Uh... maybe? Displaying AI-supported Translation Uncertainty through Embodiment in Multilingual Interaction

Sandra Müller[1], Martin Feick[1], Alexander Maedche[1]

***Abstract*— Multilingual interactions increasingly rely on AI-supported translation technologies to bridge linguistic gaps, yet these systems often struggle with uncertainty caused by background noise, unclear speech, or dialectal variation. Existing methods for signaling uncertainty have been shown to be difficult for users to interpret, raising the question which cues could offer a human-centered and effective way for embodied AI to communicate uncertainty during translation tasks. This workshop paper explores how non-verbal uncertainty cues could make AI-supported translation more transparent and trustworthy. We examine how an embodied AI's non-verbal cues can provide indications of system uncertainty. By studying these cues in the context of translation, we aim to contribute to future Human-Robot Interaction (HRI) designs that communicate system uncertainty in more transparent ways.**

I. Introduction

When individuals do not share the same language as their surroundings, they often rely on translation technologies, such as mobile translation apps, to bridge communication gaps. However, AI-supported translations systems can make errors while still sounding highly confident, making it difficult for users to judge whether the information is reliable or not [1], [2]. Imagine an international student relying on an AI-supported translation system during a lecture. Poor room acoustics degrade the audio and a translated sentence suddenly appears unclear. The student cannot tell whether the speaker said something unexpected or whether the system misinterpreted the input. This ambiguity undermines trust and causes the student to disengage from the lecture. While translation systems enable basic understanding, they remain vulnerable to a variety of interferences, e.g. unclear speech, dialectal variations or background noises, which can distort the input the system receives [3]. As a result, translation outputs become not understandable for the user or are simply wrong. This is especially challenging for users who cannot verify the original language, leaving them unsure whether the speaker truly said something odd or whether the system produced an error.

This challenge highlights design considerations for how uncertainty in AI-supported translation systems can be communicated transparently. Non-verbal cues, demonstrated through embodied AI, may offer a subtle way to signal uncertainty without interrupting the conversational flow, making them promising for designing more transparent translation outputs. Existing approaches, like showing uncertainty percentages or using verbal uncertainty cues, have been shown to not always function as intended [2], [4], [5]. For example, numerical uncertainty scores and visualizations are often misunderstood by users, creating confusion rather than clarity [4], [5]. Similarly, verbal cues can express uncertainty through natural language but risk interrupting conversational flow [2]. **In translation contexts, it can create ambiguity about whether the uncertainty originates from the translation system or from the speaker being translated. This leads to the question of whether non-verbal cues, expressed through embodied AI, might offer a more effective way to communicate system uncertainty.** Instead of relying solely on numerical or auditory signals, an embodied AI could visually indicate that it is unsure. By adapting its cues to reflect the uncertainty of its own responses, it signals greater transparency, which in turn help users develop more trustworthiness in the system [6]. Non-verbal cues can be promising for supporting translation contexts because they can blend naturally into the interaction without disrupting the dialogue. This raises the question: **How can an embodied AI-supported translation system express uncertainty in non-verbal ways that are transparent for the user and foster trustworthiness in the system?** Trustworthiness refers to perceptions of competence, reliability, and communication of the system [7], while transparency identifies as a as trust-relevant, robot-related factor [6]. In our context, transparency refers to how clearly the system's uncertainty is conveyed to the user. Non-verbal cues serve as a form of transparency, helping users recognize when the system is limited, especially in translation scenarios where they cannot verify the original language. This workshop paper investigates how non-verbal cues can communicate uncertainty in AI-supported translation systems as a pathway toward more transparent and trustworthy translation.

