

Don't Use A.I. to Cheat in School. It's Better for Studying.

Generative A.I. tools can annotate long documents, make flashcards, and produce practice quizzes.



By Brian X. Chen

June 30, 2023

Hello! We're back with another bonus edition of *On Tech: A.I.*, a pop-up newsletter that teaches you about artificial intelligence, how it works and how to use it.

Last week, I went over how to turn your chatbot into a life coach. Let's now shift into an area where many have been experimenting with A.I. since last year: education.

Generative A.I.'s specialty is language — guessing which word comes next — and students quickly realized that they could use ChatGPT and other chatbots to write essays. That created an awkward situation in many classrooms. It turns out, it's easy to get caught cheating with generative A.I. because it is prone to making stuff up, a phenomena known as “hallucinating.”

But generative A.I. can also be used as a study assistant. Some tools make highlights in long research papers and even answer questions about the material. Others can assemble study aids, like quizzes and flashcards.

One warning to keep in mind: When studying, it's paramount that the information is correct, and to get the most accurate results, you should direct A.I. tools to focus on information from trusted sources rather than pull data from across the web. I'll go over how to do that below.

Research

First, let's explore one of the most daunting studying tasks: reading and annotating long papers. Some A.I. tools, such as Humata.AI, Wordtune Read and various plug-ins inside ChatGPT, act as research assistants that will summarize documents for you.

I prefer Humata.AI because it answers your questions and shows highlights directly inside the source material, which allows you to double check for accuracy.

On the Humata.AI website, I uploaded a PDF of a scientific research paper on the accuracy of smartwatches in tracking cardio fitness. Then I clicked the “Ask” button and asked it how Garmin watches performed in the study. It scrolled down to the relevant part of the document mentioning Garmin, made highlights and answered my question.

The screenshot shows the Humata.AI interface. At the top left is the Humata logo. On the right is a 'SHARE' button. A blue box on the left contains the question: "am i correct to understand that on average, wearables were accurate at tracking cardio fitness but there were major errors on a case by case basis for some individuals?". Below the question is a text box with the answer: "Yes, you are correct. On average, wearables were found to be accurate at tracking cardio fitness at the group level. However, there were major errors on a case-by-case basis for some individuals. This means that while wearables provide reasonably accurate estimates of cardio fitness for the general population, there can be significant individual variation and errors in estimating VO2max for specific individuals. This is supported by the findings of the systematic review and meta-analysis. The study reported that wearables using exercise-based algorithms showed a high level of accuracy in estimating VO2max at the group level [5]. However, at the individual level, there was still a large estimation error, indicating the need for further improvement for sport/clinical purposes [5]. The review also highlighted the importance of considering individual factors such as sex, age, and other determinants of cardiorespiratory fitness in validation studies [1]. Therefore, while wearables can provide useful insights into cardio fitness trends for most individuals, caution should be exercised when relying on them for precise individual measurements or tracking individual changes in cardio fitness [6]. Pages: 4, 14, 17". To the right of the question and answer are two columns of highlighted text from the research paper. The first column highlights: "Wearables using exercise-based algorithms provide higher accuracy in the estimation of maximal oxygen consumption (VO2max) than those based on resting conditions. Wearables using exercise-based estimation seem to be optimal for measuring VO2max at the population level, yet the estimation error at the individual level still needs further improvement. In this article, the Towards Intelligent Health and Well-Being Network of Physical Activity Assessment (INTERLIVE) network provides best-practice recommendations to be used in future protocols to move towards a more accurate, transparent and comparable validation of VO2max derived from wearables." The second column highlights: "and optimizing health and sports performance [2, 3]. Furthermore, the omnipresence of wearables enhances digital phenotyping at a population level, which offers valuable information about physical activity and fitness levels from around the world that can be used to guide global health promotion actions [2, 4]. The most accepted measure of cardiorespiratory fitness is maximal oxygen consumption (VO2max), which has been shown to be a powerful marker of health and has recently been proposed as a clinical vital sign by the American Heart Association [5]. Furthermore, VO2max is widely known as a key indicator of endurance performance and, therefore, its measurement is of vital importance for sports performance in general [6]. The current guidelines for accurate testing of VO2max require measurement of gas exchange by indirect calorimetry usually in a laboratory during an exercise test to exhaustion [7]. These tests require expensive equipment (e.g., gas analyzer) and trained technicians to collect and interpret the data, which makes VO2max assessments less feasible for risk prediction in clinical practice and unfavorable for most recreational athletes and for the general population. Indirect estimation of VO2max by submaximal field tests overcomes some of these disadvantages and offers acceptable estimations of VO2max [8, 9]. However, the above-mentioned digital era of consumer wearable devices opens new horizons for fitness monitoring without the need for laboratory or field testing. In view of the enormous potential of these devices, wearable companies are making significant investments in ever, to the best of our knowledge, no systematic review or meta-analysis focusing on the validity of the estimated VO2max is available. Furthermore, the current science behind the validation protocols of wearable devices suffers major limitations, mainly due to a lack of consensus and guidelines ensuring good practices [17, 18]. This is precisely one of the main goals of the Towards Intelligent Health and Well-Being Network of Physical Activity Assessment (INTERLIVE) consortium, which is to develop best-practice protocols for the validation of consumer wearable fitness and activity measures. The INTERLIVE consortium has already published guidelines adapted to the nature of specific fitness/physical activity measures such as step count [19] and HR [20]. However, to date there are no specific standards guiding both manufacturers and the scientific community in the validation of estimating VO2max by consumer wearables. Therefore, in this article, INTERLIVE had two main objectives: (1) to systematically summarize previous studies investigating the validity of VO2max as estimated by consumer wearable devices based on a meta-analysis, and (2) to provide best-practice validation recommendations based on the systematic review of the literature together with an evidence-informed INTERLIVE consortium discussion." Below the highlights are two section headers: "2 Methods: Expert Statement Process and Meta-Analysis" and "2.1 The INTERLIVE Network".

