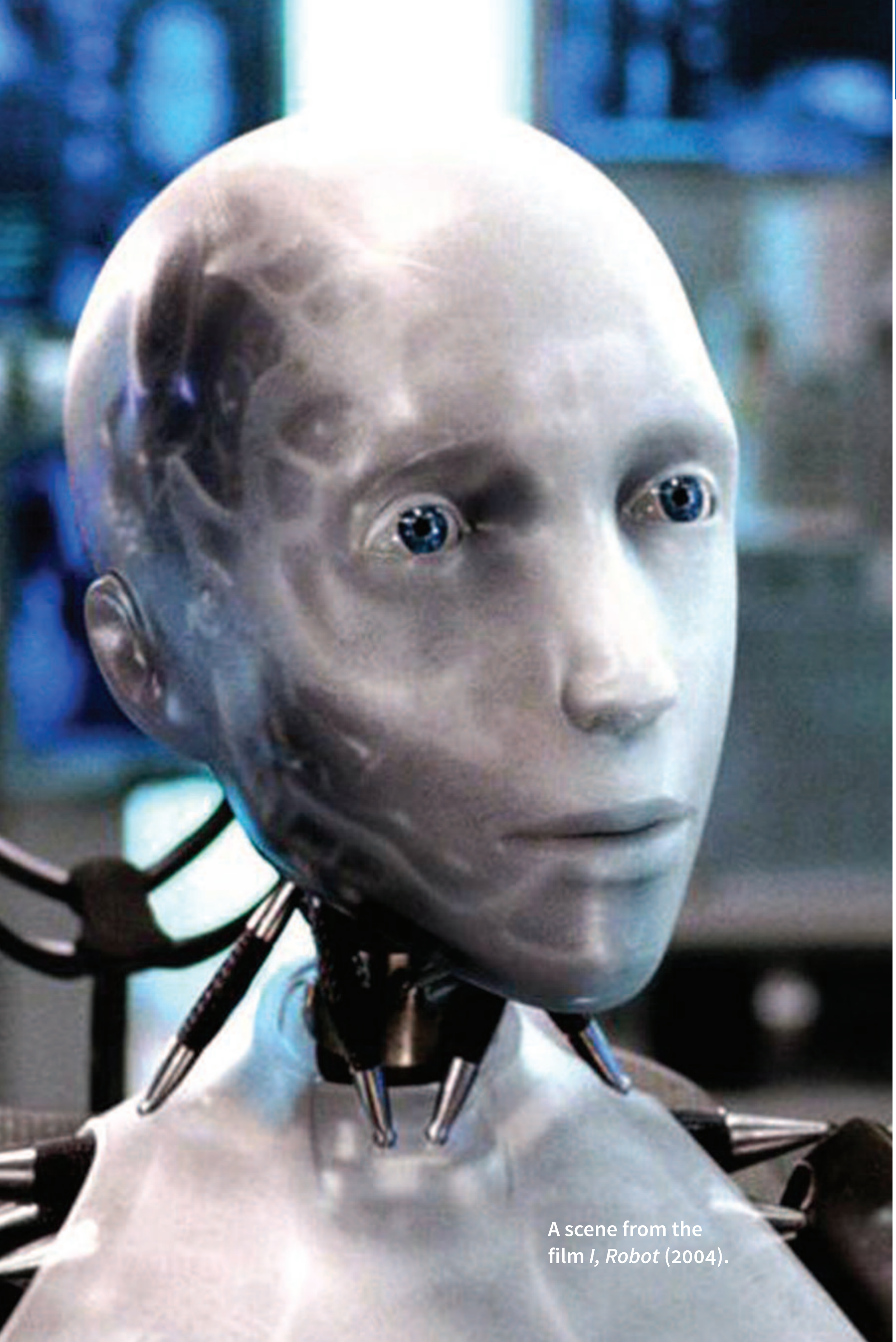


A scene from the film *I, Robot* (2004).

