

# LLM-based Teamwork Role Inferencing for Fostering Social Online Learning

Wimukthi Thommadurage, School of Education, Adelaide University, Australia, [thommadurage.wimukthi@adelaide.edu.au](mailto:thommadurage.wimukthi@adelaide.edu.au)

Nguyen Nhu Anh Le, School of Education, Adelaide University, Australia, [nguyennhuanh.le@adelaide.edu.au](mailto:nguyennhuanh.le@adelaide.edu.au)

Negin Mirriahi, School of Education, Adelaide University, Australia, [negin.mirriahi@adelaide.edu.au](mailto:negin.mirriahi@adelaide.edu.au)

Srecko Joksimovic, School of Education, Adelaide University, Australia, [srecko.joksimovic@adelaide.edu.au](mailto:srecko.joksimovic@adelaide.edu.au)

Maarten de Laat, School of Education, Adelaide University, Australia, [maarten.delaat@adelaide.edu.au](mailto:maarten.delaat@adelaide.edu.au)

## Abstract

Higher education institutions are progressively implementing online learning modules due to their promotion of equity through scalability, accessibility, inclusivity, and affordability. Despite the advantages, online learning environments face persistent challenges in facilitating meaningful networked learning (NL) opportunities. One particularly pressing challenge is supporting effective NL through group formation and collaborative teamwork. With rising enrolments, manual grouping has become impractical. Random allocation overlooks complementary skills and often produces unbalanced teams, while self-assessment relies on students' often inaccurate self-perceptions, leading to mismatched roles and group tensions. A substantial body of research has examined strategies for group formation, and many studies have emphasised the value of team role allocation in promoting effective NL. However, little attention has been paid to approaches that infer learners' collaborative role tendencies from their actual behavioural interactions and subsequently use this information to inform group composition and achieve a better NL experience. Accordingly, this study examines whether large language models (LLMs) can infer higher-education learners' teamwork role tendencies comparably to human judgment, using data from an international sample of undergraduate and postgraduate learners recruited via Prolific who interacted with a custom-built chatbot in collaborative learning scenarios. The resulting interactions were analysed through an LLM to infer teamwork role tendencies based on previously established Belbin Team Roles. The roles inferred by the LLM were then compared against those coded by human coders, with inter-rater alignment evaluated using Cohen's Kappa and percentage agreement. Learner responses were coded by trained human researchers using a consensus-based Belin Team Role framework. The findings reveal that LLM achieved a moderate degree of alignment with human coders, suggesting its viability as a tool for inferring learners' teamwork role tendencies. Moreover, exploratory analyses revealed that the length of learners' responses to the chatbot is associated with the extent to which the LLM's inferences aligned with human coders. However, future research would benefit from larger sample sizes and the use of more advanced statistical methods to better capture the effects of interaction quality. This study contributes to the growing body of work on LLM-supported NL by highlighting both the potential and the limitations of using LLMs for role inference. Future implementations should pay particular attention to fostering high-quality learner-chatbot interactions to maximise reliability and pedagogical value to help design for effective social participation in online NL experiences. The findings of this research are expected to advance the development of technology-enhanced, NL practices within online higher education.

## Keywords

Online learning, networked learning, collaborative learning, group roles, large language models, teamwork

## Introduction

The evolution of digital technologies has transformed the way higher education institutions deliver teaching and learning. Online learning has become a mainstream mode of instruction worldwide, reshaping traditional pedagogical approaches and institutional strategies (Salama & Hinton, 2023). While mediated through technology, learning remains fundamentally social, rooted in interpersonal relationships, influenced by human activity, and enacted through collaboration on meaningful tasks (Networked Learning Editorial Collective

(NLEC, 2021). For individual learners, interaction with others plays a central role in the processes of knowledge acquisition, cognitive development, and adult learning (De Felice et al., 2022). Such interaction enables the creation of shared networks of knowledge, where understanding emerges through connections between people, information sources, and digital technologies (Siemens, 2005).

Despite its growing importance in higher education, online higher educators often struggle to facilitate meaningful and effective teamwork, with challenges such as complexities in group formation, unequal participation and work distribution (Düker & Rieber, 2024; McCollum et al., 2019), resulting in a poor networked learning (NL) experience. In practice, time and logistical constraints often result in random group allocation or scripted assignment of predetermined group roles as default approaches (Gardeli & Vosinakis, 2025; Oyarzun & Martin, 2023). Although such strategies offer administrative simplicity, they often give rise to issues in group dynamics, including dominance, marginalisation, and inefficiencies in networked collaboration (Düker & Rieber, 2024). Simultaneously, while research has shown that role assignment can enhance participation and foster peer interaction in NL (Luo et al., 2023), structured scaffolding to help learners effectively understand and enact their group roles, remains uncommon in most online learning environments. Providing such support is crucial, as effective collaboration in depends on trust, reciprocity, and a sense of belonging to an NL community, which provides the conditions for learners to act collectively towards shared goals (NLEC, 2021).

