Digital innovation in EFL through inclusive AI:
The University of Macerata case study within the UNITE project

Francesca Raffi
University of Macerata
f.raffi@unimc.it

1.   Introduction

Artificial intelligence is transforming multiple domains of education, expanding opportunities for English language learning and teaching while also raising important questions about access, equity, and inclusion (Bahroun et al. 2023). Within this evolving landscape, the *Universally Inclusive Technologies to Practice English* (UNITE) project investigates how AI dialogue systems can be integrated into English as a Foreign Language (EFL) instruction when their use is guided by the Universal Design for Learning (UDL) framework (CAST 2024). The project aims to understand the pedagogical and technological conditions that enable chatbots to foster engagement, comprehension, and learner agency without introducing additional cognitive or emotional barriers (Fryer et al. 2019; Reinders and Benson 2017; Ushioda 2011; Dörnyei and Ryan, 2015).

This report focuses on the University of Macerata case study, which explores students' experiences of chatbot-mediated interaction through a structured experimental design. Participants engaged in guided conversations with ChatGPT and Pi.ai, followed by a post-task questionnaire. The analysis concentrates on two key experiential dimensions – immersion in conversational flow and perceived quality of interaction – which closely align with UDL's three principles of engagement, representation, and action/expression (CAST 2024). Among the project's main outputs, the *Chatbot Accessibility Checklist*, developed by the University of Macerata

research unit in collaboration with the Center for Applied Special Technology (CAST), the organization that originally designed the UDL framework. The checklist operationalizes these insights into UDL-based criteria for the pedagogical and technical integration of AI chatbots in higher education.

2.   Project context and rationale

UNITE is a three-year research project (2022-2025) led by the University of Bologna in collaboration with the University of Macerata and the University of Naples "L'Orientale" (https://site.unibo.it/unite/en). Funded by the European Union – NextGenerationEU through the Italian Ministry of University and Research, the project promotes inclusive digital transformation in higher education. It follows a progressive research pathway – from exploration and selection to testing and development – designed to build an evidence base for the responsible use of AI in EFL teaching. A systematic literature review and an analysis of commonly used chatbots informed the shared experimental protocol adopted by all three partner institutions (Bibauw et al. 2022; Huang et al. 2022).

The project's rationale builds on two complementary strands. First, chatbots can extend opportunities for authentic communicative practice, mitigate performance anxiety, and provide immediate, context-aware feedback (Huang et al. 2022; Fryer et al. 2019; Ortiz-Garcés et al. 2024). Second, when aligned with UDL principles, these systems can accommodate diverse learner profiles by enabling control over pacing, task complexity, and feedback modality, thereby enhancing participation and equity (Bray et al. 2024; Morgan 2024). This inclusive perspective resonates with Sustainable Development Goal 4 (SDG 4 – Quality Education), which calls for equitable access and meaningful learning opportunities for all (United Nations 2015). The University of Macerata case study contributes to this agenda by showing how the design of human–AI interaction influences learner engagement and emotional comfort in concrete educational settings.

3.   Methodological design

At the University of Macerata, data collection followed the common protocol developed within the UNITE project. Students completed two types of conversational tasks, one small talk and one role play, with either ChatGPT or Pi.ai. Sessions took place both in person and online and were supervised by a researcher, with additional support provided in larger groups. The setting was explicitly non-evaluative to reduce anxiety and promote authentic participation. While the use of external dictionaries and translation tools was not allowed, students could seek linguistic support directly from the chatbot, maintaining the conversational dynamic. This

structure reflects evidence that low-stakes, self-paced interactions enhance willingness to communicate and sustain motivation (Fryer et al. 2019; Reinders and Benson 2017). After approximately twenty-five minutes, participants completed an online questionnaire including demographic data, interaction context, self-assessed reading and writing proficiency (CEFR scale), and Likert-scale items on clarity, cognitive load, emotional comfort, accessibility, and perceived bias. Open-ended questions invited reflection on what aspects of the experience facilitated or hindered language practice. Responses were analyzed through content analysis (Bardin 2000) and triangulated with proficiency level and disability status. Two dominant themes emerged, immersion and perceived quality, which also informed the design of the *Chatbot Accessibility Checklist* (see Section 4).

