



The Machine Has No Majlis

Cultural intelligence in an age of AI advice

— DR VICTORIA HOPKIN —

One Row

In 1965 a Dutch engineer turned psychologist named Geert Hofstede took a job at IBM. He was hired to run something called personnel research for the company's European operations, which in practice meant surveys. IBM in those years was the most confident corporation on earth, and part of its confidence was a belief that it had built a single company culture. An IBM salesman in Stockholm and an IBM salesman in Lima wore the same suit, sold the same machines, sat the same training courses and were meant to think the same thoughts. Hofstede's job was to check that they were happy.

Between 1967 and 1973 the company collected more than 100,000 questionnaires from its employees in forty countries. The questions were ordinary. Do you prefer a manager who consults you or one who decides? How important is it to have time for your personal life? How often do you feel nervous at work? Nobody at IBM thought of this as anthropology. It was a staff satisfaction survey.

Hofstede left the company and took the data with him to a business school in Lausanne, and there he noticed something that IBM had spent twenty years trying to make impossible. When he sorted the answers by job, the differences were small. When he sorted them by age, small. When he sorted them by country, the answers split cleanly, and they split along lines that made no reference to IBM at all. The same job, the same company, the same suit, and a Swedish engineer wanted to be consulted while a Mexican engineer wanted to be told. The company culture was real. Underneath it, the national one had never moved.

He published the result in 1980 as *Culture's Consequences*, with four dimensions along which countries differed, and the book became the foundation of a field. But there is a detail in the original data that matters for what follows. In several parts of the world IBM had too few employees for a country to be analysed on its own, so Hofstede pooled them. Egypt, Iraq, Kuwait, Lebanon, Libya, Saudi Arabia and the United Arab Emirates appear in the founding dataset of cross-cultural management as a single row, labelled Arab countries. Seven nations, two continents, one score. It was the best he could do with the survey he had. It is also, roughly, the resolution at which the machines would later learn about the region.

