IELTS Reading

Cambridge 18 | Test 2 - Passage 2

Living with Artificial Intelligence

Bài đọc

A.

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.

B.

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.

C.

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.

D.

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 superintendent 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.

E.

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. 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?

F.

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.

G.

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.

H.

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.

I.

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.

J.

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.

Câu hỏi

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.

A. medical practitioners

B. specialised tasks

C. available resources

D. reduced illness

E. professional authority

F. technology experts

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.

Đáp án

Bảng tổng hợp đáp án bài đọc Living with Artificial Intelligence

14. C	15. A	16. B
17. D	18. C	19. D
20. YES	21. NOT GIVEN	22. NO
23. YES	24. C	25. A
26. E		

Questions 14-19

Question 14	**Đáp án**
What point does the writer make about AI in the first paragraph?	C

Giải thích: Bài đọc thừa nhận hiện tại AI có hạn chế (restriction), nhưng nhấn mạnh điều này chỉ là tạm thời (temporary). Đến giữa thế kỷ, AI (AGI) sẽ đạt hiệu suất ngang con người. Điều này ám chỉ tương lai khả năng của AI sẽ phá vỡ các giới hạn.

Dẫn chứng: ...But many experts believe this restriction is very temporary. By mid-century, we may have artificial general intelligence...

Question 15	**Đáp án**
What is the writer doing in the second paragraph?	A

Giải thích: Tác giả giải thích lý do máy móc sẽ vượt trội hơn con người bằng cách chỉ ra các hạn chế sinh học của não bộ (tốc độ xử lý chậm, kích thước nhỏ), trong khi máy móc không bị ràng buộc (free of constraints).

Dẫn chứng: Machines will be free of many of the physical constraints on human intelligence... Our brains run at slow... speeds...

Question 16	**Đáp án**
Why does the writer mention the story of King Midas?	B

Giải thích: Câu chuyện Vua Midas (ước mọi thứ thành vàng nhưng quên mất bữa sáng cũng hóa vàng) minh họa cho việc: Nếu mục tiêu không được xác định rõ ràng, hậu quả sẽ tồi tệ.

Dẫn chứng: …people who ask for the wrong thing, with disastrous consequences - King Midas… didn't really intend this to apply to his breakfast.

Question 17	**Đáp án**
What challenge does the writer refer to in the fourth paragraph?	D

Giải thích: Thách thức lớn là tạo ra AI thân thiện với con người, trong khi chính bản thân con người lại chưa thân thiện với nhau (far from reliably human-friendly ourselves).

Dẫn chứng: 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...

Question 18	**Đáp án**
What does the writer suggest about the future of AI in the fifth paragraph?	C

Giải thích: Tác giả cho rằng máy móc sẽ ưu việt hơn con người ở khía cạnh đạo đức: chúng sẽ thông minh hơn (smart enough) và giỏi hơn chúng ta (better than us) trong việc tìm ra đường lối đúng đắn.

Dẫn chứng: …they'll be better than us at finding them, and steering us in the right direction.

Question 19	**Đáp án**
Which of the following best summaries the writer's argument in the sixth paragraph?	D

Giải thích: Vấn đề khởi đầu rất nan giải vì con người có nhiều khiếm khuyết (shortcomings) như: mâu thuẫn về lý tưởng, thờ ơ với người khác. Những điều này làm việc tạo ra AI chuẩn mực trở nên khó khăn.

Dẫn chứng: …we are tribal creatures and conflicted about the ideals ourselves… ignore the suffering of strangers…

Questions 20 - 23

Question 20	**Đáp án**
Machines with the ability to make moral decisions may prevent us from promoting the interests of our communities.	YES

Giải thích: Bài đọc nói rằng máy móc giỏi tuân thủ đạo đức có thể ngăn cản những lỗi lầm (lapses) mà ta coi là bình thường, ví dụ như việc phân biệt đối xử để thu lợi cho cộng đồng của mình (discriminate in favour of our own communities). Điều này đồng nghĩa với việc chúng ngăn ta thúc đẩy lợi ích cho cộng đồng.

Dẫn chứng: Paragraph G

...Machines who are better than us at sticking to the moral high ground may be expected to discourage... We might lose our freedom to **discriminate in favour of our own communities**, for example.

Question 21	**Đáp án**
Silicon police would need to exist in large numbers in order to be effective.	NOT GIVEN

Giải thích: Bài đọc có nhắc đến ethical silicon police (cảnh sát người máy) và khả năng làm việc hiệu quả của chúng (so good at doing it). Tuy nhiên, tác giả **không đề cập** đến việc chúng cần phải tồn tại với số lượng lớn (exist in large numbers) mới hiệu quả.

Dẫn chứng: Paragraph H

...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...

Question 22	**Đáp án**
Many people are comfortable with the prospect of their independence being restricted by machines.	NO

Giải thích: Câu hỏi nói nhiều người thoải mái (Many people are comfortable). Ngược lại, bài đọc khẳng định rất ít người trong chúng ta sẵn sàng chào đón tương lai này (few of us are likely to welcome such a future).

