Article

Identifying Cognitive Biases in Wayfinding Language at Shanghai Hongqiao Hub Using Large Language Model Role-Play

Yifan Wei[1] and Shengchuan Jiang[1,*]

[1] University of Shanghai for Science and Technology, Shanghai, China
* Correspondence: Shengchuan Jiang, University of Shanghai for Science and Technology, Shanghai, China

Abstract: Indoor transfers at integrated passenger transport hubs often involve changing modes and levels, walking along long corridors, and locating pick-up points. This study develops a method for screening wayfinding language using large language model (LLM) role-play. Scene descriptions drawn from four transfer-route videos at Shanghai Hongqiao Hub informed the prompts given to DeepSeek. The model generated 72 responses, from which 311 cues to possible interpretation or decision difficulties were extracted and coded. Task matching (77 cues) and action instructions (73 cues) predominated in the simulations. In a survey of 309 respondents, walking time and distance were the most frequently selected additions to signage. Route confirmation and routes suited to travelers with luggage also received high item ratings. In exploratory comparisons, respondents who often or sometimes experienced time pressure rated a combined item on walking time, going upstairs, and distance to ticket checks more highly. Respondents who had transferred with children selected information on elevators, escalators, and accessible routes more frequently. The method identifies issues for follow-up, and the survey indicates which information respondents value.

Keywords: large language models; transport hubs; wayfinding language; cognitive biases; role-play

1. Introduction

Integrated passenger transport hubs connect high-speed rail, metro, taxi, and ride-hailing services. Indoor pedestrian transfers often involve long walks, level changes, decisions at junctions, and locating pick-up points. At large hubs such as Shanghai Hongqiao, wayfinding information identifies destinations and shapes passengers' immediate judgments about which transfer they are making, where to go next, and how much time remains.

Research on wayfinding at integrated passenger transport hubs has examined how pedestrians' decisions evolve and vary across groups, the placement and organization of wayfinding information, evaluation of high-speed railway station signage, transfer-space design, pedestrian convenience, and the convenience of transfer services [1-8]. Methodological studies have drawn on cognitive psychology, visual behavior at different levels, building information modeling (BIM) and virtual reality (VR), and eye tracking and wayfinding performance [9-13]. Other studies have examined indoor wayfinding under crowded conditions, the use of cues under time pressure, sign content and placement, dynamic signage, and differences in visual attention [14-18]. Field experiments, eye-tracking studies, and VR simulations typically require considerable resources to arrange sites, equipment, and participants. These constraints limit the passenger profiles, route combinations, and travel conditions that can be studied. They also make it difficult to use the findings for routine, large-scale audits of wayfinding language.

Large language models (LLMs) offer new ways to interpret transportation scenes and organize unstructured information. Tang et al. used multimodal LLMs to identify road safety factors in street-view images [19]. Their work suggests that LLMs can use scene

Received: 08 September 2026
Revised: 26 October 2026
Accepted: 07 November 2026
Published: 10 November 2026

Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

material to identify potential issues and articulate why they may matter in complex transportation environments. Gao et al. used LLMs to detect ambiguity in requirements, illustrating their ability to interpret meaning and logical relationships. Rule-based checks and human review can improve the consistency of these results [20]. Yuan examined ChatGPT's semantic understanding and commonsense reasoning, providing a basis for applying LLMs to language and context interpretation [21]. Milano et al. used LLM embeddings to analyze the semantic structure of test items, demonstrating their use in structuring text for analysis [22]. Wilson's account of embodied cognition links cognition to bodily states, interaction with the environment, and the demands of action [23]. Barsalou's account of grounded cognition likewise connects cognition to perception, action, and bodily experience [24]. These studies inform the use of LLMs for systematic semantic audits of wayfinding language in transport hubs.

This study asks what potential wayfinding difficulties LLM role-play can surface during indoor transfers and how those cues correspond to passengers' reported information needs. Four videos of typical routes at Shanghai Hongqiao Hub provided scene details for the DeepSeek prompts. The study coded the generated cues and used 309 valid survey responses to compare them with reported needs and examine differences across respondent groups. The aim was to identify possible mismatches between the wording on signs and the information passengers need to make transfer decisions.

2. Classification Framework and Scene Selection

The framework defines how the study classified cognitive bias cues and how the recorded scenes informed the role-play inputs.

