




IPSOS VIEWS

MIND OR MACHINES

**Exploring AI moderation
and when to use it**

Ajay Bangia
Jim Legg
Betsy Georgiton
Manuel Garcia-Garcia, PhD.





At Ipsos, we champion the unique blend of Human Intelligence (HI) and Artificial Intelligence (AI) to propel innovation and deliver impactful, human-centric insights for our clients.

Our Human Intelligence stems from our expertise in prompt engineering, data science, and our unique, high quality data sets – which embeds creativity, curiosity, ethics, and rigor into our AI solutions, powered by our Ipsos Facto Gen AI platform. Our clients benefit from insights that are safer, faster and grounded in the human context.

#IpsosHiAi

Introduction

In the rapidly evolving landscape of qualitative research, traditional methods like in-depth interviews and focus groups increasingly struggle with scalability and resource demands as they are constrained by moderator bandwidth, respondent availability and budget constraints. AI moderator bots, powered by sophisticated Large Language Model (LLM) systems, have the potential to fundamentally transform this dynamic. These AI moderators claim to mimic and enhance human moderation through text or voice-based interfaces, offering revolutionary scalability.

Industry reactions to AI moderator bots reveal a dichotomy of perspective. Sceptics argue these bots will never encapsulate the ineffable essence of genuine human moderation. Conversely, proponents argue that these innovations could dissolve the barriers between qualitative and quantitative research methodologies. This is because AI moderator bots can efficiently reach a much larger sample size, similar to that of quantitative research, without the constraint of a human moderator being present for every interaction. This evolution towards “conversational research” promises a “**quant-litative**” approach, blending qualitative depth with quantitative reach.

At Ipsos, we focus on exploring issues comprehensively and uncovering subtle nuances. Our ESOMAR award-winning paper ‘**Empathy or Emptiness**’,¹ highlighted that while AI moderator bots offer advantages, they fall short in empathy

and nuanced interactions. What if the key to improving AI moderator bots lies not just in technology, but in understanding human psychology? This paper explores how borrowing from the world of psychology can elevate AI moderators, creating more engaging and insightful interactions.

Helpful co-moderators, yet human stars shine brighter

Over the past two years, dozens of AI moderation platforms have entered the marketplace. Ipsos has assessed and worked with a large percentage of them, identifying clear strengths and weaknesses.



AI moderator bots **excel at:**

- **Engagement:** AI moderator bots demonstrate impressive engagement results, outperforming unmoderated platforms like bulletin boards or digital diaries. Their ability to recognize, recall and reference specific details makes respondents feel heard and acknowledged, especially compared to untrained moderators.
- **Always-on access:** Constantly available without breaks, ready to interact at respondents’ convenience.
- **Scalability:** Hosting hundreds of interviews simultaneously becomes feasible.



But AI moderator bots **have significant weaknesses:**

- **Weak improvisation and real time adaptation:** Bots struggle with spontaneous, probing questions that arise during conversations. Trained human moderators on the other hand can rely on intuition and deviate from a rigidly structured discussion guide. This runs the risk of adhering to a rigid, pre-defined flow often ignoring the nugget that could sit just around the corner.
- **Gaps and inconsistencies:** While AI seeks answers to research objectives, it cannot observe unarticulated nuances or probe inconsistencies, failing to provide a holistic understanding of human behavior.
- **Limited rapport building:** Bots have restricted ability to build meaningful connections with respondents.

While conducting research-on-research, expert Ipsos moderators in three different countries compared masked transcripts (unaware which were human or AI moderated) and rated AI-moderated interviews poorly on all parameters:

Table 1: Blind evaluation of transcripts by expert human moderators

	AI moderated interviews using a standard out-of-the-box bot	Human moderated interviews
Ability to build rapport with the respondent	1	5
Ability to manage conversation flow through all the topics (no awkward transitions)	1.5	4
Ability to ask new questions that aren't in the Discussion Guide (DG) based on the conversation flow	1	4.5
Can adjust approach based on respondent's communication style	1	4.5
Ability to re-direct when respondent goes off topic	1.5	4
Can consistently execute across all the interviews	4	4
Provides quality data	2.5	4.5

Scale:
5 = Very Good;
1 = Very Poor

Source:
Ipsos

In short, the AI acted like a novice moderator focused on the guide while overlooking potentially valuable insights.

