Nota. *Estimados lectores, estamos de regreso. Apenas llevamos una semana de clases del nuevo semestre y todavía estamos gozando la gratísima sensación de volvernos a ver. Cumplimos un año de clases presenciales. Increíble, ¿no? El regreso ha tenido sus complicaciones y peculiaridades. Estamos, todavía, en un proceso de transición que exige de todos un esfuerzo extra. Sin embargo, no tenemos ninguna duda, estamos felices de haber regresado a nuestras instalaciones. En esta nueva normalidad (llamémosla así por un momento), hay cambios muy interesantes. El mundo es sutilmente distinto a como era en marzo de 2020.*
Entre los múltiples detalles que ahora forman parte de nuestra vida cotidiana está ese "pequeño salto" de la Inteligencia Artificial (I.A.). Es cierto, la I.A. ha estado entre nosotros desde hace ya muchos años.
De todos modos hay algo novedoso en los "A.I. Chatbots" que atrae nuestra atención. Reproducimos a continuación fragmentos del artículo

When A.I. Chatbots Hallucinate

escrito por Karen Weise y Cade Metz.
Lo encontramos en el The New York Times *en la edición del día 8*
de mayo de 2023.
La verdad nos pareció súper interesante.
La descripción de lo que los autores llaman
"alucinaciones" de los Chatbots
es realmente sorprendente.
Da la impresión de que en verdad estamos
en una nueva etapa de la I.A., y de que
el estudio de las consecuencias de su
desarrollo promete ser muy interesante.
Cerramos con un emotivo saludo
a nuestros nuevos estudiantes.
Generación 2024:
¡Bienvenidos a la Facultad de Ciencias!

When A.I. Chatbots Hallucinate

Karen Weise and Cade Metz

When did The New York Times first report on "artificial intelligence"?

According to ChatGPT, it was July 10, 1956, in an article titled "Machines Will Be Capable of Learning, Solving Problems, Scientists Predict" about a seminal conference at Dartmouth College. The chatbot added:

CHATGPT
This conference is now considered to be the birthplace of artificial intelligence as a field of study, and the article mentions the term "artificial intelligence" several times in the context of the discussions and presentations that took place at the conference.

The 1956 conference was real. The article was not. ChatGPT simply made it up. ChatGPT doesn't just get things wrong at times, it can fabricate information. Names and dates. Medical explanations. The plots of books. Internet addresses. Even historical events that never happened.
When ChatGPT was recently asked how James Joyce and Vladimir Lenin first met -an encounter that has never been confirmed- this is how it responded:

CHATGPT
James Joyce and Vladimir Lenin met in Zurich, Switzerland in 1916. Both men were living in exile in Zurich during World War I. Joyce was a writer and Lenin was a revolutionary. They met at the Cafe Odéon, a popular gathering place for artists and intellectuals in Zurich.

Fabrications and definitive statements on uncertain history like these are common. Figuring out why chatbots make things up and how to solve the problem has become one of the most pressing issues facing researchers as the tech industry races toward the development of new A.I. systems.
Chatbots like ChatGPT are used by hundreds of millions of people for an increasingly wide array of tasks, including email services, online tutors and search engines. And they could change the way people interact with information. But there is no way of ensuring that these systems produce information that is accurate.
The technology, called generative A.I., relies on a complex algorithm that analyzes the way humans put words together on the internet. It does not decide what is true and what is not. That uncertainty has raised concerns about the reliability of this new kind of artificial intelligence and calls into question how useful it can be until the issue is solved or controlled.
The tech industry often refers to the inaccuracies as "hallucinations." But to some researchers, "hallucinations" is too much of a euphemism. Even researchers within tech companies worry that people will rely too heavily on these systems for medical and legal advice and other information they use to make daily decisions.

A New Generation of Chatbots

A brave new world. A new crop of chatbots powered by artificial intelligence has ignited a scramble to determine whether the technology could upend the economics of the internet, turning today's powerhouses into has-beens and creating the industry's next giants. Here are the bots to know:

ChatGPT, the artificial intelligence language model from a research lab, OpenAI, has been making headlines since November for its ability to respond to complex questions, write poetry, generate code, plan vacations and translate languages. GPT-4, the latest version introduced in mid-March, can even respond to images (and ace the Uniform Bar Exam).

Bing. Two months after ChatGPT's debut, Microsoft, OpenAI's primary investor and partner, added a similar chatbot, capable of having open-ended text conversations on virtually any topic, to its Bing internet search engine. But it was the bot's occasionally inaccurate, misleading and weird responses that drew much of the attention after its release.