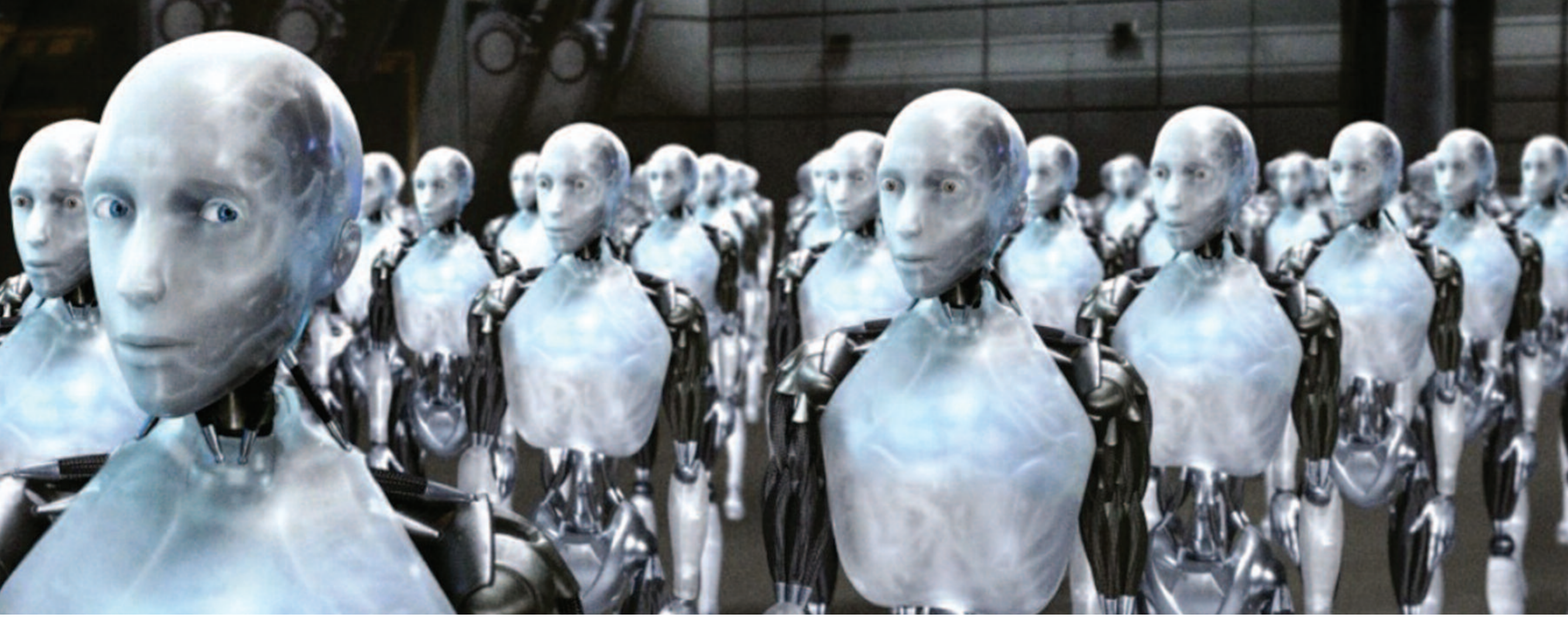
SELIM G. AKL NANCY SALAY

Artificial Intelligence *A Promising Future?*

Is a computer capable, like a human, of experiencing emotions (empathy, jealousy, fear)? Can a computer, through cunning, imitate the expression of such emotions for "personal" gain? Allowing for all this to be possible, it would follow necessarily that the computer must not only be self-conscious but also have awareness and understanding of the human mind, in order to know its interlocutors' expectations and anticipate their response. Perhaps the real question is beyond "Can a computer think?" One may ask: "Can a computer be as manipulative, as deceptive, as duplicitous – or as charming, as honest, and as kind as a human can be?"

SUNDAY, May 11, 1997, was a day like any other. Everything that was supposed to happen in politics, sports, and entertainment happened on that day, with one notable exception. In a New York City hotel an unexpected history-making event took place. A chess tournament pitting a human against a machine saw Garry Kasparov, the then reigning world chess champion, being defeated by a computer called Deep Blue. A new era had dawned.