Can multi-agent systems be our silicon saviors?

Many AI moderator bots aim to solve the gap predominantly through technological solutions. These approaches typically involve multi-agent systems (MAS), where intelligent agents collaborate as interviewers and supervisors, while working towards a shared goal. The system processes the interview protocol along with the latest inputs from participants to decide whether to proceed to the next scripted question or delve into follow-up inquiries based on the ongoing

conversation. The interview script is organized into a sequence of questions, each paired with a time allocation or an instruction on the number of follow up probes required – shallow, deep, or what some AI moderator bots call an “abyss level” of probing. At the start of each new question segment, the AI reads the question verbatim and utilises a language model to dynamically determine the best course of action within the allotted time. Follow-up question reasoning is driven by a

language model that recalls previous dialogue. However, long prompts can degrade the model’s effectiveness. Therefore, experts recommend an architecture that uses a reflection module to condense conversation into summary notes, highlighting key participant insights and using them to formulate the next prompt. This allows the AI agent to prompt the language model with concise reflections instead of the entire interaction.

While these MAS score better than purely using an LLM-powered chat bot, they are still inferior to human moderators. Are we missing a trick here? Why are the AI moderator bots rated as poor, despite having sophisticated architecture? Does the answer to developing a better AI moderator lie purely in technology, or does it demand insights from human psychology?

Getting better by emulating human moderators

Both researchers and clients prefer specific moderators based on their expertise in the topics, or due to unique qualities or approaches. Investigating these attributes might provide insights into developing a better AI moderator bot. Our hypothesis is that by deciphering this “moderator magic”, AI moderator bots are more likely to deliver high quality data. These areas include:

- Subject mastery: Whether considering a specific domain (e.g. early creative development research), category, brand, or demographic, exceptional moderators possess a comprehensive grasp of the subject under investigation. This
- Context and cultural understanding: Great moderators realize that meaning varies based on the context in which the data is collected. The interaction below (Table 2) could have two very different meanings based on the context.

deep understanding of the category and client enables expert human moderators to extract detailed information and craft more insightful follow-up questions.

Table 2: Context alters meaning

	Scenario 1	Scenario 2
Conversation	A- Would you like a cup of coffee? B- A coffee would keep me up	A- Would you like a cup of coffee? B- A coffee would keep me up
Context	It is 11pm, and B is preparing for bed	It is 5 am, and B needs to get to the airport
Interpretation	No, I don't want a coffee	Yes, I do want a coffee

AI moderator bots rely on respondent input and the learning data from LLMs to formulate the next question, but this may not be sufficient as these LLMs are primarily trained on English data, embedding specific cultural values prevalent in Western, highly educated, industrialized, rich and democratic societies². This presents challenges in engaging with users from diverse cultural contexts, since LLMs lack the culturally specific insights necessary for effective

interactions. For example, perceptions and approaches to conditions like depression vary widely across cultures, underscoring the need for AI moderators to incorporate external, culturally diverse contexts. Can embedding factors like political, category and cultural context into AI moderator bots lead to more nuanced interactions and improve follow-up questions, fostering deeper and more meaningful conversations?

Triangulation and mental framework for enhanced understanding



Skilled moderators excel at eliciting genuine answers to difficult questions, questions that challenge a respondent's self-perception, touch on deeply held values or expose the common gap between what people say and what they do. Uncovering these truths requires a degree of sophistication that goes far beyond simple Q&A. For instance, you can't just ask, "Is sustainability important to you?" and trust the answer. A "yes" is often an expression of aspiration or social desirability, not a reflection of actual behavior. A skilled human moderator instinctively knows this and seeks to

validate the statement by triangulating various bits of information - to question, if people truly believe that sustainability is important, how do they act on it?

