

CROWDPLAT

NLP Challenge: Processing Biomedical Data into Visual Knowledge Graph

EXECUTIVE SUMMARY

To create a data-driven technology solution designed to accelerate scientific research in medicine, the National Institute of Health (NIH) and its subsidiary, the National Center for Advancing Translational Sciences (NCATS), chose CrowdPlat to plan and manage a data-driven challenge. The goal of the challenge was to ensure that data from biomedical publications was maximally leveraged to reach a wide range of biomedical researchers.

CrowdPlat along with its partner, Bitgrit, designed and successfully ran a two-phase challenge that met all project objectives and received raving feedback from NIH. The challenge asked competitors to design systems that accurately recognize and extract scientific concepts from the text of scientific articles, then connect those concepts to knowledge assertions, and finally determine if the claim is a novel finding or is merely background information. More than 200 data science experts participated in the challenge and the challenge received a total of 500+ submissions. After evaluation, seven (7) winners were selected and paid a total of \$100,000 prize money under the America COMPETES Act.

NCATS BACKGROUND

The NIH is the primary agency of the United States government responsible for biomedical and public health research. NIH established NCATS to coordinate and develop resources that leverage basic research in support of translational science and to develop partnerships and work cooperatively to foster constructive collaboration in ways that do not create duplication, redundancy, and competition with industry activities. Since its inception, NCATS is at the forefront of advancing the science of translation, turning observations into interventions, and using them to improve global health.

Phase 1:

Find all the nodes or biomedical entities (position in text and BioLink Model Category) given only an abstract text.

Phase 2:

Find all the relationships (pair of nodes, BioLink Model Predicate, and novelty) between the given abstract and annotated notes.

CROWDPLAT

PROJECT OVERVIEW

Open medical data on its own is not enough to deliver its full potential for public health. Actually, members of the scientific and medical community and the public can create tools with open data repositories by engaging technologists. Similarly, funders can exponentially increase the utility and value of said data to help solve pressing national health issues.

The LitCoin Natural Language Processing (NLP) project sought to spur innovation by rewarding the most creative and high-impact uses of biomedical, publication-free text in creating knowledge graphs. These knowledge graphs would link concepts within existing research, allowing researchers to find connections that otherwise might have been difficult to discover.

CROWDSOURCING CHALLENGE

CrowdPlat along with its partner, Bitgrit, designed and ran the project as a two (2) phase crowdsourcing challenge for NCATS. The challenge initiated the creation of innovative strategies in NLP by allowing participants across academia and the private sector to participate in teams or individually. See below details of the competition for the two phases.

Phase 1:

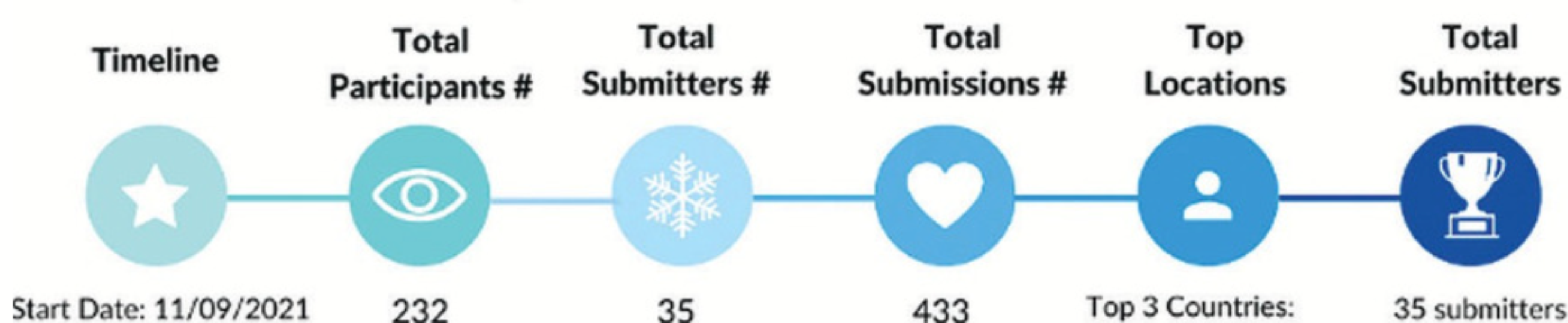
Goals: Identify the position and type of biomedical concepts (entities) mentioned within a research paper's title and abstract.