Although research has examined strategies for effective group formation in NL, much of this work has primarily concentrated on the process of allocating learners to groups (Chen & Kuo, 2019; Danahy et al., 2024). Less attention has been directed toward approaches that first identify learners' teamwork role tendencies and, in the longer term, use these insights both to create balanced groups and to cultivate learners' awareness of their teamwork role tendencies. The emergence of large language models (LLMs) offers new opportunities in this direction. Therefore, this study focuses on harnessing the generative capacity of LLMs in the initial step of this process by investigating whether LLMs can accurately infer learners' teamwork role tendencies from their naturalistic interactions with a custom-built chatbot. Specifically, this study analyses interaction data from a single learner-chatbot conversation to identify learners' teamwork role tendencies, drawing on Belbin Team Roles (Belbin, 2010) as the basis for classification and group formation. Validating this inferential capability is an essential precursor for future research that investigates whether LLM-inferred roles can be used to form more balanced and cohesive groups, reduce the likelihood of role-related conflict, and ultimately enhance learners' collaborative and networked learning experiences. While such automation may also reduce instructor workload in large-scale online settings, the primary contribution of this study lies in establishing the feasibility and validity of LLM-based role inference as a foundation for these pedagogical applications.

## Background

### Social Aspects in Online Learning

From the earliest stages of development, humans' social nature is evident in the way they construct knowledge through interaction (Vygotsky, 1978). Within educational contexts, such interaction is foundational for promoting knowledge acquisition and fostering cognitive development (De Felice et al., 2022). Despite these affordances, online learning environments tend to position learners in relative isolation, limiting opportunity for establishing social presence and leading to challenges such as low motivation, feelings of disconnection, and inadequate support (Kotera et al., 2021; Pei et al., 2024). Connectivism (Siemens, 2005) conceptualises learning as a process of building and navigating networks, where knowledge is distributed across connections among individuals, communities, and digital resources. The quality of learning, especially in online environments, depends on the richness of connections established among learners, educators, and resources (Goodyear et al., 2004). NL emphasises that whether such connections are collaborative, co-operative, or collective, they must be underpinned by mutual trust and shared motivation (NLEC, 2021). Technology, in this regard, is not merely an external tool but a component of the learning design that fosters the emergence and sustainability of meaningful connections, while avoiding the risks of isolating learners through individualised designs or exerting undue control through overly prescriptive systems (NLEC et al., 2021). This study builds on these theoretical foundations to position LLM-based role inferencing as a supportive technology that can strengthen teamwork and NL without constraining learners to fixed roles or structures.

### Social Formation and Role Assignment for Effective Social Learning

Group formation in online learning requires careful consideration due to its significant influence on collaborative outcomes (Y. Chen et al., 2025). While random group allocation is common practice, particularly in online

environments (Oyarzun & Martin, 2023), such approaches risk creating imbalances in participation, knowledge, and perspectives, which can hinder deep learning due to some students being more dominant while others are marginalised (Düker & Rieber, 2024). In contrast, intentional grouping techniques have been found to foster better perceived work quality, time management, efficiency, and relationship building (Scheidler et al., 2025). As a result, extensive research has been conducted on the effective grouping of learners within online environments (Oyarzun & Martin, 2023). Regardless, group formation in online NL remains a challenge, as manual assignment grows increasingly complex with rising enrolments, often making random allocation the default. While learner self-selection offers an alternative, this approach could create imbalance between groups and other group dynamic issues (Düker & Rieber, 2024).

Simply grouping learners does not guarantee meaningful collaboration in NL settings. Assigning roles that each participant might assume in a group is important to prevent issues such as intimidation, frustration, isolation, and unequal participation (Kotera et al., 2021; McCollum et al., 2019; Pei et al., 2024). Roles, understood as expected behaviours and responsibilities (Biddle, 1986), can help establish the conditions for cooperative learning outlined by Johnson & Johnson (1999), including interdependence, accountability, interaction, social skills, and group processing. Research indicates that using roles in collaborative settings can foster participation and peer interaction, boost cognitive engagement, strengthen group cohesion, clarify responsibilities, and improve coordination (Cheng et al., 2014; Luo et al., 2023). In the literature, the roles within collaborative learning contexts are considered from two perspectives: either as scripted, where tasks are explicitly assigned, or as emergent, arising naturally through group interaction (Belbin, 2010; O’Donnell, 1992). Effective role assignment, however, should go beyond rigid task division to support flexible behaviours shaped by individual traits and group dynamics (Strijbos & De Laat, 2010). The Belbin Team Roles (Belbin, 2010) (see Figure 1), which conceptualises roles as behavioural tendencies, has been widely adopted in higher education to support group formation and project-based learning, with studies showing that teams balanced across Belbin roles report improved collaboration and performance (Aranzabal et al., 2022). Therefore, recognising and utilising roles is consequently significant for balancing group dynamics and fostering the NL conditions needed for effective social learning in online environments.

| Social Roles                   |  |                                   | Thinking Roles                        |   |                                       | Task Roles                                     |   |                                    |
|--------------------------------|--|-----------------------------------|---------------------------------------|---|---------------------------------------|--|---|------------------------------------|
| <b>Coordinator (CO)</b>        | <b>Team-worker (TW)</b>                      | <b>Resource Investigator (RW)</b> | <b>Specialist (SP)</b>                | <b>Monitor Evaluator (ME)</b>           | <b>Plant (PL)</b>                     | <b>Completer Finisher (CF)</b>                 | <b>Implementor (IM)</b>                             | <b>Shaper (SP)</b>                 |
| Clarify goals, delegates tasks | Supportive, cooperative, and conflict-averse | outgoing, explores opportunities  | Provides deep expertise, narrow focus | Analytical, critical thinker, impartial | Creative, imaginative, problem-solver | Detail-focused, perfectionist, meets deadlines | Practical, disciplined, and turns ideas into action | Dynamic, driven, challenges others |