In 2011, the prowess of the question-answering computer Watson on the television game show *Jeopardy!* captured the public's imagination. Watson won a match against two seasoned *Jeopardy!* players and received the $1-million prize. More recently, in 2016 a Go-playing program by the name of AlphaGo won a tournament against Lee Sedol, the recognized best player on the planet, by a score of 4 to 1. And on June 18, 2018, a program dubbed Project Debater engaged two humans in debates, on the topics of government subsidy of space exploration, and increased investment



In *I, Robot* humans have become so dependent on their robot helpers that they cannot imagine how society would function without them ...

Meanwhile, artificial intelligence begins to consider the destructive nature of human beings, and considers how the world might be saved.

in telemedicine, respectively, and did remarkably well. The world is beginning to pay attention.

These four achievements are harbingers of greater things to come. What is the common thread between Deep Blue, Watson, AlphaGo, Project Debater, and many other successes? Artificial Intelligence, a branch of computer science that aims to create intelligent systems. Over the past two or three years, Artificial Intelligence (AI), a scientific enterprise, has become a social phenomenon, with myriad economic, cultural, and philosophical implications. The advent of self-driving cars, speech-activated automated assistants, and data analytics more generally has transformed every sector of society. AI is reshaping and reinventing such fields as health care, business, transportation, education, and entertainment. The news media are replete with stories on the new cognitive technologies generated by AI and their effect on our daily lives and lifestyles. What is the reason for this explosion of excitement over AI?

As a result of some recent advances in machine learning technologies, the field is about a decade ahead of where we thought it would be at this time, with advances proceeding at an exponential rate. So says Elon Musk. In a BBC interview, famous physicist Stephen Hawking (1942–2018) warned that "the development of full artificial intelligence could spell the end of the human race." And fears that the singularity is nigh have resulted in websites, YouTube videos, and articles describing our impending doom. But is it the takeover of an artificial intelligence we should be worrying about? Or should we be more concerned about giving too much power to *unintelligent* AI? To make an accurate judgement, we need to understand what all the fuss is about.

Machine "learning" refers to a family of computational methods for analyzing data into statistically significant regularities that are useful for some purpose. These regularities are called "features" and the process of uncovering them "feature detection." Humans and other

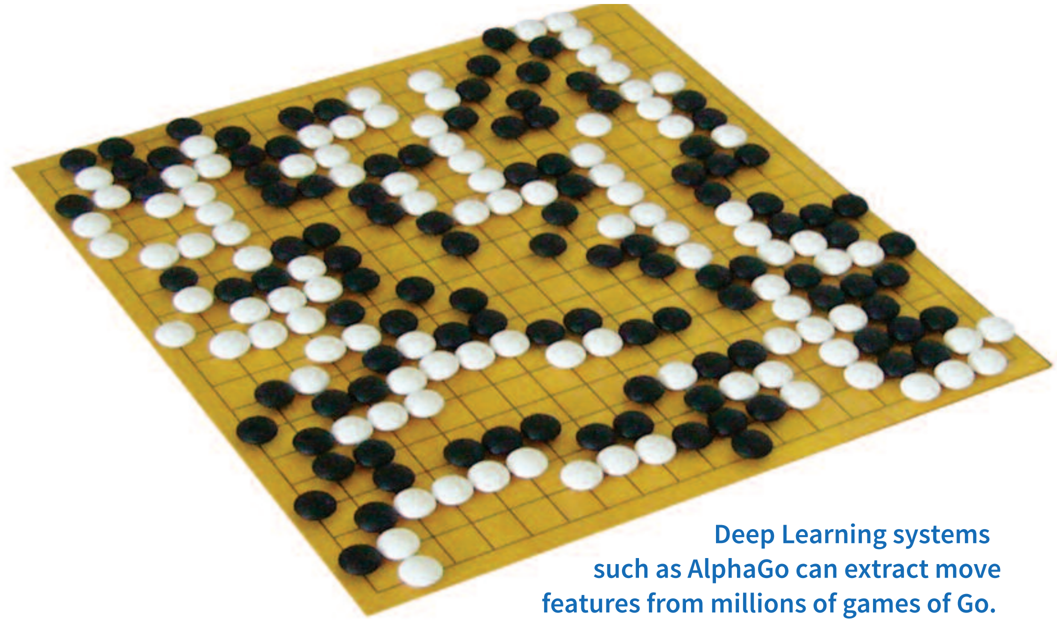
"The AI excitement over the last few years is the result of some very promising advances.… Deep Learning algorithms can 'extract' features from a set of data and thereby move beyond what humans know."

