READING PASSAGE 2

*You should spend about 20 minutes on **Questions 14–26**, which are based on Reading Passage 2 below.*

Living with artificial intelligence

Powerful artificial intelligence (AI) needs to be reliably aligned with human values, but does this mean AI will eventually have to police those values?

This has been the decade of AI, with one astonishing feat after another. A chess-playing AI that can defeat not only all human chess players, but also all previous human-programmed chess machines, after learning the game in just four hours? That's yesterday's news, what's next? True, these prodigious accomplishments are all in so-called narrow AI, where machines perform highly specialised tasks. But many experts believe this restriction is very temporary. By mid-century, we may have artificial general intelligence (AGI) – machines that can achieve human-level performance on the full range of tasks that we ourselves can tackle.

If so, there's little reason to think it will stop there. Machines will be free of many of the physical constraints on human intelligence. Our brains run at slow biochemical processing speeds on the power of a light bulb, and their size is restricted by the dimensions of the human birth canal. It is remarkable what they accomplish, given these handicaps. But they may be as far from the physical limits of thought as our eyes are from the incredibly powerful Webb Space Telescope.

Once machines are better than us at designing even smarter machines, progress towards these limits could accelerate. What would this mean for us? Could we ensure a safe and worthwhile coexistence with such machines? On the plus side, AI is already useful and profitable for many things, and super AI might be expected to be super useful, and super profitable. But the more powerful AI becomes, the more important it will be to specify its goals with great care. Folklore is full of tales of people who ask for the wrong thing, with disastrous consequences – King Midas, for example, might have wished that everything he touched turned to gold, but didn't really intend this to apply to his breakfast.

So we need to create powerful AI machines that are 'human-friendly' – that have goals reliably aligned with our own values. One thing that makes this task difficult is that we are far from reliably human-friendly ourselves. We do many terrible things to each other and to many other creatures with whom we share the planet. If superintelligent machines don't do a lot better than us, we'll be in deep trouble. We'll have powerful new intelligence amplifying the dark sides of our own fallible natures.

For safety's sake, then, we want the machines to be ethically as well as cognitively superhuman. We want them to aim for the moral high ground, not for the troughs in which many of us spend some of our time. Luckily they'll be smart enough for the job. If there are routes to the moral high ground, they'll be better than us at finding them, and steering us in the right direction.

However, there are two big problems with this utopian vision. One is how we get the machines started on the journey, the other is what it would mean to reach this destination. The 'getting started' problem is that we need to tell the machines what they're looking for with sufficient clarity that we can be confident they will find it – whatever 'it' actually turns out to be. This won't be easy, given that we are tribal creatures and conflicted about the ideals ourselves. We often ignore the suffering of strangers, and even contribute to it, at least indirectly. How then, do we point machines in the direction of something better?

As for the 'destination' problem, we might, by putting ourselves in the hands of these moral guides and gatekeepers, be sacrificing our own autonomy – an important part of what makes us human. Machines who are better than us at sticking to the moral high ground may be expected to discourage some of the lapses we presently take for granted. We might lose our freedom to discriminate in favour of our own communities, for example.

Loss of freedom to behave badly isn't always a bad thing, of course: denying ourselves the freedom to put children to work in factories, or to smoke in restaurants are signs of progress. But are we ready for ethical silicon police limiting our options? They might be so good at doing it that we won't notice them; but few of us are likely to welcome such a future.

These issues might seem far-fetched, but they are to some extent already here. AI already has some input into how resources are used in our National Health Service (NHS) here in the UK, for example. If it was given a greater role, it might do so much more efficiently than humans can manage, and act in the interests of taxpayers and those who use the health system. However, we'd be depriving some humans (e.g. senior doctors) of the control they presently enjoy. Since we'd want to ensure that people are treated equally and that policies are fair, the goals of AI would need to be specified correctly.

We have a new powerful technology to deal with – itself, literally, a new way of thinking. For our own safety, we need to point these new thinkers in the right direction, and get them to act well for us. It is not yet clear whether this is possible, but if it is, it will require a cooperative spirit, and a willingness to set aside self-interest.

Both general intelligence and moral reasoning are often thought to be uniquely human capacities. But safety seems to require that we think of them as a package: if we are to give general intelligence to machines, we'll need to give them moral authority, too. And where exactly would that leave human beings? All the more reason to think about the destination now, and to be careful about what we wish for.

Questions 14–19

Choose the correct letter, ***A****,* ***B****,* ***C*** *or* ***D****.*

Write the correct letter in boxes 14–19 on your answer sheet.

14 What point does the writer make about AI in the first paragraph?

A It is difficult to predict how quickly AI will progress.
B Much can be learned about the use of AI in chess machines.
C The future is unlikely to see limitations on the capabilities of AI.
D Experts disagree on which specialised tasks AI will be able to perform.

15 What is the writer doing in the second paragraph?

A explaining why machines will be able to outperform humans
B describing the characteristics that humans and machines share
C giving information about the development of machine intelligence
D indicating which aspects of humans are the most advanced

16 Why does the writer mention the story of King Midas?

A to compare different visions of progress
B to illustrate that poorly defined objectives can go wrong
C to emphasise the need for cooperation
D to point out the financial advantages of a course of action

17 What challenge does the writer refer to in the fourth paragraph?

A encouraging humans to behave in a more principled way
B deciding which values we want AI to share with us
C creating a better world for all creatures on the planet
D ensuring AI is more human-friendly than we are ourselves

18 What does the writer suggest about the future of AI in the fifth paragraph?

A The safety of machines will become a key issue.
B It is hard to know what impact machines will have on the world.
C Machines will be superior to humans in certain respects.
D Many humans will oppose machines having a wider role.

19 Which of the following best summarises the writer's argument in the sixth paragraph?

A More intelligent machines will result in greater abuses of power.
B Machine learning will share very few features with human learning.
C There are a limited number of people with the knowledge to program machines.
D Human shortcomings will make creating the machines we need more difficult.

Questions 20–23

Do the following statements agree with the claims of the writer in Reading Passage 2?

In boxes 20–23 on your answer sheet, write

YES	*if the statement agrees with the claims of the writer*
NO	*if the statement contradicts the claims of the writer*
NOT GIVEN	*if it is impossible to say what the writer thinks about this*

20 Machines with the ability to make moral decisions may prevent us from promoting the interests of our communities.

21 Silicon police would need to exist in large numbers in order to be effective.

22 Many people are comfortable with the prospect of their independence being restricted by machines.

23 If we want to ensure that machines act in our best interests, we all need to work together.

Questions 24–26

Complete the summary using the list of phrases, ***A–F****, below.*

Write the correct letter, ***A–F****, in boxes 24–26 on your answer sheet.*

Using AI in the UK health system

AI currently has a limited role in the way **24** .... are allocated in the health service. The positive aspect of AI having a bigger role is that it would be more efficient and lead to patient benefits. However, such a change would result, for example, in certain **25** .... not having their current level of **26** .... It is therefore important that AI goals are appropriate so that discriminatory practices could be avoided.

A medical practitioners

B specialised tasks

C available resources

D reduced illness

E professional authority

F technology experts