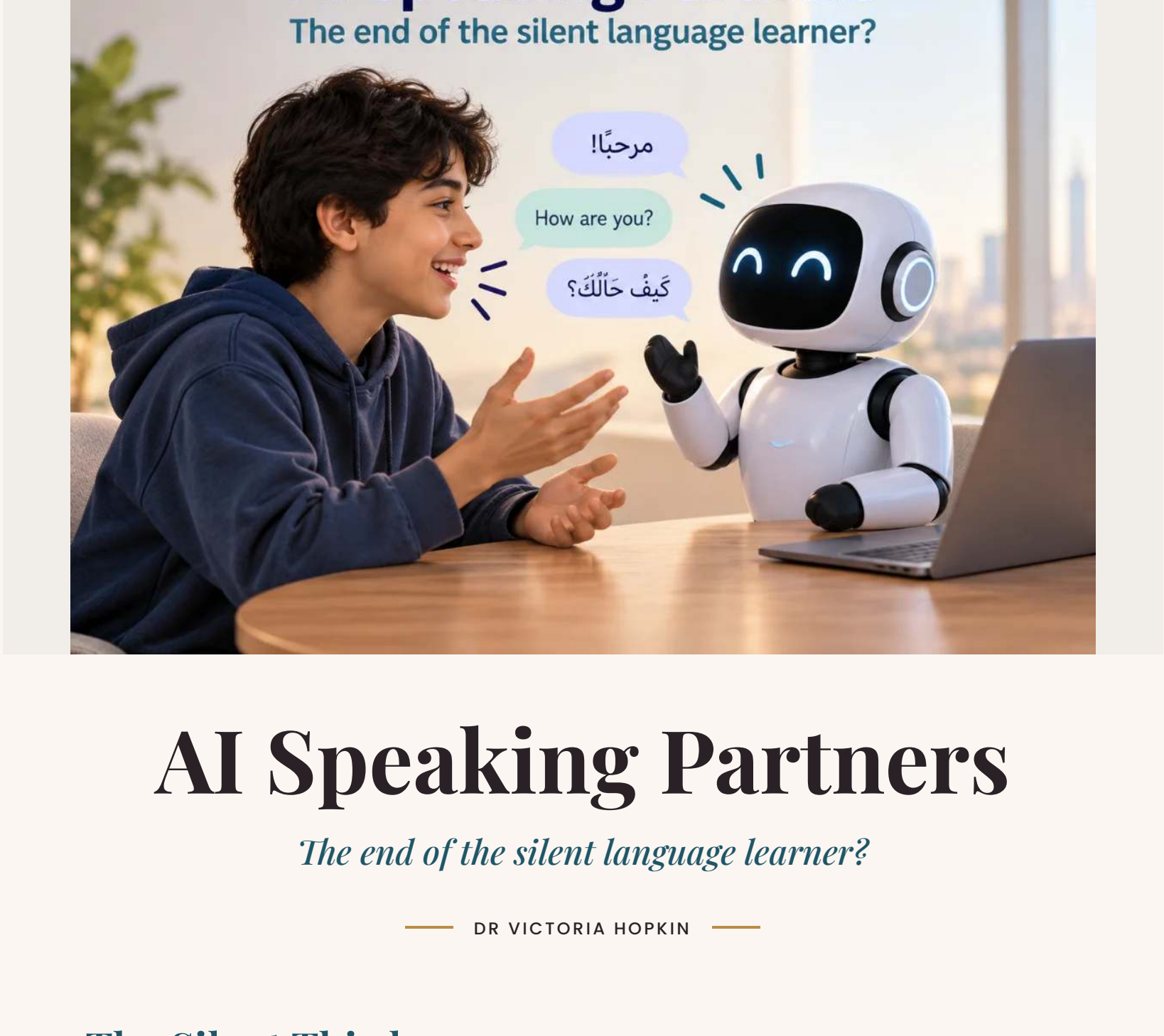




AI Speaking Partners

The end of the silent language learner?

