ISBN: 978-84-1091-189-5 (pdf) / 978-84-1091-190-1 (Impreso POD)
DOI: https://doi.org/10.14201/0AQ0394139143

ADAPTATION, ETHICAL TRACEABILITY AND REGULATORY COMPLIANCE IN CONVERSATIONAL ARTIFICIAL INTELLIGENCE AGENTS: AN INCLUSIVE RESEARCH FRAMEWORK FOR COGNITIVE AND SENSORY DIVERSITY

ADAPTACIÓN, TRAZABILIDAD ÉTICA Y CUMPLIMIENTO NORMATIVO EN AGENTES DE INTELIGENCIA ARTIFICIAL CONVERSACIONAL: UN MARCO DE INVESTIGACIÓN INCLUSIVO PARA LA DIVERSIDAD COGNITIVA Y SENSORIAL

Teresa Celestino-Buentello[1] and Guillermo Hernández[2]

[1] *Air Institute. Salamanca, Castilla y León, Spain*
✉ mtcelestino@air-institute.com
https://orcid.org/0000-0003-3984-7623

[2] *Universidad de Salamanca. Salamanca, Castilla y León, Spain*
✉ guillehg@usal.es
https://orcid.org/0000-0002-7481-5961

Abstract: This doctoral research proposes the development of an inclusive framework for conversational Artificial Intelligence that integrates functional adaptation, ethical traceability, and regulatory compliance. The research design addresses persistent accessibility gaps affecting users with speech, hearing, and cognitive disabilities through three empirical studies: the evaluation of speech recognition performance for disordered speech, the assessment of comprehension of synthetic versus natural speech among hearing-impaired users, and the analysis of cognitive adaptation in generative conversational agents. By combining quantitative and qualitative methods, the project establishes standardized metrics and ethical protocols aligned with the EU AI Act, WCAG 2.2, and the GDPR. By embedding accessibility and transparency throughout the AI lifecycle, this research advances evidence-based design principles for trustworthy, human-centered conversational systems, contributing to a more equitable and sustainable digital society.

Keywords: conversational AI; accessibility; trustworthy AI; AI act; cognitive diversity; sensory impairment; ethical compliance.

Resumen: Esta investigación doctoral propone el desarrollo de un marco inclusivo para la inteligencia artificial conversacional que integre la adaptación funcional, la trazabilidad ética y el cumplimiento normativo. El diseño de la investigación aborda las persistentes deficiencias de accesibilidad que afectan a los usuarios con discapacidades del habla, auditivas y cognitivas a través de tres estudios empíricos: la evaluación del rendimiento del reconocimiento del habla para el habla desordenada, la evaluación de la comprensión del habla sintética frente a la natural entre los usuarios con discapacidad auditiva y el análisis de la adaptación cognitiva en los agentes conversacionales generativos. Mediante la combinación de métodos cuantitativos y cualitativos, el proyecto establece métricas estandarizadas y protocolos éticos alineados con la Ley de IA de la UE, las WCAG 2.2 y el RGPD. Al incorporar la accesibilidad y la transparencia en todo el ciclo de vida de la IA, esta investigación promueve principios de diseño basados en la evidencia para sistemas conversacionales fiables y centrados en el ser humano, lo que contribuye a una sociedad digital más equitativa y sostenible.

Palabras clave: IA conversacional; accesibilidad; IA fiable; ley de IA; diversidad cognitiva; discapacidad sensorial; cumplimiento ético.

1. INTRODUCTION

Conversational Artificial Intelligence (AI) systems have become a dominant interface for mediating interactions between individuals and digital infrastructures in domains such as healthcare, education, and public administration. While recent advances in automatic speech recognition, natural language processing, and dialogue management have significantly improved overall system performance [1-3], most conversational agents continue to operate under a one-size-fits-all paradigm [4]. This design assumption privileges statistically normative communication patterns and, as a result, systematically excludes users whose speech, auditory perception, or cognitive processing diverge from these norms, thereby reinforcing existing digital inequalities. Empirical research has consistently shown that recognition accuracy for dysarthric or otherwise atypical speech remains substantially lower than for typical speakers [5-6], while synthetic or generated speech may introduce additional comprehension barriers for users with hearing or cognitive impairments [7].