Humata.AI responds to questions and makes relevant highlights inside documents. Brian X. Chen

Most interesting to me was when I asked the bot whether my understanding of the paper was correct — that on average, wearable devices like Garmins and Fitbits tracked cardio fitness fairly accurately, but there were some individuals whose results were very wrong. “Yes, you are correct,” the bot responded. It followed up with a summary of the study and listed the page numbers where this conclusion was

mentioned.

Studying

Generative A.I. can also help with rote memorization. While any chatbot will generate flashcards or quizzes if you paste in the information that you're studying, I decided to use ChatGPT because it includes plug-ins that generate study aids that pull from specific web articles or documents.

(Only subscribers who pay \$20 a month for ChatGPT Plus can use plug-ins. We explained how to use them in a previous newsletter.)

I wanted ChatGPT to create flashcards for me to learn Chinese vocabulary words. To do this, I installed two plug-ins: Link Reader, which let me tell the bot to use data from a specific website, and MetaMentor, a plug-in that automatically generates flashcards.

In the ChatGPT dashboard, I selected both plug-ins. Then, I wrote this prompt:

Act as a tutor. I am a native English speaker learning Chinese. Take the vocabulary words and phrases from this link and create a set of flashcards for each: <https://preply.com/en/blog/basic-chinese-words/>

About five minutes later, the bot responded with a link where I could download the flashcards. They were exactly what I asked for.

Next, I wanted my tutor to quiz me. I told ChatGPT that I was studying for the written exam to get my motorcycle license in California. Again, using the Link Reader plug-in, I pasted a link to the California D.M.V.'s latest motorcycle handbook (an important step because traffic laws vary between states and rules are occasionally updated) and asked for a multiple-choice quiz.

The bot processed the information inside the handbook and produced a quiz, asking me five questions at a time.

Finally, to test my grasp of the subject, I directed ChatGPT to ask me questions without presenting multiple-choice answers. The bot adapted accordingly, and I aced the quiz.

I would have loved having these tools when I was in school. And probably would have earned better grades with them as study companions.

What's next?

Next week, in the final installment of this how-to newsletter, we'll take everything we've learned and apply it to enriching the time we spend with our families.

Brian X. Chen is the lead consumer technology writer for The Times. He reviews products and writes Tech Fix, a column about the social implications of the tech we use. Before joining The Times in 2011, he reported on Apple and the wireless industry for Wired. More about Brian X. Chen

A version of this article appears in print on , Section B, Page 5 of the New York edition with the headline: A.I. Can Aid Cheating. It's Better for Studying.