

# AI for content generation: CNET's dilemma

Christoph Schneider

## Introduction

"I just want to reassure everybody: this will pass. It's uncomfortable, we will get through it, the news cycle will move on."<sup>1</sup> In a staff call on Friday, January 20, 2023, Lindsey Turrentine, EVP of content and audience of the popular technology-focused news website *CNET*, attempted to assuage *CNET* employees' fears. About a week earlier, *Futurism*, a news website focused on "tomorrow's topics," had revealed that *CNET* was using AI tools to generate content, a story that was picked up by outlets ranging from *CNN*<sup>2</sup> to the *Washington Post*,<sup>3</sup> giving rise to widespread criticism.<sup>4</sup>

Using automated tools to generate content was not unusual in the industry. The automated content creation process at *CNET* entailed that an editor would create an outline and then edit (or add to) the AI-generated content before the article's publication. However, the (proprietary) AI tool used by *CNET* to generate a total of 77 articles—often short explainers of financial topics—between the fall of 2022 and mid-January 2023 was found to have plagiarized sections and/or introduced factual errors, such as misrepresenting the amount of interest a person would earn on a deposit or misstating basic facts about mortgage types.<sup>5</sup> While there had apparently been problems with the content creation process, the issues were not limited to the use of AI content-generation tools. As revealed by *Futurism*, *CNET* only provided an innocuous disclosure statement about the content's origins when a user clicked on the article's byline, which read "CNET Money Staff."<sup>6</sup> While criticism mounted, *CNET*'s editor-in-chief maintained the outlet's innocence: "We didn't do it in secret. We did it quietly."<sup>7</sup>

## CNET and Red Ventures

Launched in 1994 as a producer of technology-focused television content, *CNET* had been acquired by the media company Red Ventures in 2020. Focusing on consumer advice on a range of topics, *CNET*'s parent company also owned brands such as *Bankrate* (personal finance tools), *CreditCards.com* (consumer credit card recommendations), *The Points Guy* (credit card and

---

This case was prepared by Professor Christoph Schneider. July 2024.

IESE cases are designed to promote class discussion rather than to illustrate effective or ineffective management of a given situation.

Copyright © 2024 IESE. No part of this publication may be reproduced, stored in a retrieval system, made available on any LLM (i.e., ChatGPT), or transmitted in any form or by any means - electronic, mechanical, photocopying, recording, or otherwise - without the permission of IESE.

Document only to be used in IESE programs.

Last edited: 22/7/24



frequent flyer point advice), *Reviews.com* (insurance advice), *BestColleges.com* (education advice), *Slumber Yard* (mattress reviews), and the well-known travel guide publisher *Lonely Planet*.<sup>8</sup>

Sites under the Red Ventures umbrella typically generated revenues not through display advertising but primarily through affiliate marketing.<sup>9</sup> For example, a *CNET* article about research on biofuels (<https://www.cnet.com/science/power-from-plants-newly-discovered-photosynthesis-leak-means-more-juice>) was accompanied by a review of the “best solar generators,” which contained links to purchase these at Amazon. Whenever a customer purchased a product through such an affiliate link, the site earned a referral fee from the vendor. Whereas the company earned a single-digit percentage of the transaction for referrals to Amazon, a referral to a high-end credit card (such as the Sapphire Reserve card from Chase) could bring in as much as US\$300 to US\$900.<sup>10</sup>

To make this business model work, Red Ventures's brands needed to provide trusted content and attract web traffic. In the words of Courtney Jeffus, the president of the company's financial services division, “Brand and trust are at the core of everything that we do. If you lose brand trust, then you don't have a business.”<sup>11</sup> While trusted content in itself can help attract visitors, the brands needed to excel at search engine optimization—through relevant and up-to-date content written to rank highly in search engines' algorithms. In other words, the content needed to be frequently updated and targeted at what was sometimes referred to as “high-intent” queries—that is, search queries that signal that a user's readiness to take a particular action, such as making a purchase.<sup>12</sup> Such queries typically included transactional key words, such as “best,” “purchase,” “comparison,” “vs.,” or “where to buy.”

## The rise of generative AI

AI is a broad field of study that aims to create intelligent machines that can perform tasks that typically require human intelligence, such as reasoning, learning, perception, and problem-solving.<sup>13</sup> AI includes various techniques, such as machine learning (ML), natural language processing, computer vision, and robotics, to achieve this goal.