DR VICTORIA HOPKIN

The Silent Third

In the early 1980s Elaine Horwitz, a young professor at the University of Texas at Austin, started noticing the same student in her office. Not literally the same one. A succession of them, with the same complaint, which they found hard to put into words and which their language teachers tended to dismiss. They could read the language. They could write it. They did the homework. And when the teacher turned to them in class and asked a question in Spanish or French, something happened in their bodies that stopped them speaking. Their minds went blank. They heard their own accent and could not go on. Some of them had dropped the course. Some had changed degree.

Horwitz did what a psychologist does with a pattern nobody has named. She set up a support group, at the university's learning skills centre, for students who were anxious about learning a foreign language, and she listened. Then, with two colleagues, she turned what she heard into a questionnaire of thirty-three statements. I never feel quite sure of myself when I am speaking in my foreign language class. I start to panic when I have to speak without preparation. It embarrasses me to volunteer answers. She gave it to students in beginning Spanish, and found that roughly a third of them agreed with statements like these. Not a few nervous individuals. A third of the room.

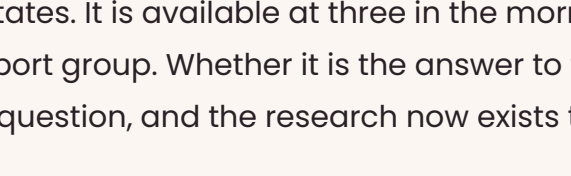


Fig. 1 – Roughly a third of a beginning language class agreed with statements like 'I start to panic when I have to speak without preparation' (Horwitz, Horwitz & Cope, 1986). Not a few nervous individuals. A third of the room.

The paper appeared in the *Modern Language Journal* in 1986 under the title *Foreign Language Classroom Anxiety*, and the questionnaire, the FLCAS, became the most used instrument in the field. What Horwitz had established was that the silence in a language classroom was not laziness or shyness or lack of aptitude. It was a specific, measurable fear, concentrated almost entirely on one activity, which was being asked to speak in front of other people.

Forty years later there is a device that will let a learner speak for as long as she likes to something that is not a person. It never sighs, never glances at the clock, never turns to the next student when she hesitates. It is available at three in the morning. It is, on the face of it, the answer to Horwitz's support group. Whether it is the answer to the problem the support group was for is a different question, and the research now exists to ask it.

Fryer's Curve

The first person to try was Luke Fryer, an Australian who was teaching English at a university in Japan and who in 2006 published a short paper with a man called Rollo Carpenter. Carpenter was not an academic. He was a programmer who had built a chatbot called Jabberwacky, later Cleverbot, which learned to talk by remembering everything anyone had ever said to it. Fryer put his students in front of it and asked them how it felt. The answer that stayed with him was that they felt more comfortable talking to the bot than to a person. They were not afraid of it. Horwitz's third of the room had found somewhere to speak.

A decade later Fryer ran the experiment properly, and it is the one every school considering these tools should know about. With colleagues in Australia and Japan he took 122 university students and gave them a series of speaking tasks over twelve weeks. For some tasks the partner was the chatbot; for others it was a fellow student. After each task he measured their interest in continuing.

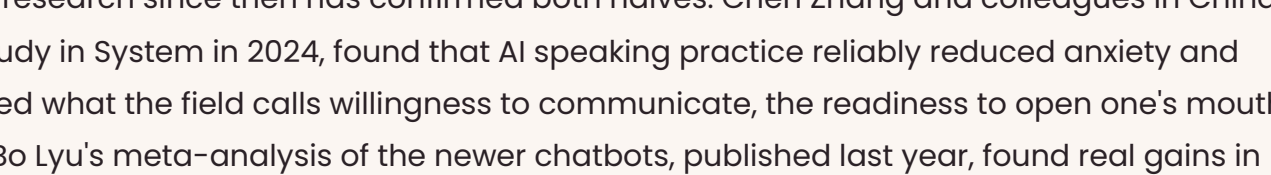


Fig. 2 – Fryer's twelve-week experiment (2017), shown schematically. The chatbot won the first round: task by task, interest fell. Interest in the human partner held. By the end, the students who had begun by preferring the machine were choosing the person.