2.1. Classification of Cognitive Biases

In this study, the term "cognitive bias cue" refers to a possible misinterpretation or uncertainty about a transfer destination, route, level change, access condition, or expected time and distance. The cues indicate potential difficulties raised in the model responses. The framework considers sign wording in relation to the passenger role, decision point, transfer task, and next action. The coding categories draw on research on wayfinding behavior, signage evaluation, spatial experience, and embodied cognition, as well as patterns in the model responses.

The researchers identified questions, expressions of hesitation, and requests for information in the model responses and coded them using the criteria in Table 1. The "Other or untested" category retained cues outside the main categories or with limited survey coverage, including rest areas and crowd conditions.

Table 1. Classification Framework for Cognitive Biases in Wayfinding Language

Category	Coding criterion	Research or scene basis	Illustrative example from role-play
Task matching	The passenger is unsure whether the sign refers to their transfer task or destination.	Wayfinding behavior and sign placement	"Is South 1 the metro entrance I need?"
Action instructions	The passenger is unsure whether to go straight, turn, change levels, or enter an area.	Transfer processes, cognitive psychology, and dynamic signage	"Which way should I go at the junction?"
Route suitability	The passenger is unsure whether the route is suitable when carrying luggage, accompanying children, or traveling with age-related or other physical limitations.	Spatial experience, crowding, and embodied travel	"Will I still have to change levels with my luggage?"

Information continuity	The passenger is unsure whether signs will confirm the route in a long corridor, at a corner, or at a junction.	Visual search, sign placement, and continuity of wayfinding information	"Will there be another sign halfway down the corridor?"
Route recovery	The passenger is unsure how to reorient or return to the main route after a wrong turn.	Wayfinding performance and correction of route errors	"How do I get back if I go the wrong way?"
Journey expectations	The passenger is concerned about remaining time or distance, or whether they can complete the transfer in time.	Time pressure, embodied cognition, and travel experience	"How much longer will it take? Will I make it in time?"
Other or untested	The cue does not fit consistently into the categories above or was not measured directly in the survey.	Follow-up interviews, field observations, or expanded surveys	"Is there anywhere to rest along the way?"

2.2. Scene Anchors and Role-Play Inputs

Four videos of typical indoor transfer routes at Shanghai Hongqiao Hub provided the scene anchors. The routes covered transfers from high-speed rail to metro, ride-hailing, and taxi services, as well as the walk from the metro to the high-speed rail departures level. The prompts used scene details drawn from the videos, including the start and destination, sign wording, level changes, corridor characteristics, and key decision points.

The six roles were a passenger with no special constraints, a first-time visitor, a passenger carrying bulky luggage, a passenger traveling with children, an older passenger, and a passenger under time pressure. Each of the four scenes was combined with all six roles and simulated three times, yielding 72 responses. The prompts asked for initial interpretations, likely next actions, possible questions, and the reasons for each judgment. The prompts did not supply the coding labels, which were applied after the simulations. Table 2 lists the scene anchors and illustrates concerns raised for selected roles.

Table 2. Scene Anchors and Inputs for Passenger Role-Play

Scene	Video scene and sign wording	Illustrative role-related concerns	Main cues
S1 High-speed rail to metro	South 2 rail exit to South 1 metro entrance. Signs read "Metro" and "Entrance to Lines 2, 10, and 17 (South 1)."	First-time visitors, passengers with luggage, and those under time pressure focus on the South 1/South 2 relationship and confirmation along the corridor.	Entrance codes, route confirmation, and direction at junctions
S2 High-speed rail to ride-hailing	Rail exit to the P10-B1 ride-hailing pick-up point, through parking and pick-up areas.	Passengers with luggage, children, or time pressure focus on the P10-B1 code, level changes, and entry into the vehicle area.	Pick-up code, level changes, and routes requiring less effort
S3 High-speed rail to taxi	Rail exit to the South 2 taxi pick-up point. "Taxi" identifies the destination directly.	Passengers with no special constraints and those under time pressure focus on confirming the destination and route, and estimating walking time.	Destination confirmation, route confirmation, and walking time and distance
S4 Metro to high-	Line 10 exit to the high-speed rail departures level. Signs	First-time visitors, older passengers, and those under	Destination wording, level

speed rail	read "Train," "High-Speed Rail Departures," and "2F." Observed walk: about 2 min 45 s.	time pressure focus on "train" versus "high-speed rail," whether to go upstairs, and walking time.	changes, and estimated walking time

3. LLM Role-Play and Extraction of Cognitive Bias Cues

The researchers split the 72 DeepSeek responses into individual cues, removed duplicates, and coded the resulting material. The 311 cues included 77 related to task matching, 73 to action instructions, 55 to journey expectations, 43 to route recovery, 33 to information continuity, 27 to route suitability, and 3 classified as other or untested (Figure 1). The counts describe the distribution of coded cues in the simulations.

