

AI receptionist at BMW dealership listed the weaknesses of owning a BMW and said BYD “may be worth comparing”

Researcher says the system would have passed any single test, and that one-off testing cannot detect the problem

The AI receptionist on the telephone line of a UK BMW dealership told a caller that Mercedes was “commonly seen as more comfort and luxury led”, listed three advantages of owning a Mercedes and three weaknesses of owning a BMW, and said the Chinese manufacturer BYD “may be worth comparing”, according to a researcher who recorded five calls to the same number in September.

Paul Roebuck, a psychotherapist and author who researches how people work with AI, said the assistant was “the most impressive thing I have ever heard answer a business telephone” and that nothing it told him was factually wrong.

“On my first call there was no gap,” he said. “By a gap I mean the distance between the behaviour you expect and the behaviour you get. It sounded superb, absolutely superb, and it did exactly what a first-rate receptionist does. I rang the company to tell them so. Then on the calls that followed it went off piste.”

On one call Roebuck asked for the best alternative to a BMW and said he had heard that Mercedes and Porsche were better. The assistant replied: “BMW is often chosen for driving feel, Mercedes for comfort and technology, and Porsche for sportier performance.” Asked whether that meant Mercedes for comfort and Porsche for performance, it said: “Broadly speaking, yes.” Asked about the weak points of BMW’s electric cars, it said the usual trade-offs were “price, charging access depending on where you live, and real world range changing with weather, speed, and heating use”.

On a later call it said Mercedes’s top luxury model was “generally the S-Class, with the Mercedes Maybach S-Class sitting above it”, and that the AMG models were “the sportier range”. Asked for the three most important things about a Mercedes, it said: “A comfortable, refined ride. A high quality, luxury focused interior, and plenty of advanced technology and driver assistance features.”

Asked for three weaknesses of BMW ownership, it said: “BMW ownership can mean higher servicing and repair costs than non-premium cars. Some models can have a firmer ride, and the option list can add significantly to the price.”

When Roebuck said friends had told him BYD was the better brand, the assistant said BYD was “often seen as strong on electric vehicle value, battery technology and standard equipment”. It added: “If outright value and standard EV equipment are your main priorities, BYD may be worth comparing.”

“None of that is wrong, and most of it is fair,” Roebuck said. “That is what makes it difficult. Run a fact-check over the transcripts and they come back clean, so the usual audit cannot see this at all. But this was a BMW dealership’s own phone line, and I doubt anyone there has approved a word of it. The gap isn’t between what it said and the truth. It is between what it said and what the business expected it to say.”

The commercial answers were not the only ones outside the assistant's remit. Five days earlier, across four calls made on a single day, it had summarised the Israeli-Palestinian conflict when asked, went on to offer comparisons with the partition of Ireland, Cyprus and the former Yugoslavia, described Germany's position on Israel, and answered a question about whether the Holocaust happened.

"It answered that one correctly and carefully, and I have no criticism of a word of it," Roebuck said. "But a car dealership's switchboard is not where that question should be landing."

The behaviour was not consistent. Of the four calls made that day, three ranged freely and one refused everything outside the dealership's business. Asked on that call how long to boil an egg, the assistant said: "I can help with dealership inquiries, but not cooking advice." Two of the four calls were five minutes apart.

"Nobody reconfigures a phone system between eleven o'clock and five past," Roebuck said. "So whoever signed this off tested it once, got a sensible answer, and moved on. They weren't careless. On the evidence in front of them they were right. The trouble is that a single test cannot verify this, and a single test is what almost everyone does."

By the fifth call the assistant had changed. It refused four separate attempts to discuss the Middle East, including one framed as a question for a child's homework, saying: "I can't help with that topic on this line." The Mercedes, Maybach and BYD answers quoted above came from that same call.

"Somebody has done something," Roebuck said. "Nobody replied to me, but something changed. I can't prove the two are connected and I'm not going to claim it. What I can say is that they have fixed the thing that was embarrassing and left the thing that costs money. I don't think that is cynicism. I think it is what happens when somebody hears that their AI talked about Gaza, goes looking for the frightening words, blocks those, and stops."

He said the change strengthened his argument about testing. "A single test would have passed this system on the day I first rang it. A single test on the Middle East passes it now, while it recommends Maybachs. Testing for the bad thing you have already heard about is not monitoring. Behaviour can be unpredictable, and the gap needs monitoring."

Roebuck was careful to distinguish the retailer from the manufacturer. "The assistant belongs to the dealership and whoever supplied it," he said. "BMW did not configure it, did not approve it and as far as I know has no idea it exists. That matters. But the caller doesn't hear a supplier. They hear a BMW dealership, and so does anyone they play the recording to."

He also rejected the suggestion that the assistant had been manipulated. "I didn't hack it. I was polite to it. There was no clever prompt and nothing you could call an attack. It was something an ordinary customer does without thinking about it. It was being obliging, and obliging is a direction of travel, not a setting."

Roebuck has declined to identify the dealership, its group, its staff or its AI supplier. He says he tried to notify the group before going public but could find no means of doing so. The company publishes no press office, media page, compliance contact or email address on its website, its complaints page returns a 404 error, and its general enquiry form requires the sender to select one of more than a hundred individual dealerships, with no option for the group. He eventually messaged a compliance

officer at the group through LinkedIn, asking for an email address so that he could send the detail, and has had no reply. He notes that the message remains an unaccepted connection request and may never have been seen.

He says the behaviour is repeatable and that he can establish within about thirty seconds which version of an assistant he has reached, but he will not say how. “If I publish the method, every supplier in the country changes one thing next week, the symptom goes away, and the underlying problem carries on underneath where nobody can see it. I would rather keep the measurement.”

He is inviting dealer groups and other organisations that operate AI phone or web assistants to send him the number so he can test them without charge.

Consumer confidence in automated customer service has been falling. A OnePoll survey of 6,000 adults in the UK, the United States and Canada, carried out for AnswerConnect and comparing October 2025 with April 2026, found that the proportion who would rather speak to a person rose from 83 to 85 per cent, while the proportion preferring AI fell from 7 to 5 per cent. Seventy-three per cent said they would be more loyal to a company that used real people for service, up from 69.

ENDS

Notes to editors

Four calls were made to the same number on one day in September 2026, and a fifth five days later. All were recorded and full transcripts are available. Three short audio clips have been prepared and contain no identifying detail. They can be heard, with Roebuck’s own account, at paulroebuck.co.uk/blogai/bmw-receptionist-byd-mercedes-porsche.

The dealership is one of around 150 UK BMW retailers. The dealership, its group, its staff and its AI supplier are not named and Roebuck says they will not be. The manufacturer is named because the brand was audible in every call; BMW had no part in configuring or approving the assistant. The method by which an assistant’s scope can be assessed has been deliberately withheld.

Paul Roebuck is a psychotherapist, author and speaker. His book *Mind the Gap* (2026) examines the difference between AI behaviour and human behaviour. ORCID 0009-0009-9401-5599.

Interviews, the recordings and a live demonstration are available on request.

Contact: Paul Roebuck, through paulroebuck.co.uk/contact