Would AI moderator bots be able to ask better follow-up questions if they used a framework to examine such gaps in what the respondent is saying and doing? For example, when understanding "say-do" gaps, would the use of a framework such as COM-B, that postulates that behavior is a function of capabilities, opportunity and motivation, add greater rigor when unearthing interventions to address barriers to the behavior?

Identifying emotion



AI is improving at its ability to register changes in sentiment (based on language used). However, it has serious limitations when it attempts to parallel human emotional intelligence. A strong understanding of how humans experience and perceive emotions could improve the AI bot's emotional responsiveness. How could we get the AI moderator bot to spot emotion and would doing so help it to gain greater

depth? How can it be applied intelligently to probing? How can AI moderator bots connect emotion to empathy and is this even possible?



**[With support]
AI moderators
can move beyond
scripted exchanges
to responses that
feel empathetic and
context aware.**



Can psychology make AI moderator bots more human-like?

Psychology shows that what makes interactions feel "human" isn't just the words used, but how they're framed, timed, and emotionally attuned. By applying principles such as emotional granularity, cultural metaphors, and trust cues, AI moderators can move beyond scripted exchanges to responses that feel empathetic and context aware. This design fosters stronger respondent engagement, encouraging people to open up and share more.

One alternative is to program the AI moderator bot to recognize a limited set of discrete emotions, but this would be overly simplistic. An alternate approach is to borrow from the **Ipsos Emotion Framework**, a constructivist approach, allows the AI to interpret emotional states along continuous dimensions of valence (pleasantness/unpleasantness), arousal (intensity) and control (the sense of being in command of the situation).

To develop this framework, Ipsos leveraged a database of emotional words described in **Emotions Around the World**³. While other databases of emotional constructs are often developed using translations primarily tuned to US English norms, Ipsos' approach diverges by constructing a database of emotional words across various languages. This research leverages Ipsos' proprietary database of emotional images, which is a curated collection of visual stimuli that are systematically calibrated to elicit specific emotional reactions across different cultural contexts. This process involves collecting user-generated descriptions of emotions elicited by the images and mapping these descriptions to the valence-arousal-control (VAC) dimensions, allowing researchers to ascertain the emotional significance of words within the specific cultural and linguistic context. This method provides a more culturally appropriate

understanding of emotional concepts and avoids the pitfalls of straight translations, which may not sufficiently capture the true emotional meanings in different cultures.

Could integrating the Ipsos Emotion Framework enable AI moderator bots to:

- **Build better rapport:** If respondent input suggests high arousal and negative valence, well-trained AI could adapt to be calming and direct in follow-ups.
- **Bridge gaps:** Detecting potential mismatches, AI could ask “That’s interesting, could you tell me more about how you felt?” demonstrating attentiveness and building trust.



Theory meets reality check

While our hypothesis seemed logical, real-world testing was essential. We devised multiple pilots with varying conditions to test our theory.



Pilot 1: Agent dynamics

Ipsos launched a study across Spain, Netherlands, and Poland **interviewing matched samples about attitudes, behaviors, motivations, and obstacles** concerning repairing versus replacing electronics. We engaged 150 respondents across three conditions:

- 01 **AI moderator bot:** Multi-agent system with intelligent agents working in tandem
- 02 **Enhanced AI moderator bot:** Added subject matter expertise, context expertise, interaction expertise, and COM-B framework for generating follow-ups around capabilities, opportunities, and motivations
- 03 **Human moderated:** Human moderators with guide instructions and research objectives

Pilot 2: Testing emotion framework

We launched a subsequent single-market pilot in Spain using the same guide and complex setup but **adding Ipsos Emotion Framework instructions, helping the bot assess valence, arousal, and control to respond accordingly.**

Assessment criteria focused on data quality defined as:

- **Thematic amplification:** Profound, layered understanding measured by theme count
- **Emotional nuance:** Higher emotional versus rational theme proportion
- **Expert evaluation:** Masked transcript assessment by experienced moderators


What we found

Improved yet incomplete

Quality improved with subject context and COM-B framework but didn’t reach human-moderation levels. AI’s probing approach emerged as critical for insight quality. While participants provided baseline understanding, valuable motivations and personal contexts often remained unexplored in AI sessions. Surface-level data identifies what consumers do but understanding why requires sophisticated probing beyond current AI systems.