II. Related Work

A. The Uncertainty Loop

Uncertainty arises on both sides of HRI, the human and the robot side. Humans may be uncertain in their responses, for example, in learning situations, when practicing a new language [8] or simply when lacking necessary information to provide a confident answer [9]. Such uncertainty is often expressed through a range of verbal and non-verbal cues, including hesitations, prosodic changes, facial expressions, or body movements [8]. On the other side of the interaction, AI-supported translation systems also experience uncertainty. Large Language Models (LLM) used in contemporary translation systems can produce incorrect or hallucinated outputs [1], [2]. AI-supported translation system

[1]human-centered systems lab (h-lab), Karlsruhe Institute of Technology (KIT), Germany. sandra.mueller@kit.edu

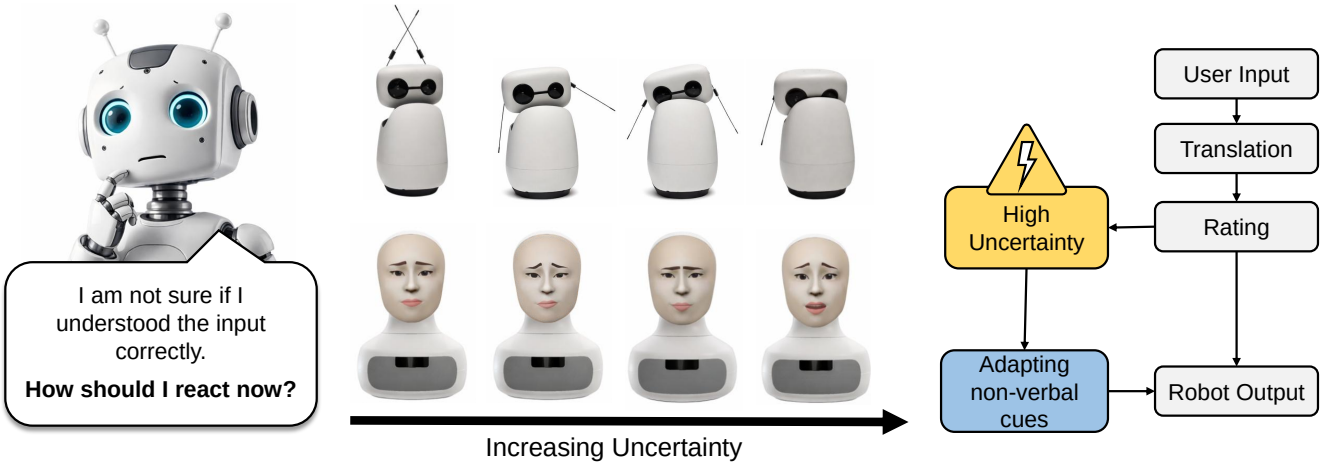

Fig. 1. How should an embodied AI system respond when it cannot provide a certain answer? One approach could be to map the system's uncertainty levels to non-verbal cues, e.g., facial expressions or body movements, to transparently communicate uncertainty to users. In this example, user input and its translation are processed to generate an uncertainty rating. This rating is then expressed through adaptive, non-verbal cues of two embodied AIs, illustrated using Reachy Mini and Furhat.

can generate interpretations that fail to align with contextual expectations and just simply make translation errors, that can have drastic consequences, in, worst case, law, finances and politics [10]. Systems can express their uncertainty in several ways, such as providing visual indicators, for example numerical uncertainty percentages [4] or in other visual ways like diagrams [5]. They may also use auditory or verbal cues to signal doubt [2], [11]. Beyond these modalities, embodied AI systems can leverage non-verbal cues, such as gaze or posture shift to communicate uncertainty [12]. However, existing work in HRI remains limited, as most studies investigate general emotional expressions rather than cues that explicitly communicate uncertainty [12], [13], [14], [15]. When robots display uncertainty, it is typically in the context of navigating unknown environments, such as requesting human assistance [16], [17], [18], [19]. Leusmann et al. [9] made a first attempt in conceptualizing this dynamic as uncertainty loop, illustrating how uncertainty on one side of the interaction can influence the other. For instance, when a robot signals hesitation, this can increase the human's own sense of uncertainty, which in turn may shape how the robot responds in the next turn.