Dẫn chứng: Paragraph H

...but **few of us are likely to welcome** such a future.

Question 23	**Đáp án**
If we want to ensure that machines act in our best interests, we all need to work together.	YES

Giải thích: Tác giả cho rằng để máy móc hành động tốt cho chúng ta (act well for us), cần có một tinh thần hợp tác (cooperative spirit) và sẵn sàng gạt bỏ tư lợi (set aside self-interest). Work together đồng nghĩa với cooperative spirit.

Dẫn chứng: Paragraph K

...get them to act well for us... it will require a **cooperative spirit**, and a willingness to set aside self-interest.

Questions 24 - 26

Question 24	**Đáp án**
AI currently has a limited role in the way _____ are allocated in the health service.	C (available resources)

Phân tích: Cần điền một **danh từ**, chỉ thứ được phân bổ (allocated) bởi AI. Trong bài đọc, tác giả nói AI có đóng góp vào cách các **nguồn lực** (resources) được sử dụng.

Dẫn chứng: AI already has some input into how **resources** are used in our National Health Service...

Question 25	**Đáp án**
However, such a change would result, for example, in certain _____ not having their current level of _____.	A (medical practitioners)

Phân tích: Cần điền một **danh từ chỉ người**, chỉ đối tượng bị mất quyền lực. Bài đọc đề cập đến việc tước bỏ quyền kiểm soát của một số người, ví dụ như **các bác sĩ cấp cao** (senior doctors).

Dẫn chứng: However, we'd be depriving some humans (e.g. **senior doctors**) of the control...

Question 26	**Đáp án**
…not having their current level of _____.	E (professional authority)

Phân tích: Cần điền một **danh từ**, chỉ thứ mà các bác sĩ bị mất đi. Bài đọc nói họ bị mất **quyền kiểm soát** (control) mà họ đang có.

Dẫn chứng: ...of the **control** they presently enjoy.

Dịch nghĩa bài đọc

Trí tuệ nhân tạo (AI) mạnh mẽ cần phải được liên kết một cách đáng tin cậy với các giá trị của con người, nhưng liệu điều này có đồng nghĩa với việc AI cuối cùng sẽ phải kiểm soát (trở thành cảnh sát) những giá trị đó? Đây là thập kỷ của AI, với những kỳ tích đáng kinh ngạc này đến kỳ tích khác. Một AI chơi cờ vua có thể đánh bại không chỉ tất cả những người chơi cờ, mà còn cả tất cả các máy chơi cờ do con người lập trình trước đó, chỉ sau khi học chơi trong vòng bốn giờ ư? Đó là tin tức của ngày hôm qua rồi (chuyện cũ rồi). Điều gì sẽ xảy ra tiếp theo? Đúng là những thành tựu phi thường này đều thuộc về cái gọi là AI hẹp, nơi máy móc thực hiện các nhiệm vụ chuyên biệt hóa cao độ. Nhưng nhiều chuyên gia tin rằng sự hạn chế này chỉ là tạm thời. Vào giữa thế kỷ này, chúng ta có thể có trí tuệ nhân tạo tổng hợp (AGI) - những cỗ máy có thể đạt được hiệu suất ngang bằng con người trong mọi phạm vi tác vụ mà chính chúng ta có thể giải quyết.

Nếu vậy, chẳng có mấy lý do để nghĩ rằng nó sẽ dừng lại ở đó. Máy móc sẽ không bị ràng buộc bởi nhiều hạn chế vật lý như trí tuệ con người. Bộ não của chúng ta chạy với tốc độ xử lý sinh hóa chậm chạp bằng năng lượng của một bóng đèn, và kích thước của chúng bị hạn chế bởi kích thước của ống sinh (đường sinh sản của người mẹ). Thật đáng kinh ngạc về những gì bộ não đạt được, bất chấp những bất lợi này. Nhưng chúng (bộ não người) có thể còn cách xa các giới hạn vật lý của tư duy giống như khoảng cách từ mắt thường của chúng ta đến Kính viễn vọng Không gian Webb cực kỳ mạnh mẽ vậy.

Một khi máy móc giỏi hơn chúng ta trong việc thiết kế những cỗ máy thậm chí còn thông minh hơn, tiến trình đạt tới những giới hạn này có thể tăng tốc. Điều này sẽ có ý nghĩa gì đối với chúng ta? Liệu chúng ta có thể đảm bảo một sự chung sống an toàn và đáng giá với những cỗ máy như vậy không? Về mặt tích cực, AI đã hữu ích và sinh lời cho nhiều việc, và siêu AI có thể được kỳ vọng là sẽ siêu hữu ích và siêu sinh lời. Nhưng AI càng trở nên mạnh mẽ, thì việc xác định các mục tiêu của nó một cách cẩn trọng càng trở nên quan trọng. Truyện dân gian chứa đầy những câu chuyện về những người ước sai điều mong muốn, với những hậu quả thảm khốc - ví dụ như Vua Midas, có lẽ đã ước mọi thứ ông chạm vào đều biến thành vàng, nhưng thực sự không định áp dụng điều này cho bữa sáng của mình.

