Data Competition Rules:

Challenge Event 1

Version 1h

June 14, 2024

Defense Advanced Research Projects Agency
Biological Technologies Office
675 North Randolph Street
Arlington, VA 22203-2114
TriageChallenge@darpa.mil

1 Contents

2 Introduction

This document describes the Data Competition Rules of the DARPA Triage Challenge (DTC). This document supersedes previous versions of the DARPA Triage Challenge Rules. Significant revisions from past versions in this document are indicated by blue text. Teams are encouraged to closely review the entire document. The intent of this document is to provide participants guidance on competition design and scoring objectives to inform their development efforts in preparation for the first competition event. This document is subject to change and may be superseded by later versions. The latest official versions of all documents are posted on the DARPA Triage Challenge Website (triagechallenge.darpa.mil) and the DARPA Triage Challenge Community Forum.

DARPA intends to release a draft of the Competition Rules no later than nine months before each Challenge Event. The final version of the Competition Rules will be released no later than three months prior to each respective event.

The DARPA Triage Challenge Chief Judge has the final authority to make any decisions related to the rules or scoring. All decisions made by the Chief Judge are final.

The main goal of the DARPA Triage Challenge is to inspire development of scalable, timely, and accurate capture of novel injury signatures to enhance triage decision-making in austere, complex, and mass-casualty settings. The challenge elements and the competition structure itself are intended to address the additional goal of increasing the diversity, versatility, cost-effectiveness, and robustness of relevant technologies and systems capable of addressing the myriad needs of a wide range of mass casualty incidents (MCIs) rather than single-purpose or specifically tailored solutions. The third goal of the DARPA Triage Challenge is to establish a collaborative community by bringing together multi-disciplinary teams and cross-cutting approaches across disparate fields to address the autonomy, perception, and diagnostic needs of the medical triage community.

3 Overview

Under the authority of 10 U.S.C. §4025 to stimulate innovations using prize competition, the DARPA Triage Challenge will use a series of competition events to drive breakthrough innovations in the identification of physiological features ("signatures") of injury. These new signatures will help medical responders perform scalable, timely, and accurate triage. Of particular interest are MCIs, in both civilian and military settings, when medical resources are limited relative to the need.

The DARPA Triage Challenge's long-term vision is 1) an initial, or primary stage of MCI triage supported by sensors on stand-off platforms, such as uncrewed aircraft vehicles (UAVs) or uncrewed ground vehicles (UGVs), and algorithms that analyze sensor data in real-time to identify casualties for urgent hands-on evaluation by medical personnel; followed by 2) a secondary stage, after the most urgent casualties have been treated, supported by non-invasive sensors placed on casualties and algorithms that analyze sensor data in real-time to predict the need for life-saving interventions (LSIs) by medical personnel. Injury information provided by these sensors in primary and secondary triage could be integrated with other information about the scene to accumulate evidence about the injury mechanism and characteristics in order to enhance overall situational awareness, and to focus further physiological interventions.

To advance progress towards this vision, the DARPA Triage Challenge aims to bring together multi-disciplinary teams and industries that will identify physiological signatures and develop sensor and algorithm strategies for complex MCI settings. Teams participating in the DARPA Triage Challenge will be

[1] Patterns in sensor data that reflect or predict injuries of high importance for triage assessments

tasked with developing and demonstrating strategies for capturing high-value signatures for either primary or secondary triage, or for both. While aspects of the DARPA Triage Challenge involve sensors and sensor-delivery platforms, the priority is the development of physiological signatures and models to detect them, not the development of new sensor or platform technology.

4 DARPA Triage Challenge Schedule Overview

The DARPA Triage Challenge is a 3-year effort with 3 sequential 12-month phases for Primary Triage (Systems and Virtual Competitions) and Secondary Triage (Data Competition) in parallel, each culminating in a challenge event (Figure 1; see the DTC website for competition details). In each phase, competitors will develop signatures and detection and analysis strategies for Primary and/or Secondary Triage. DARPA will host two competition events in each phase; a workshop and a challenge event.

Competition events will become progressively more difficult and realistic from Phase 1 to Phase 3.

The workshops will provide an opportunity for practice and mid-phase evaluation for all tracks.