"The machine had not stopped being a safe place to speak. It had stopped being a reason to."

The chatbot won the first round. Interest was high, anxiety was low, the students enjoyed themselves. Then, task by task, the interest in the chatbot fell. Interest in the human partner held steady. By the end of the twelve weeks the students who had begun by preferring the machine were choosing the person. Fryer called it a novelty effect, and the phrase is now standard, but it undersells what he found. The machine had not stopped being a safe place to speak. It had stopped being a reason to.

The research since then has confirmed both halves. Chen Zhang and colleagues in China, in a study in System in 2024, found that AI speaking practice reliably reduced anxiety and raised what the field calls willingness to communicate, the readiness to open one's mouth at all. Bo Lyu's meta-analysis of the newer chatbots, published last year, found real gains in proficiency, and then listed what the chatbots lacked: the cues of a face, the emotional response of someone who is actually listening. And nearly every study in the literature, Zhang's included, lasted between two and six weeks, which is to say inside the window before Fryer's curve turns down.

Another Mind

So the machine gets the silent learner talking. The question is what she is learning to do, and here two studies from the testing world are unusually clear.

Gary Ockey runs language assessment research at Iowa State, and in 2021 he and Evgeny Chukharev-Hudilainen did something simple. They had the same students take a paired speaking test twice: once with a human partner, once with a computer partner that had been programmed to discuss the same topic. Then they scored not the grammar but the interaction: whether the student asked for clarification, built on what the partner said, signalled that she was following, repaired a misunderstanding when it happened. With the human partner the students did all of these things. With the computer they did far fewer. They were the same students, with the same ability. Faced with a partner that would not misunderstand them, they stopped doing the things people do when they might be misunderstood.

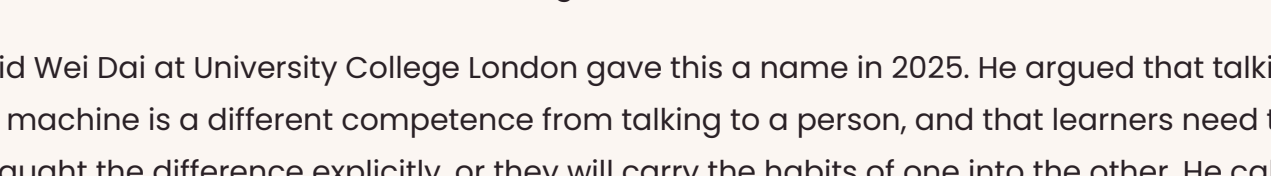


Fig. 3 – The same students, the same ability, two partners (Ockey & Chukharev-Hudilainen, 2021), shown schematically. Faced with a partner that would not misunderstand them, they stopped doing the things people do when they might be misunderstood.

Veronica Timpe-Laughlin at ETS in Princeton had found the same pattern a year earlier from a different angle. She watched learners make requests of an automated agent and found them blunter and less polite than they would be with a person. In July this year she published, in *Language Learning*, the fullest review yet of what conversational agents do and do not develop, and the answer is consistent with everything above. They develop fluency, confidence and vocabulary. They do not develop what the field calls interactional competence, because interactional competence is the skill of managing another mind, and the machine does not have one to manage.

David Wei Dai at University College London gave this a name in 2025. He argued that talking to a machine is a different competence from talking to a person, and that learners need to be taught the difference explicitly, or they will carry the habits of one into the other. He called it critical interactional competence. The point is not that the machine is a bad partner. It is that it is a partner of a kind that has never existed before, and nobody has ever learned a language from one.

