KEN FIREMAN: I'm Ken Fireman, the managing editor for SAGE Business Researcher. And I'm speaking with Hannah Kim, who has written our report on the ethical issues surrounding artificial intelligence, or AI. Hello, Hannah.

HANNAH KIM: Hi.

KEN FIREMAN: You write in your report that investment in AI is surging, and the uses to which AI is being put are also growing rapidly. Can you describe some of those uses?

HANNAH KIM: Sure. So often, when people refer to AI, they're talking about machine learning, which is basically automated statistical analysis. Machine learning allows a system to learn from examples, find patterns, make predictions, and improve its own performance.

So the way that I've made sense of it is machine learning is like a wheel. A wheel can go into a bicycle. It can go into machinery. It can go into a microwave. It's essentially an enabler. In the same way, machine learning can be applied to other technologies, such as for companies to optimize projections for advertising placement, to improving the performance of robots, to improving the accuracy of medical diagnosis.

KEN FIREMAN: Well, you can certainly see why people would be interested in AI, given those potentials. But as I understand it, one of the main concerns about AI is that it can introduce bias into decision-making. What kinds of bias are we talking about?

HANNAH KIM: So what's happening now is that while AI research began more than six decades ago, the field is still considered early in its development. So when AI systems were introduced to aid in decisions in hiring, for example, researchers were optimistic that AI systems could actually provide objectivity and mitigate human biases.

However, researchers started realizing that AI systems actually could reflect and reinforce historic human biases. So in the hiring example, if a group of people were historically put at advantage, the AI system would continue to put those categories of people at advantage.

KEN FIREMAN: I see. What can be done to mitigate or even eliminate these kinds of biases?

HANNAH KIM: So one of the main issues that AI researchers are addressing now is how to create methods of accountability. Many AI systems work as what they call black boxes, which means that the systems are so complex that the developers themselves can't really tell you why a machine is

making a decision.

So there are many conversations being made to bring transparency to the black box. This can mean to create a standard of documentation for when AI systems are tested and made, and also to allow for auditing. There are recommendations to have more of a multi-disciplinary approach in the deployment of AI systems to include ethicists and social scientists.

KEN FIREMAN: Mm-hmm. Is transparency going to be enough, do you think?

HANNAH KIM: Well, one of the interesting things that I was learning was that even if an AI system can explain its decision, it won't necessarily be in a way that would make sense to the humans. So there is also efforts to create tracking in how these AI systems are impacting social structures, and creating a system of documentation to be able to understand their effect in that way.

KEN FIREMAN: Mm-hmm. So as you write in your report, these concerns have been raised very forcefully by a number of people. What has been the tech industry's reaction?

HANNAH KIM: So there are many efforts being made in the tech industry to address issues of bias. Since around 2014, there have been numerous organizations in the industry and academia that are formed to address the impact of AI systems in our culture.

One of the main things I've discovered is that the people who are actually best to address these issues of bias are the engineers themselves. So there have been industry initiatives for ethics trainings for data scientists, engineers, to become aware of these issues, and also conversations surrounding increasing diversity within the AI industry itself to reflect a broader array of perspectives.

KEN FIREMAN: So I mean, do you think the industry is embracing this, and moving to address these concerns?

HANNAH KIM: I think that there is awareness of that these concerns are very serious, and that there is a lot being made to do as much as they can.

KEN FIREMAN: Mm-hmm. I see. OK. Let me ask you this-- what's the most surprising thing you learned in the course of researching this report?

HANNAH KIM: The key takeaway for me has been the sense of trust that people had in AI systems when they were first deployed. So you know, I heard many instances of people trusting an AI system or

just referring to a system as being right, without really being able to question or knowing that they could question a system's decision.

So right now, I think the trend is moving toward people understanding that AI systems are just a tool. And as they're being increasingly integrated into our lives, what we as humans can learn to work with AI systems and also how to question them.

KEN FIREMAN: Mm-hmm. Well, that's really interesting. And your whole report is really interesting. And I commend it to our readers. I've been speaking with Hannah Kim, who has written the Business Researcher report on AI and the ethical issues that exist around it. Thank you very much for joining us, Hannah.

HANNAH KIM: Thank you.