 0 faster than Random, while the Low-to-high baseline converges more slowly than Random. In the all-flooded case, the Low-to-high baseline approaches 1 faster than the proposed method, whereas the High-to-low baseline requires more steps than the other methods. These results show that fixed-order selection can be advantageous when its ordering aligns with the true flooding condition, but it can also increase the number of accesses required for convergence when the ordering is mismatched. By contrast, the proposed method updates its selection based on the observations obtained so far, and therefore achieves consistently fast convergence with a relatively small number of accesses in both cases.

While the current evaluation is conducted in a simplified simulation setting with a static flooding state and noise-free, binary observations, these assumptions were chosen to verify the fundamental convergence properties of the proposed method. A key advantage of the active inference approach is its inherent extensibility; sensor-specific reliability and noise can be directly incorporated into the observation likelihood $P(o_\tau \mid s_\tau)$, and spatiotemporal dynamics, such as the gradual expansion of a flood, can be represented by refining the transition model $P(s_{\tau+1}^{\text{flood}} \mid s_\tau^{\text{flood}}, a_\tau)$. These extensions can be accommodated while preserving the same iterative cycle of source selection and belief update.

V. Conclusion

In this paper, we proposed an active-inference-based method for collecting supporting evidence for disinformation verification and implemented it as one of the modules within the disinformation countermeasure platform. By formulating flooding-state estimation and source selection as a unified process of free energy minimization, the system systematically mimics human-like cognitive processes to reduce uncertainty. The simulation results, based on a real-data-driven scenario, demonstrate that the proposed method efficiently collects useful evidence with significantly fewer source accesses than the baseline methods. As future work, we plan to evaluate the system in more complex, operational environments by incorporating heterogeneous sensor noise and dynamic flooding patterns.

Acknowledgment

This paper is based on results obtained from a project, JPNP22007, commissioned by the New Energy and Industrial Technology Development Organization (NEDO).

References

[1] Ministry of Internal Affairs and Communications, Japan, “WHITE PAPER Information and Communications in Japan,” 2025, (Accessed: 2026-02-19). [Online]. Available: https://www.soumu.go.jp/johotsusintokei/whitepaper/r07.html

[2] S. Hilberts, M. Govers, E. Petelos, and S. Evers, “The impact of misinformation on social media in the context of natural disasters: Narrative review,” *JMIR Infodemiology*, vol. 5, p. e70413, Jul. 2025.

[3] R. Smith, K. J. Friston, and C. J. Whyte, “A step-by-step tutorial on active inference and its application to empirical data,” *Journal of Mathematical Psychology*, vol. 107, pp. 1–60, Feb. 2022.

[4] World Economic Forum, “Global risks report 2025,” 2025, (Accessed: 2025-12-04). [Online]. Available: https://www.weforum.org/publications/global-risks-report-2025/

[5] H. Allcott and M. Gentzkow, “Social media and fake news in the 2016 election,” *Journal of Economic Perspectives*, vol. 31, no. 2, pp. 211–236, May 2017.

[6] M. S. Rana, M. N. Nobi, B. Murali, and A. H. Sung, “Deepfake detection: A systematic literature review,” *IEEE Access*, vol. 10, pp. 25 494–25 513, Jan. 2022.

[7] Z. Xu, J. Liu, W. Lu, B. Xu, X. Zhao, B. Li, and J. Huang, “Detecting facial manipulated videos based on set convolutional neural networks,” *Journal of Visual Communication and Image Representation*, vol. 77, pp. 1–9, May 2021.

[8] L. Blue, K. Warren, H. Abdullah, C. Gibson, L. Vargas, J. O’Dell, K. Butler, and P. Traynor, “Who are you (I really wanna know)? Detecting audio DeepFakes through vocal tract reconstruction,” in *Proceedings of 31st USENIX Security Symposium (USENIX Security 22)*. USENIX Association, Aug. 2022, pp. 2691–2708.

[9] Fujitsu Limited, “Fujitsu to combat fake news in collaboration with leading Japanese organizations,” 2024, (Accessed: 2026-02-20). [Online]. Available: https://www.fujitsu.com/global/about/resources/news/press-releases/2024/1016-01.html

[10] C. Heins, B. Millidge, D. Demekas, B. Klein, K. Friston, I. D. Couzin, and A. Tschantz, “pymdp: A python library for active inference in discrete state spaces,” *Journal of Open Source Software*, vol. 7, no. 73, p. 4098, May 2022.

[11] The Geospatial Information Authority of Japan, “GSI Tiles – Tile List,” (Accessed: 2026-02-20). [Online]. Available: https://maps.gsi.go.jp/development/ichiran.html

[12] Ministry of Land, Infrastructure, Transport and Tourism, Japan, “MLIT Hazard Map Portal Site,” (Accessed: 2025-12-17). [Online]. Available: https://disaportal.gsi.go.jp/