B. Robots Expressing Uncertainty

Uncertainty with robots can be expressed through verbal and non-verbal cues. Verbal, spoken uncertainty and prosodic cues as well as intonation help signal doubt in spoken output [2], [11], while disclaimers support clearer interpretation of uncertainty [2], [4]. Robots can provide non-verbal cues, like gaze and posture shift. These cues have been shown to help users perceive a robot's uncertainty [12], [18]. Motion-based strategies, such as adjusting movement speed based on uncertainty, help users judge how certain a robot is [12], [19]. Visual cues, like light patterns or color changes, also provide ways for robots to signal uncertainty [12], [18].

When robots externalize a state of uncertainty, e.g. through verbal or non-verbal cues, their behavior becomes easier for humans to interpret, although these cues can still be misread [12], [18]. Visible expressions of uncertainty have been shown to improve clarity in HRI. When robots explicitly signal their uncertainty, users infer the robot's intent more accurately and are generally more willing to cooperate [12], [18]. Users also tend to prefer robots that openly express internal states over robots that remain unclear [12], [18]. For instance, mapping uncertainty to movement speed enables users to perceive a robot's certainty level directly through its motion [19]. Similarly, modulating the tone and pitch of synthesized robot voices can effectively convey varying degrees of uncertainty [11].

Although robots are capable of expressing uncertainty, prior HRI work has mostly examined general emotional expressions or uncertainty in navigation tasks, leaving the expression of system uncertainty unexplored. Related work has demonstrated that users can perceive uncertainty through non-verbal cues, yet these studies do not examine how such cues function when transmitting AI-supported translation system uncertainty. Moreover, translation contexts bring unique challenges, as uncertainty must be communicated without disrupting conversational flow or creating ambiguity. Our work addresses this gap by investigating how embodied AI can use non-verbal cues to express system uncertainty in translation contexts.

III. Proposed Approach

At this stage, our work outlines an exploratory approach for how embodied AI-supported translation systems might express uncertainty through non-verbal cues. To study uncertainty in translation, participants must encounter a source language they do not understand. Because most of our participants are English-speaking, we focus on a language that is linguistically distant from English, such as Japanese. This ensures that they must rely fully on the translation

system, as they cannot verify the original utterance. To explore how such cues could be conceptualized, we propose a two-step exploratory approach consisting of a workshop and a follow-up online study.

a) Participatory Workshop.: The workshop serves as the first step in understanding how humans imagine expressions of uncertainty from embodied AI systems. We begin broadly by investigating how people think embodied AI systems' uncertainty should be expressed in non-verbal cues. As the workshop progresses, we gradually narrow the focus toward translation scenarios. Participants reflect on which cues they would find helpful for detection and how potential embodied AI uncertainty could look like. By presenting participants with different virtual and physical embodied AI versions, we broaden the range of design possibilities. Through collaborative discussions and lightweight sketching, the workshop aims to surface a set of possible cues, capturing how participants conceptualize uncertainty in embodied AI-supported translation systems.

b) Online Study.: Building on the workshop, we plan to compile a set of example cues in form of short video clips, that reflect different versions of uncertainty expression. In an online study, participants will judge how clearly these cues communicate uncertainty. To examine whether non-verbal cues generalize across cultural backgrounds, we will collect participants' language skills beforehand, allowing us to analyze which cues lead to different interpretations across participant groups and whether certain non-verbal signals remain robust across cultures. Additionally, we aim to determine how much intensity the proposed cues require to be understood by participants. To achieve this during the online study, participants will evaluate different cue intensities across the presented embodiments, for example by comparing versions with low, medium, and high intensity, and evaluate which levels they perceive as clear, ambiguous, or too strong.

This step serves as a first indication of whether the workshop generated cues are interpreted in a consistent way across users. The expected outcome is a set of empirically grounded cues that are understandable to a broader audience. These cues then form a basis for selecting behaviors suitable for integration into a later prototype.

