AcademyHealth

# AI for HSR:
## Use Cases for AI across the Research Lifecycle

**Table 1. Study Design**

| Explanation | Examples |
|---|---|
| Algorithms can be used to predict clinical trial outcomes, anticipate trial success, and propose study design refinements that would increase likelihood of success. | Researchers at the **University of Illinois Urbana-Champaign** developed the **HINT** (hierarchical interaction network) algorithm and related **SPOT** (sequential predictive modelling of clinical trial outcomes) system, to predict the likelihood of clinical trial success. Predictions—used to inform needed alterations to the study design—are generated based on:<br>• the results and recency of other trials included in the training data set, and<br>• details about the drug molecule, target disease, and patient eligibility criteria for the trial at hand. |
| | **SEETrials**, developed by the company Intelligent Medical Objects, quickly pulls and synthesizes "safety and efficacy information" from clinical trials abstracts (sourced using ChatGPT) to see what past trial designs and corresponding outcomes have been. |

**Table 2. Recruitment and Retention**

| Explanation | Examples |
|---|---|
| AI can streamline and expedite the processes both of developing research eligibility criteria, and of identifying eligible participants. | Developed by James Zou and a group at Stanford, **Trial Pathfinder** optimizes study eligibility criteria by assessing how adjustments to inclusion criteria might impact trial hazard ratios. Analyzing training data from completed clinical trials, the tool can suggest appropriate criteria for potentially eligible trial participants—without increasing the likelihood of negative incidents. One study found that, in doing so, "the system…reduced harmful outcomes because it made sicker people – who had more to gain from the drugs—eligible for treatment." |
| | **Criteria2Query** helps researchers identify matching candidates by querying patient databases, based on inputs such as trial identification numbers or in/exclusion criteria presented in natural language. Also developed by the lab of Chunua Weng at Columbia University, **DqueST** functions as an interactive questionnaire that helps patients identify clinical trials for which they match eligibility criteria. |
| AI can manufacturer realistic and likely viable data, to help researchers generate statistically significant and generalizable results even when sample size is limited. | San Francisco based startup, **Unlearn**, creates a digital "twin" for each patient enrolled in a clinical trial. Using that patient's baseline data as a reference point, the system anticipates how the patient would have progressed in the control versus experimental groups. This approach is believed to offer a solution that both lowers the likelihood an enrolled patient will receive a placebo (and thus risk not benefitting from any form of needed treatment), and the number of patients needed for a control group by 20-50%. |
| AI tools and applications can support retention by predicting which enrollees will likely be lost to attrition, or by offering tailored communication to keep participants engaged. | **Patient Connect,** a mobile app developed by **AiCure**, utilizes AI and machine learning to remotely monitor patient dosing behavior through smartphones and computers. This software provides automated reminders and alerts to reduce dropouts and enhance retention. This software can help researchers anticipate participant attrition, and take proactive steps to keep patients engaged and adherent to clinical trial guidelines. |

| Table 3. Data Collection and Curation | |
|---|---|
| **Explanation** | **Examples** |
| Numerous applications of AI can support data collection and curation. Beyond traditional or straightforward applications, some can extract and translate unstructured data (e.g., images, lab results, natural language excerpts) into formats readily interoperable and usable for research. | Developed by Zou's group at Stanford University, **PLIP** functions as a search engine that helps users extract unstructured patient data (e.g., billing codes from medical records, text from provider notes, pathology photos or other images) captured in different formats or stored in different databases. |
| | **SDQ**, an AI-enabled tool developed by the company **Saama**, was used to rapidly clean data for over 30,000 patients as part of Pfizer's COVID-19 vaccine trial. Using machine learning approaches, the tool quickly surfaced issues with the trial (e.g., anomalous or duplicate data) so that these could be corrected. |
| Generative AI can also complete missing data points or generate and anticipate new "rows" of data. This "synthetic data" is artificially generated information created by algorithms or simulations, which can be used to supplement data or replace it when privacy concerns or data limitations prevent the use of actual data points. | In a joint case study with **Illumina**, **Gretel.ai** utilized generative neural networks to create synthetic genomic datasets: generated by algorithms or computer simulations and used to augment limited datasets, or as a substitute when real data are either not available or not sharable. This allowed researchers to collaborate and test ideas using data based on real-world sequences, without compromising patient privacy. |

| Table 4. Data and Evidence Synthesis | |
|---|---|
| **Explanation** | **Examples** |
| AI aids in developing and refining a strategy for article sourcing and synthesis: locating relevant articles, screening and extracting data, and identifying themes or insights. | **SWIFT-Active Screener** assists data analysts by using an "Active Learning" model that predicts the relevance of unscreened documents and continuously improves its accuracy with each reviewed article. Additionally, a separate model estimates the number of potentially relevant unscreened documents quickly and efficiently, which saves researchers time by filtering out irrelevant articles that may otherwise meet eligibility criteria. |
| | **System Pro** assists researchers with finding, synthesizing, and contextualizing relevant articles and surfacing new insights. This tool uses AI to synthesize millions of peer-reviewed studies on PubMed and set a new industry standard by citing all sources used while synthesizing. |
| AI can convert natural language inputs into visual charts and dashboards—enabling users to easily generate insights from raw, qualitative, or unstructured data. | **Data Formulator** invites data analysts to articulate concepts they want reflected in their data visualizations, even if these are not obviously apparent in the raw data. As users describe these concepts in the form of natural language queries, the LLM generates coding language to program the system. The code "instructions" then program the system to create new data categories that reflect these concepts. The system can then organize data points into those new categories, and translate the information into a visualization. |

| Table 5. Translation and Dissemination | |
|---|---|
| **Explanation** | **Examples** |
| AI can facilitate research translation and dissemination in many ways, including by translating complex scientific texts into multiple languages or by tailoring research insights into language and formats meaningful for diverse audiences. | Developed by Aaron Patzer and Justin Schrager, **Vital**'s **Doctor-to-Patient Translator** uses AI and language models to convert complex medical terminology into plain language at a fifth grade reading level, making medical information more accessible and understandable for patients. This tool distills clear, actionable insights from medical notes, test results, and discharge summaries. This translator can help researchers by simplifying complex medical information, to improve the quality of translated data, as well as the accuracy of analyses based on them. |
| | **Repustate** can assist researchers by capturing real-time sentiments through text and other news, reviews, or social media sources. This tool supports sentiment analysis in 23 languages and uses machine learning to stay up to date on trends, slang, and industry jargon. This can support researchers seeking to use or link social media data to support their analysis (e.g., using online posts or search records to anticipate a medical diagnosis), or to inform their understanding and interpretation of research results impacted by broader health care and public health influences. |
| | **ChatGPT**, an open access generative AI tool, can assist researchers in creating plain language summaries of academic articles to make complex findings accessible for non-academic audiences. It can also assist in writing social media posts, blog drafts, or press releases that communicate research findings for public audiences. |