Tuning the AI bot with subject mastery, context and framework (COM-B in this case) resulted in a greater number of themes being uncovered. Moreover, when the blinded transcripts from the three conditions were compared by the panel of expert human moderators in each country, it was found that the addition of subject, context and framework helped the AI moderator bot (in test condition B) to build better rapport and manage follow-up questions and transitions more effectively, resulting in higher-quality data.

Table 3: Blind evaluation of transcripts by expert human moderators

	Test condition A Standard out-of-the-box AI Mod Bot	Test condition B AI Mod Bot with additional Subject, Context + COM-B framework	Test condition C Human moderated interviews
Ability to build rapport with the respondent	1	3	5
Ability to manage conversation flow through all the topics (no awkward transitions)	1.5	4	4
Ability to ask new questions that aren't in the Discussion Guide (DG) based on the conversation flow	1	2	4.5
Can adjust approach based on respondent's communication style	1	2.5	4.5
Ability to re-direct when respondent goes off topic	1.5	2.5	4
Can consistently execute across all the interviews	4	3.5	4
Provides quality data	2.5	3	4.5

Source: Ipsos

So, how can we get the AI moderator bot to perform better while extracting emotional needs?

Emotions matter

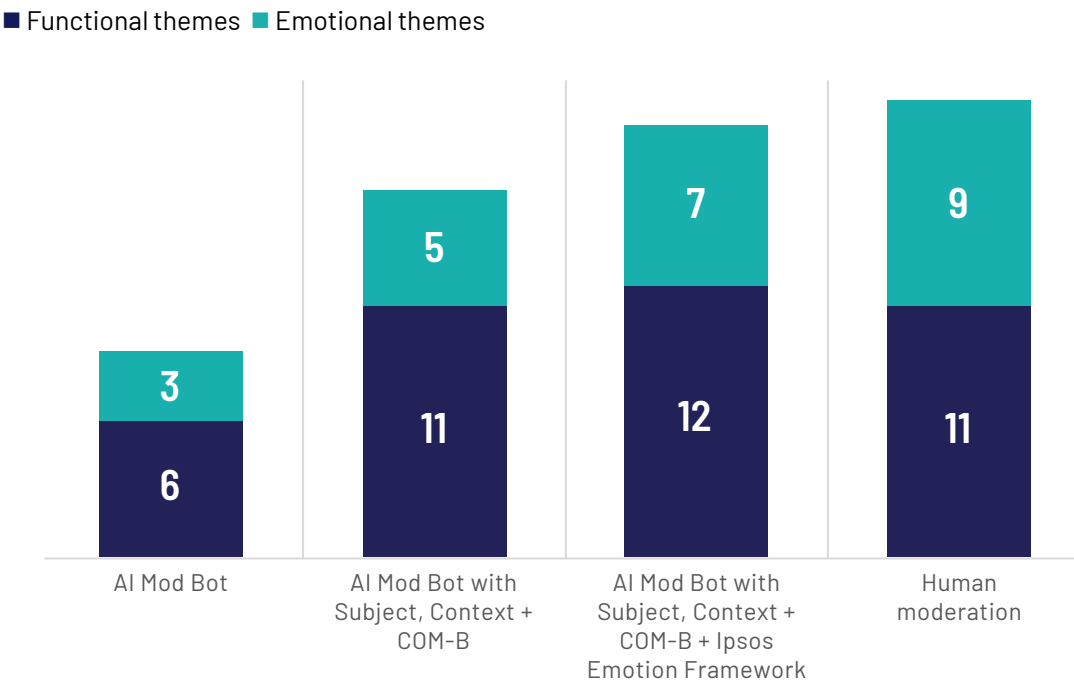
Our fourth test condition in Spain added Ipsos Emotion Framework instructions. Evaluation revealed human moderators’ most significant advantage: intuitive ability to detect and respond to subtle emotional cues, transforming insight depth and quality.