3.1. Participants and main findings

The analysis focuses on 102 valid questionnaires out of 108 collected at the University of Macerata, excluding six incomplete responses. Most participants were first-year students from non-language degree programs within the Department of Education, Cultural Heritage, and Tourism, where English is part of the curriculum but not a specialized subject. The sample was predominantly female (90 female, 11 male, 1 non-binary), with 14% reporting a disability. Since all interactions were text-based, the analysis concentrated on reading and writing proficiency, which most respondents rated as intermediate, with smaller groups at basic and advanced levels. Most students described the interaction as low-pressure and non-judgmental, correlating with a strong willingness to continue using chatbots for EFL practice. A smaller subset reported discomfort or perceived unfair treatment, including a few students with specific learning disorders. Dissatisfaction did not cluster by chatbot type, proficiency, or disability, suggesting that accessibility outcomes depend on interaction features rather than platform identity.

3.2. Immersion in conversational flow

Students reported higher immersion when the chatbot's responses remained contextually relevant, balanced in length, and responsive to user input. Small talk facilitated elaboration and self-expression, while role play proved effective when roles and expectations were clearly defined. Immersion weakened when replies were excessively long or disconnected from previous turns, breaking the illusion of genuine dialogue. These dynamics reflect broader findings on authenticity and engagement in L2 learning (Fredricks et al. 2004; Dörnyei and Ryan 2015; Ushioda 2011). Within the UDL framework, immersion serves as an indicator of engagement: it strengthens when learners can adjust difficulty, request

paraphrases, or reset the topic fluidly, without disrupting conversational flow.

3.3. Perceived quality of interaction

Students assessed interaction quality primarily through consistency, clarity, and supportiveness. Fast, coherent replies that maintained topic continuity were appreciated, as were short, actionable corrections embedded in the conversation rather than lengthy metalinguistic explanations. Some learners, particularly in role play, perceived an imbalance when chatbot responses became overly extended, diminishing their sense of control. Emotional tone also played a key role: most participants found the exchange non-judgmental and encouraging, while a minority experienced moment of detachment or discomfort. These perceptions align with studies linking perceived human-likeness and emotional safety to user satisfaction and learning engagement (Lee and Hahn 2024; Huang et al. 2022; Fryer et al. 2019).

4. The Chatbot Accessibility Checklist

The *Chatbot Accessibility Checklist* represents one of the key outputs developed by the University of Macerata within the UNITE project, in collaboration with CAST. This partnership was crucial to ensure both theoretical rigor and practical usability. The checklist was not conceived as a fixed set of standards, but as a dynamic evaluative tool adaptable to different contexts, institutional goals, and learner profiles. Its purpose is not to judge inclusivity as an absolute quality, but to help educators, designers, and researchers reflect systematically on how accessibility and equity are operationalized in AI-mediated learning. Grounded in the three UDL pillars – engagement, representation, and action and expression – the checklist converts empirical evidence from the UNITE studies into concrete, actionable descriptors. Each criterion is assigned a 1-5 priority score based on its potential impact on accessibility, as derived from learner data collected across the partner universities and validated through consultation with CAST experts.

In the engagement dimension, the checklist verifies whether learners can actively manage the interaction: by changing topics, switching between small talk and role play, regulating tone, or capping the chatbot's reply length through visible controls. These elements directly sustain motivation and agency, echoing findings that low-anxiety, self-paced environments support willingness to communicate (Fryer et al. 2019; Reinders and Benson 2017).

The representation dimension focuses on the accessibility and clarity of language. It includes features such as on-demand paraphrasing, concise summaries, and feedback that is non-technical yet informative. It also

incorporates interface-level checks like screen-reader compatibility, keyboard navigation, and logical focus order. The goal is to ensure that information is presented in multiple perceivable and cognitively manageable ways, in line with UDL's emphasis on flexible representation (CAST 2024).