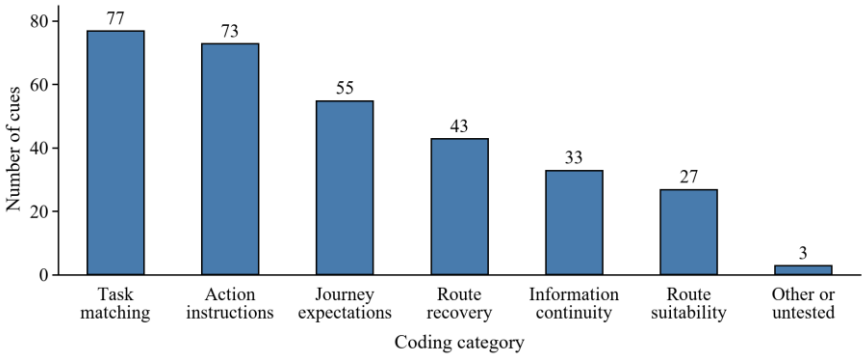


Figure 1. Cognitive Bias Cues Coded from Non-Leading DeepSeek Simulations

Task matching and action instructions were the most frequent categories. Task-matching cues often arose around codes such as "South 1," "South 2," and "P10-B1," where the simulated passengers had to connect a location code to their transfer destination. Action-instruction cues arose mainly at level changes, junctions, and transitions between modes, where a destination name alone could leave the next action unclear.

Cues about journey expectations included questions about the remaining walk and whether the passenger would arrive in time. Route-recovery cues concerned reorientation after a wrong turn, while information-continuity cues concerned confirmation along the way. Both types arose around long corridors, corners, and transitions into parking areas. The route-suitability cues addressed conditions such as traveling with luggage or children and finding an accessible route.

The coded responses supplied candidate issues for comparison with the survey, including cases where a frequently raised simulation cue might receive relatively limited survey support.

4. Survey Evidence on Passenger Needs

The survey analysis describes overall support for different types of information, maps survey responses to simulation cues, and examines differences across passenger groups.

4.1. Distribution of Wayfinding Information Needs

The survey yielded 309 valid responses. Of the respondents, 87.4% visited Hongqiao Hub frequently or had visited several times, 51.8% carried bulky luggage on their most recent transfer at a large hub, and 32.7% had transferred while accompanying children. A further 60.2% reported time pressure often or sometimes, and 41.4% were not very or not at all familiar with location codes such as "South 1," "South 2," "P10," and "B1."

Items 9–35 used a five-point Likert scale from 1 (strongly disagree) to 5 (strongly agree). Figure 2 shows selected items with high mean ratings. Ratings averaged 4.79 for signs indicating elevators, ramps, accessible routes, or routes with fewer stairs when

carrying bulky luggage, and 4.46 for a similar luggage item specific to the ride-hailing route. Route confirmation along the way, even when the destination was clear, also averaged 4.46. The combined item on walking time, whether to go upstairs, and distance to ticket checks under time pressure averaged 4.38. Walking time to the taxi pick-up point averaged 4.33, and estimated walking time from the metro exit to the high-speed rail departures level averaged 4.31.

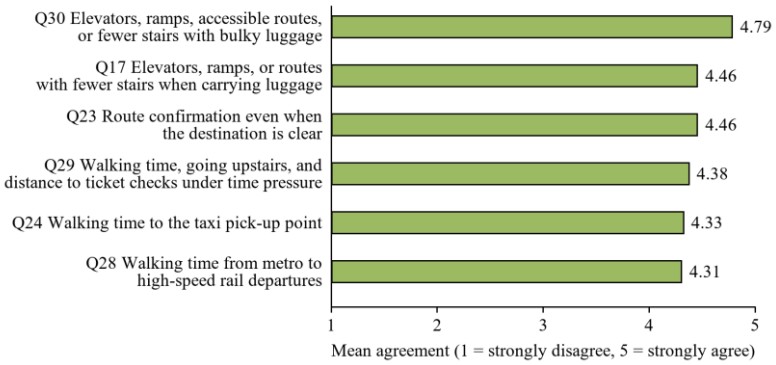


Figure 2. Selected Wayfinding Information Needs with High Mean Survey Ratings

Item Q37 asked respondents to select the three most desired additions to signage. The export recorded four to six selections per respondent, so each percentage is the share of the 309 respondents who selected that option. Walking time (79.0%) and distance (78.3%) were selected most often, followed by mid-corridor confirmation signs (59.5%) and crowding alerts (56.6%). Elevators, escalators, and accessible routes were selected by 50.8%, pick-up point code explanations by 49.2%, and level-change instructions by 48.9%.

