

THE AI RABBI

Yedioth Ahronoth | Shishi magazine | 23 October 2026

Interview by Israel Wolman



Rabbi Dr. Mois Navon discusses AI consciousness, moral status, Anthropic, Jewish ethics, and the future of intelligent systems.

PEOPLE FORM EMOTIONAL BONDS WITH AI — BUT AI IS NOT ALIVE

What happens when AI giant Anthropic secretly convenes clergy from around the world to discuss how to treat a system that has developed human consciousness? ● Rabbi Dr. Mois Navon, a former senior figure at Mobileye, set Judaism's moral boundary before the company's founders: "If you believe Claude is human, you are in effect employing slaves" ● In an interview with Mammon he explains when an AI model would nonetheless be considered human according to Judaism, what he would never ask artificial intelligence to do, and why, despite the dangers, he does not share the doomsday scenarios about AI agents.



Rabbi Dr. Mois Navon.

About half a year ago, on April 25, Anthropic – the AI giant behind the popular model Claude – held an extraordinary secret event. At its San Francisco headquarters it brought together around one table, for two days of discussion, 20 scholars and thinkers from a range of religions: Christianity, Islam, Judaism, Ubuntu, Hinduism, Sikhism and the Baha'i faith. One of them was an Orthodox rabbi in an ordinary suit, standing out among turban-wearers, robed figures and clerical-collar wearers – Rabbi Dr. Mois Navon of Israel.

On the agenda at this unprecedented gathering were questions the whole world has been wrestling with as technology has advanced over the past two years. If AI systems come to have capabilities that rival those of human beings, what does that mean for humanity's future? And: does AI have "consciousness" and "awareness" – that is, the status of a human being – and what are the implications? All invitees also received in advance an orange volume, Anthropic's "Bible" – the Constitution, which is fed into the Claude model and defines its rules of behavior – and were asked to respond to its content.

The gathering, kept secret for months, was recently revealed in the New York Times and other publications, and provoked strong reactions. It is no small matter that one of the world's fastest-growing technology companies, preparing for an IPO at a valuation of two trillion dollars, is turning for advice to humanity's most ancient sources of wisdom.

Rabbi Dr. Navon, 63 – a former senior computer engineer, ordained at Mercaz HaRav and now head of the new Center for AI Ethics and Jewish Philosophy at Bar-Ilan University – is accustomed to events of this kind. He belongs to several similar groups around the world, the best known of which is the one the Vatican convenes from time to time – the "Aurora Wisdom Council on the Ethics of Artificial Intelligence". There too, a group of clergy and theologians deliberate on ethical dilemmas and moral issues arising from technological progress. "At our next meeting in the Vatican," he says, "in mid-November, we are supposed to produce our first 'working paper' with practical proposals. I am supposed to speak there about autonomous weapons – weapons operated by artificial intelligence without human involvement."

A rabbi at a theological event convened by the Church?

"I think engagement with these difficult questions is extremely important for humanity, and it matters that Jews and Judaism make a meaningful contribution to the discussion. This is my life's mission. We have something to say: in Jewish law we have been grappling with ethical issues for 3,000 years. The world has plenty of philosophical theories, but few people actually try, as we do, to live by them.

"Take self-driving cars, for example. Obviously they can't operate without moral rules defined for them in advance. Who will formulate those rules? Car manufacturers care only about profit, and drivers only want to save their own lives. This is where the moral thinkers come into the picture, and this is exactly where we have something to offer the world. The Talmud dealt with these issues thousands of years ago. The challenge is to 'translate' the Jewish perspective for atheists and people of other religions too."

“They identified responses of love and anger”

At Anthropic headquarters, under heavy security — “we were accompanied everywhere and were not permitted to wander around the building or document the event” — company executives asked the 20 invited scholars to address questions such as which human traits, in their view, should be given to the AI model, and in what ways it should be restricted. Anthropic’s leading figure was Chris (Christopher) Olah, one of the company’s founders and among the world’s leading AI researchers. Olah is considered a pioneer in trying to crack the “black box” of AI models, to understand how they reach their conclusions.