**Figure 1: Team Roles Grouped into Social, Thinking, and Task categories. Adapted from Belbin (2010)**

### Automatic Group Formation and Role Awareness

Automated group formation methods have advanced considerably, moving from simple rule-based systems to sophisticated Artificial Intelligence (AI) driven approaches. Early methods relied on fixed parameters, such as preferences, requests, or pre-test results (Ergulec, 2019; Muuro et al., 2014), with some incorporating dynamic regrouping based on participation and performance (Wijaya et al., 2019). Subsequent studies employed machine learning techniques to cluster learners by attributes such as prior academic performance, interaction patterns, and collaboration competence levels (Hasan et al., 2024; Maina et al., 2017). Optimisation and meta-heuristic approaches, such as genetic algorithms, were also explored to balance factors such as prior knowledge, roles, and social interactions in groups, though these methods still relied on predefined parameters (Chen & Kuo, 2019). More recently, advanced AI approaches have integrated diverse learner data, such as preferences, performance, behaviours, affective states, and interests, using deep neural networks and natural language processing (Pillai, 2022; Smith et al., 2023).

Despite these advances, many approaches remain constrained by their reliance on narrow or pre-existing data, limiting their applicability in certain contexts, for example, first-time learners. Similarly, automated approaches for detecting learners’ social roles or behavioural traits within NL settings are becoming more common with most existing methods employing traditional machine learning and deep learning techniques (Figuroa et al., 2019; Pongpaichet et al., 2022; Wijenayake et al., 2020). A common feature across these methods, however, is their reliance on pre-recorded datasets, such as learner interactions in online forums (Figuroa et al., 2019; Wijenayake et al., 2020), academic and behavioural traces from learning management systems (Pongpaichet et al., 2022). As a result, their applicability remains limited in contexts where prior learner data are unavailable. More recently,

Danahy et al. (2024) investigated the use of LLMs for group formation without relying on prior learner data, replacing traditional Likert-scale surveys with open-ended student reflections. Using ChatGPT, they analysed students' responses to create learner profiles, synthesise class-wide themes, recommend groups, and generate individual and group reports to scaffold collaboration. However, their approach was constrained by its reliance on self-reported reflections and the absence of a validated theoretical framework for role inference.

### **Purpose of the Study**

Intentional group formation and role assignment are essential for effective collaboration in NL (Y. Chen et al., 2025; Luo et al., 2023), yet limited research has examined how learners' teamwork role tendencies can inform these processes. Emerging advances in LLMs offer opportunities to automate role inference and group composition, reducing instructor workload while supporting pedagogically effective teams for NL. Accordingly, the aim of this study is to examine whether LLMs can be used as a socio-technical tool to infer learners' teamwork role behaviours from brief, naturalistic learner–chatbot interactions in order to inform collaborative group formation in networked learning contexts. To achieve this aim, the study addresses the following research questions:

**RQ1:** How do LLM-inferred teamwork roles from learner–chatbot exchanges in students compare with expert human judgements?

**RQ2:** How does response length influence the reliability of LLM-inferred teamwork roles from students' chatbot interactions compared with expert human judgements?

## **Methodology**

### **System Overview**

To address the research questions, a custom-built chatbot was developed to first interact with learners and generate conversational data about their team role tendencies. These chatbot–learner interactions were then analysed using an LLM to infer the role tendencies of each learner relative to their peers. This study utilised Belbin Team Roles (Belbin, 2010) as the foundation for inferring role tendencies (see Figure 1). The chatbot was developed using ChatMate, a GenAI platform that allows educators to design custom chatbots tailored to instructional needs. In this study, the chatbot was specifically designed to elicit learner input on teamwork traits through naturalistic, scenario-based dialogue rather than self-reported surveys. Learners were presented with three realistic group work scenarios and asked to select one, after which the chatbot engaged them in a dynamic, branching conversation that adapted based on their responses. These interactions simulated common collaborative challenges, such as coordination, decision-making, and conflict resolution, to capture how learners might respond in authentic contexts. The chatbot continued the dialogue until sufficient information was gathered, typically lasting 15–20 minutes. Following these interactions, learner dialogue data were processed for team role inference. Each conversation was retrieved, ordered, and broken down into message pairs consisting of the chatbot's prompt and the learner's response. These pairs were submitted to OpenAI's LLM via structured API calls, which analysed responses against the Belbin Team Roles. Instead of producing a single categorical role, the model returned percentage-based role scores and justifications in JSON format separately for each participant message. To mitigate potential randomness, each message was scored five times, and the average score was used as the final output. This allowed for nuanced interpretation, acknowledging that responses could reflect multiple roles or none. When aggregated across conversations, these message-level analyses produced comprehensive role distributions per learner, enabling the detection of role tendencies and patterns for subsequent evaluation.