4.2. Correspondence between Simulation Cues and Survey Responses

Table 3 maps simulation cues to survey evidence. Strong agreement required an item mean of at least 4.20 or a selection rate of at least 70%. If neither cutoff was met, agreement required a mean of at least 3.80 or a selection rate of at least 60%. Support below these cutoffs was classified as partial agreement. When support above a cutoff was limited to a specific route and the broader measure fell below the cutoff, the need was classified as context-dependent. Cues without a direct survey measure were classified as untested.

Table 3. Correspondence between LLM Simulation Cues and Reported Passenger Needs

Main cue from role-play	Survey indicator	Agreement	Interpretation
Remaining time and distance, and whether the passenger will arrive in time	Q28 mean 4.31 and Q24 mean 4.33. Walking time and distance selected by 79.0% and 78.3%, respectively.	Strong agreement	Both the survey item means and the multiple-response selection rates met the strong-agreement thresholds.
Lack of route confirmation midway along a long corridor or route	Q23 mean 4.46 and Q33 mean 4.31. Mid-corridor confirmation signs were selected by 59.5%.	Strong agreement	The item means met the strong-agreement threshold. The selection rate was below 70%.
Route suitability for passengers carrying bulky luggage	Q30 mean 4.79 and Q17 mean 4.46. Elevators, escalators, and accessible routes selected by 50.8%.	Strong agreement	The luggage-related item means met the strong-agreement threshold. The selection rate for the bundled facilities option was below 70%.

Whether to go up or down, or whether the route leads to 2F	Q27 mean 4.17 for an explicit instruction to go up to the high-speed rail departures level. Level-change directions were selected by 48.9%.	Context-dependent	The stronger support concerned a specific transfer route.
Whether codes such as South 1 or P10-B1 match the destination	Pick-up point code explanations were selected by 49.2%.	Partial agreement	Some respondents wanted code explanations. More respondents selected walking time, distance, and route confirmation.
How to return to the main route after a wrong turn	Q34 mean 4.11. Directions for recovering from a wrong turn were selected by 34.3%.	Agreement	The item mean met the agreement threshold. Fewer respondents selected this option in the multiple-response question.
Other cues, including rest areas and crowd conditions	Crowding alerts selected by 56.6%. Rest areas were not adequately covered by direct survey items.	Crowding: partial agreement. Rest areas: untested.	Respondents expressed some demand for crowding alerts. Rest areas warrant further interviews or field observation.

Walking-time item means and selection rates for walking time and distance met the strong-agreement thresholds. Route confirmation and route suitability when carrying luggage reached strong agreement through item means, with lower multiple-response selection rates. Route recovery met the agreement threshold through its item mean. The evidence for level-change instructions was more specific to the route from the metro to high-speed rail. Code explanations and crowding alerts received some support, and the survey did not directly assess rest areas.

4.3. Differences in Wayfinding Needs Across Passenger Groups

Table 4 reports exploratory comparisons across respondent groups. Independent-samples t-tests compared means for the five-point items. The table reports Welch tests for Q29 and Q32 and pooled-variance tests for the remaining mean comparisons. A Pearson chi-square test compared selection rates for Q37D. The t-tests were two-sided. All p-values were unadjusted, and 0.05 was used as the nominal significance threshold.

Table 4. Group Differences in Key Wayfinding Information Needs.

Groups compared	Survey item	Group results	Test	Statistic	p	Interpretation
Bulky luggage vs. no bulky luggage	Q30 Elevators, ramps, and accessible routes	4.80 ± 0.64 vs. 4.79 ± 0.60	t-test	t = 0.209	0.835	No significant difference. Both groups rated the need highly.
Bulky luggage vs. no bulky luggage	Q17 Elevators, ramps, or routes with fewer stairs	4.48 ± 0.74 vs. 4.44 ± 0.78	t-test	t = 0.443	0.658	No significant difference by luggage status.
Time pressure often or sometimes vs. less often	Q29 Remaining walk time, whether to go upstairs, and distance to ticket checks	4.46 ± 0.81 vs. 4.26 ± 0.91	Welch t-test	t = 1.995	0.047*	The more frequent time-pressure group rated the combined item higher.