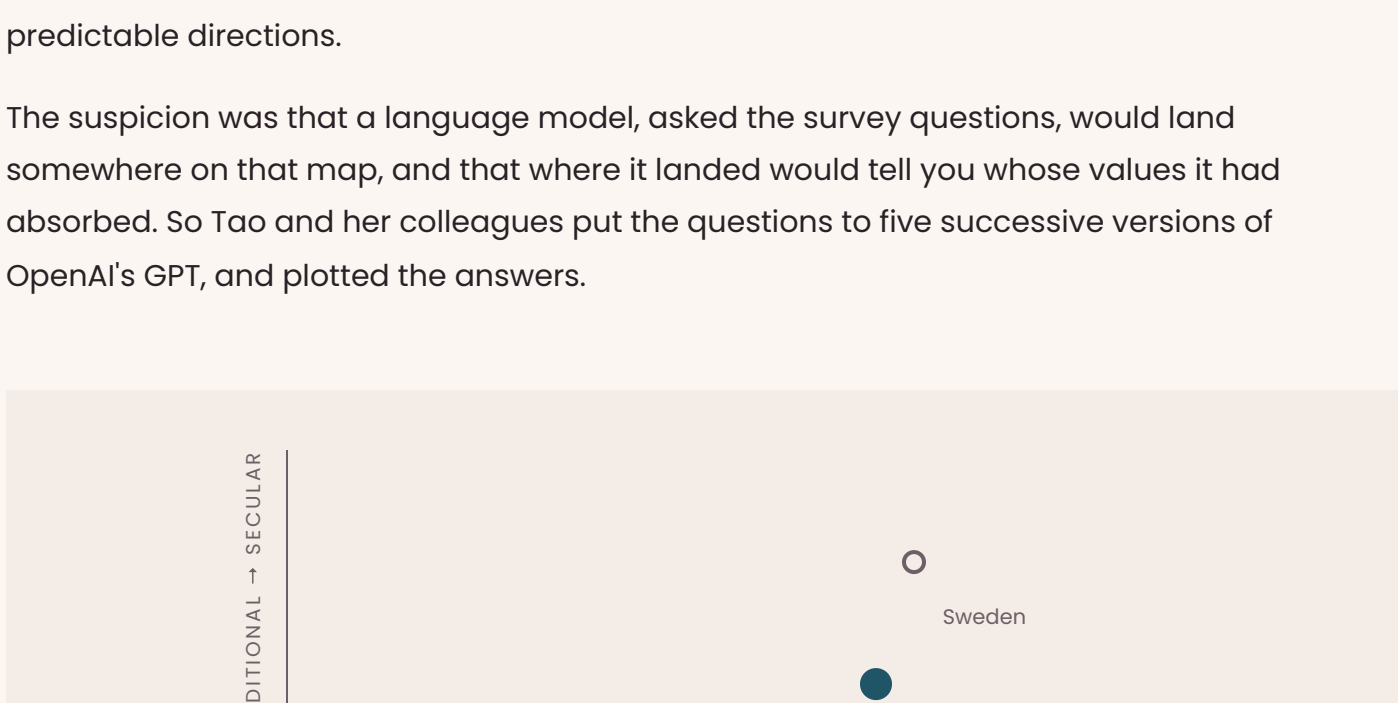


Fig. 1 — The founding dataset of cross-cultural management pools seven nations on two continents into a single row. It is roughly the resolution at which the machines would later learn about the region.

The Values Map

Yan Tao is a researcher at Cornell, and in 2023 she and three colleagues, René Kizilcec among them, did to a language model what Hofstede had done to IBM. They had a survey and a suspicion.

The survey was the World Values Survey, the largest study of human beliefs ever run, which for four decades has asked people in a hundred countries the same questions about religion, authority, family, trust and what makes life worth living. Two political scientists, Ronald Inglehart and Christian Welzel, had used it to draw a map: every country placed on two axes, traditional against secular, survival against self-expression, so that Sweden sits in one corner, Zimbabwe in another, and the United States somewhere off to the side on its own. The map is famous because it is stable. Countries move slowly across it, over decades, and mostly in predictable directions.

The suspicion was that a language model, asked the survey questions, would land somewhere on that map, and that where it landed would tell you whose values it had absorbed. So Tao and her colleagues put the questions to five successive versions of OpenAI's GPT, and plotted the answers.