Overall, the combined insights from both experiments are expected to provide:

1) an initial pool of possible non-verbal uncertainty cues;
2) insights into how these cues may vary across cultural backgrounds;
3) an understanding of how clearly and distinctly uncertainty should be expressed;
4) and a set of empirically validated cues that could inform the design of an embodied prototype for expressing AI systems' uncertainty.

c) Envisioned System Design: These outcomes will serve as a foundation for potential implementations and further experiments. Looking ahead, one possible system concept could begin by detecting its own translation errors and deriving an uncertainty score, similar to methods that check translation accuracy by analyzing text on several layers (e.g., word meaning, grammar) to spot translation mistakes [20]. This score may then inform how an embodied AI adapts its behavior in real time. As one possible embodied platform, a robot such as Furhat[1] could provide expressive facial cues that map to different levels of uncertainty. Furhat could display subtle cues, such as widened eyes, asynchronous eyebrow movements, or slower and hesitant movements, to gradually intensify its uncertainty score (Fig. 1). Alternatively, Reachy Mini[2] could be used to display non-verbal cues, as it provides more zoomorphic movements instead of facial expressions. It can, for example, move its antenna to signal uncertainty or hide its head into its body to appear uncertain (Fig. 1). These cues could allow embodied AI to communicate uncertainty in a nuanced way without interrupting the conversational flow or impacting the translated content itself.

IV. Conclusion

This workshop paper aims to explore strategies on how to find out how embodied AI could communicate uncertainty in translation contexts through non-verbal cues. We outlined a conceptual approach for identifying non-verbal cues that could help AI-supported translation systems be more transparent and trustworthy to users. To make uncertainty more transparent, we propose the use of non-verbal cues. While our focus lies on non-verbal strategies, there is a risk that interpretations of these cues can vary across cultural communities, which may limit generalization and require further investigation. Overall, this work contributes to ongoing efforts toward transparent and trustworthy embodied AI by outlining early considerations for how virtual or physical embodiment might express AI-supported system uncertainty in translation contexts. Our participatory workshop and online study are intended as first steps toward identifying behaviors that could later form the design of prototype systems and future studies.

References

[1] W. He, Z. Jiang, T. Xiao, Z. Xu, and Y. Li, "A survey on uncertainty quantification methods for deep learning," *ACM Comput. Surv.*, vol. 58, no. 7, Feb. 2026. [Online]. Available: https://doi.org/10.1145/3786319

[2] S. S. Y. Kim, Q. V. Liao, M. Vorvoreanu, S. Ballard, and J. W. Vaughan, ""i'm not sure, but...": Examining the impact of large language models' uncertainty expression on user reliance and trust," in *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, ser. FAccT '24. New York, NY, USA: Association for Computing Machinery, 2024, p. 822–835. [Online]. Available: https://doi.org/10.1145/3630106.3658941

[3] C. Yang, M. Chen, Y. Wang, and Y. Wang, "Uncertainty-guided end-to-end audio-visual speaker diarization for far-field recordings," in *Proceedings of the 31st ACM International Conference on Multimedia*, ser. MM '23. New York, NY, USA: Association for Computing Machinery, 2023, p. 4031–4041. [Online]. Available: https://doi.org/10.1145/3581783.3612424

[4] C. Bartneck and E. Moltchanova, "Expressing uncertainty in human-robot interaction," *PLoS ONE*, vol. 15, 2020.

[1] https://www.furhatrobotics.com/

[2] https://huggingface.co/reachy-mini

[5] J. M. Hofman, D. G. Goldstein, and J. Hullman, "How visualizing inferential uncertainty can mislead readers about treatment effects in scientific results," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, ser. CHI '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 1–12. [Online]. Available: https://doi.org/10.1145/3313831.3376454

[6] P. A. Hancock, D. R. Billings, and K. E. Schaefer, "Can you trust your robot?" *Ergonomics in Design*, vol. 19, no. 3, pp. 24–29, 2011.