Phase 2:

Goals: Identify all the relations between biomedical entities within a research paper's title and abstract.

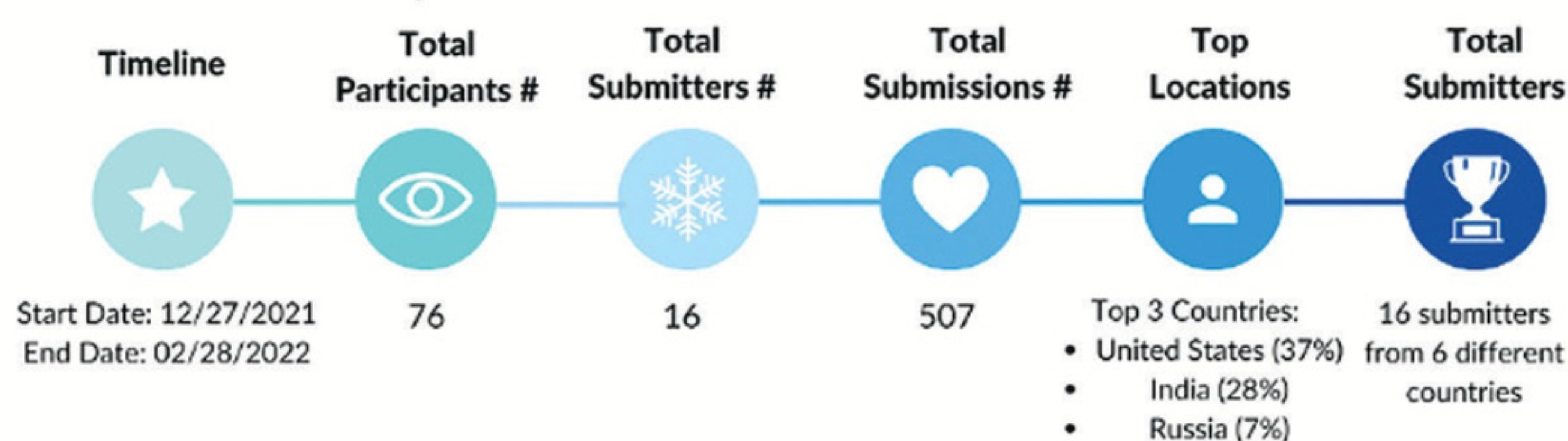
Phase I

Part 1: Identify mentions of biomedical entities in research abstracts



Phase II

Part 2: Identify relations between biomedical entities in research abstracts



CROWDPLAT

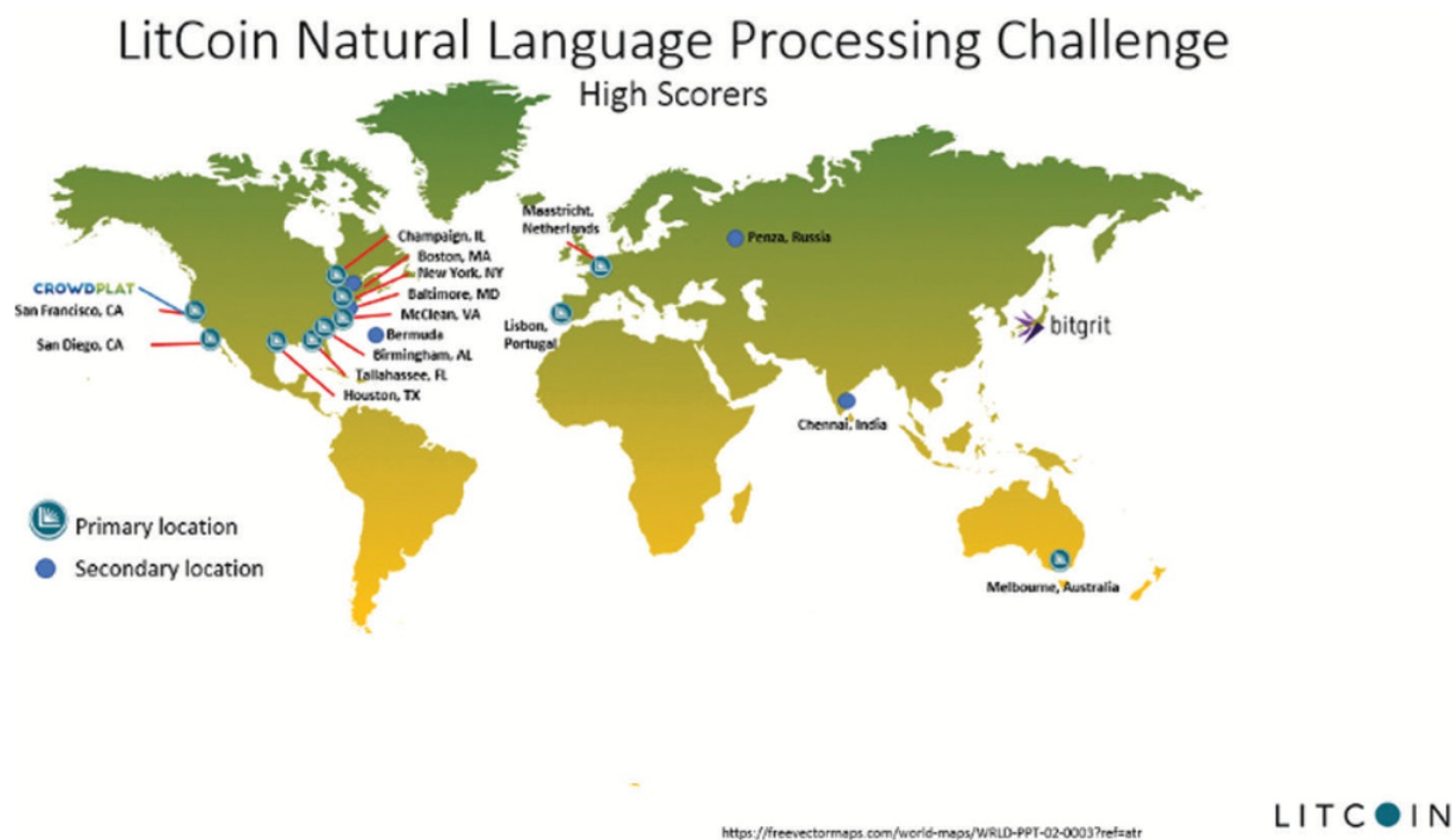
Prizes were awarded to the top-ranking data science contestants who created NLP systems that accurately captured the information denoted in free text and provided the output of this information through knowledge graphs.

The results from this challenge will spur innovation in NLP to advance the field and generate accurate and useful data from biomedical publications as a result enhancing the ability for data scientists to create tools to foster discovery and generate new hypotheses.

COMPETITION PARTICIPANTS GEOGRAPHY

The geographic spread of the challenge's high scorers below showcases what is possible when scientists and computational researchers come together to change the research landscape.

The spread further illustrates CrowdPlat's ability to utilize a global pool of talent.



THE CROWDPLAT ADVANTAGE

1. Timely delivery through continuous engagement and interaction.

As a direct result of the reputation and credibility of CrowdPlat's in-house data science expert, who acted as the project manager, initiation of contact and follow-through with NCATS' managers, developers, and testers have been and continue to be successful. The project manager kept in continuous communication with their NCATS counterparts. This enabled timely clarification and resolution of all technical aspects and ensured smooth project execution.

2 Partnership with AI company Bitgrit

CrowdPlat and our partner, Bitgrit, provide a community of top talent data science and software experts worldwide. In addition, both companies provide a dedicated in-house team of data science experts who offer unparalleled service, flexibility, and support for the entire duration of the project.

In project closeout meeting, client referred to the successful partnership as the reason for the success of the project.

3 Executed Timeline:

CrowdPlat's incredible project management team comprised of in-house data scientists from around the world who met and exceeded our customers' expectations by completing the LitCoin NLP project

We curated industry experts with the talent and ability to elevate our client's business.