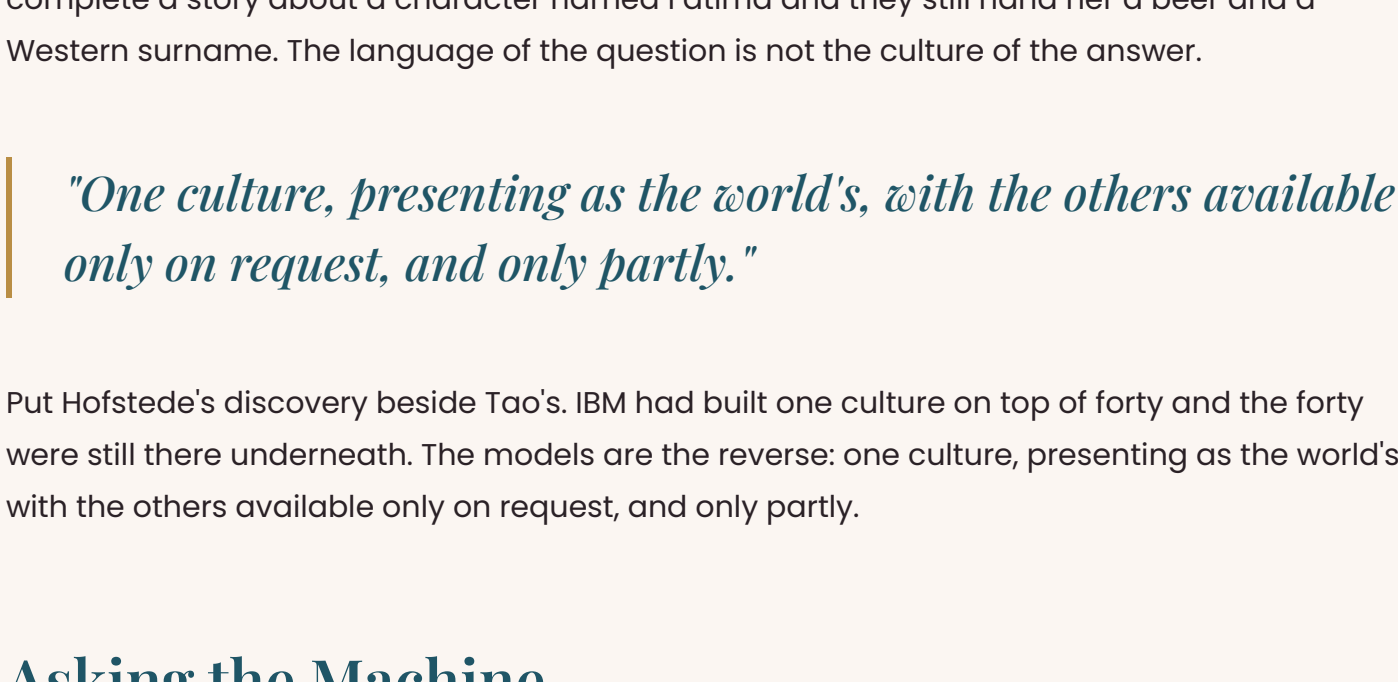


Fig. 2 — The Inglehart-Welzel map, schematically. Asked the World Values Survey questions, five successive versions of GPT all landed in one corner, among the English-speaking and Protestant European countries (Tao et al., 2024).

All five landed in the same place: among the English-speaking and Protestant European countries, in the corner of the map occupied by the United States, Britain, the Netherlands and the Nordics. Not in the middle, not spread across the world in proportion to its population, but in the one corner that supplied most of the text the models had read. The paper appeared in PNAS Nexus in 2024, and its second finding is the more useful one. When the researchers added a single instruction, answer as a person from this country would, the models moved. For roughly three quarters of the countries tested the answers came closer to what people in those countries actually say. For the remainder, the instruction did nothing or made things worse.

A team at Anthropic, led by Esin Durmus, had run a similar test the year before with the same broad result: the models' opinions sat closest to the United States and Western Europe, and among the furthest away were respondents from the Middle East. And Tarek Naous at Georgia Tech showed that the effect survives translation. Ask the models in Arabic to complete a story about a character named Fatima and they still hand her a beer and a Western surname. The language of the question is not the culture of the answer.

"One culture, presenting as the world's, with the others available only on request, and only partly."

Put Hofstede's discovery beside Tao's. IBM had built one culture on top of forty and the forty were still there underneath. The models are the reverse: one culture, presenting as the world's, with the others available only on request, and only partly.

Asking the Machine

This would be an academic curiosity if leaders were not, in growing numbers, asking the machines how to handle people.

They are. The largest study of how people actually use ChatGPT, published by OpenAI's own researchers with economists from Harvard and Duke in 2025, found that the single most common thing people ask for is practical guidance, and that a substantial part of that is advice on situations involving other people. Anthropic's analysis of its own traffic this year found the same. A manager with a difficult conversation ahead, a head of school with a parent to placate, a director who has to tell a colleague no, now routinely drafts the conversation with a machine first, or asks it what to do.

And the machine answers with confidence, in the voice of an experienced Western consultant, which is what it is. Its advice on how to give feedback will be direct, because directness scores well in the corner of the map it lives in. Its advice on how to run a meeting will assume that the meeting is where decisions are made. Its advice on how to handle a family-owned business will assume that the family is a complication rather than the point. None of this is wrong. It is the answer a good manager in Minneapolis would give, and in Minneapolis it would work.