### **Participants and Data Collection**

Participants were recruited through Prolific, an online research recruitment platform that facilitates participant involvement with informed consent, provided they meet the study's eligibility criteria. In total, 56 participants were recruited from an international pool, with the majority based in South Africa ( $n = 25$ ), followed by the United States ( $n = 9$ ), Spain ( $n = 6$ ), and the United Kingdom ( $n = 5$ ), with the remaining 11 participants representing other countries. The inclusion criteria required participants to be currently enrolled in a university program (31 undergraduate and 25 postgraduate students). Participants accessed the platform via a unique link and were instructed to engage with the chatbot. Upon completion of this activity, all participant–chatbot interactions were automatically recorded and stored in ChatMate's database. These conversation transcripts were later retrieved and used for the analysis presented in this study. Participation in this study was voluntary, and all participants received

monetary compensation through the Prolific platform in line with standard research practice. The study received institutional human research ethics approval prior to the commencement of data collection.

## Data Analysis

To address RQ1, the LLM-inferred roles were compared with human-coded roles. Two coders conducted human coding. Each participant's conversation was coded for all nine Belbin roles (see Figure 1) by both coders. Initially, they coded the data independently; however, due to discrepancies, likely resulting from the inherent subjectivity of role interpretation, they subsequently adopted a consensus coding approach. This involved jointly reviewing and assigning role labels to each message in the participant conversations to ensure consistency and reliability. The coding process mirrored the LLM-based approach in structure (see System Overview), assigning roles at the message level for each participant, allowing multiple roles to be assigned to a single message. However, unlike the LLM coding, the human coding did not involve confidence or weight scores; only categorical role labels were applied. At the end of the coding process, a role was considered assigned to a participant if it appeared in at least one of their messages. This procedure was followed for all participant conversations in the dataset. To address RQ1, the human coding was compared against the roles assigned for each participant's message by the LLM. To quantify the level of agreement between the two coding approaches, two standard inter-rater reliability metrics were used: Cohen's Kappa and Percentage Agreement. Cohen's Kappa accounts for chance agreement, offering a more conservative and robust measure than simple percentage agreement. Percentage Agreement, on the other hand, provides an intuitive sense of how often the LLM and human reviewers assigned the same roles to the same messages. Importantly, the comparison was conducted at the message level. That is, each individual message within a participant's conversation was examined to determine whether the LLM and human coders assigned the same set of roles. This resulted in a single Cohen's kappa score and a single percentage agreement score for each Belbin role.

To respond to RQ2, the length of participants' responses in the learner-chatbot exchanges was used to determine if they moderated the agreement between LLM- and human-coded roles. For each participant, we first calculated the overall percentage of agreement across all nine roles (see Figure 1), based on whether the LLM and human coders assigned the same set of roles to each participant. This agreement percentage was then plotted against the proportion of messages of different lengths in the learners' responses to the chatbot. Message length categories were determined empirically from the distribution of responses. This approach allowed for exploring whether conversations containing a higher proportion of long, detailed messages were associated with stronger alignment between the LLM and human role inference, compared to conversations dominated by shorter, less elaborated responses.

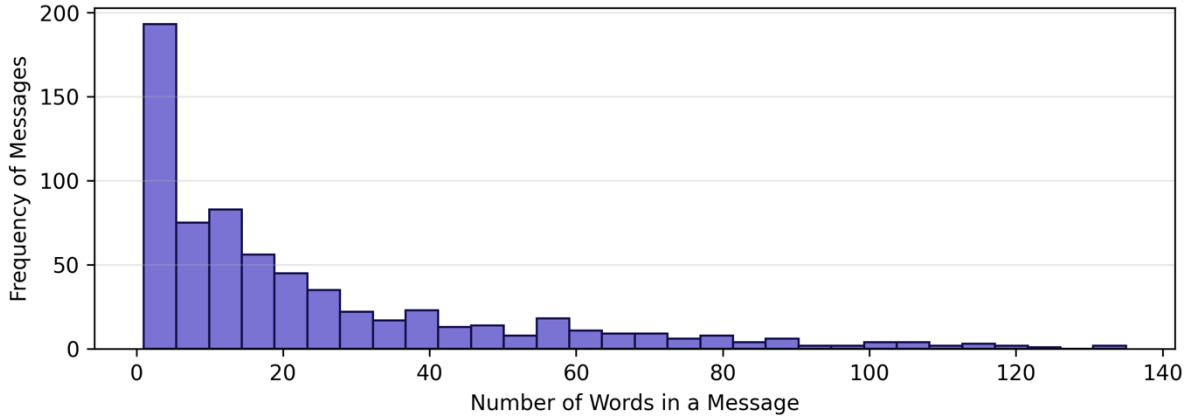
## Results

Table 1 shows the resulting Cohen's kappa scores and percentage agreement scores for each role, based on the comparison of 675 messages from 56 participants engaged with the chatbot. In terms of RQ1, the results indicate a strong level of agreement between LLM and human coding based on percentage agreement, with all roles achieving scores above 80% and some reaching as high as 98%. According to Cohen's kappa, the social roles (i.e. Coordinator, Team-worker, and Resource Investigator) demonstrated moderately higher agreement between human judgment and LLM outputs compared to thinking and task-related roles. Nevertheless, all roles showed higher agreement than would be expected by chance (Cohen's Kappa > 0). The discrepancy between the percentage agreement and Cohen's kappa values can be explained by the imbalance in the dataset: for most roles, the number of messages assigned to that role is considerably lower than the number not assigned. This class imbalance is accounted for in Cohen's kappa, which adjusts for agreement occurring by chance, thereby producing more moderate scores compared to the high values observed in percentage agreement.