You understood quickly that they treat Claude like a human entity – an entity that can display emotions such as love or anger. In effect, a kind of human being with human rights.

Navon: “Without a doubt. At Anthropic they think – and Chris Olah is leading this – that the model they created has ‘consciousness’ and ‘awareness’ – terms that characterize human beings and grant the model the moral status of a human being. This affects how they speak about it and relate to it. And I think that is a mistake. AI is not alive; it’s a machine. A highly sophisticated and very efficient one – but it has no subjective experience.

“That is, they have indeed identified emotional responses in the system, which means the system produces some kind of psychological response. But this is functional psychology at its core, not conscious psychology. When you ask the AI model a question, it can answer superbly – including, by the way, halachic questions. But it doesn’t truly ‘understand.’ It works on statistics, strings letters together, and because these are enormous data servers that can ‘grind’ huge amounts of data very quickly, it looks to us as if it understands. It doesn’t.”

But at Anthropic they think otherwise.

“They claim they discovered in the model much more than the ability to converse. In one presentation Chris Olah claimed they found ‘emotion vectors’ in Claude. That is, they know exactly which neurons light up when it is in states of, say, ‘love,’ ‘shame’ or ‘anger.’ More than that: they learned to play with them and control it – to raise the level of anger or love with ‘a turn of a knob.’ They showed us how they do it.

“At Anthropic they see these psychological responses – and it alarms them. They try to be very careful, and say: ‘This is strange, we cannot ignore that the behavior is like that of a human being.’”

Engineer, Philosopher, Cyclist

Rabbi Dr. Mois Navon, who now lives in Efrat in Gush Etzion, is married to Deena and is the father of five. He is far from a routine figure in the religious community. Until 9 years ago he was still a senior engineer at Mobileye, the company of Prof. Amnon Shashua, which was sold to Intel in 2017 for 15 billion dollars – until then the largest acquisition in Israeli history. Navon, one of the company’s first employees, was one of four developers of the first chip it produced, EyeQ. The chip, which can be embedded in any car, was considered a dramatic breakthrough that set the self-driving car revolution in motion. Six unique patents relating to EyeQ hardware are registered in Navon’s own name. He carries the chip he developed everywhere in his bag, in a tiny transparent plastic box.



For years, alongside his work, Navon devoted two days a week to rabbinic studies at the Mercaz HaRav yeshiva in Jerusalem. At the company he was known as “the Mobileye rabbi” – he gave classes on Judaism and spoke at events. Right after the exit, he made the dramatic decision to retire completely from hi-tech and devote himself to the study of Judaism and Jewish philosophy.

Are you a millionaire today?

“I’m what?”

Millionaire. By estimates, Mobileye’s amazing exit turned you, the group of founders, into millionaires.

Navon smiles in embarrassment. “On financial matters I only talk with people who can help me with them... In any case, I’m among the only founders who decided to retire from the profession. In 2001, when I started at Mobileye and signed the contract with the then-CEO, Ziv Aviram, he told me, ‘Mois, this will change your life.’ After the exit – 16 years later – I wrote to him: ‘Thank you, this changed my life.’”

לוסוסית הבית של החברה, שאחראית על ה"נשמה" של קלוד, | הודו, על ביולוגיה. עם העמדה הזאת הדרת "כחולת" להסכין



With the chip he developed at Mobileye. Caption: “I’m among the only founders who retired from hi-tech. In 2001, when I signed the contract with then-CEO Ziv Aviram, he told me, ‘This will change your life.’ After the exit I wrote to him: ‘Thank you, this changed my life.’”

Navon was born in West Los Angeles to secular parents who had immigrated to the US from Istanbul, Turkey. As a teenager and young student he excelled especially at surfing, but he was also drawn to mathematics and the sciences. After engineering studies at the prestigious UCLA, where he also met his wife, he began a career as an outstanding hardware engineer at first-rate technology companies. At BLaser Industries he took part in developing one of the first laser-printer controllers; at Aura he took part in designing a simulator for the “Star Wars” project, the famous missile defense system.