There is a further subtlety, and it comes from Walcir Cardoso at Concordia in Montreal, who in 2020 tested whether Amazon's Alexa could understand learners with accents. Often it could not. So it is not quite right to say the machine never misunderstands you. It misunderstands you constantly, at the level of sounds, and a learner who wants to be understood by it learns to speak more slowly, more clearly, and more like the recordings it was trained on. That is a real skill. It is also the thing the first piece in this series warned about: the learner adjusting herself toward the machine's idea of the language. Human misunderstanding is social, and repairing it is the competence. Machine misunderstanding is acoustic, and working around it is accommodation.

The Media Equation

Why is the machine so comfortable to talk to in the first place? The answer was worked out before any of these tools existed, by two Stanford researchers who were studying something else.

Byron Reeves and Clifford Nass spent the 1990s running experiments in which people interacted with ordinary computers, and they kept finding that people treated the computers as if they were people. They were polite to them. They were flattered by them. They socially ascribed to them. Nass later showed that a computer with a voice was treated more disloyally still. Reeves and Nass called this the media equation, and their explanation was that the human brain had evolved in a world where anything that talked was a person, and had not caught up.

Gale Lucas at the University of Southern California then showed the other side. In 2014 she had people talk to a virtual interviewer, and told half of them that a human was operating it and the other half that it was fully automated. The people who believed no human was listening disclosed more, and more honestly. The absence of judgement was not a feature of the software. It was a belief about who was on the other end.

"It is human enough to talk to and not human enough to be afraid of."

Put those two findings together and you have the whole appeal of the AI speaking partner in one sentence: it is human enough to talk to and not human enough to be afraid of. That is exactly what Horwitz's students needed. It is also, by Fryer's and Ockey's findings, exactly why it stops being enough.

The Second Layer

For Arabic in Dubai the picture has a second layer, and it is the one the vendors do not mention.

A speaking partner has to do two things: hear you, and answer. On the first, Bashar Talafha and colleagues at the University of British Columbia tested Whisper, the speech recognition engine under most of these tools, on Arabic in 2023 and found its accuracy dropping sharply the moment the speech moved from formal Arabic to dialect. On the second, as the previous piece in this series described, the language models understand dialect and will not speak it, sliding back into the Arabic of the news broadcast whenever they answer.

So an eight-year-old in Al Barsha practising with an AI partner will be heard poorly, because she speaks the Arabic she hears around her, and answered in an Arabic that nobody around her speaks. She will experience, in miniature, both halves of the machine problem at once: acoustic misunderstanding on the way in, and a partner who cannot be misunderstood socially on the way out, because it has no social register to get wrong. She will still feel safer than she does with the teacher. Horwitz's finding does not go away. But what she practises will be a language that exists only on the screen.

No study has yet tested an AI speaking partner with Arabic learners of any age. Everything above about Arabic is inference from the English studies plus the two Arabic benchmarks. At LEAP we are designing the first, with the two-school panel: the same pupils, an AI partner and a human partner, and Ockey's measures rather than a grammar score. Until it runs, a school should treat the claims on the box as claims.

The On-Ramp

What, then, is the honest use of these tools? Three things, in an order that matters.

The first is to use the machine for the on-ramp, and to know that it is an on-ramp. Horwitz's third of the room needs somewhere to speak that will not look back, and the evidence that the machine provides it is as strong as anything in the field. Give the anxious pupil the tool for the first hundred conversations. Zhang's study and a dozen like it say her anxiety will fall and her willingness will rise. Then look at the calendar, because Fryer's curve turns down somewhere between the sixth and the twelfth week, and a tool that is still the main speaking partner at that point is holding her where she started.



Fig. 4 – The on-ramp. The move from machine to human has to be taught as a move: with a person, you will be misunderstood, and fixing that is the skill.

The second is to hand over to people deliberately, and to teach the handover. Ockey's students did not lose the ability to ask for clarification; they stopped using it because the partner did not need it. So the move from machine to human has to be taught as a move: here is what you have been doing with the tool, here is what a person will need from you that the tool did not. Dai's critical interactional competence is a mouthful, but in a classroom it is a single lesson: with a person, you will be misunderstood, and fixing that is the skill.