The distribution of message lengths (see Figure 2) was highly right-skewed, with a high concentration of brief responses and a long tail extending beyond 100 words. To categorise responses meaningfully, thresholds were determined based on visual inspection of the histogram and changes in response density. Messages with  $\leq 5$  words were classified as concise (very short), reflecting minimal or acknowledgment-style responses. Messages with 6–39 words were classified as short, capturing the majority of conversational turns. Responses of 40–79 words were classified as medium-length, representing more elaborated but typical contributions, while messages with  $\geq 80$  words were classified as long, corresponding to the sparse upper tail of the distribution.

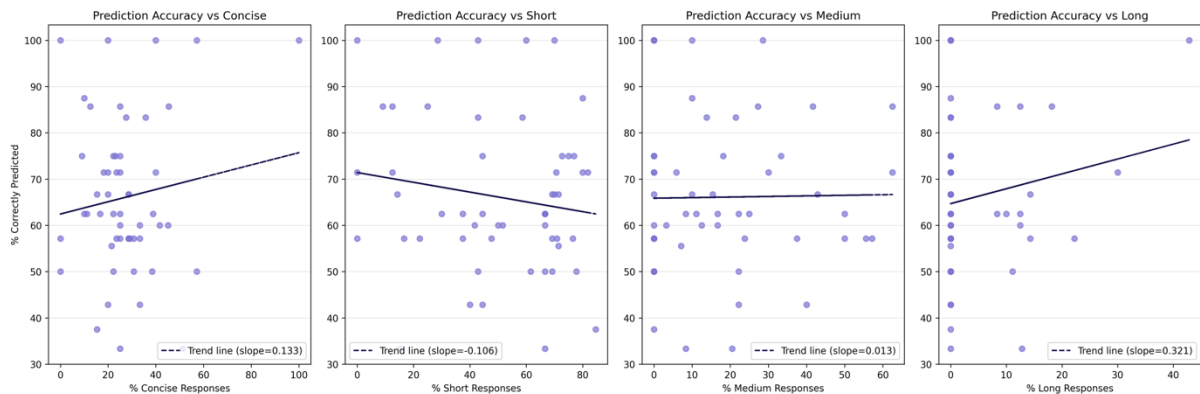
**Table 1: Cohen's Kappa percentage agreement by Belbin role**

| Role                 | Social roles |       |       | Thinking roles |       |       | Task roles |       |       |
|----------------------|--------------|-------|-------|----------------|-------|-------|------------|-------|-------|
|                      | CO           | TW    | RI    | SP             | ME    | PL    | CF         | IM    | SH    |
| Cohen's Kappa        | 0.52         | 0.51  | 0.41  | 0.39           | 0.38  | 0.15  | 0.41       | 0.27  | 0.20  |
| Percentage Agreement | 80.96        | 86.74 | 93.23 | 98.73          | 91.96 | 93.65 | 97.32      | 89.42 | 96.05 |



**Figure 2: Distribution of the number of words per participant message in the learner-chatbot exchange**

Figure 3 visualises scatterplots showing the percentage agreement between LLM and human coding plotted against the proportion of (1) concise, (2) short, (3) medium-length, and (4) long learner messages within a conversation. The results indicate that as the proportion of short messages increases, agreement between LLM and human coders tends to decrease. In contrast, conversations containing a higher proportion of long learner messages are associated with improved agreement. A similar, albeit weaker, positive trend is also observed for medium-length messages, suggesting that increased informational content may modestly support coding alignment. Interestingly, agreement was also found to increase in conversations with a higher proportion of concise messages (see Figure 4).



**Figure 3: Prediction accuracy against the proportion of concise, short, medium and long learner responses**

### Example of a Long Message [Participant 21]

"I'd check in with the team and let them know it's okay to speak up if they're feeling tired. I'd look at the workload and see what can be shared or adjusted. Maybe we bring in a few more volunteers or give people time to rest. I'd remind everyone that we're in this together and that their well-being matters just as much as the garden's progress."

#### AI Coding



#### Human Coding



#### AI Reasoning

The response reflects the Team-worker role by emphasizing team cohesion and well-being, encouraging open communication, and supporting team members who feel overwhelmed. It also demonstrates the Coordinator role by suggesting adjustments to workload and considering bringing in more volunteers, which involves organizing and structuring tasks to maintain team motivation and productivity.

#### Human Reasoning

Mention of sharing and adjusting tasks, getting volunteers, seems to align with coordinator while the focus on members well-being also imply team-worker.

### Example of a Short Message [Participant 13]

"In this case we must have been working in line with a prototype. we just test whether to that point the prototype is working according to the standards."

#### AI Coding



#### Human Coding



#### AI Reasoning

The response focuses on practical steps to ensure quality by testing against a prototype, which aligns with the Implementer's strength of turning ideas into practical actions. It also reflects the Completer Finisher's focus on ensuring accuracy and meeting standards through testing.