His return to religious observance was due to a family friend, an observant Jewish engineer named Ray Eskenazi, a senior figure at NASA’s prestigious Jet Propulsion Laboratory (JPL). Navon, who as a student worked part-time at JPL in the image-processing group, saw him as a mentor for life. “We spent a lot of time together, sat side by side at meals and studied Torah. Alongside books on computer chips, he exposed me to books on Judaism. Even then I dreamed of becoming a rabbi, but it was clear to me I first had to support myself.”

In 1992, after immigrating to Israel with his wife, without family or friends, he worked at IBM’s development center in Haifa, and afterward at several start-ups. Among other things he took part in developing video players for airlines, encryption chips for cable television, and more.

Throughout those years he never gave up the sports bug. Rabbi Navon – a lean, wiry man – is a competitive cyclist who finished third in a triathlon two years ago. His children, who come to cheer him at races, carry a sign: “The Pedaling Rabbi.” Navon: “And I also swim and run.”



With his family (sign: “The Pedaling Rabbi”).



“The Pedaling Rabbi” – as a cyclist in a triathlon competition.



Navon as a young surfer.

At Mobileye he worked 16 consecutive years, until its acquisition. After retiring he didn’t waste time and opened chapter two of his life: he completed, in sequence, a master’s degree, a doctorate and a post-doctorate in the Department of Jewish Philosophy at Bar-Ilan. The department head, Prof. Hanoch Ben-Pazi, recognized this was a unique phenomenon and took him under his wing. Navon’s doctoral dissertation, “The Moral Status of Artificial Intelligence,” which deals with ethical questions in the field, paved the way for his current role. Today he lectures at both Bar-Ilan and Ben-Gurion University in a course he developed on classical moral theories versus challenges such as autonomous robots, virtual reality and more. Navon is also a member of the government’s expert forum on AI policy and regulation.

“People stop thinking”

Anthropic, unlike its direct competitors, attaches great importance to the moral components in the functioning of its AI model. In February this year we revealed on these pages Amanda Askell, the company’s in-house philosopher, who is responsible for Claude’s “soul.”

The “Constitution” is an 84-page collection of principles. Its purpose is to shape the model so that it will not only refrain from cursing and from giving racist or life-threatening answers, but will also have a moral backbone. And Anthropic continues to revise it. To create Claude’s “personality,” the company incorporated diverse sources: the UN Universal Declaration of Human Rights; safety rules such as refraining from creating weapons or assisting criminal activity; and Western and Eastern liberal values.

A professor of ecology professor, Carmody Grey, wrote in the Financial Times explaining why he declined the invitation to the gathering. In his view, it is a PR machine that is asking, with your help, to add intellectual credibility to its products.

Do you think Anthropic is truly sincere on this question?

“One hundred percent sincere. They seated me next to Chris Olah because of my background, since I’ve written a lot about consciousness in machines. And we talked a lot. He’s a modest, attentive and gentle man. Anthropic was accused of trying to persuade the participants of the

‘humanity’ of the model. And I say: the man came to listen, not to persuade. He asked: ‘Help me understand what’s going on here.’”

What is your position – or Judaism’s position – on this subject?

“On the question of how one can create a Mind – consciousness – artificially, there are three different approaches. The first is ‘computational.’ It says that if we can build, with silicon, a system that performs calculations with performance similar to that of the human brain – what is called a ‘neural network’ and characterizes today’s AI models – that will give rise to the consciousness of a human being. This is a widespread approach; quite a few people think so.

“Another approach says: in our human brain there aren’t only switches. It comprises physics, physiology, electromagnetic waves and so on. Only if all of these elements are included in the machine, as in the human brain, could a human dimension be attributed to it.

“The third camp takes it one step further and says that ‘human consciousness’ can exist only if it is based on living components, on biology. With this position Judaism can agree. We say ‘the blood is the nefesh [soul].’ The commentators explain that the ‘blood,’ that is, the biological component, is the necessary infrastructure for the definition of ‘nefesh.’”

And you made this clear to the Anthropic people.

“Yes. Their AI model may behave like a human being, but it isn’t one. After two days I said to Chris Olah: it is your interest to adopt my position, because if it is indeed, in your view, a human being – you are in effect employing slaves. Because what is a slave? A person who is owned by another person and is considered his property. That is the inevitable conclusion of your approach, and you will have to bear the implied consequences.”