In 1959, IBM computer scientist Arthur Samuels popularized the term “machine learning.” However, in the years since the field's inception, despite the foundations having been laid, visible progress in the area has been slow. Recently, harnessing algorithms' increasing capabilities of tremendous increases in computing power (of desktop as well as mobile devices), organizations have started including numerous ML-enabled features in their products, and consumers have gotten used to tools such as autocomplete on their smartphones and voice interfaces (think Alexa or Siri).

In 2022, in what felt like a seismic shift, tools such as DALL·E 2, Stable Diffusion, and ChatGPT were released to the public. These tools, collectively referred to as “generative AI,” could be used for a variety of novel purposes. For example, DALL·E 2, released to test users in May 2022 by OpenAI,<sup>14</sup> and Stable Diffusion, released in August 2022 by Stability AI,<sup>15</sup> encompassed a type of generative AI referred to as text-to-image models, which allowed images to be created based on text prompts. Although the output was initially often crude, rapid advances in the performance of the underlying models have led to the creation of AI-generated images of at times impressive quality. Released to much fanfare in November 2022 was ChatGPT,<sup>16</sup> a large language model (LLM) that allowed for natural language conversations with users and provided answers on just about any topic (limited to restrictions and safeguards built into the system). As so-called “foundation models,” these tools, trained on vast amounts of data, had the ability to be adapted to novel tasks, often in very different domains.<sup>17</sup>



Clearly, in late 2022/early 2023, the race was on. In February 2023, Facebook released LLaMa, its foundational LLM. In March 2023, Google released Bard, a conversational AI engine powered by its LLM, LaMDA. At the same time, Microsoft announced a planned investment of US\$10 billion in OpenAI, an amount that some considered a large bet (OpenAI ran on Microsoft's Azure cloud infrastructure).<sup>18</sup> Only a few weeks after its announcement, Microsoft had already integrated into its Bing search engine the technology, which combined AI-generated answers to user queries with Internet search. Introducing Bing Chat, Yusuf Mehdi, Microsoft's corporate vice president and consumer chief marketing officer, said, "To empower people to unlock the joy of discovery, feel the wonder of creation and better harness the world's knowledge, today we're improving how the world benefits from the web by reinventing the tools billions of people use every day, the search engine and the browser."<sup>19</sup> Shortly after the launch of Bing Chat, Bing had reached 100 million daily active users—not a small feat for a comparably small player in the search business. This also gave Microsoft's Edge browser a much-needed leap in popularity in the browser market, which was dominated by Google's Chrome browser. Microsoft CEO Satya Nadella proudly proclaimed, "Google is the 800-pound gorilla in search. I want people to know that we made them dance."<sup>20</sup>

## What are large language models?

LLMs were designed to understand and generate human language by using artificial neural networks, a type of ML algorithm intended to process information in a way that mimics how our brains work to analyze large amounts of data and identify patterns and relationships. To do so, a neural network uses layers of interconnected nodes, or "neurons," that receive input data, apply a mathematical function to that input, and then pass the output on to the next layer of neurons. The output of the final layer is the output of the network.<sup>21</sup>

During training, a neural network learns to adjust the connections between its neurons to better predict the correct output. This is done by feeding the network a large dataset of input–output pairs. The model then adjusts the weights of the connections between neurons to minimize the discrepancy between the predicted output and the actual output. In an LLM, the input layer receives a sequence of words or characters as input; these are passed through the next layers, which perform complex calculations to transform the input into a numerical representation that captures the semantic and syntactic relationships between the words. Finally, the output layer generates a sequence of words or characters that corresponds to the predicted text.

In the case of ChatGPT, once trained, the algorithm can generate text using the following process. First, it processes the input prompt and—based on patterns learned from the training data—predicts the probabilities of different words or phrases that could come next. Then, given the input prompt and/or the prior output, ChatGPT generates text by sampling (i.e., choosing at random) from this probability distribution of likely next words or phrases. The higher the probability of a word, the more likely the system is to choose that word. The resulting text is likely to be grammatically correct and coherent; the sampling introduces variability in the output, which is intended to make it more human-like but also introduces unpredictability. In other words, after ChatGPT outputs each word, it asks itself what the most likely next word would be, given the prior output, over and over again.<sup>22</sup> For example, if you type, "Hello, how are you today?" into ChatGPT, the model will analyze the input text and, based on the learned probability distribution of likely words, and predict what the most likely response would be, such as "I'm good, thanks for asking!"