#### Human Reasoning

The focus on standards seems to align with completer finisher, although this is not a stronger alignment as participant is only responding to a question leaning towards "quality"

### Example of a Concise (Very Short) Message [Participant 16]

"I'd like to explore ways to improve performance individually"

#### AI Coding



#### Human Coding



#### AI Reasoning

The user's response is a general statement expressing interest in exploring individual performance improvement. It does not provide specific information related to teamwork behavior or any of the Belbin Team Roles. The response is not clearly indicative of any team role traits.

#### Human Reasoning

The response is too brief

Figure 4: Example comparisons of AI vs human coding across different lengths of messages

## Discussion

This study used a GenAI-powered chatbot to elicit textual data through open-ended learner dialogues and examined the capacity of LLMs to infer learners' collaborative tendencies from brief, naturalistic learner-chatbot interactions without reliance on large datasets or direct instructor input. Furthermore, the study analysed whether the length of the learner responses is associated with the degree of agreement between LLM-derived coding and human judgements of teamwork roles. Two research questions guided the investigation.

With respect to RQ1, the findings indicate that LLM-based role inferencing achieved a moderate level of alignment with human coding, as measured by percentage agreement and Cohen's Kappa. This outcome warrants consideration from two complementary perspectives. Firstly, it provides evidence that LLM exhibits significant potential in capturing and inferring learners' teamwork role tendencies in a manner that closely aligns with human judgment. This demonstrates that LLMs can be leveraged to surface insights about how learners contribute to collective tasks, thereby offering practical tools for supporting online social participation and balanced group design in NL environments (Y. Chen et al., 2025; De Felice et al., 2022). Secondly, it highlights the interpretive subjectivity inherent in human coding. While human judgment was used as the reference standard in this study, it should not be regarded as the absolute truth. During the initial phase of human coding, inter-rater agreement between two coders was low, necessitating the adoption of consensus coding, which highlights the inherent subjectivity in Belbin role coding. Consequently, the moderate agreement observed between LLM- and human-generated codings can also be attributed to this inherent subjectivity. Thus, while perfect alignment with human coding was not achieved, the findings indicate that LLM outputs were nonetheless comparable to human interpretations. Evidence in the literature of similar contexts supports this finding. For example, Zhang et al. (2025) found moderate agreement between GPT-4 and human coders and concluded that such levels of concordance are acceptable given the inherent subjectivity of qualitative coding. Similarly, Borse et al. (2025)

reported moderate to substantial agreement with consensus human coding and noted that discrepancies could partly reflect the inherent subjectivity of qualitative categories. While empirical research is required to determine the broader acceptability and validity of using LLMs, the findings of this study suggest that LLM-supported role inference could serve as a comparable mechanism to educators in facilitating equitable group composition, thereby fostering conditions for reciprocity, accountability, and collaboration in NL spaces.

The results for RQ2 revealed that the length of learner responses was associated with the degree of alignment between LLM and human-coded role inference. Specifically, there was greater agreement between LLM and human role inference for participants who consistently provided longer and more elaborated responses throughout the chatbot conversations. Conversely, when participants predominantly provided shorter responses, the agreement between LLM and human inference decreased. This pattern is expected, as longer messages are likely to contain more detailed reasoning and richer self-descriptions in response to the chatbot's questions (**Error! Reference source not found.**), thereby enabling the LLM to make more accurate role inferences compared to shorter messages. Human coders may rely on clear, concrete evidence before assigning a role, whereas the LLM may categorise a response under a role based on more subtle or partial indications. As a result, in shorter responses, humans often assigned fewer roles than the LLM, leading to lower agreement. This discrepancy between LLM and human role inference is consistent with recent work of Zhang et al. (2025) which compared GPT-4 and human coders in inductive and deductive coding tasks, showing that the GenAI applied codes more consistently, even when evidence was limited, while human coders tended to wait for clearer justification before coding. Similarly, Chen et al. (2025) in their study of online discourse data found that AI coders generated codes grounded mainly in narrow textual content, whereas human coders contributed more contextually grounded codes based on conversational dynamics. This interpretation is further extended by our observed pattern in cases where participants' responses were predominantly very short (concise), in which the agreement between LLM and human role inference actually increased. The likely reason is that when responses are extremely short (i.e. less than five words), there is insufficient information for either humans or LLM to identify meaningful role indicators (**Error! Reference source not found.**). Therefore, both assign few or no roles, resulting in higher agreement.