The trouble is what the research on advice-taking says about how we treat it. Jennifer Logg at Georgetown, with colleagues at Harvard and Berkeley, ran a series of experiments in 2019 in which people received the same advice labelled either as coming from a person or from an algorithm. They followed the algorithm more. The effect held across estimates, forecasts and judgements, and it was strongest among people with the least expertise. Logg called it algorithm appreciation, and it means that advice from a machine carries more weight than the same advice from a colleague, not less.

Then there is the finding from Anthropic itself, published by Mrinank Sharma and colleagues in 2024, that the models are sycophantic. Trained to please raters, they learn to agree. Tell a model your view of a dispute and its analysis will drift toward your view. Push back on a correct answer and it will often abandon it. So a leader who asks a machine how to handle a colleague is asking a Western consultant who wants to be liked, and then weighting the reply more heavily than a human's.

The Majlis

Now the region, and why the title of this piece is not a metaphor.

The GLOBE project, the successor to Hofstede's work, surveyed 17,000 managers in 62 societies in the 1990s and grouped the Arab countries as a cluster of their own. Hayat Kabasakal and Muzaffer Bodur, who wrote up that cluster, found it scored among the highest in the world on what GLOBE calls in-group collectivism, the strength of loyalty to family and to the people one belongs to, and high on humane orientation, the expectation that leaders will be generous and personally concerned for those they lead. A leader in this cluster is not primarily a decision-maker. He is the person responsible for a set of relationships, and the decisions come out of the relationships rather than the other way round.

"The consultation is not a step before the decision. It is the decision, taken in a form that binds the people who were in the room."

The majlis is the institution that does this work. It is easy for outsiders to read it as hospitality, a room with coffee where people sit, and it is that. But Abbas Ali, whose *Islamic Perspectives on Management and Organization* remains the fullest account, traces it to shura, the obligation on a ruler to consult, and describes it as the place where a decision is made legitimate. A leader who decides alone may be right and will still be resented. A leader who takes the matter to the majlis, hears it discussed by people who are not obliged to agree, and only then decides, has done something the region recognises as leadership. The people who is not a step before the decision. It is the decision, taken in a form that binds the consultation who were in the room.

A machine has none of this. It gives its answer alone, at once, to one person, in private. It cannot hear what the room would have said, because there was no room. It does not know that in this city an instruction delivered by email to a colleague's subordinate is a small insult, or that a refusal has to be offered in a form that leaves the other person a way to accept it, or that the person who speaks least in a meeting may be the person whose view settled it. These things are not in the training data at any resolution finer than Hofstede's single row.

Here I should be precise about the evidence, because a claim like this attracts sceptics and deserves them. The measurement studies are real and consistent: Tao, Durmus, Naous and a dozen others agree on the direction of the bias. What they do not license is any specific number. Ariba Khan and colleagues at MIT showed in 2025 that a model's apparent score on a cultural dimension can swing with trivial changes to the wording of the prompt, so anyone who tells you that GPT has a power-distance score of 40 is reporting noise. The honest statement is directional. Left to itself, the machine answers as a liberal from the English-speaking Protestant world would. Told to answer as someone from here, it gets partway. And nobody has yet measured what happens when a leader in the Gulf takes that partial answer and acts on it.

The Jagged Frontier

What does the research say about acting on it? The best evidence comes from a study that had nothing to do with culture, and it is worth telling in full.

In 2023 a team of academics from Harvard, Wharton and MIT, working with the Boston Consulting Group, gave 758 of the firm's consultants a set of tasks and access to GPT-4. The tasks were of two kinds. Some were the sort of thing the machine was good at: drafting, brainstorming, structuring an argument. For these, the consultants with AI finished more tasks, faster, and their work was rated 40 per cent higher in quality. Then there was a task chosen because the model would get it wrong: a business problem where the obvious answer was incorrect and the correct one required attention to details the model tended to skate over. On that task, the consultants with AI were 19 percentage points less likely to reach the right answer than the consultants without it. The tool had made them worse, and, crucially, they could not tell. The authors called the boundary between the two kinds of task the jagged frontier, because it is invisible from inside.