Under action and expression, the checklist evaluates how learners can regulate the tempo and complexity of interaction: for instance, by setting pacing preferences, toggling between inline and post-turn corrections, or adjusting scaffolding levels. Defaults favor short, actionable feedback that can be expanded if needed. Such controls allow students to demonstrate understanding through varying strategies and to adapt the interaction to their own rhythm and processing needs.

Finally, a cross-cutting safety and bias section addresses the ethical dimension of inclusion. It ensures the presence of clear "flag and reset" mechanisms, conservative defaults for sensitive scenarios, and transparent communication of system limitations. These safeguards respond directly to participants' feedback from the University of Macerata case study, in which a small subset of students reported occasional discomfort or misrecognition.

What makes the checklist distinctive is its dual nature: it is both diagnostic and developmental. Teachers can use it to plan and monitor inclusive use of chatbots in language courses; developers can adopt it as a design brief to integrate accessibility from the outset; and students can use it as a literacy tool to understand how and why certain configurations support their learning better than others. Importantly, the checklist is conceived as an evolving instrument which is open to revision as technologies, learner profiles, and educational contexts continue to change.

5. Final remarks

The University of Macerata case study illustrates how inclusive design principles can be translated into practice using AI-powered chatbots in EFL education. The results indicate that learners benefit most when they are granted control over key aspects of interaction (topic, pacing, feedback style) and when systems maintain contextual sensitivity and emotional balance. Participants consistently valued the opportunity to practice in a non-judgmental, low-pressure environment that enabled spontaneity, reflection, and experimentation with language. At the same time, engagement dropped when chatbot responses became verbose, generic, or detached from the ongoing conversational flow.

Viewed through the UDL framework, these findings reaffirm that accessibility is not a static feature but a dynamic relation between the learner, the tool, and the context. Flexible topic management, adaptive feedback, and transparent controls all contribute to lowering affective

barriers while sustaining cognitive challenge. Likewise, multiple modes of representation (clear language, paraphrases, summaries) help learners manage comprehension and memory load. Together, these dimensions create the conditions for learner agency, which is central to both inclusion and long-term motivation.

From a pedagogical standpoint, the University of Macerata experience suggests that chatbot use in EFL courses should follow a progressive, scaffolded sequence. Initial sessions can focus on open small-talk tasks to build comfort and familiarity with AI tools; later stages can introduce structured role plays that balance communicative freedom with explicit objectives. Regular debriefs, where students reflect on what felt authentic or frustrating, strengthen metacognitive awareness and help tailor future interactions. In this process, teachers remain pivotal. They frame the purpose of AI use, guide students in managing interaction controls, and ensure that technology amplifies rather than replaces human mediation.

The *Chatbot Accessibility Checklist* consolidates these insights into a concrete evaluative framework. By grounding design recommendations in empirical data and validating them with CAST, it aims to offer higher education institutions a structured yet flexible pathway for inclusive digital innovation. Aligning chatbot functionality with learner diversity directly advances Sustainable Development Goal 4 – Quality Education, promoting participation, accessibility, and well-being across learning communities.

Looking ahead, UNITE's next phase will test the checklist across new contexts and modalities, including multimodal and spoken interactions, to determine whether the same accessibility priorities (turn-length control, contextual coherence, and bias mitigation) also predict sustained motivation and measurable language gains. In this sense, the checklist is not an endpoint but a foundation for continued dialogue between research, pedagogy, and technology design: an evolving instrument that embodies the very principles of Universal Design for Learning it seeks to operationalize.

Acknowledgments

References:

Ait Baha, Tarek; El Hajji, Mohamed; Es-Saady, Youssef and Fadili, Hammou. 2023. The Power of Personalization: A Systematic Review of Personality-Adaptive Chatbots. In: *SN Computer Science*. 4(661).

ANVUR & CNUDD. 2020. *Indagine Disabilità, DSA e accesso alla formazione universitaria*. https://disabilita.governo.it/media/1553/risultati-indagine-anvur-disabilit-e-dsa.pdf Accessed on: 14 November 2025.

Bardin, Laurence. 2000. *Análise de conteúdo*. Lisboa: Edições 70.