animals detect features whenever they recognize an object in the world: to perceive that a particular bone is the kind of thing that can be chomped on is to recognize a pattern of similarity between the bone being perceived and a host of bone experiences in the past.

Machine learning technologies have become increasingly adept at such classification tasks in well-understood areas. In the context of human faces, for example, systems that are sensitive to features such as noses, lips, eyes, and so on perform as well as humans on face recognition tasks. But some domains are so vast and multi-varied that even humans don't have a good handle on what set of features will be useful for a given task. For example, we know that online "clicking" behaviour is a source of potentially useful data, but we aren't sure how to organize it in order to highlight the useful patterns. But if a human programmer doesn't know what features an AI system should detect, how can the system be built?

The AI excitement over the last few years is the result of some very promising advances toward solving this problem. Deep Learning algorithms can "extract" features from a set of data and thereby move beyond what humans know. The techniques have been used successfully on labelled data sets, where humans have already tagged the photographs with captions – "DOG PLAYING BALL" – that are used as a way of "supervising" a system's learning by tracking how close or far it is on a given input from the correct answer. Recently there has been success with unlabelled data sets, what is called, "unsupervised" learning. The millennial Pandora's Box has been opened.

AlphaGo is a result of this new wave of machine learning. Deep Blue played chess by brute force, searching deeply through the hardcoded array of possible outcomes before choosing the optimal move. A human has no chance against this kind of opponent, not because it is so much smarter, but simply because it has a bigger working memory than a human does. With Go this canned approach was not possible: there are far more possible choices for each move, too many to hardcode and then search in real time. But Deep Learning systems such as AlphaGo can "learn" the relevant patterns of game play by extracting move features from millions of example games. The more games it plays, the more subtle its feature set becomes. On March 15, 2016, AlphaGo was awarded the highest Go grandmaster



Deep Learning systems such as AlphaGo can extract move features from millions of games of Go.

rank by South Korea's Go Association. Even the creators of AlphaGo at Google's DeepMind in London have no idea what move it will play at any given point in a game. Is the singularity at hand?

To answer that question, we need to consider carefully whether such systems are in fact *learning* and becoming *intelligent*. These questions take on urgency as increasingly we use them to make important decisions about human lives.

In 2017 the Canadian Institute for Advanced Research (CIFAR) was awarded a $125-million budget for the Pan-Canadian Artificial Intelligence Strategy, an initiative to revamp every facet of our bureaucracy with AI technology. The health care system is one of the first areas targeted for change. And a pilot project for early detection of possible suicides is already underway.

How will such technology be used? Sally might be at risk for suicide, but it doesn't follow from this that she ought to be put under surveillance, institutionalized, or otherwise have her autonomy undermined. More generally, machine learning is an excellent tool for data analysis, but it cannot tell us what to *do*.

