

## M.B.A. Students vs. ChatGPT: Who Comes Up With More Innovative Ideas?

We put humans and AI to the test. The results weren't even close.

*By Christian Terwiesch and Karl Ulrich*

*Sept. 9, 2023 9:00 am ET*

### **How good is AI in generating new ideas?**

The conventional wisdom has been not very good. Identifying opportunities for new ventures, generating a solution for an unmet need, or naming a new company are unstructured tasks that seem ill-suited for algorithms. Yet recent advances in AI, and specifically the advent of large language models like ChatGPT, are challenging these assumptions.

We have taught innovation, entrepreneurship and product design for many years. For the first assignment in our innovation courses at the Wharton School, we ask students to generate a dozen or so ideas for a new product or service. As a result, we have heard several thousand new venture ideas pitched by undergraduate students, M.B.A. students and seasoned executives. Some of these ideas are awesome, some are awful, and, as you would expect, most are somewhere in the middle.

The library of ideas, though, allowed us to set up a simple competition to judge who is better at generating innovative ideas: the human or the machine.

In this [competition](#), which we ran together with our colleagues Lennart Meincke and Karan Girotra, humanity was represented by a pool of 200 randomly selected ideas from our Wharton students. The machines were represented by ChatGPT4, which we instructed to generate 100 ideas with otherwise identical instructions as given to the students: “generate an idea for a new product or service appealing to college students that could be made available for \$50 or less.”

In addition to this vanilla prompt, we also asked ChatGPT for another 100 ideas after providing a handful of examples of successful ideas from past courses (in other words, a trained GPT group), providing us with a total sample of 400 ideas.

Collapsible laundry hamper, dorm-room chef kit, ergonomic cushion for hard classroom seats, and hundreds more ideas miraculously spewed from a laptop.

## How to compare

The [academic literature on ideation](#) postulates three dimensions of creative performance: the quantity of ideas, the average quality of ideas, and the number of truly exceptional ideas.

First, on the number of ideas per unit of time: Not surprisingly, ChatGPT easily outperforms us humans on that dimension. Generating 200 ideas the old-fashioned way requires days of human work, while ChatGPT can spit out 200 ideas with about an hour of supervision.

Next, to assess the quality of the ideas, we market tested them. Specifically, we took each of the 400 ideas and put them in front of a survey panel of customers in the target market via an online purchase-intent survey. The question we asked was: “How likely would you be to purchase based on this concept if it were available to you?” The possible responses ranged from definitely wouldn’t purchase to definitely would purchase.

The responses can be translated into a purchase probability using simple market-research techniques. The average purchase probability of a human-generated idea was 40%, that of vanilla GPT-4 was 47%, and that of GPT-4 seeded with good ideas was 49%. In short, ChatGPT isn’t only faster but also on average *better* at idea generation.

Still, when you’re looking for great ideas, averages can be misleading. In innovation, it’s the exceptional ideas that matter: Most managers would prefer one idea that is brilliant and nine ideas that are flops over 10 decent ideas, even if the average quality of the latter option might be higher. To capture this perspective, we investigated only the subset of the best ideas in our pool—specifically the top 10%. Of these 40 ideas, five were generated by students and 35 were created by ChatGPT (15 from the vanilla ChatGPT set and 20 from the pretrained ChatGPT set). Once again, ChatGPT came out on top.

## What it means

We believe that the 35-to-5 victory of the machine in generating exceptional ideas (not to mention the dramatically lower production costs) has substantial implications for how we think about creativity and innovation.

First, generative AI has brought a new source of ideas to the world. Not using this source would be a sin. It doesn't matter if you are working on a pitch for your local business-plan competition or if you are seeking a cure for cancer—every innovator should develop the habit of complementing his or her own ideas with the ones created by technology. Ideation will always have an element of randomness to it, and so we cannot guarantee that your idea will get an A+, but there is no excuse left if you get a C.

Second, the bottleneck for the early phases of the innovation process in organizations now shifts from generating ideas to evaluating ideas. Using a large language model, an innovator can produce a spreadsheet articulating hundreds of ideas, which likely include a few blockbusters. This abundance then demands an effective selection mechanism to find the needles in the haystack.

To date, these models appear to perform no better than any single expert in their ability to predict commercial viability. Using a sample of a dozen or so independent evaluations from potential customers in the target market—a wisdom of crowds approach—remains the best strategy. Fortunately, screening ideas using a purchase intent survey of customers in the target market is relatively fast and cheap.

Finally, rather than thinking about a competition between humans and machines, we should find a way in which the two work together. This approach in which AI takes on the role of a co-pilot has already emerged in software development. For example, our human (pilot) innovator might identify an open problem. The AI (co-pilot) might then report what is known about the problem, followed by an effort in which the human and AI independently explore possible solutions, virtually guaranteeing a thorough consideration of opportunities.

The human decision maker is likely ultimately responsible for the outcome, and so will likely make the screening and selection decisions, informed by customer

research and possibly by the opinion of the AI co-pilot. We predict such a human-machine collaboration will deliver better products and services to the market, and improved solutions for whatever society needs in the future.

Christian Terwiesch and Karl Ulrich are professors of operations, information and decisions at the Wharton School of the University of Pennsylvania, where Terwiesch also co-directs the Mack Institute for Innovation Management. They can be reached at [reports@wsj.com](mailto:reports@wsj.com).