Bard. Google's chatbot, called Bard, was released in March to a limited number of users in the United States and Britain. Originally conceived as a creative tool designed to draft emails and poems, it can generate ideas, write blog posts and answer questions with facts or opinions.

Ernie. The search giant Baidu unveiled China's first major rival to ChatGPT in March. The debut of Ernie, short for Enhanced Representation through Knowledge Integration, turned out to be a flop after a promised "live" demonstration of the bot was revealed to have been recorded.

"If you don't know an answer to a question already, I would not give the question to one of these systems," said Subbarao Kambhampati, a professor and researcher of artificial intelligence at Arizona State University.

ChatGPT wasn't alone in erring on the first reference to A.I. in The Times. Google's Bard and Microsoft's Bing chatbots both repeatedly provided inaccurate answers to the same question. Though false, the answers seemed plausible as they blurred and conflated people, events and ideas.

Google's Bard said:

The earliest mention of artificial intelligence (AI) in The New York Times was in 1956, when the term was coined by John McCarthy at a workshop at Dartmouth College. The workshop brought together some of the brightest minds in computer science to discuss the possibility of creating machines that could think like humans.

The Times reported on the workshop in a front-page article, calling it "a new field of science" that was "aimed at creating machines that can think like humans." The article quoted McCarthy as saying that the goal of AI was to "build machines with minds, capable of learning and adapting to new situations.

NEW YORK TIMES ANALYSIS

Mr. McCarthy is credited with creating the term artificial intelligence in 1955, in a proposal for the Dartmouth conference.The Times did not cover the Dartmouth workshop at all, let alone on the front page. More recent online articles refer to the conference as starting "a new field of science."

The Times could not find a reference to Mr. McCarthy ever saying this phrase, though it includes common ways to describe A.I.

MICROSOFT'S BING

Based on web search results, the earliest mention of artificial intelligence in the New York Times seems to be from a book review published on March 5, 1956. The book is called "The Human Use of Human Beings" by Norbert Wiener, and the reviewer is John R. Pierce. The review mentions artificial intelligence as one of the topics covered by Wiener, who is described as "a pioneer in cybernetics.: ***https://www.nytimes.com/1956/03/05/archives/the-human-use-of-human-beings-cybernetics-and-society-by-norbert.html***

NEW YORK TIMES ANALYSIS

The Times published a review of a book by Mr. Wiener one day earlier, on March 4, 1956.

This seminal book was published in 1950, not 1956. The March 4 review was of Mr. Wiener's book "I Am A Mathematician: The Later Life of a Prodigy."

The March 4 review was not written by John R. Pierce, a noted electrical engineer. It was written by John G. Kemeny, a mathematician.

The format of this URL is consistent with Times archives for articles in this period, but this exact webpage never existed.

According to The Times's archives, all the chatbots were wrong. They cited articles that did not exist. And while coverage of early research on thinking machines dated to the 1930s, it wasn't until 1963 that The Times first published an article with the phrase "artificial intelligence."

"We released Bard as an experiment and want to be as transparent as possible about well documented limitations," Jennifer Rodstrom, a spokeswoman for Google, said. "These are top of mind for us as we continue to fine tune Bard."

Like Google, Microsoft and OpenAI say they are working to reduce hallucinations.

The new AI. systems are "built to be persuasive, not truthful," an internal Microsoft document said. "This means that outputs can look very realistic but include statements that aren't true."

The chatbots are driven by a technology called a large language model, or L.L.M., which learns its skills by analyzing massive amounts of digital text culled from the internet.

By pinpointing patterns in that data, an L.L.M. learns to do one thing in particular: guess the next word in a sequence of words. It acts like a powerful version of an autocomplete tool. Given the sequence "The New York Times is a ____," it might guess "newspaper."

Because the internet is filled with untruthful information, the technology learns to repeat the same untruths. And sometimes the chatbots make things up. They produce new text, combining billions of patterns in unexpected ways. This means even if they learned solely from text that is accurate, they may still generate something that is not.

Because these systems learn from more data than humans could ever analyze, even A.I. experts cannot understand why they generate a particular sequence of text at a given moment. And if you ask the same question twice, they can generate different text.

Bard said in another chat:

GOOGLE’S BARD
The earliest mention of artificial intelligence in the New York Times was in a 1955 article titled “Machine That Thinks”. The article described a computer program that could play checkers at a human level. The program, developed by Arthur Samuel, was able to learn from its mistakes and improve its play over time. The article was written by Herbert Goldstine, a mathematician who worked on the Manhattan Project...The article was published on July 21, 1955, and it was met with mixed reactions.