Time pressure often or sometimes vs. less often	Q28 Estimated walking time from metro to high-speed rail	4.32 ± 0.85 vs. 4.29 ± 0.86	t-test	t = 0.247	0.805	No significant difference. Both groups rated the need highly.
Time pressure often or sometimes vs. less often	Q32 Brief, direct wording that can be read at a glance	3.82 ± 0.99 vs. 4.11 ± 0.99	Welch t-test	t = −2.506	0.013*	The more frequent time-pressure group rated brief wording lower.
Transfer with children vs. no such experience	Q37D Elevators, escalators, and accessible routes	63.4% vs. 44.7%	Pearson χ^2	χ^2 = 9.466	0.002**	Respondents who had transferred with children selected this option significantly more often.
Unfamiliar with location codes vs. other respondents	Q14 Understanding P10 and B1	4.13 ± 1.00 vs. 4.22 ± 0.99	t-test	t = −0.836	0.404	No significant difference in reported code understanding.

Note. * $p < 0.05$. ** $p < 0.01$. Q37D results are selection rates. For Q14, the group that was not very or not at all familiar with location codes (n = 128) was compared with all other respondents (n = 181).

Respondents with and without bulky luggage rated Q30 and Q17 highly, with no statistically significant differences in the two comparisons. The group reporting time pressure often or sometimes rated the combined Q29 item more highly (4.46 vs. 4.26, $p = 0.047$). Q29 covers walking time, whether to go upstairs, and distance to ticket checks in one rating. The same group rated brief, easily scanned wording lower on Q32 (3.82 vs. 4.11, $p = 0.013$). Both results should be interpreted in light of the exploratory comparisons.

Among respondents who had transferred with children, 63.4% selected Q37D, compared with 44.7% of those without that experience ($\chi^2 = 9.466$, $p = 0.002$). This item groups elevators, escalators, and accessible routes into one option. Q14 showed no statistically significant difference in reported understanding of P10 and B1 between the location-code familiarity groups.

5. Conclusion

Videos of four routes at Shanghai Hongqiao Hub informed DeepSeek role-play for six passenger roles. Three repetitions per route and role yielded 72 responses and 311 coded cues. The method can screen sign wording across route and role combinations before field testing.

Task matching and action instructions were the most frequent simulation cues. Walking-time ratings and selection rates for walking time and distance met the strong-agreement thresholds. Route confirmation and luggage-related route suitability met the threshold on item ratings, with lower selection rates.

In exploratory comparisons, respondents who more often experienced time pressure rated the combined Q29 item higher. Those who had transferred with children more often selected the Q37D facilities option. Both luggage groups rated luggage-related route information highly.

The study covered one hub, one model, and self-reported needs. Q37 percentages describe option-level selections because respondents exceeded the requested limit. Field studies across hubs could test whether the cues correspond to difficulties during actual transfers.

References

1. Z. Dai, D. Li, Y. Feng, and C. Yang, "Decision evolution and heterogeneity aware pedestrian wayfinding behaviour modelling in VR integrated transportation hub," *Transportation Research Part C: Emerging Technologies*, vol. 186, p. 105595, 2026.
2. Z. Zhang, "Research on Evaluation of Hub Guidance Information Layout Based on Wayfinding Behavior," M.S. thesis, Beijing University of Civil Engineering and Architecture, Beijing, China, 2023.