The third is specific to Arabic. Before any school buys a speaking tool for Arabic B, a Head of Arabic should speak to it in the Arabic of the corridor and listen to what it does. If it fails to hear the dialect, or answers in the Arabic of the sermon, it will teach a child to speak to screens. That may still be worth doing for the anxious pupil. It is not what the box says it does, and it is not what the KHDA policy is buying 200 minutes a week to achieve.

The Empty Car

Horwitz's speaking partner met in a room, and that was the point of it. The students who came were not looking for somewhere to speak without being judged. They could have found that in an empty car. They were looking for other people who would let them speak badly and stay in the room, which is a different and more difficult thing, and the only thing that ever cured the fear.

The machine offers the empty car, with a voice in it. For the student who has never dared open her mouth, that is not nothing; it may be the first hundred sentences she ever says aloud. But Fryer's students, given twelve weeks, chose the person, and Ockey's students, given a machine, stopped doing the things that make speech a conversation. The silent learner may well stop being silent. Whether she becomes a speaker depends on who is listening when the novelty wears off.

Sources referred to in this piece

Dai, D.W., Hua, Z. and Chen, G. (2025) 'How does interaction with LLM-powered chatbots shape human understanding of culture? The need for critical interactional competence (CritIC)', *Annual Review of Applied Linguistics*, 45, pp. 28–49.

Fryer, L. and Carpenter, R. (2006) 'Bots as language learning tools', *Language Learning & Technology*, 10(3), pp. 8–14.

Fryer, L.K., Ainley, M., Thompson, A., Gibson, A. and Sherlock, Z. (2017) 'Stimulating and sustaining interest in a language course: an experimental comparison of chatbot and human task partners', *Computers in Human Behavior*, 75, pp. 461–468.

Horwitz, E.K., Horwitz, M.B. and Cope, J. (1986) 'Foreign language classroom anxiety', *Modern Language Journal*, 70(2), pp. 125–132.

Lucas, G.M., Gratch, J., King, A. and Morency, L.-P. (2014) 'It's only a computer: virtual humans increase willingness to disclose', *Computers in Human Behavior*, 37, pp. 94–100.

Lyu, B. (2025) 'Effectiveness of chatbots in improving language learning: a meta-analysis of comparative studies', *International Journal of Applied Linguistics*.

Moussalli, S. and Cardoso, W. (2020) 'Intelligent personal assistants: can they understand and be understood by accented L2 learners?', *Computer Assisted Language Learning*, 33(8), pp. 865–890.

Ockey, G.J. and Chukharev-Hudilainen, E. (2021) 'Human versus computer partner in the paired oral discussion test', *Applied Linguistics*, 42(5), pp. 924–944.

Reeves, B. and Nass, C. (1996) *The media equation: how people treat computers, television, and new media like real people and places*. Stanford, CA: CSLI Publications.

Robinson, N.R., Abdelmoneim, S., Marchisio, K. and Ruder, S. (2025) 'AL-QASIDA: analyzing LLM quality and accuracy systematically in dialectal Arabic', *Findings of ACL 2025*, pp. 22048–22065.

Talafha, B., Waheed, A. and Abdul-Mageed, M. (2023) 'N-shot benchmarking of Whisper on diverse Arabic speech recognition', *Proceedings of Interspeech 2023*.

Timpe-Laughlin, V. and Dombi, J. (2020) 'Exploring L2 learners' request behavior in a multi-turn conversation with a fully automated agent', *Intercultural Pragmatics*, 17(2), pp. 221–257.

Timpe-Laughlin, V. (2026) 'Conversational agents in practicing and assessing interactional competence: current research and future directions', *Language Learning*.

Zhang, C., Meng, Y. and Ma, X. (2024) 'Artificial intelligence in EFL speaking: impact on enjoyment, anxiety, and willingness to communicate', *System*, 121, 103259.