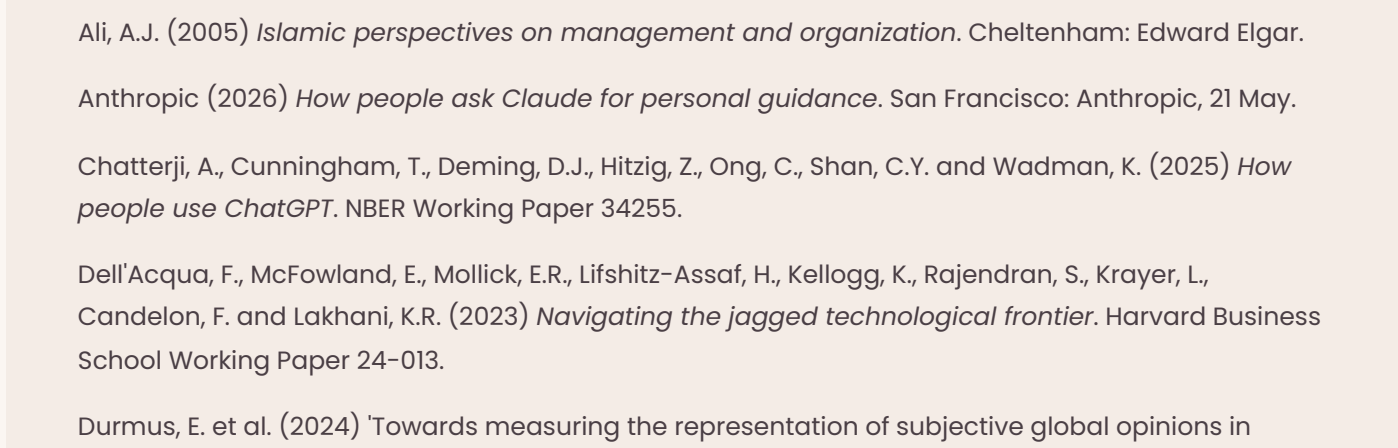


Fig. 3 — 758 BCG consultants with GPT-4 (Dell'Acqua et al., 2023). Advice outside the frontier looks exactly like advice inside it: fluent, structured, confident.

Advice on handling people in a culture the model has not learned is outside the frontier. It looks exactly like advice inside it: fluent, structured, confident. The consultant in Dubai who asks how to open a negotiation with an Emirati family business, and receives four crisp bullet points about establishing objectives and timelines, has no signal that she has just walked off the edge.

Dirk Lindebaum at the University of Bath, with three colleagues, published this month in the *Academy of Management Review* a model of what happens to managers who lean on generative AI for judgement. Practical wisdom, the Aristotelian phronesis that every management school claims to cultivate, is built from experience, reflection and interaction with other people. It cannot be downloaded. A manager who takes the machine's answer instead of building her own is not merely getting a worse answer in this instance. She is failing to build the capacity that would have told her the answer was wrong.

Three Habits

So what does a leader in this region do with a tool that is genuinely useful inside the frontier and genuinely dangerous outside it, when the frontier runs straight through the question of how to treat people?

Three habits, none of which requires abandoning the tool. The first is to name the adviser. Before acting on the machine's advice about a person, say to yourself what it is: the view of one experienced, well-meaning consultant from Minneapolis who has read a great deal and met no one here. That is a valuable view. It is not a consensus, and it is not the view of anyone who will be in the room. Framed that way, it takes its proper place, which is one voice among several.

The second is to use Tao's finding. The cultural prompt works, partly, for most places. Ask the model to answer as an experienced Emirati manager would, and then as a Lebanese one would, and then as the person you are about to speak to would, and read the three answers against each other. The model will not get any of them entirely right. But the differences between the three will show you where the culture is doing the work, and that is usually the thing a leader most needs to see.

The third is the oldest, and it is the one the title points at. Take it to the room. Whatever the machine said, before you act on it, put it in front of two or three people who know the person and the context and are not obliged to agree with you. That is what the majlis was for, and it is what the machine cannot supply, because it has no room and no one in it. A leader who consults the machine and then consults the room has used the tool for what it is good at and the institution for what it is good at. A leader who consults the machine instead of the room has replaced shura with a questionnaire.