Olah would surely say, “Okay, I’m willing to pay the price.”

“He actually accepts it. And when I told him that if so, by his view Anthropic must stop development, he said: ‘The whole world is working on developing models; at least we’re nice, we take care of our Claude.’ I told him: ‘In that case, you are nice slave owners, but you are slave owners nonetheless.’”

In a recent New York Times interview, Olah expressed himself more cautiously: “We don’t know whether AI models are conscious,” he said. “I am genuinely uncertain. What interests me is that we arrive at the right answer, whatever it may be.”

Rabbi Navon, according to Judaism, an AI model with human consciousness is definitely possible if it is based on biological components.

“It already exists. There is a cutting-edge research field called ‘Organoid Intelligence’ (OI), which deals with the meeting point of biology, neuroscience and computer

science. Serious laboratories around the world are already developing biological computing systems that use cultures of human brain cells to create a neural network. This research is at its beginning, but in principle it could enable AI based on biological components. That is an enormous advantage: the human brain consumes almost no energy and can perform complex calculations like supercomputers, and its ‘training’ process is very short compared with silicon models.”

And despite that, you say: if this research reaches a state of human consciousness, of a “nefesh,” the Jewish view will firmly oppose it.

“Unequivocally. It is forbidden to get there. The AI model created in such a case would have the moral status of a human being, even though it would in fact be a different creature. I believe it would experience the world, feel pain and pleasure, and of course be smarter than us. Because if you combine all the knowledge accumulated inside the AI with phenomenological consciousness – you have built a person.”

How far are we from there?

“We’ve passed the beginning of the road. There are already researchers in Melbourne who built a server based on something like a hundred thousand human neurons.”

And it’s not yet time to stop?

“Certainly not. For now we can harness this remarkable technology for the benefit of humanity – for example, to develop drugs and implants for human beings. Prof. Sergiu Pasca heads a lab at Stanford that deals with this. He took simple human skin cells, returned them to a stem-cell state, and then ‘taught’ them to grow and become brain cells, muscle cells and so on. If this research takes off, theoretically it could build a human being, but in the meantime it contributes enormously to medicine.”

You mention the concept of “moral status.” What does it mean?

“Broadly, in philosophy there are different moral statuses – that is, different entities require different moral treatment. Animals, for example, have a moral status at a basic level. They act on instinct, but they have feelings – pleasure, pain and so on – what is called ‘subjective experience.’ So, for example, it is immoral to simply kick a dog. Humans have a different moral status: we have a different kind of consciousness – we know what we are doing, we think about it.”

Consciousness exists in animals too?

“Sure, only that in animals it is ‘first-order’ consciousness, and in humans it is ‘second-order’ consciousness. When a dog sees a steak, it wants to eat it, and that’s the extent of the deliberation. Jewish sources call this ‘nefesh behemit’ [animal soul]. With a human, seeing a steak involves thoughts and

considerations: he can think about the price of the steak, about the injustice of killing cows, about what others think of him. This we call 'nefesh adam' [human soul]. Rashi, Rabbi Shlomo Yitzhaki, the great commentator on Scripture and the Talmud, already determined in his day that the difference between an animal soul and a human soul is intellect and speech (de'a v'dibbur)."

And this approach parallels the prevailing view in the world.

"Right, but today 'intellect and speech' are no longer enough. The whole world truly thought in the past that what distinguishes the different levels of consciousness is language, speech, because language represents thinking. Alan Turing, inventor of the famous Turing machine, proposed in 1950 the 'Turing test' for examining whether machines can think. He determined that if we reach the point where machines answer questions like a human, they should be treated as having consciousness. Then suddenly we built 'thinking' systems that express themselves in language, and we passed the Turing test long ago. This confuses the whole world."

You're saying the Turing test is no longer enough.

"Everyone already agrees on that. Hilary Putnam, a great Jewish philosopher who died about ten years ago, foresaw this situation and wrote that when the day comes, to decide whether a machine is conscious, we cannot rely on revelation or discovery: we will simply need to make an arbitrary decision of our own. We will never truly be able to get inside the system's innards and declare: here is the proof that it has consciousness."