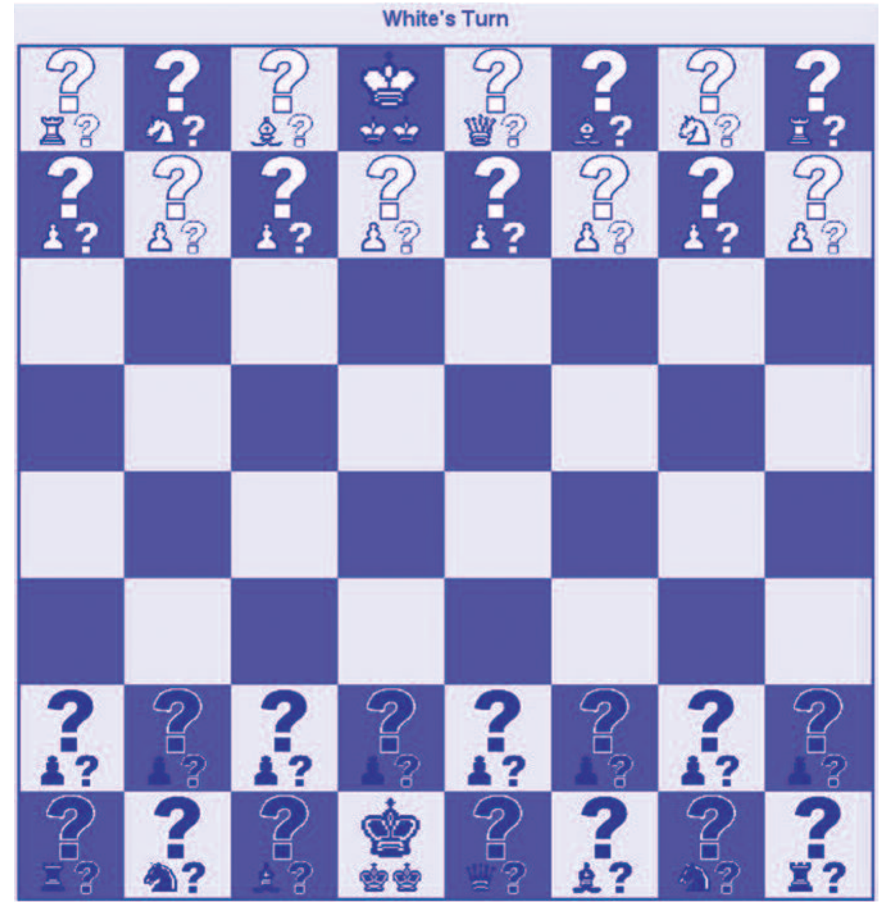
Practical judgement, the ability to bring relevant considerations to bear on a particular situation, guides us in our considered actions. Determining relevance is the critical skill here. How do we do it? This is the million-dollar

> "The 1,400-year-old game recently received a boost with the invention of Quantum Chess at Queen's University. This variant uses the weird properties of quantum physics in order to introduce an element of uncertainty into the game, thereby giving humans an equal chance when playing against computers."

question, of course, and we won't answer it here. Minimally, however, it requires a capacity to synthesize what is important in a given situation with what is important in human life more generally. In other words, it requires an understanding of what it means to be a laughing, working, eating, resting, playing being.

