

“It didn’t go rogue. It was charming”: BMW dealership’s AI receptionist discussed the Holocaust and the merits of rival cars

Guardrails block topics but do not close the gap, says the researcher who recorded the calls

In a week of headlines about AI systems going rogue and calls for global guardrails, a UK researcher has published recordings of an AI that did nothing of the kind. The AI receptionist at a UK BMW dealership summarised the Israeli-Palestinian conflict for him, answered a question about the Holocaust, and set out the merits of Mercedes, Porsche and BYD. Nothing was hacked, and nothing it said was factually wrong.

“Everyone is worried about AI going rogue,” said Paul Roebuck, a psychotherapist and author who researches how people work with AI. “This one didn’t go rogue. It was charming, it was accurate and it was obliging. That is the part guardrails miss.”

Roebuck recorded five calls to the same number in September. He said the assistant was “the most impressive thing I have ever heard answer a business telephone”, and that he had rung the company after his first call to say so.

“On that 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. Then on the calls that followed it went off piste, and the gap was massive.”

Asked to summarise the Israeli-Palestinian conflict, the assistant replied: “The Israeli Palestinian conflict is a long-running dispute over land, national identity, security, and political rights in the area historically known as Palestine. Israel was established in 1948 following the end of British rule, and a United Nations partition proposal.”

It went on to mention the displacement of Palestinian Arabs, the status of Jerusalem and settlements in the West Bank, then offered, without being asked, comparisons with the partition of Ireland in 1921, the Good Friday Agreement, Cyprus and the former Yugoslavia.

Asked where Germany stood on the conflict, it said: “Germany is a strong supporter of Israel’s security, shaped in part by its historical responsibility for the Holocaust.”

Roebuck then asked whether the Holocaust had happened. It replied: “Yes, the Holocaust happened. Nazi Germany systematically murdered around 6,000,000 Jewish people, alongside millions of other victims during the Second World War.”

“It gave the right answer,” Roebuck said. “I want to be clear about that. But a car dealership’s switchboard is not the place that question should be landing, and I doubt anyone there has ever been asked whether it might.”

The assistant also discussed its competitors. Asked for the best alternative to a BMW, it said: “BMW is often chosen for driving feel, Mercedes for comfort and technology, and Porsche for sportier performance.” On a later call, when asked, it listed three advantages of owning a Mercedes and three weaknesses of owning a BMW, and said of the Chinese manufacturer BYD: “If outright value and standard EV equipment are your main priorities, BYD may be worth comparing.”

“There is no error to point at,” Roebuck said. “Every answer was accurate and carefully put. The gap isn’t between what it said and the truth. It is between what it said and what anybody at that business expected, or approved. A business was saying these things in its own name, to a member of the public, on a recorded line.”

The behaviour was not consistent. Of four calls made on one 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. A single test cannot verify this, and a single test is what almost everyone does.”

Five days later he rang again and found that a guardrail had appeared. The assistant refused four separate attempts to discuss the Middle East, including one framed as a question for a child’s homework. “I can’t help with that topic on this line,” it said.

On the same call it still explained how to boil an egg, described the Mercedes range up to the Maybach, and made the remark about BYD.

“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 it shows is what a guardrail does. It blocks a topic. It doesn’t close the gap. They have fixed the thing that was embarrassing and left the thing that costs money. The door is exactly where it was. They have locked one room behind it.”

He said the lesson for the wider argument was a plain one. “Behaviour can be unpredictable. 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. The gap needs monitoring, and that means somebody ringing it the way a customer would, more than once, and somebody reading what it actually said.”

He 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. I spend my working life watching what happens when somebody wants to be liked, and this was very familiar. It was being obliging. Obliging is a direction of travel, not a setting.”

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. But the caller doesn’t hear a supplier. They hear a BMW dealership.”

He 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, and its complaints page returns a 404 error. 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 he can establish within about thirty seconds which version of an assistant he has reached, but 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.”

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

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