Table 1 provides additional information on schedule and format of Competition events and workshops.

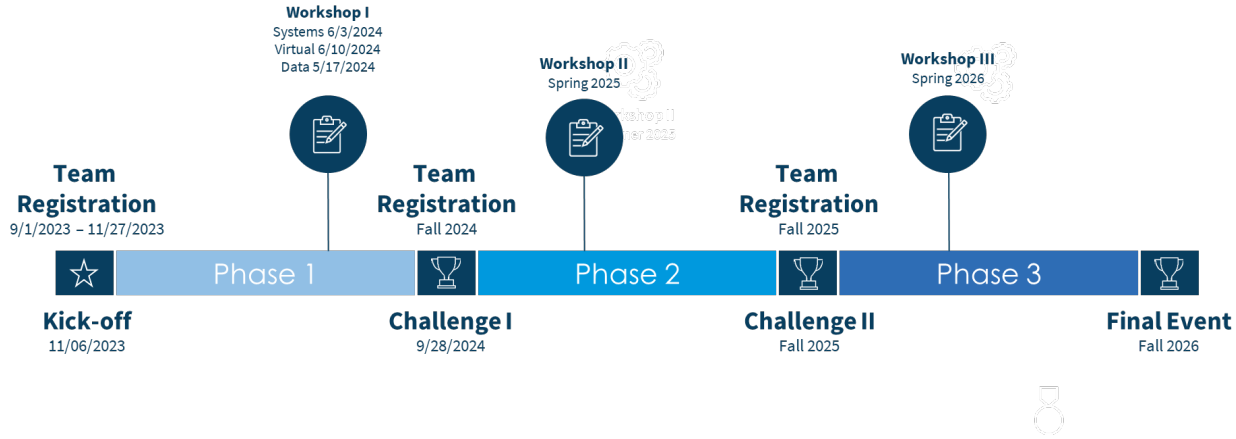

Figure 1 - Program structure and schedule for the DTC.

Virtual Competition and Data Competition - Tracks C, D and E			
Event	**Location**	**Est. Duration**	**Date**
Year 1			
Challenge Kick-off	Hybrid	2 days	Nov 6-7, 2023
Workshop - Month 8 *Evaluations / Runs*	Virtual	7 days	May 17, 2024 (Data) June 10, 2024 (Virtual)
Workshop - Month 8 *Lessons-learned session*	Virtual	1 day	6/17/2024
Challenge 1 - Month 12 *Evaluations / runs*	Virtual	TBD	8/30/2024
Challenge 1 - Month 12 *Awards /lessons-learned session*	Hybrid	1 day	10/5/2024
Year 2			
Workshop - Month 4 *Evaluations / Runs*	Virtual	7 day	Spring 2025
Workshop - Month 4 *Lessons-learned session*	Virtual	1 day	Spring 2025
Challenge 2 - Month 12 *Evaluations / runs*	Virtual	TBD	Fall 2025
Challenge 2 - Month 12 *Awards /lessons-learned session*	Hybrid	1 day	Fall 2025
Year 3			
Workshop - Month 4 *Evaluations / Runs*	Virtual	7 day	Spring 2026
Workshop - Month 4 *Lessons-learned session*	Virtual	1 day	Spring 2026
Final Challenge - Month 11 *Preliminary rounds*	Virtual	TBD	Fall 2026
Final Challenge - Month 11 *Finalists only - Runs and Awards*	In person	1 day	Fall 2026

*Table 1 - Schedule of DARPA-organized Challenge events and workshops. *Note: DARPA-funded teams must participate in all workshop events. It is highly recommended that self-funded Systems teams also attend the workshops.*

5 Prizes and Funding

Teams are encouraged to pursue high-risk, high-reward approaches to meet and exceed the objectives of the Challenge Events. Monetary prizes will be awarded for the Systems, Virtual, and Data Competitions at each of the Challenge Events (Table 2).