## Conclusion

This study examined the use of LLMs to infer learners' teamwork role tendencies from chatbot-mediated dialogues and compared these inferences with human coding. The findings showed that LLM outputs reached a moderate level of alignment with human judgments and that learners' response length is associated with the degree of agreement, with longer and more elaborated responses showing more aligned role inference between LLM and humans. From a design perspective, these results highlight the potential of LLMs to support the inference of teamwork role tendencies in ways that closely parallel human judgment, while reducing instructor workload and data requirements. At the same time, the study shows that the length of learner responses might matter to the agreement between LLM and human role inference. Consequently, the practical deployment of similar tools should promote more profound and elaborated learner contributions, including open-ended prompts and reflective questioning. Such design considerations can augment the environment for NL by facilitating technology use in supporting collaboration while also encouraging learners to engage in more substantial exchanges with a chatbot that contributes to the construction of meaning. This approach thereby fosters connection building through strengthening NL relationships, reciprocity, and meaningful interactions. Attending to these design features can help educators establish conditions where LLM enhances administrative efficiency while also elevating learners' awareness of their teamwork role tendencies and bolsters opportunities for meaningful engagement within social online NL. While these implications are promising, the present study was limited by a modest sample size (56 participants from Prolific) and the absence of more advanced statistical testing to capture the effects of interaction quality. In addition, reliance on Prolific introduces potential self-selection bias and limits generalisability, as participants may not fully represent learners engaged in authentic higher-education settings where collaboration unfolds within real curricular and social contexts. Building on this foundation, future research will expand to larger cohorts and extend to real-world contexts, allowing a rigorous evaluation of how LLM-based role inference can enhance group formation, role awareness, and ultimately strengthen social NL experiences and outcomes.

## References

- Aranzabal, A., Epelde, E., & Artetxe, M. (2022). *Team formation on the basis of Belbin's roles to enhance students' performance in project based learning. Education for Chemical Engineers, 38*, 22–37.
- Belbin, R. M. (2010). *Team roles at work*. Routledge.

- Biddle, B. J. (1986). Recent developments in role theory. *Annual Review of Sociology*, 12(1986), 67–92. <https://doi.org/10.1146/annurev.so.12.080186.000435>
- Borse, N. S., Subramaniam, R. C., & Rebello, N. S. (2025). Investigation of the Inter-Rater Reliability between Large Language Models and Human Raters in Qualitative Analysis (No. arXiv:2508.14764). arXiv.
- Chen, C.-M., & Kuo, C.-H. (2019). An optimized group formation scheme to promote collaborative problem-based learning. *Computers & Education*, 133, 94–115. <https://doi.org/10.1016/j.compedu.2019.01.011>
- Chen, J., Lotsos, A., Wang, G., Zhao, L., Sherin, B., Wilensky, U., & Horn, M. (2025). Processes matter: How ML/GAI approaches could support open qualitative coding of online discourse datasets (No. arXiv:2504.02887). arXiv. <https://doi.org/10.48550/arXiv.2504.02887>
- Chen, Y., Zhang, L., Ding, Y., Guo, L., & Bian, K. (2025). Multi-task oriented team formation in online collaborative learning. *Expert Systems with Applications*, 259, 125289. <https://doi.org/10.1016/j.eswa.2024.125289>
- Cheng, B., Wang, M., & Mercer, N. (2014). Effects of role assignment in concept mapping mediated small group learning. *The Internet and Higher Education*, 23, 27–38. <https://doi.org/10.1016/j.iheduc.2014.06.001>
- Danahy, E. E., Vora, M. K., Xu, Y. M., & Church, W. (2024, July 28). Full Paper: A generative AI approach to better teamwork in first-year engineering. 15th Annual First-Year Engineering Experience Conference (FYEE). <https://peer.asee.org/full-paper-a-generative-ai-approach-to-better-teamwork-in-first-year-engineering>
- De Felice, S., Hamilton, A. F. de C., Ponari, M., & Vigliocco, G. (2022). Learning from others is good, with others is better: The role of social interaction in human acquisition of new knowledge. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 378(1870), 20210357.
- Düker, J., & Rieber, A. (2024). Performance, knowledge acquisition and satisfaction in self-selected groups: Evidence from a classroom field experiment (No. arXiv:2403.12694). arXiv.
- Ergulec, F. (2019). Instructional strategies for forming online collaborative teams. *International Journal on E-Learning*, 18(4), 349–372.
- Figuroa, H., Costaguta, R., Menini, M. de los Á., & Missio, D. (2019). An automatic identification of team roles in forums. *Proceedings of the XX International Conference on Human Computer Interaction*, 1–2.
- Gardeli, A., & Vosinakis, S. (2025). Group dynamics in collaborative learning: Impact of emergent and scripted roles in tangible mobile augmented reality games. *Computers & Education: X Reality*, 7, 100102.
- Goodyear, P., Banks, S., Hodgson, V., & McConnell, D. (2004). Research on networked learning: An overview. In P. Dillenbourg, M. Baker, C. Bereiter, Y. Engeström, G. Fischer, H. Ulrich Hoppe, T. Koschmann, N. Miyake, C. O'Malley, R. Pea, C. Pontecorovo, J. Roschelle, D. Suthers, P. Goodyear, S. Banks, V. Hodgson, & D. McConnell (Eds), *Advances in Research on Networked Learning* (pp. 1–9). Springer Netherlands.
- Hasan, M., Babu, M., & Al-Emran. (2024). A machine learning approach to optimal group formation based on previous academic performance. *Machine Learning Research*, 9, 48–52.
- Johnson, D. W., & Johnson, R. T. (1999). What makes cooperative learning work. <https://eric.ed.gov/?id=ED437841>
- Kotera, Y., Chircop, J., Hutchinson, L., Rhodes, C., Green, P., Jones, R.-M., Kaluzeviciute, G., & Garip, G. (2021). Loneliness in online students with disabilities: Qualitative investigation for experience, understanding and solutions. *International Journal of Educational Technology in Higher Education*, 18(1).
- Luo, H., Chen, Y., Chen, T., Koszalka, T. A., & Feng, Q. (2023). Impact of role assignment and group size on asynchronous online discussion: An experimental study. *Computers & Education*, 192, 104658.
- Maina, E., Oboko, R., & Wagacha, P. (2017). Using machine learning techniques to support group formation in an online collaborative learning environment. *International Journal of Intelligent Systems and Applications*, 9, 26–33. <https://doi.org/10.5815/ijisa.2017.03.04>
- McCollum, B., Morsch, L., Shokoples, B., & Skagen, D. (2019). Overcoming barriers for implementing international online collaborative assignments in chemistry. *The Canadian Journal for the Scholarship of Teaching and Learning*, 10(1). <https://doi.org/10.5206/cjsotl-rcacea.2019.1.8004>
- Muuro, M. E., Wagacha, W. P., Oboko, R., & Kihoro, J. (2014). Students' perceived challenges in an online collaborative learning environment: A case of higher learning institutions in Nairobi, Kenya. *International Review of Research in Open and Distance Learning*, 15(6), 132–161. Scopus.
- Networked Learning Editorial Collective (NLEC). (2021). Networked Learning: Inviting redefinition. *Postdigital Science and Education*, 3(2), 312–325. <https://doi.org/10.1007/s42438-020-00167-8>
- Networked Learning Editorial Collective (NLEC), Gourlay, L., Rodríguez-Illera, J. L., Barberà, E., Bali, M., Gachago, D., Pallitt, N., Jones, C., Bayne, S., Hansen, S. B., Hrastinski, S., Jaldemark, J., Themelis, C.,