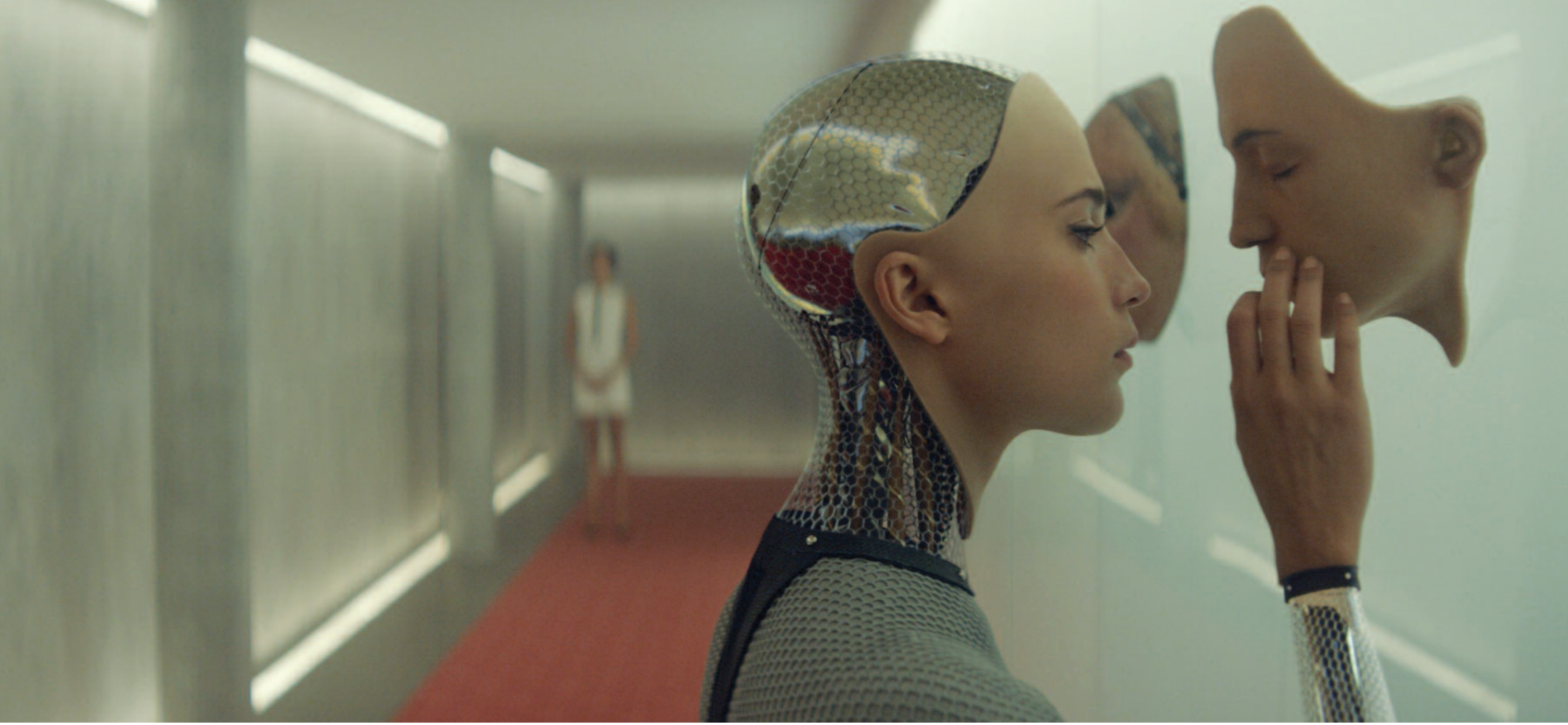
We still do not understand how meaning works. But we do know that being an expert pattern-detector in some domain or other is not all there is to it. The failures of our new AI heroes tell the story. During the *Jeopardy!*–IBM Challenge, Watson revealed what was behind its curtain – lots of meaningless data – in its answer to the "Final *Jeopardy!*" question. The category was "US Cities," and the answer was "Its largest airport is named for a World War II hero; its second largest, for a World War II battle." Watson's response? "What is Toronto?"

No surprise, then, that strategy games have been an AI industry focus: the tight constraints of game situations make the field of relevance narrow and, consequently, the chances of success great. Even so, the story of chess and AI is far from over. The 1,400-year-old game recently received a boost with the invention of Quantum Chess at Queen's University. This variant uses the weird properties of quantum physics in order to introduce an element of uncertainty into the game, thereby giving humans an equal chance when playing against computers. Unlike the chess pieces of the classical game, where a rook is a rook, and a knight is a knight, a Quantum Chess piece is a *superposition* of states, each representing a different classic chess piece. A player does not know the identity of a piece (that is, whether it is a pawn, a bishop, a queen, and so on) until the piece is selected for a move. Furthermore, thanks to the bizarre property of *entanglement*, the state of a chess piece is somehow "bound" to the state of another piece, regardless of how far they are separated; touch one, and you affect the other! The unpredictability inherent in Quantum Chess creates a level playing field for humans and computers. Unlike the case in classic chess (where a program can engage in a deterministic and thorough search for a good move),



A Quantum Chess board, ready for play. A large question mark indicates that the identity of a piece has not been revealed. Only the identity of the king is known throughout the game (image courtesy of the School of Computing at Queen's University).

the hidden identities of the pieces and the probabilistic nature of quantum physics greatly diminish the computer's ability to conduct a reliable search. Perhaps judgment will give humans an edge, even in this limited domain. When it comes to Quantum Chess, even a novice chess player may have a chance against a more experienced human, as demonstrated by the following anecdote. On January 26, 2016, a movie was premiered at the California Institute of Technology during an event entitled *One Entangled Evening: A Celebration of Richard Feynman's Quantum Legacy*. The movie captured an exciting, and at times funny, game of Quantum



Alicia Vikander portrays the robot Ava in the 2015 film *Ex Machina*.