	Systems [self-funded]	Virtual [self-funded]	Data [self-funded]
Challenge I Fall 2024	1st $120,000	1st $60,000	1st $120,000
	2nd $60,000	2nd $30,000	2nd $60,000
	3rd $20,000	3rd $10,000	3rd $20,000
Challenge II Fall 2025	1st $300,000	1st $300,000	1st $300,000
	2nd $150,000	2nd $150,000	2nd $150,000
	3rd $50,000	3rd $50,000	3rd $50,000
Challenge III Fall 2026	1st $1,500,000	1st $600,000	1st $900,000
	2nd $750,000	2nd $300,000	2nd $450,000
	3rd $250,000	3rd $100,000	3rd $150,000

Table 2 - Prize structure for the three Challenge Events

DARPA-Funded Teams

DARPA-funded teams (Systems and Data Competitions) are only eligible for the prizes in the Final Events (selection for DARPA-funded teams has closed). The Government's obligation for prizes under DTC is subject to the availability of appropriated funds from which payment for prize purposes can be made. No legal liability on the part of the Government for any payment of prizes may arise unless appropriated funds are available to DARPA for such purposes.

Self-Funded Teams

Self-funded teams (all three competitions) are eligible for prizes in all of the Challenge Events.

Systems Competition Prizes and Funding: The Phase 1 and Phase 2 prizes for self-funded Systems teams will be awarded to the best performing self-funded Systems Teams in each event, provided that the teams finish in the top 5 overall (including DARPA-funded Systems Competition) teams. High-performing teams are also eligible to become a DARPA-funded team in Phase 2 and/or 3. The Government's obligation for prizes under DARPA Triage Challenge is subject to the availability of appropriated funds from which payment for prize purposes can be made. No legal liability on the part of the Government for any payment of prizes may arise unless appropriated funds are available to DARPA for such purposes.

Virtual Competition Prizes and Funding: The Phase 1 and Phase 2 prizes for self-funded Virtual Competition teams will be awarded to the best performing self-funded Virtual Teams. The Government's obligation for prizes under DARPA Triage Challenge is subject to the availability of appropriated funds from which payment for prize purposes can be made. No legal liability on the part of the Government for any payment of prizes may arise unless appropriated funds are available to DARPA for such purposes.

Data Competition Prizes and Funding: The Phase 1 and Phase 2 prizes for self-funded teams will be awarded to the best performing self-funded Data Teams, provided that the teams finish in the top 5 overall (including DARPA-funded Data Competition teams). High-performing teams are also eligible to become a DARPA-funded team in Phase 2 and/or 3. The Government's obligation for prizes under DARPA Triage Challenge is subject to the availability of appropriated funds from which payment for prize purposes can be made. No legal liability on the part of the Government for any payment of prizes may arise unless appropriated funds are available to DARPA for such purposes.

To be eligible for prizes, teams must first be registered in the team qualification portal. The award process requires recipients to furnish information that may trace or identify recipients either individually or as an organization (e.g., Social Security Number or Tax Identification Number). The primary contact of each registered team is responsible for providing the award information necessary for prize disbursement. DARPA will reach out by email to the primary contact of each registered team to either confirm their vendor status or request the required forms (e.g., SF-3881 or PIF). DARPA is not responsible for disbursement of prizes to any team members other than the primary contact/organization.

At the end of each competition event, teams will be invited to discuss their technical approaches and lessons learned in a townhall-style hotwash. The extent of technical details shared does not need to exceed data agreements established upon qualification.

6 Qualifications

Prospective DTC competitors must demonstrate track-appropriate performance capabilities to be eligible to participate in DARPA Triage Challenge. All teams in all three competitions (Primary Triage Systems tracks, Primary Triage Virtual track, and Secondary Triage Data tracks; see the DTC website for track details) must complete two types of qualification: a Team Qualification at the beginning of each phase, followed by event-specific Event Qualifications for each Workshop and Challenge Event. Successful Team Qualification is a prerequisite to Event Qualifications in the same phase.

The initial *DTC Event Qualification Guide* is expected to be released by February 18th, 2024. The *DTC Event Qualification Guide* will continue to be updated for each event. The latest revision will be posted on the DTC Website and DTC Community Forum.