In parallel, the European regulatory landscape has evolved toward a stricter conception of trustworthy AI. The Artificial Intelligence Act [8] identifies conversational systems deployed in sensitive domains as high-risk applications, explicitly requiring transparency, human oversight, and accessibility, in alignment with complementary frameworks such as WCAG 2.2 [9], EN 301 549, and the GDPR. These instruments emphasize that accessibility cannot be treated as a superficial interface feature but must be embedded into the design, evaluation, and governance of AI systems. Grounded in the social model of disability [10], which locates barriers in system design rather than in individual impairments [11-12], this doctoral research addresses the intersection of functional adaptation, ethical traceability, and legal compliance. It proposes an integrated methodological framework [13] to empirically evaluate and support inclusive conversational AI, addressing key gaps in accessibility

metrics, adaptive system behavior, and the responsible use of generative AI for cognitive and sensory diversity.

1.1. Objectives and Research Questions

The doctoral project aims to develop and validate an inclusive conversational AI framework that integrates functional adaptation, ethical traceability, and compliance with European regulatory standards.

General Objective: To design, implement, and empirically evaluate an adaptive conversational AI system focused on spoken interaction, ensuring accessible, transparent, and lawful communication for users with cognitive and sensory diversity.

Specific Objectives: (1) To empirically evaluate the performance of automatic speech recognition (ASR) systems for speakers presenting phonological, motor, or articulatory impairments, establishing standardized metrics for accessibility and performance comparison. (2) To assess the comprehension and perceptual experience of synthetic versus natural speech among users with hearing impairment, identifying perceptual barriers and adaptation requirements for inclusive conversational interfaces. (3) To evaluate the cognitive adaptability of generative conversational AI systems (specifically GPT-based models) when supporting task execution and information comprehension for users with mild cognitive disabilities, ensuring compliance with accessibility and ethical standards.

2. METHODOLOGICAL FRAMEWORK

The research adopts a mixed-methods, multi-study design integrating quantitative and qualitative approaches. The empirical sequence comprises three interlinked studies addressing distinct gaps identified in the literature.

Study 1: Evaluation of automatic speech recognition performance in speakers with altered speech. This study evaluates the transcription accuracy of open-source ASR models, such as OpenAI's Whisper, using speech samples from participants with phonological, motor, or articulatory impairments (e.g., dysarthria, apraxia). Speech samples (both read and semi-spontaneous) are recorded and analyzed against a control group. Quantitative metrics include Character Error Rate (CER), substitution and omission rates, and systematic phoneme confusion, providing a detailed view of accuracy and error patterns in atypical speech. Statistical comparisons (t-tests, ANOVA) identify performance gaps and error typologies.

Study 2: Comprehension of synthetic versus natural speech in individuals with hearing impairment. This study evaluates auditory comprehension and listening effort when participants are exposed to natural versus text-to-speech (TTS) outputs. Individuals with mild-to-moderate hearing loss listen to paired messages identical in content but differing in source. Objective metrics (accuracy, response time) and subjective measures (perceived effort via adapted NASA-TLX, preference ratings) are analyzed using nonparametric tests where appropriate.

Study 3: Evaluation of cognitive adaptation in generative conversational agents. This study focuses on users with mild cognitive disabilities performing guided conversational tasks with generative conversational AI models (e.g., GPT-based systems), enabling comparison across architectures and adaptive behaviors. Linguistic complexity is controlled through distinct prompt configurations. Quantitative indicators are complemented by semi-structured interviews, with qualitative data analyzed through thematic analysis to ensure traceability and methodological rigor.

All studies include informed consent procedures, accessible materials, ethical review, and compliance with EU data protection regulations and research standards for vulnerable populations.

2.1. Expected Impact and Ethical Implications

This doctoral research generates scientific, social, and regulatory contributions through the integration of empirical evidence across three complementary studies. Scientifically, it advances the field of Inclusive and Trustworthy AI by proposing standardized metrics and methodological protocols for evaluating accessibility in conversational systems spanning speech recognition, speech synthesis, and cognitive interaction. These results support a more evidence-based understanding of how accessibility mechanisms influence usability, comprehension, trust, and fairness.

From a social perspective, the research addresses persistent exclusion in digital communication by improving access for users with speech, hearing, and cognitive disabilities. By supporting intelligible voice interfaces and cognitively accessible interactions, the project aligns with European priorities on digital inclusion and equal participation in education, healthcare, and public services.