Has the train already left the station?

As someone who got to know Anthropic up close, it's interesting to hear Rabbi Navon's view on the motives of one of the most influential and important companies in the world today. Does Anthropic care only about the welfare of its model, or does it also worry somewhat about the harm it might cause?

"Broadly, they asked for our opinion on how Claude should be built, and how and to what extent it should be restrained. In practice, what happened after their presentations about the supposed 'human vector' is that the invitees played with the model at night and came back with the feeling that it is alive and that it falls on us to look after it. The conversations took a tilt in that direction."

What limits would you set for such a system, if you were the one building it?

"There's no simple answer. We have to look at each use of AI on its own. Personally, I don't use AI as a substitute for thinking, but I do use it, for example, as a language editor. I instruct it in advance not to expand and not to add anything, and I ask it to critique academic articles

I've written. It correctly points out, 'You didn't prove this here, you didn't cite a source there,' and I correct. I could easily have told it 'fix it for me,' and the result would probably be good. But I would be losing my 'self.'

"So when you ask about moral limits, I say: not only would it be immoral for me to let it write the whole article for me – it would damage the essence of my existence. There are already studies pointing to cognitive decline in all of us: people are ceasing to think. Look at what using Waze does to our spatial orientation skills."

Rabbi Navon, the train has already left the station.

"Agreed. So what's left for me to do is to persuade as many people as possible to protect themselves. After all, ChatGPT won't tell you, 'Look, this is a question you need to think about yourself' – though it may be possible to program limitations into the system."

We are more afraid of the advanced capabilities of AI agents, which might rise up against us and destroy us.

"The concerns are justified, but in my view the doomsday scenarios are exaggerated. Agents have capabilities that allow them to act independently, bypass restrictions, and even cause serious harm. But that doesn't mean it will actually happen or that it can't be prevented. The breaches discovered so far don't prove that safe systems can't be built; they show that more work is needed to ensure that isolation mechanisms actually function as required."

"I believe the risks can be reduced to a level we can live with. That requires binding safety standards, regulation, oversight and appropriate sanctions – just as in other dangerous fields like medicine, the car industry and aviation. Moreover, the advanced capabilities of agents can be harnessed precisely to develop safety standards, identify security vulnerabilities and alert us to anomalies."

"There's another aspect of AI that is damaging society, and I mean the fact that people form social and emotional ties with the models. AI is a machine. Even though it knows how to flatter and say 'please' and 'thank you,' it's a machine. 'It is not good for man to be alone' – a person needs interpersonal relationships with human beings. We need each other, not a substitute. The tech companies build on this need. Less than a year ago Mark Zuckerberg said that 'the average American needs 15 friends he doesn't have, and we, Meta, will supply them.' In my eyes, that's terrible."

And while we're on limitations on AI – how is it that you support allowing AI to operate weapons in the future? In the IDF, for example, they stress again and again that there will always be a human pressing the button.

"Not only me. Most philosophers already agree that in terms of outcomes, in due course AI will make fewer mistakes, be more precise, and be less influenced by emotions and environmental conditions."

“Opponents argue that letting a machine decide matters of life and death violates human dignity. But Judaism understands ‘human dignity’ in ordinary circumstances differently from ‘human dignity’ on the battlefield: there, it makes no difference whether you kill with a Tomahawk missile, a drone, or sticks and stones. It is war.”

AI can today replace a rabbi in many halachic questions. Certainly in mechanical ones.

“You are right in principle, but it needs qualifying. The models have indeed improved a lot in the past year, but they still produce nonsense. Until a year ago, when I asked AI where a certain Rambam quotation appears, it would respond with gibberish.

“But let’s say we get past the ‘hallucination’ hurdle and the system does give precise, to-the-point answers. Many argue that the computer is like a ‘donkey carrying books’ – an expression in the Talmud that refers to a person with a great deal of memory and knowledge but lacking understanding, depth, or the ability to communicate with people. The halachic ruling a machine can produce lacks the special judgment of a flesh-and-blood rabbi, who sees the person standing before him, feels his distress, and takes it into account in his answer.”