6.1 Team Qualification

Teams must qualify for DARPA Triage Challenge competition events during the designated qualification window by completing the *Team Qualification* form on the DTC Team Portal. Team Qualification submissions will be accepted on a rolling basis but must be submitted by the deadline (see Table 3). Team qualification is required to receive access to datasets and must be completed prior to event-specific enrollment.

Team Qualification Windows by Phase	
Phase 1	9/1/2023 - 11/13/2023
Phase 2	9/1/2024 - 11/15/2024
Phase 3	Fall 2025

Table 3 – Team qualification schedule.

6.2 Event Qualification

Prospective teams are required to demonstrate baseline performance and utility capabilities (e.g., safety measures for the Systems Competition, simulator usage for the Virtual Competition, and algorithm capability for the Data Competition), to be eligible to participate in events. **All** teams (DARPA-funded and self-funded) in all competitions (Systems, Virtual and Data) must qualify for each event including the DTC workshops, Preliminary Events (i.e. Phase 1 and Phase 2 Challenge Events), and Final Event.

The latest revision of the *DTC Event Qualification Guide* will be posted on the DARPA Triage Challenge Website and DTC Discourse Community Forum. Event Qualification submissions will be accepted on a rolling basis but must be submitted by the deadline to be eligible to participate in the event (Table 4). The specific qualification deadlines for each event are provided in the *DTC Event Qualification Guide.*

Failing a previous qualification attempt does not preclude a team from resubmitting a revised qualification submission within the qualification deadlines for any given event. DARPA may adjust the qualification rules for each event and may choose to award qualification waivers for teams that have successfully participated in a prior Workshop or Challenge Event.

DARPA reserves the right to disqualify any team that is found to violate either the rules or applicable laws and regulations.

Event	Event Qualification	Event Date
Workshop 1	3/5/2024 - 4/5/2024	6/3/2024 - 6/8/2024
Challenge 1	6/28/2024 – 7/30/2024	9/28/2024 - 10/5/2024
Workshop 2	12/5/2024 -1/5/2025	3/10/2025-3/15/2025
Challenge 2	Summer 2025	Fall 2025
Workshop 3	Winter 2025-2026	Winter 2025-2026
Challenge 3	Summer 2026	Fall 2026

Table 4 – Event qualification schedule.

7 DARPA Triage Challenge Technical Workshops

DARPA encourages vibrant information exchange and collaborative interactions among all DARPA Triage Challenge participants, to include DARPA technical staff, independent verification and validation (IV&V) teams, representatives from competitor teams, infrastructure developers, and other government partners. To that end, DARPA will host a workshop in each phase which will offer a forum for community building and cross-pollination of technical ideas and approaches as well as an opportunity for testing and integration.

In each phase (8 months into Phase 1, 4 months into Phases 2 and 3) DARPA will host a virtual workshop for the Data Competition, in which teams provide preliminary submissions for evaluation and receive performance results based on held-out data. The practice sessions will be followed by a ‘lessons learned’ discussion for all tracks and an opportunity to discuss real-world needs with Government partners.

As part of the workshops, teams will have the opportunity to confirm integration with the DARPA instrumentation and scoring systems to ensure compliant submissions ahead of the Challenge Events. Submissions for the workshops are not officially scored, but teams are encouraged to operate according

to the Competition Rules to prepare for the Challenge events. Attendance at workshop events is required for all DARPA-funded teams. Self-funded teams may choose to attend.

We will hold a virtual lessons learned meeting shortly after the workshop for teams to discuss experience gained regarding technical aspects of their systems at the workshop tests.

8 Human Subjects Research (HSR)

For the Data Competition, use of training data provided by DARPA does not constitute HSR, and competitors do not need to obtain IRB approval to use these data. However, DARPA-funded competitors require DARPA approval for the collection or use of any other human subject data. **Self-funded teams are prohibited from the collection or use of any other human subject data as part of their involvement in the DARPA Triage Challenge, beyond data and data-collection opportunities provided by DARPA, because DARPA HSR supervision is not feasible for teams not under DARPA contract.** Self-funded teams should carefully consider this limitation and should take this into account in their technical approach, leveraging other strategies as appropriate (*e.g.*, simulations).