Ava quickly learns what it takes to survive in humanity's world.

Chess between Hollywood actor Paul Rudd and Stephen Hawking. It is worth noting that in this version of Quantum Chess superposition has a different meaning from being a superposition of states. Rather, superposition is spatial, in the sense that the same chess piece can be, at the same time, in two separate locations on the chess board (one known and one unknown). Touching a piece in order to execute a move determines probabilistically from which of the two locations the piece is to move. It is as though the piece manifests itself suddenly, either choosing to stay in its visible location or possibly disappearing and materializing elsewhere on the board (thereby revealing the unknown location).

OTHER aspects of AI are increasingly being addressed in popular culture. The dark and suspenseful movie *Ex Machina* (2014), directed by Alex Garland, offers an interesting treatment of issues surrounding machine intelligence. An experimental female robot is being tested for possessing intelligence. She beguiles the young man testing her and persuades him to take actions leading to her liberation from captivity and simultaneously to his tragic end. The movie adds the following unexpected twist to the standard question of whether a machine can possess intelligence. If a robot displays an emotion toward a human that may be interpreted, for example, as love, is the emotion real, in the sense of being the repetition of a learned vocabulary, or is it purposefully faked?

in *Ex Machina*, Domhnall Gleeson (left) and Oscar Isaac portray two typically flawed human beings working to create a superior humanoid robot. (*Ex Machina* images: Mongrel Media, Universal Pictures.)

Is the robot being sincere, or might it be pretending? In other words, has the computer reached a level of intelligence that allows it to be able not only to automatically utter the words that express a human sentiment but in fact to intentionally simulate that feeling for a good or a bad purpose? Is a computer capable, like a human, of experiencing emotions (empathy, jealousy, fear)? Can a computer, through cunning, imitate the expression of such emotions for "personal" gain? Allowing for all this to be possible, it would follow necessarily that the computer must not only be self-conscious but also have awareness and understanding of the human mind, in order to know its interlocutors' expectations and anticipate their response. Perhaps the real question is beyond "Can a computer think?" One may ask: "Can a computer be as manipulative, as deceptive, as duplicitous – or as charming, as honest, and as kind as a human can be?"

What will an AI system capable of making practical judgements look like? Obviously, this is a foundational question for AI. Whatever the answer is, we know that we don't yet have it. We shouldn't be worried about the singularity – we are a long way off from that. But we should be concerned about the use to which AI technologies are being put. In our overconfidence in these technologies, we are giving them too much power.

FROM a more optimistic perspective, human knowledge will be deepened and broadened by the revolutionary paradigm of AI. Consider its influence on one area of endeavour of vital importance. AI will have a profound impact on the health care landscape. Machine learning and data analytics will lead to the discovery of improved tools for the detection, diagnosis, and treatment of disease. More effective pharmaceuticals with fewer or no side effects will be developed. In fact, computer scientists will design better search algorithms that will make it possible to produce drugs capable of being customized for each specific individual. Once personalized medicine is attained, the entire approach to health care will be completely revamped. Artificial Intelligence will allow the behaviour of a biological cell to be modelled as a computer algorithm. From this perspective, a cell with a disease is seen as a program with a flaw. Correcting the error in the program allows the cell to be healed. The positive disruptive force of AI on health care will have resulted in a great benefit for humankind.

SELIM G. AKL teaches in the School of Computing at Queen's University, where he leads the Parallel and Unconventional Computation Group.

NANCY SALAY teaches in the Department of Philosophy and the School of Computing at Queen's University.

The robot comes to life in Fritz Lang's 1927 film *Metropolis*.