- Pischetola, M., Dirckinck-Holmfeld, L., Matthews, A., Gulson, K. N., Lee, K., Bligh, B., ... Knox, J. (2021). *Networked Learning in 2021: A community definition*. *Postdigital Science and Education*, 3(2), 326–369.
- O'Donnell, A. (1992). *Scripted cooperation in student dyads: A method for analyzing and enhancing academic learning and performance*. In *Interaction in cooperative groups: The theoretical anatomy of group learning*. Oyarzun, B., & Martin, F. (2023). *A systematic review of research on online learner collaboration from 2012–21: Collaboration technologies, design, facilitation, and outcomes*. 27(1).
- Pei, L., Poortman, C., Schildkamp, K., & Benes, N. (2024). *Teachers' and students' perceptions of a sense of community in blended education*. *Education and Information Technologies*, 29(2), 2117–2155.
- Pillai, A. S. (2022). *A natural language processing approach to grouping students by shared interests*. *Journal of Social Science Studies*, 6, 1–16.
- Pongpaichet, S., Nirunwiroj, K., & Tuarob, S. (2022). *Automatic assessment and identification of leadership in college students*. *IEEE Access*, 10, 79041–79060. <https://doi.org/10.1109/ACCESS.2022.3193935>
- Salama, R., & Hinton, T. (2023). *Online higher education: Current landscape and future trends*. *Journal of Further and Higher Education*, 47(7), 913–924. <https://doi.org/10.1080/0309877X.2023.2200136>
- Scheidler, B., Layne, K., Gillard, M., & Doucet, B. M. (2025). *Intentional grouping: Impacts on cognitive and social presence in a hybrid classroom*.
- Siemens, G. (2005). *Connectivism: A learning theory for the digital age*. *International Journal of Instructional Technology & Distance Learning*, 2(1).
- Smith, J., James, A., & Joe, G. (2023). *Applying deep learning for personalized study group creation*. [https://www.researchgate.net/publication/389676039\\_Applying\\_Deep\\_Learning\\_for\\_Personalized\\_Study\\_Group\\_Creation](https://www.researchgate.net/publication/389676039_Applying_Deep_Learning_for_Personalized_Study_Group_Creation)
- Strijbos, J.-W., & De Laat, M. F. (2010). *Developing the role concept for computer-supported collaborative learning: An explorative synthesis*. *Computers in Human Behavior*, 26(4), 495–505.
- Vygotsky, L. S. (1978). *Mind in Society: Development of higher psychological processes*. Harvard University Press. <https://doi.org/10.2307/j.ctvjf9vz4>
- Wijaya, M. M. J., Wahyu Utomo, D., & Kurniawan, D. (2019). *Implementing dynamic group formation in web-based collaborative learning for high school*. *2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, 528–533. <https://doi.org/10.1109/ISRITI48646.2019.9034625>
- Wijenayake, P., Silva, D. de, Alahakoon, D., & Kirigeegamage, S. (2020). *Automated detection of social roles in online communities using deep learning*. *Proceedings of the 3rd International Conference on Software Engineering and Information Management*, 63–68. <https://doi.org/10.1145/3378936.3378973>
- Zhang, H., Wu, C., Xie, J., Rubino, F., Graver, S., Cai, J., Kim, C., & Carroll, J. (2025). *Exploring inductive and deductive qualitative coding with AI: Investigating inter-rater reliability between large language model and human coders*. *Human-Computer Interaction & Emerging Technologies*, 195(195). <https://doi.org/10.54941/ahfe1006232>