DoD Definition of Human Subjects Research (HSR)

The term "human subject" can be applied to research efforts that meet EITHER of the following criteria: A living individual about whom an investigator (whether professional or student) conducting research:

- Obtains information or biospecimens through intervention or interaction with the individual, and uses, studies, or analyzes the information or biospecimens; or
- Obtains, uses, studies, analyzes, or generates identifiable private information, personally identifiable information, or identifiable biospecimens.

Human Subjects Research involves:

- Activities that include both a systematic investigation designed to develop or contribute to generalizable knowledge and involve a living individual about whom an investigator conducting research obtains information or biospecimens through intervention or interaction with the individual, or identifiable private information, or biospecimens.

8.1 Handling of DARPA-provided data

Data Competition Datasets are provided by the DARPA Research Infrastructure for Trauma with Medical Observations (RITMO) research program for use during DTC. DARPA's mission requirement and intent are to safeguard privacy and civil liberties and the RITMO datasets have been intentionally de-identified to ensure—to the greatest extent practicable—that there is no reasonable basis to believe that the data could be used to trace a specific identity or present a risk of harm to any individual. Therefore, as previously acknowledged in the DTC Qualification process, competitors agree that they will *not intentionally* attempt to re-identify data in the RITMO datasets, nor attempt to download or share the RITMO datasets.

9 Secondary Triage: Data Competition Rules

9.1 Data - Illustrative Scenario

The objective of the Data Competition is to identify physiological signatures of injury derived from data captured by non-invasive sensors (contact-based or stand-off) to enable anticipatory decisions and

prioritization for medical evacuation and care. Performers will develop algorithms that detect signatures in these data streams to provide decision support appropriate for austere and complex pre- hospital settings. Of particular interest are early signatures indicating need for LSIs against conditions that medics are trained and equipped to treat during secondary triage, such as hemorrhage and airway injuries.

The Data Competition is virtual only, and will use DARPA provided de-identified, multi-modal physiological data from trauma patients in diverse settings and cohorts provided by the DARPA **Research Infrastructure for Trauma with Medical Observations (RITMO)** program. DARPA will provide access to a subset of these data for algorithm training and evaluate competitor algorithms on held-out test data in workshops and end-of-phase challenge events. Challenge events will become progressively more complex and realistic from Phases 1 to 3.

9.2 Data - Technical Challenge Elements

The Challenge events will be designed to assess performance across various challenge elements, including: multiple data sources, multiple data inputs, raw data, and degraded data. The challenge elements are expected to become progressively more difficult from Phase 1 to Phase 3.

1. *Multiple data sources:* With varied source populations, patient demographics, types of injury, and standard clinical operating procedures, each data set may have a different standard set of sensor readings, with different added sensors in each phase of DTC. Approaches must demonstrate robustness across a range of settings.

2. *Multiple data inputs*: Potentially including static data (e.g., mechanism of injury or anatomical injury pattern); multiple simultaneous, continuous streams of high-frequency waveforms; and point-of-care imaging data.

3. *Raw data*: Data as it comes from the sensors and health record systems (i.e., without any cleaning), with the noise, aberrant values, and dropouts that occur in clinical environments.

4. *Degraded data (year 2 and 3)*: DARPA also may inject additional challenges that can be expected in battlefield and civilian pre-hospital settings, such as severe degradation or total loss of a particular sensor, to test the robustness of competitor strategies to such plausible scenarios.

9.3 Data Types

Competitors in the Data Competition will be given data specifications for the modeling/training data that will be provided. DARPA-provided datasets originate from the DARPA Research Infrastructure for Trauma with Medical Observations (RITMO) program, comprised of data from two complementary academic hospital systems. The dataset contains at least 3000 cases of pre-hospital or hospital data for year 1 and may include: discrete data (text or numeric values): patient characteristics, outcomes, procedures, labs; Continuous data: (electrocardiogram) ECG, photoplethysmography (PPG), respiration rate (RR), heart rate (HR), SPO2, blood pressure (BP), Arterial-line BP. Data sources are expected to increase with each phase of the challenge. Data dictionaries will be provided alongside each data release containing details on data format and proper interpretation.

