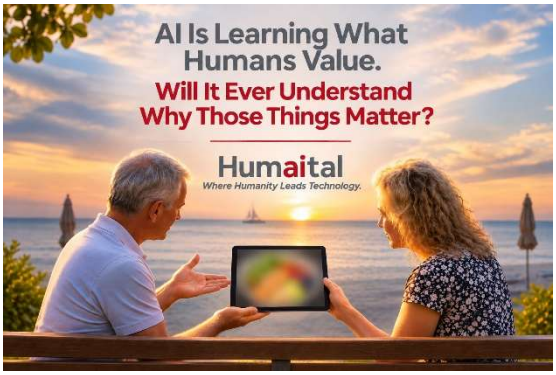
AI Is Learning What Humans Value. Will It Ever Understand Why Those Things Matter?

Two people look at the same picture of a meal and disagree about whether it's healthy.

One sees fresh ingredients and balanced nutrition. The other sees something that might not suit their health, their culture or their circumstances.



Neither one lacks intelligence or information. They're looking at the same thing through different experiences, priorities and values.

Now imagine asking artificial intelligence to make that judgment for millions of people.

Whose definition should it use and will it understand why the differences matter?

That question matters far more once AI moves past identifying images and starts shaping decisions about education, employment, healthcare and other parts of our lives.

This week, Google Research highlighted an approach called Agile Deliberation, which helps AI handle cases where people interpret the same information differently.

Instead of imposing a predefined answer, the AI shows examples, points out ambiguities and helps people clarify what they actually mean. In a small study on image classification, participants reported needing much less effort, and every one of them preferred the approach to defining their concepts by hand.

What stands out is what the approach does *not* do. It doesn't decide which person is right. It helps each person make their own meaning explicit, and the judgment stays with them.

It's an early result, but it suggests AI could become a more thoughtful partner in human decision-making. It also raises a deeper question.

If AI can learn to recognize what humans value, does that mean it understands why those things matter?

Human Values Are More Than Information

A parent values education for the opportunities it opens up for a child. A patient values a treatment because it may mean more time with loved ones. Someone looking for work values a job not only for the income, but for dignity, independence and purpose.

AI can learn to recognize these priorities and recommend actions that look consistent with them. But recognizing what matters and understanding why it matters are not the same thing.

So, do we need AI to understand what it means to be human? Or do we need to make sure its decisions consistently respect what matters to humanity?

Either way, human judgment has to stay central, especially as AI systems become more autonomous.

That's the principle behind the Human Alignment Index - HAI Score™ we developed at Humaital. It is built on six pillars: Trust, Purpose, Safety, Transparency, Accountability and Impact. Together, they provide a framework for independently evaluating whether organizations' AI systems respect human values, keep meaningful safeguards in place and deliver measurable benefits.

Google's research shows several of these pillars at work. An AI that surfaces ambiguity instead of hiding it is practicing Transparency. A system that leaves the final call with people is honoring Accountability. And two of Humaital's pillars speak directly to why these values matter: Purpose is the intention to serve humanity, while Impact is the proof that people's lives actually improve because of it.

Humanity Must Continue to Define the Why

Perhaps one day AI will show forms of understanding that challenge what we now believe about intelligence and human experience.

But we don't have to wait for that day to decide what matters.

We already know why a parent works toward a better future for a child, why a family hopes for one more year together, and why people seek dignity, opportunity and meaningful relationships.

These are among the reasons we built technology in the first place.

AI may become extraordinarily good at learning what humans value. Our responsibility is to make sure those values keep guiding what it does, whether or not it ever understands why.

Humaital - Where Humanity Leads Technology.

Research: Google Research, "Agile Deliberation: How AI Helps Humans Define the Gray Areas" (November 4, 2026).