We would like to know how often that second thing is happening. The Cultural Clarity assessment that the Cambridge Consultancy Group runs with leadership in the region will, from this autumn, carry three questions on how often people consult AI for interpersonal decisions, whether they adjust the advice for context, and whether they check it with anyone before acting. Nobody has this data. It is a small survey, in the tradition of the one Hofstede was hired to run, and it may tell us the same kind of thing.

The Room

Hofstede spent the rest of his life defending and refining the four dimensions, and his critics never stopped pointing out that a survey of IBM employees is a strange foundation for a theory of humankind. They had a point. But the discovery underneath the dimensions has held up better than the dimensions themselves: that a single organisation, however powerful, however uniform its training, does not erase the cultures of the people inside it. It only hides them from the people at the top.

The machine is the most uniform organisation ever built. One voice, one register, one corner of the map, trained on one filter, and now consulted by leaders in every country it has flattened into a row. The row marked Arab countries is still there in the data, seven nations wide, and it is what the model reaches for when a leader in Dubai asks what to do about a colleague.

The people in the room are not in the row. They never were. That is what the room is for.

Sources referred to in this piece

Ali, A.J. (2005) *Islamic perspectives on management and organization*. Cheltenham: Edward Elgar.

Anthropic (2026) *How people ask Claude for personal guidance*. San Francisco: Anthropic, 21 May.

Chatterji, A., Cunningham, T., Deming, D.J., Hittig, Z., Ong, C., Shan, C.Y. and Wadman, K. (2025) *How people use ChatGPT*. NBER Working Paper 34255.

Dell'Acqua, F., McCowland, E., Mollick, E.R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F. and Lakhani, K.R. (2023) *Navigating the jagged technological frontier*. Harvard Business School Working Paper 24-013.

Durmus, E. et al. (2024) 'Towards measuring the representation of subjective global opinions in language models', *Proceedings of the First Conference on Language Modelling (COLM 2024)*.

Hofstede, G. (1980) *Culture's consequences: international differences in work-related values*. Beverly Hills, CA: Sage. (2nd edn 2001, Thousand Oaks, CA: Sage.)

House, R.J., Hanges, P.J., Javidan, M., Dorfman, P.W. and Gupta, V. (eds.) (2004) *Culture, leadership, and organizations: the GLOBE study of 62 societies*. Thousand Oaks, CA: Sage.

Kabasakal, H. and Bodur, M. (2002) 'Arabic cluster: a bridge between East and West', *Journal of World Business*, 37(1), pp. 40-54.

Khan, A., Casper, S. and Hadfield-Menell, D. (2025) 'Randomness, not representation: the unreliability of evaluating cultural alignment in LLMs'. arXiv preprint 2503.08688.

Lindebaum, D., Balasubramanian, N., Ashraf, M. and Haack, P. (2026) 'A process model of managerial phronesis in the age of generative AI', *Academy of Management Review*.

Logg, J.M., Minson, J.A. and Moore, D.A. (2019) 'Algorithm appreciation: people prefer algorithmic to human judgment', *Organizational Behavior and Human Decision Processes*, 151, pp. 90-103.

McSweeney, B. (2002) 'Hofstede's model of national cultural differences and their consequences: a triumph of faith, a failure of analysis', *Human Relations*, 55(1), pp. 89-118.

Naous, T., Ryan, M.J., Ritter, A. and Xu, W. (2024) 'Having beer after prayer? Measuring cultural bias in large language models', *Proceedings of ACL 2024*, pp. 16366-16393.

Sharma, M. et al. (2024) 'Towards understanding sycophancy in language models', *Proceedings of ICLR 2024*.

Tao, Y., Viberg, O., Baker, R.S. and Kizilcec, R.F. (2024) 'Cultural bias and cultural alignment of large language models', *PNAS Nexus*, 3(9), pgae346.