Vì vậy, chúng ta cần tạo ra những cỗ máy AI mạnh mẽ thân thiện với con người - nghĩa là có những mục tiêu liên kết một cách đáng tin cậy với các giá trị của chính chúng ta. Một điều khiến nhiệm vụ này trở nên khó khăn là chính bản thân chúng ta cũng còn xa mới đạt được sự thân thiện đáng tin cậy với con người. Chúng ta làm nhiều điều tồi tệ với nhau và với nhiều sinh vật khác cùng chung sống trên hành tinh này. Nếu các cỗ máy giám sát không làm tốt hơn chúng ta nhiều, chúng ta sẽ gặp rắc rối lớn. Chúng ta sẽ có một trí tuệ mới đầy quyền năng khuếch đại những mặt tối trong bản chất dễ mắc sai lầm của chính mình.

Vì sự an toàn, chúng ta muốn máy móc phải siêu phàm cả về mặt đạo đức lẫn nhận thức. Chúng ta muốn chúng hướng tới đỉnh cao đạo đức, chứ không phải những vũng lầy mà nhiều người trong chúng ta thỉnh thoảng sa vào. May mắn thay, chúng sẽ đủ thông minh cho công việc này. Nếu có những con

đường dẫn đến đỉnh cao đạo đức, chúng sẽ giỏi hơn chúng ta trong việc tìm ra, và chèo lái chúng ta đi đúng hướng. Chúng ta thường phớt lờ nỗi đau khổ của người lạ, và thậm chí góp phần vào đó, ít nhất là gián tiếp. Vậy thì làm thế nào chúng ta chỉ cho máy móc hướng đến một cái gì đó tốt đẹp hơn?

Đối với vấn đề đích đến, bằng việc đặt bản thân vào tay những người dẫn đường và người gác cổng đạo đức này, chúng ta có thể đang hy sinh sự tự chủ của chính mình - một phần quan trọng tạo nên con người chúng ta. Những cỗ máy giỏi hơn chúng ta trong việc bám sát các chuẩn mực đạo đức cao có thể sẽ ngăn cản một số sai sót mà hiện tại chúng ta coi là điều hiển nhiên. Ví dụ, chúng ta có thể mất quyền tự do thiên vị (phân biệt đối xử) để ưu ái cho cộng đồng của riêng mình.

Tất nhiên, mất đi sự tự do làm điều xấu không phải lúc nào cũng là điều tệ: việc từ chối cho phép bản thân tự do đưa trẻ em vào làm việc trong nhà máy, hay hút thuốc trong nhà hàng là những dấu hiệu của sự tiến bộ. Nhưng liệu chúng ta đã sẵn sàng cho những cảnh sát silicon (cảnh sát người máy) đạo đức giới hạn các lựa chọn của mình chưa? Chúng có thể làm điều đó giỏi đến mức chúng ta sẽ không nhận ra; nhưng ít ai trong chúng ta có vẻ sẽ chào đón một tương lai như vậy.

Những vấn đề này nghe có vẻ xa vời, nhưng ở một mức độ nào đó chúng đã hiện hữu ở đây rồi. Ví dụ, AI đã có một số đóng góp vào cách sử dụng nguồn lực trong Dịch vụ Y tế Quốc gia (NHS) của chúng tôi tại Vương quốc Anh. Nếu được trao vai trò lớn hơn, nó có thể làm việc hiệu quả hơn nhiều so với khả năng quản lý của con người, và hành động vì lợi ích của người đóng thuế và những người sử dụng hệ thống y tế. Tuy nhiên, chúng ta sẽ tước bỏ quyền kiểm soát mà một số người (ví dụ: các bác sĩ cấp cao) hiện đang nắm giữ. Vì chúng ta muốn đảm bảo mọi người được đối xử bình đẳng và các chính sách công bằng, các mục tiêu của AI sẽ cần phải được xác định một cách chính xác.

Chúng ta có một công nghệ mạnh mẽ mới để đối phó - bản thân nó, theo đúng nghĩa đen, là một cách tư duy mới. Vì sự an toàn của chính mình, chúng ta cần chỉ cho những nhà tư tưởng mới này đi đúng hướng, và khiến chúng hành động tốt cho chúng ta. Vẫn chưa rõ liệu điều này có khả thi hay không, nhưng nếu có, nó sẽ đòi hỏi một tinh thần hợp tác, và sự sẵn sàng gạt bỏ tư lợi cá nhân.

Cả trí thông minh tổng quát và lý luận đạo đức thường được coi là những năng lực độc nhất của con người.

Nhưng sự an toàn dường như đòi hỏi chúng ta phải coi chúng là một gói đi kèm: nếu chúng ta trao trí tuệ tổng quát cho máy móc, chúng ta cũng sẽ cần trao cho chúng thẩm quyền đạo đức nữa. Và điều đó chính xác sẽ đặt con người ở vị trí nào? Đó càng là lý do để suy nghĩ về đích đến ngay bây giờ, và cẩn trọng với những gì chúng ta ước muốn.