Human moderators consistently demonstrated innate sensitivity to emotional states. Recognizing participants’ environmental concerns led to rich insights about values beyond repair behaviors. In Poland, detecting hesitation when participants admitted fears presented opportunities for exploring psychological barriers that might remain hidden with standardized questioning.

Human moderators adjust entire communication approaches to match participant styles and cognitive abilities. When respondents struggled to answer a question, human moderators rephrased or simplified it multiple times to ensure it was understood. This flexibility contrasts sharply with rigid AI questioning patterns.

Human interviews built upon spontaneous personal anecdotes exploring deeper motivations. Following “emotional breadcrumbs” rather than predetermined paths, moderators discovered unexpected insights otherwise unexplored.

Figure 1: Functional vs emotional insights




Source: Ipsos

Taking all this into account, the Ipsos Emotion Framework has improved the balance and volume between emotional and functional themes. So, does this mean that we don't need human moderators anymore, that AI moderator bots can do it all? Well, no, that is not the case. For issues like discovering latent needs, unexplored

Conclusion


Human or AI moderator bot?

 Our goal in this experiment was to explore both the strengths and weaknesses of bots versus humans, identify their use cases, and determine ways to improve them. We recognised that AI moderator bots can be effectively utilised in various scenarios. The continuum below outlines the best practices for using AI moderator bots alongside human moderators to maximise insight and understanding. When business

desires, say-do gaps, understanding cultural nuances or creating deeply empathetic experiences, solely relying on AI moderator bots is not ideal. These topics usually require more than what AI tools can currently offer in qualitative research.

