ACS Study Finds Artificial Intelligence Aids in Data Gathering for Cancer Research

New research led by American Cancer Society (ACS) scientists found that large language model (or LLM) -based systems, a form of artificial intelligence, can offer a time-efficient and accurate solution to abstract key data for use in cancer research studies. Clinical data elements that are crucial to oncologic research are often only captured in patients' unstructured medical records, which can contain hundreds of pages of clinical notes. To maximize the value of these data, medical records have traditionally been manually abstracted by oncology data specialists, which is a time and labor-intensive effort. To modernize and accelerate data collection in the large ACS population studies, ACS explored the use of LLM-based systems. The findings are to be presented today at this year's annual meeting of the American Association for Cancer Research (AACR) in Chicago, April 25-30, 2025.

Researchers in Population Science at ACS, led by associate scientist Jillian Nelson, MPH, analyzed oncologic data abstracted from breast cancer-related medical records collected during routine follow-up of participants in the ACS's longitudinal Cancer Prevention Study 3, which initially enrolled 300,000 cancer-free adults nationwide from 2006-2013. Medical records and associated data, abstracted by Oncology Data Specialists, were used to develop and test the performance of the LLM-based platform *Distill* to abstract seven data elements: cancer behavior, laterality, neoadjuvant therapy status, and presence of key biopsy or surgery with associated procedure dates. Development began with aligning the *Distill* platform with 200 medical records and associated data. Abstraction guidelines and decision trees following the guidance of the North American Association of Central Cancer Registries were used as part of the development process. Den E Bloodworth, an ACS oncology data specialist, adjudicated disagreements between human and LLM-based abstraction during development.

The *Distill* platform was then tested using the remaining 100 medical records, and the platform completed abstraction for this test set within 1 day. Accuracy across all of the abstracted data elements matched or exceeded human abstraction, ranging from 94% -100% accuracy.

# # #

About the American Cancer Society
The American Cancer Society is a leading cancer-fighting organization with a vision to end cancer as we know it, for everyone. For more than 110 years, we have been improving the lives of people with cancer and their families as the only organization combating cancer through advocacy, research, and patient support. We are committed to ensuring everyone has an opportunity to prevent, detect, treat, and survive cancer. To learn more, visit cancer.org or call our 24/7 helpline at 1-800-227-2345. Connect with us on Facebook, X, and Instagram.

For further information: FOR MORE INFORMATION, CONTACT: American Cancer Society, Anne.Doerr@cancer.org

---

Additional assets available online: Photos (1)