Bahroun, Zied; Anane, Chiraz; Ahmed, Vian and Zacca, Andrew. 2023. Transforming Education: A Comprehensive Review of Generative Artificial Intelligence in Educational Settings through Bibliometric and Content Analysis. In: *Sustainability*. 15(17): pp. 12983.

Bibauw, Serge; François, Thomas; Van den Noortgate, Wim and Desmet, Piet. 2022. Dialogue Systems for Language Learning: Chatbots and Beyond. In: Thomas, Michael; Reinders, Hayo and Warschauer, Mark (eds.). 2022. *The Routledge Handbook of Second Language Acquisition and Technology*. London/New York: Routledge. pp. 121–135.

Bray, Aibhin et al. 2024. What Next for Universal Design for Learning? A Systematic Literature Review of Technology in UDL Implementations at Second Level. In: *British Journal of Educational Technology*. 55: pp. 113–138.

CAST. 2024. *Universal Design for Learning Guidelines Version 3.0*. CAST. https://udlguidelines.cast.org/static/udlg3-graphicorganizer-digital-numbers-a11y.pdf. Accessed on: 14 November 2025.

Dörnyei, Zoltán. 2005. *The Psychology of the Language Learner: Individual Differences in Second Language Acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates.

Dörnyei, Zoltán and Ryan, Stephen. 2015. *The Psychology of the Language Learner Revisited*. London/New York: Routledge.

Fredricks, Jennifer A.; Blumenfeld, Phyllis C. and Paris, Allison H. 2004. School Engagement: Potential of the Concept, State of the Evidence. In: *Review of Educational Research*. 74(1): pp. 59–109.

Fryer, Luke K., Nakao, Kaori and Thompson, Andrew. 2019. Chatbot Learning Partners: Connecting Learning Experiences, Interest, and Competence. In: *Computers in Human Behavior*. 93: pp. 279–289.

Gierhart, Aaron R.; Shefferly, Nancy; Li, Yue and Speetzen, Erin. 2025. Generative Artificial Intelligence and Postsecondary Education: Rethinking Policy and Course Design. In: Gierhart, Andrea (ed.). 2025. *Cases on Enhancing P–16 Student Engagement with Digital Technologies*. Hershey, PA: IGI Global Scientific Publishing. pp. 31–54.

Huang, Weijiao, Hew, Khe Foon and Fryer, Luke K. 2022. Chatbots for L2 Learning—Are They Really Useful? A Systematic Review of AI Chatbot-Supported L2 Learning. In: *Journal of Computer Assisted Learning*. 38(1): pp. 237–257.

Lee, Inju and Hahn, Sowon. 2024. On the Relationship Between Mind Perception and Social Support of Chatbots. In: *Frontiers in Psychology*. 15: pp. 1282036. https://doi.org/10.3389/fpsyg.2024.1282036

Morgan, Angel. 2024. Leveraging Generative Artificial Intelligence to Expedite UDL Implementation in Online Courses. In: Bartlett, Melissa and Ehrlich, Sarah M. (eds.). 2024. *Unlocking Learning Potential with Universal Design in Online Learning Environments*. Hershey, PA: IGI Global Scientific Publishing. pp. 171–195.

Ortiz-Garcés, Iván; Govea, Javier; Andrade, Roberto O. and Villegas-Ch, William. 2024. Optimizing AI Chatbot Effectiveness Through Advanced Syntactic Analysis: A Comprehensive Study in Natural Language Processing. In: *Applied Sciences*. 14(5): pp. 1737.

Reinders, Hayo and Benson, Phil. 2017. Research Agenda: Language Learning Beyond the Classroom. In: *Language Teaching*. 50(4): pp. 561–578.

UNITE. 2022-2025. *UNiversally Inclusive Technologies to practice English*. https://site.unibo.it/unite/en. Accessed on: 14 November 2025.

United Nations. 2015. *Transforming Our World: The 2030 Agenda for Sustainable Development (A/RES/70/1)*. New York: United Nations General Assembly.

Ushioda, Ema. 2011. Language Learning Motivation, Self and Identity: Current Theoretical Perspectives. In: *Computer Assisted Language Learning*. 24(3): pp. 199–210.