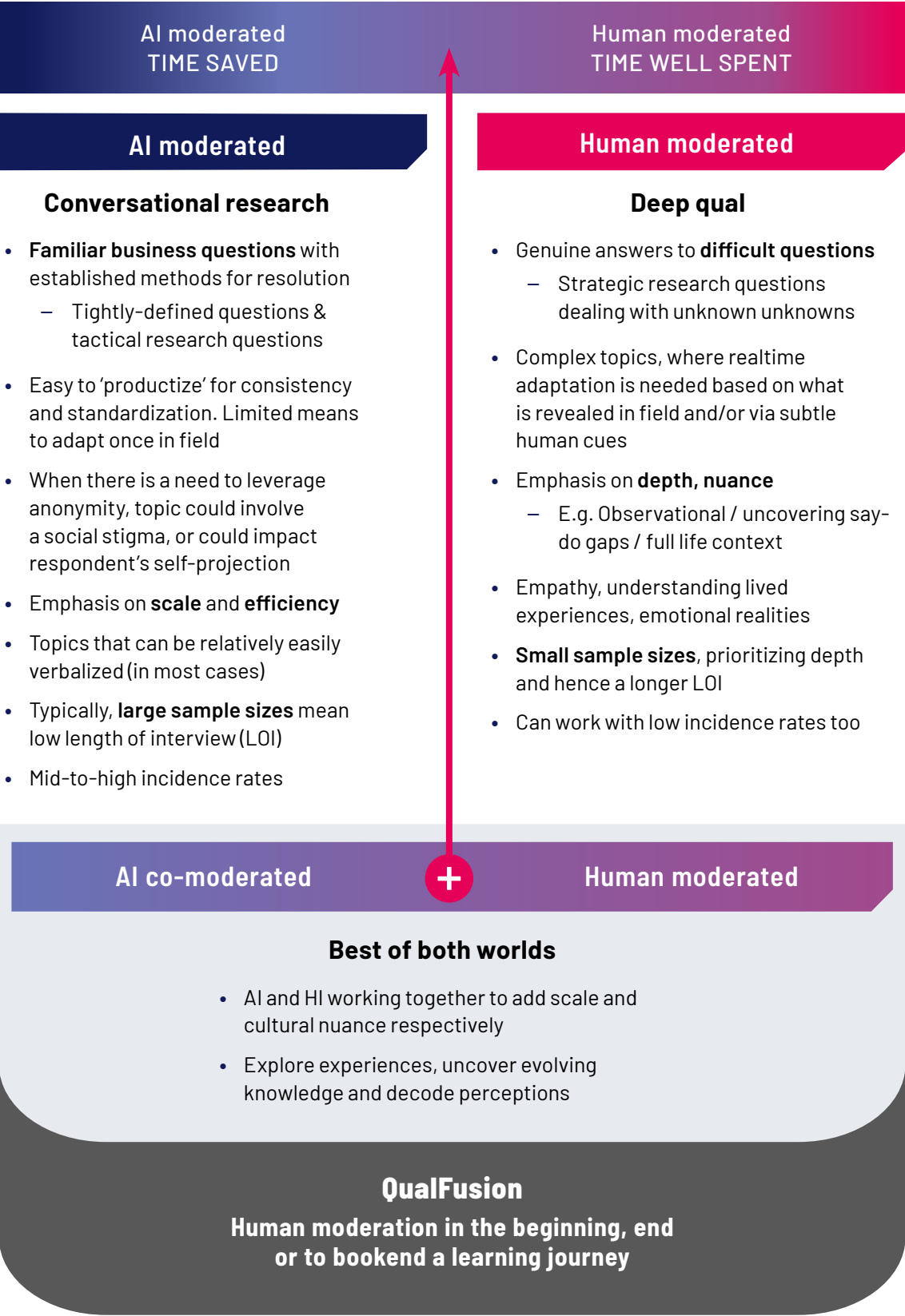
decisions are low-risk, tactical and familiar, and speed is prioritised over depth – where “good enough” suffices – AI moderator bots are promising. Conversely, when research requires culturally nuanced interactions or deeply empathetic engagements, particularly with loosely defined business questions, human involvement becomes essential (see Figure 2).

Human AND AI moderator bot! AI moderators as co-moderators:

 Despite the terminology suggesting that an AI moderator bot or MAS can independently gather extensive data and possibly replace human moderators, this perception does not reflect reality. It plays more of a role of a co-moderator, than the lead moderator. Human qualitative experts play a crucial role in enhancing the moderation process by translating business issues into specific questions or activities, or providing additional subject expertise, context and frameworks that enable the bot to ask more pertinent questions. Expert

human moderators could be invaluable while setting up AI moderator bots. Our research clearly indicates that increasing complexity in setup – through elements like subject context, the COM-B model and the Ipsos emotion framework – yields more robust data. Leveraging expert human moderators can significantly improve data quality. Importantly, this is not a matter of choosing between a human or an AI moderator bot; the answer lies in integrating both approaches.

Figure 2: Integrating expertise and technology for richer, more nuanced data collection



Source:
Ipsos



A well-designed AI moderator must be trained to be aware of its own limitations. Instead of making assumptions based on incomplete data, the AI can be programmed to ask clarifying follow-up questions.

Psychology AND technology build better AI moderator bots:



Psychology and technology together build a better AI moderator bot. We suggest that developing truly effective AI moderators is less an engineering challenge and more a psychological one, requiring these multi-agent systems to be imbued with an understanding of human cognition, emotion and behavior. A human-centric approach is vital, as the goal of any interview or moderated discussion is to foster a connection that encourages open and honest sharing, transforming the interaction from a simple data extraction exercise into a meaningful dialogue. A significant advantage of training an AI moderator with a deep understanding of human psychology is the ability to create a more engaging and comfortable environment for respondents. This is where a constructivist approach to emotion becomes particularly valuable. Instead of programming an AI to recognise a limited set of discrete emotions, which can be overly simplistic, a constructivist framework allows the AI to interpret emotional states along continuous

dimensions of valence (pleasantness/unpleasantness), arousal (intensity) and control (the sense of being in command of the situation). This more nuanced understanding enables the AI to tailor its responses in real-time.