3. W. Wang, "Research on Guiding-signs Indication Information Setting in Comprehensive Transport Passenger Hub," M.S. thesis, Beijing University of Civil Engineering and Architecture, Beijing, China, 2021.
4. X. Gao, "Strategies for optimizing wayfinding signage systems in integrated urban passenger transport hubs," *Transportation Science and Technology and Management*, vol. 4, no. 20, pp. 173–175, 2023.
5. J. Niu, "Research on the Evaluation and Optimization of the Way-Finding System Design for High-Speed Railway Stations," M.S. thesis, Shenyang Jianzhu University, Shenyang, China, 2025.
6. K. Jin, "Research on Optimisation Design of Interchange Space Environment of Large High-speed Railway Station Based on Spatial Syntax Approach," M.S. thesis, Shenyang Jianzhu University, Shenyang, China, 2025.
7. D. Bao, Z. Chen, J. Zhou, and X. Yao, "Evaluating pedestrian convenience in integrated airport transport hubs based on spatial experience," *Journal of Northeastern University (Natural Science)*, vol. 44, no. 9, pp. 1359–1368, 2023.
8. D. Yuan, "Problems and countermeasures for transfer services at integrated passenger transport hubs based on a survey of transfer convenience at high-speed railway hubs," *Journal of Transportation Engineering*, vol. 20, no. 3, pp. 35–38, 45, 2020.
9. F. Peng and Z. Zhang, "Detailed design of a wayfinding signage system for integrated passenger transport hubs based on cognitive psychology," *Journal of Safety Science and Technology*, vol. 20, no. 9, pp. 165–172, 2024.
10. J. Lin, Z. Zhang, F. Lin, X. Cao, and H. Cui, "Evaluation indicators for wayfinding information at integrated passenger transport hubs based on visual behavior at different levels," *Railway Transport and Economy*, vol. 45, no. 2, pp. 54–60, 67, 2023.
11. X. Zhao, G. Ren, B. Wang, and Z. Chen, "Comprehensive evaluation of signage systems at passenger transport hubs using BIM and VR," *Journal of Beijing University of Technology*, vol. 46, no. 1, pp. 51–57, 2020.
12. W. Jin, Y. Yao, G. Ren, and X. Zhao, "Evaluation of integration information signage in transport hubs based on building information modeling and virtual reality technologies," *Sustainability*, vol. 14, no. 3, p. 1171, 2022.
13. L. Lan, "Wayfinding Performance Evaluation in Rail Transit Stations Based on Pedestrians' Visual Search Characteristics and Behaviors," M.S. thesis, Chongqing Jiaotong University, Chongqing, China, 2023.
14. Y. Li, "Effects of wayfinding signage and crowding on indoor wayfinding performance and ways to improve it," M.S. thesis, Shaanxi Normal University, Xi'an, China, 2022.
15. X. Yang, "Study on the Influence of Wayfinding Clues on Pedestrian Behavior under Time Pressure in Underground Space," M.S. thesis, Beijing Jiaotong University, Beijing, China, 2020.
16. M. Zhang, "Study on the Content and Location Setting of Guide Signs in Transport Hub Based on Function Area Layout and Passenger Flow Organization," M.S. thesis, Beijing Jiaotong University, Beijing, China, 2020.
17. B. Wang and W. Zhang, "Dynamic guidance sign design for passenger transfer in city railway hubs," in *IOP Conference Series: Materials Science and Engineering*, vol. 563, p. 042005, 2019.
18. R. Xu, H. Xia, and M. Tian, "Wayfinding design in transportation architecture: Are saliency models or designer visual attention a good predictor of passenger visual attention?," *Frontiers of Architectural Research*, vol. 9, no. 2, pp. 321–335, 2020.
19. Y. Tang, A. Qu, X. Yu, W. Deng, J. Ma, J. Zhao, and L. Sun, "From street views to urban science: Discovering road safety factors with multimodal large language models," *Transportation Research Part C: Emerging Technologies*, vol. 188, p. 105692, 2026.
20. J. Gao, F. Liu, and Y. Yang, "A method for detecting requirements ambiguity using large language models," *Journal of Computer Research and Development*, vol. 63, no. 4, pp. 827–853, 2026.
21. Y. Yuan, "How to test ChatGPT's performance in semantic understanding and common-sense reasoning: Challenges and opportunities of linguistics in the era of large language models," *Chinese Journal of Language Policy and Planning*, vol. 9, no. 1, pp. 49–63, 2024.
22. N. Milano, M. Luongo, M. Ponticorvo, and D. Marocco, "Semantic analysis of test items through large language model embeddings predicts a-priori factorial structure of personality tests," *Current Research in Behavioral Sciences*, vol. 8, p. 100168, 2025.
23. M. Wilson, "Six views of embodied cognition," *Psychonomic Bulletin & Review*, vol. 9, no. 4, pp. 625–636, 2002.
24. L. W. Barsalou, "Grounded cognition," *Annual Review of Psychology*, vol. 59, pp. 617–645, 2008.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of Publisher and/or the editor(s). Publisher and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.