NEW YORK TIMES ANALYSIS
Mr. Samuel started programming his checkers player in 1952, and it debuted on television in 1956. News articles in The Times are written by journalists, not subjects of the news. Herbert Goldstine appears to be an inaccurate reference to Herman Goldstine, a noted computer scientist.

Companies like OpenAI, Google and Microsoft have developed ways to improve the accuracy. OpenAI, for instance, tries to refine the technology with feedback from human testers.

As people test ChatGPT, they rate the chatbot’s responses, separating useful and truthful answers from those that are not. Then, using a technique called reinforcement learning, the system spends weeks analyzing the ratings to better understand what it is fact versus fiction.

A newer version of ChatGPT called ChatGPT Plus, which is available for a $20 monthly subscription, consistently avoided answering the question about the first mention of artificial intelligence in The Times. This could be the result of reinforcement learning or other changes to the system applied by OpenAI.

CHATGPT PLUS
I am unable to browse the internet or access specific archives directly. However, I can provide you with some general historical context on the topic of artificial intelligence.

“The New York Times”, being a prominent newspaper, likely covered the topic soon after its inception, with articles in the late 1950s or early 1960s. Early mentions of artificial intelligence in the paper would have discussed the hopes and challenges of AI research, as well as the potential impact of AI on society.

Microsoft built its Bing chatbot on top of OpenAI’s underlying technology, called GPT-4, and has layered on other ways to improve accuracy. The company uses GPT-4 to compare the chatbot’s responses with the underlying data and rate how the model is performing. In other words, Microsoft uses the A.I. to make the A.I. better.

The company also tries to improve the chatbot’s responses with help from its traditional internet search engine. When you type a query into the Bing chatbot, Microsoft runs an internet search on the same subject and then folds the results into the query before sending it on to the bot. By editing the query, said Sarah Bird, a leader in Microsoft’s responsible A.I. efforts, the company can push the system to produce better results.

Microsoft does not check the bot’s responses for accuracy in real time, Ms. Bird said, though it is researching how to do that. It checks the accuracy of a small portion of results after the fact and then uses that analysis.

Los chatbots no deberían hablar en primera persona

Jordi Pérez Colomé

La profesora Emily M. Bender tiene una misión: quiere que sepamos que la aparente maravilla de ChatGPT es más bien un loro. No un loro cualquiera, sino un “loro estocástico”. “Estocástico” significa que escoge las combinaciones de palabras según un cálculo de probabilidades, pero no entiende nada de lo que dice. Es difícil conversar con ChatGPT o con Bing y ser consciente de que es un loro y sólo un loro. Pero para Bender de esa conciencia dependen muchas cosas malas: “Estamos en un momento frágil”, dice. Y advierte: “Estamos interactuando con una tecnología nueva y el mundo entero necesita equilibrar rápido su alfabetización para saber cómo tratar bien con ella”. Su mensaje, en resumen, es: por favor, es una máquina que hace muy bien una cosa, pero nada más.

Esa falsa humanidad tiene varios problemas: “Nos provocará confianza. Y no asume responsabilidad. Tiene tendencia a inventarse cosas. Si muestra un texto que es cierto, es por casualidad”, asegura. “Nuestras sociedades son un sistema de relación y confianza. Si comenzamos a poner esa confianza en algo que no tiene responsabilidad, hay riesgos. Como individuos que interactuamos con esto, debemos tener cuidado con lo que hacemos con nuestra confianza. Las personas que la construyen deben dejar de hacerla parecer humana. No debería estar hablando en primera persona”, añade.

Bender, lingüista computacional de la Universidad de Washington, intuía que esto podía suceder desde 2021, cuando publicó un artículo académico ahora célebre sobre “los peligros de los loros estocásticos”: “No dijimos que esto iba a pasar. Dijimos que esto podría pasar y que deberíamos tratar de evitarlo. No era una predicción. Era una advertencia. Allí sólo hablamos un poco de lo peligroso que es hacer algo que parezca humano. Es mejor no imitar el comportamiento humano porque eso puede llevar a problemas”.

“Si creemos que existe una inteligencia artificial real, también seremos más propensos a creer que claro que podemos hacer sistemas automáticos de decisión que estén menos sesgados que los humanos cuando en realidad no podemos”, dice Bender.

“Nos parece mágico que una máquina pueda ser tan humana, pero en realidad es la máquina creando la ilusión de ser humana”, dice Bender. “Si alguien está en el negocio de vender tecnología, cuanto más mágica parezca, más fácil será venderla”, añade.

Diario *El País*, 17 de marzo de 2023.