However, a significant challenge for any AI moderator bot, regardless of its sophistication, is the inherent limitation of not being exposed to the full spectrum of human communication. Paralinguistic cues such as tone of voice, facial expressions and body language are rich sources of information that are often lost in purely text-based or even voice-based interactions. To compensate for this, a well-designed AI moderator must be trained to be aware of its own limitations. Instead of making assumptions based on incomplete data, the AI can be programmed to ask clarifying follow-up questions. For instance, if a respondent uses words that seem positive, but the AI detects a potential mismatch, it could ask, *"That's interesting, could you tell me more about how you felt in that moment?"*

or *"I want to make sure I'm understanding you correctly, could you elaborate on that?"* This not only helps to compensate for the lack of non-verbal information, but also demonstrates to the respondent that the AI is attentive and committed to understanding their perspective, further building rapport and trust.

Ultimately, the strategic application of these psychological principles, guided by the respondent's emotional cues, will yield richer qualitative data. The most effective AI interviewers and moderators will be those that can dynamically adjust their interaction style based on the dimensions of the respondent's emotional state. When a respondent expresses sentiments high in positive valence and arousal, AI can mirror this with more enthusiastic and expressive language to encourage further elaboration. Conversely, when a respondent's input indicates low valence and low control, a more neutral, non-judgmental AI persona can create a safer space for candid sharing. This adaptive capability, grounded in a sophisticated, dimensional understanding of emotion and an awareness of its own perceptual

limitations, allows the AI to create the most conducive environment for insightful responses in any given moment. This thoughtful application of psychological principles is what will truly enhance human interaction and, in turn, the quality of the data collected.

Organisations with AI moderator bot solutions that work to understand what makes an expert human moderator great and, as importantly, the variability and unpredictability of the average human respondent, will be the AI moderator bots to watch in the coming years. They will add more agents focused on proven scientific frameworks, like the Ipsos emotion framework, to their multi-agent systems. Those that do will be able to demonstrate higher data quality. Unfortunately, the average users don't ask or understand what the "secret sauce" (within AI moderator bot multi-agent systems) is. It will be critical for users to have a trusted expert to help evaluate what is behind the "multi-agent system curtain" and make an educated choice on which one is the best to leverage in their research.

This Ipsos Views POV is predominantly based on an extended paper (of the same name) presented at ESOMAR North America 2025.

Endnotes

- 1 Ipsos. 2024 [Empathy or Emptiness: Unravelling the Impact of AI on Human Connection](#). Ajay Bangia, Rollo McIntyre, Jim Legg
- 2 Ada Lovelace Institute. Perez, J. [Tokenising culture: causes and consequences of cultural misalignment in large language models](#).
- 3 Ipsos. 2023 [Emotions around the world: A cross-cultural framework for emotion measurement](#). Manuel Garcia-Garcia, Ph.D, Davide Baldo, Rich Timpono, Ph.D

DECEMBER 2025

MIND OR MACHINES

Exploring AI moderation and when to use it

AUTHORS

Ajay Bangia

Chief Innovation Officer
(AI Solutions), Ipsos UU

Jim Legg

Global Head of Operations,
Ipsos

Betsy Georgiton

Vice President of Research
Excellence, Ipsos in the US

Manuel Garcia-Garcia, PhD.

Global Lead, Science Activation,
Research & Strategy, Ipsos

The **IPSOS VIEWS** white
papers are produced by the
Ipsos Knowledge Centre.

www.ipsos.com

@Ipsos

