Crowdsourcing: Innovation in Information Collection

In 1714, after exhausting the expertise of English scientists, the British government challenged the public to create a reliable system for calculating the longitude of a ship at sea for a prize. This challenge is cited as one of the first instances of crowdsourcing. Crowdsourcing is the leveraging of collective intelligence for a single purpose. The creation of the Oxford English Dictionary and many marketing decisions were made through crowdsourcing, but now the term is associated with web-based collections of knowledge, with Wikipedia and Trip Advisor as leading examples.

Today, crowdsourcing is used as a tool to break up highly monotonous tasks into tiny puzzles for web users and it can also be used to create a dataset for teaching machine learning. Through artificial intelligence, computers can "learn" how to properly identify images and texts or do other tasks that, due to their complexity, were previously only able to be done by hand. For a computer to learn a complicated task, it must repeatedly attempt to solve a problem. For example, if a computer is hoping to properly identify the image of a boat, it would be shown thousands of images and would attempt to flag each photo as either a boat or not a boat. Crowdsourcing helps create this dataset of right and wrong answers. By answering small puzzles like CAPTCHAs, everyday web-users are helping to formulate datasets and are "teaching" computers to do human-like tasks.

Luis Von Ahn, the co-creator of CAPTCHAs and reCAPTCHAs and co-founder of Duolingo has created tools that benefit users and machine learning simultaneously. reCAPTCHA, is an anti-fraud system that many people encounter on a daily basis when browsing on websites. The website is able to tell that the user is a real person through the proper identification of words and images. Simultaneously, the user is helping computers that are learning to digitize books by properly spelling out words the computer is unable to read. Similarly, Duolingo teaches users new languages and uses their responses to accurately translate outside webpages. Von Ahn's creations create a holistically beneficial service that protects users and websites from fraud while also teaching, digitizing, and translating the web.

By Logan Cort '22, Clarke Forum Student Project Manager

Information gathered from Abhigna, Soni, Dixit's "Crowdsourcing – A Step Towards Advanced Machine Learning" in *ScienceDirect*, vol. 132, 2018; "Crowdsourcing from its Beginnings to the Present," Clickworker (https://www.clickworker.com/customer-blog/evolution-of-crowdsourcing/); Daren Brabham's *Crowdsourcing* (MIT Press, 2013); *"*Information Systems" in *Encyclopedia Britannica* (Encyclopedia Britannica, Inc., 2020); Steven Zeitchik's "Former Google Scientist Says the Computers that Run our Lives Exploit Us" in *The Washington Post*, January 17, 2022.