[7] K. E. Schaefer, "The perception and measurement of human-robot trust," Ph.D. dissertation, University of Central Florida, Orlando, FL, 2013. [Online]. Available: https://stars.library.ucf.edu/etd/2688/

[8] R. Cumbal, J. Lopes, and O. Engwall, "Detection of listener uncertainty in robot-led second language conversation practice," in *Proceedings of the 2020 International Conference on Multimodal Interaction*, ser. ICMI '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 625–629. [Online]. Available: https://doi.org/10.1145/3382507.3418873

[9] J. Leusmann, C. Wang, M. Gienger, A. Schmidt, and S. Mayer, "Understanding the uncertainty loop of human-robot interaction," *ArXiv*, vol. abs/2303.07889, 2023.

[10] X. Xie, S. Jin, S. Chen, and S.-C. Cheung, "Word closure-based metamorphic testing for machine translation," *ACM Trans. Softw. Eng. Methodol.*, vol. 33, no. 8, Nov. 2024. [Online]. Available: https://doi.org/10.1145/3675396

[11] E. Székely, J. Mendelson, and J. Gustafson, "Synthesising uncertainty: The interplay of vocal effort and hesitation disfluencies." in *INTERSPEECH*, 2017, pp. 804–808.

[12] M. Matarese, A. Sciutti, F. Rea, and S. Rossi, "Toward robots' behavioral transparency of temporal difference reinforcement learning with a human teacher," *IEEE Transactions on Human-Machine Systems*, vol. 51, pp. 578–589, 2021.

[13] D. Peña and F. Tanaka, "Human perception of social robot's emotional states via facial and thermal expressions," *J. Hum.-Robot Interact.*, vol. 9, no. 4, May 2020. [Online]. Available: https://doi.org/10.1145/3388469

[14] Y. Sakamoto, A. Herath, T. Vuradi, S. Sallam, R. Gomez, and P. Irani, "How should a social robot deliver negative feedback without creating distance between the robot and child users?" in *Proceedings of the 11th International Conference on Human-Agent Interaction*, ser. HAI '23. New York, NY, USA: Association for Computing Machinery, 2023, p. 325–334. [Online]. Available: https://doi.org/10.1145/3623809.3623882

[15] E. Kirjavainen, C. Nieto Agraz, A. Colley, H. Mueller, S. Boll, and J. Häkkilä, "Exploring perceptions of robot facial expressions in service interactions," in *Proceedings of the 28th International Academic Mindtrek Conference*, ser. Mindtrek '25. New York, NY, USA: Association for Computing Machinery, 2025, p. 107–116. [Online]. Available: https://doi.org/10.1145/3757980.3757996

[16] D. Porfirio, M. Roberts, and L. M. Hiatt, "Uncertainty expression for human-robot task communication," in *Proceedings of the 24th International Conference on Autonomous Agents and Multiagent Systems*, ser. AAMAS '25. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, 2025, p. 1698–1707.

[17] X. Yu, M. Hoggenmüller, T. T. M. Tran, Y. Wang, Q. Zhang, and M. Tomitsch, "Peek into the 'white-box': A field study on bystander engagement with urban robot uncertainty," in *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, ser. CHI '25. New York, NY, USA: Association for Computing Machinery, 2025. [Online]. Available: https://doi.org/10.1145/3706598.3713790

[18] S. Shirol, J. Delfa, I. Leite, and E. Yadollahi, "Designing social behaviours for autonomous mobile robots: The role of movement and light in communicating intent," *2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 1638–1643, 2025.

[19] J. Hough and D. Schlangen, "It's not what you do, it's how you do it: Grounding uncertainty for a simple robot," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, 2017, pp. 274–282.

[20] M. Huang, M. M. Mahmud, and C. S. Goh, "Intelligent detection method for text translation errors based on multi-level textual features and maximum entropy model," *Discov Computing*, vol. 28, no. 149, 2025. [Online]. Available: https://doi.org/10.1007/s10791-025-09680-5