Ethically and regulatorily, the project operationalizes key principles of the EU AI Act and the GDPR through ethical traceability, transparency, and human oversight. Privacy-by-design and human-in-the-loop supervision ensure accountability, data minimization, and risk mitigation in the use of generative AI. Overall, the research positions accessibility as a core requirement of trustworthy AI, offering a replicable framework for compliant, human-centered, and socially sustainable conversational systems.

3. ACKNOWLEDGMENTS

This research is part of the International Chair Project on Trustworthy Artificial Intelligence and the Demographic Challenge, within the framework of the National Artificial Intelligence Strategy (ENIA) and the European Recovery, Transformation and Resilience Plan. Reference: TSI-100933-2023-0001. This project is funded by the Secretary of State for Digitalization and Artificial Intelligence and by the European Union (Next Generation EU).

4. REFERENCES

1. Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2023). *Robust speech recognition via large-scale weak supervision* (arXiv:2212.04356). arXiv. https://doi.org/10.48550/arXiv.2212.04356
2. Hannun, A., Case, C., Casper, J., Catanzaro, B., Diamos, G., Elsen, E., Prenger, R., Satheesh, S., Sengupta, S., Coates, A., & Ng, A. Y. (2014). *Deep Speech: Scaling up end-to-end speech recognition* (arXiv:1412.5567). arXiv. https://doi.org/10.48550/arXiv.1412.5567
3. Gulati, A., Qin, J., Chiu, C.-C., Parmar, N., Zhang, Y., Yu, J., Han, W., Wang, S., Zhang, Z., Pang, R., & Wu, Y. (2020). Conformer: Convolution-augmented transformer for speech recognition. In *Interspeech 2020 Proceedings* (pp. 5036-5040). International Speech Communication Association. https://doi.org/10.21437/Interspeech.2020-3015
4. Vereschak, O., Alizadeh, Fatemeh., Bailly, G., & Caramiaux, B. (2024). Trust in AI-assisted decision making: Perspectives from those behind the system and those for whom the decision is made. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*. ACM. https://doi.org/10.1145/3613904.3642018
5. Calvo, I., Tropea, P., Viganò, M., Scialla, M., Cavalcante, A. B., Grajzer, M., Gilardone, M., & Corbo, M. (2021). Evaluation of an automatic speech recognition platform for dysarthric speech. *Folia Phoniatrica et Logopaedica, 73*(5), 427-435. https://doi.org/10.1159/000511042
6. Patel, T., Hutiri, W., Ding, A. Y., & Scharenborg, O. (2025). *How to evaluate automatic speech recognition: Comparing different performance and bias measures* [Manuscrito presentado para publicación].
7. Fu, B., Hadid, A., & Damer, N. (2025). Generative AI in the context of assistive technologies: Trends, limitations and future directions. *Image and Vision Computing, 154*, Artículo 105347. https://doi.org/10.1016/j.imavis.2024.105347
8. European Commission. (n.d.). *AI Act*. Retrieved June 23, 2025, from https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence
9. Kirkpatrick, A., O Connor, J., Campbell, A., & Cooper, M. (Eds.). (2025). *Web Content Accessibility Guidelines (WCAG) 2.1.* W3C Recommendation.
10. Sharma, T., Stangl, A., Zhang, L., Tseng, Y. Y., Xu, I., Findlater, L., Gurari, D., & Wang, Y. (2023). Disability-first design and creation of a dataset showing private visual information collected with people who are blind. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM. https://doi.org/10.1145/3544548.3580922
11. Bunton, K. (2008). Speech versus nonspeech: Different tasks, different neural organization. *Seminars in Speech and Language, 29*(4), 267-275. https://doi.org/10.1055/s-0028-1103390
12. Yorkston, K., & Beukelman, D. (1999). Staging interventions in progressive dysarthria. *Perspectives on Neurophysiology and Neurogenic Speech and Language Disorders, 9*(4), 7-12. https://doi.org/10.1044/nnsld9.4.7
13. Schlosser, R. (2009). Social validation of interventions in augmentative and alternative communication. *Augmentative and Alternative Communication, 15*(4), 234-247. https://doi.org/10.1080/07434619912331278775