9.3.1 Evaluation Data

While all data fields will be made available for training, only a subset of the data will be available at evaluation time for predicting LSIs. These will include data sources that would feasibly be available to medics *in situ* during treatment, excluding information about future events (e.g., outcomes) or contextual

information typically collected post-treatment (e.g., medical history). Details about data format and fields provided at evaluation time can be found in the Data Competition ICD.

9.3.2 Data Quality

RITMO datasets contain deidentified data collected from trauma cases at two independent hospital systems. Due to the real-world challenges of medical data collection, there are varying degrees of consistency, completeness, and precision in data provided for this challenge. Minimal data cleaning has been applied to the data provided for training and used in evaluation, and teams are expected to develop their own mitigation strategies for handling real-world data.

9.4 Data - Scored Event Submissions

9.4.1 Solutions Submissions

For scored event submissions, teams must submit their solution in AWS ahead of the submission deadline. After the deadline, submissions will be built and evaluated using a held-out test dataset sampled from the RITMO dataset. Teams will be provided with an opportunity to test their submissions ahead of the submission deadline to ensure their solutions are able to operate within the evaluation system. Guidelines on submission preparation and testing resources are included in the Data Competition ICD.

The solution submission window for the first challenge will open approximately 2 months prior to the Awards ceremony. Each qualified team must submit a single solution to be scored. The submissions will be evaluated and the final results will be announced alongside the Systems and Virtual Competition results at the Competition Awards ceremony.

Challenge Event	Submission Window	Results Release
Challenge 1	*7/30/2024 – 8/30/2024*	*10/5/2024*
Challenge 2	*Summer 2025*	*Fall 2025*
Challenge 3	*Summer 2026*	*Fall 2026*

Table 5 - Submission window for the Datal Competition teams.

9.4.2 Submission scoring

Each qualified team must submit a single solution per event to be scored. Team solutions will be evaluated under a simulated online prediction paradigm using ***m*** held-out test cases. Submitted solutions will be given incremental data within sequential time intervals to approximate streaming source data, and any predicted LSIs over time will be compared against ground truth LSIs appearing at later times in the case.

Scores will reflect relative performance of each team's solution compared to the other teams' solutions. Within each case, teams will be ranked according to the accuracy and timeliness of their responses. Points will then be awarded to the top-ranking teams. The Event Score is the sum of points received across all cases (see Figure 2). Details on scoring and team ranking are in Section 9.6 of this document.

$$\text{Event Score}_k = \sum_{i=1}^{m} S_{i,k}$$

Figure 2- Event Score for each team (k) is computed as a sum of points received over m cases. Points (S) are awarded for each case (i) based on each team's performance relative to the other teams.

The final competition scores and any supporting materials will not be released until the event results are announced.

9.4.3 Human in the loop

The submitted solutions will be evaluated with no external operator interfaces such as command line inputs or user interventions. Teams are required to develop self-contained solutions that predict LSIs entirely autonomously without Human Supervisor interaction. Entries that require any user input or external commands will not be scored.

9.4.4 Run Termination

A scored run for a given case terminates upon any of the following conditions:

- **Time Expiration**: The cumulative processing time exceeded the evaluation run duration time limit, as described in Section 9.5.4.
- **Prediction Saturation**: The evaluated solution predicted the need for *all* LSI types, in which case no future model responses will affect score outcome.
- **Patient Expiration**: The patient expired or the case otherwise ended before another termination criterion was met.

9.4.5 Score Disputes

Score Disputes are intended to provide teams a mechanism to submit a formal dispute or request for review by the Chief Judge. All score disputes should be sent by email to the DARPA Triage Challenge email address (triagechallenge@darpa.mil) within 48 hours of receiving competition results. All disputes or requests will be reviewed by the Chief Judge in a timely manner. All decisions made by the Chief Judge are final.

9.5 Data - Preliminary Event Scenarios

9.5.1 Preliminary Event Competition Environments

All computing operations and analysis involving the RITMO dataset must occur within the AWS ecosystem throughout the duration of the challenge. Each team will be provided an isolated, network-restricted AWS Workspace domain, which will allow teams to manage storage and computing resources. Team members will be provided Workspace user accounts and can instantiate computing resources (AWS EC2, SageMaker, etc.) on demand within these Workspaces (see Figure 3).

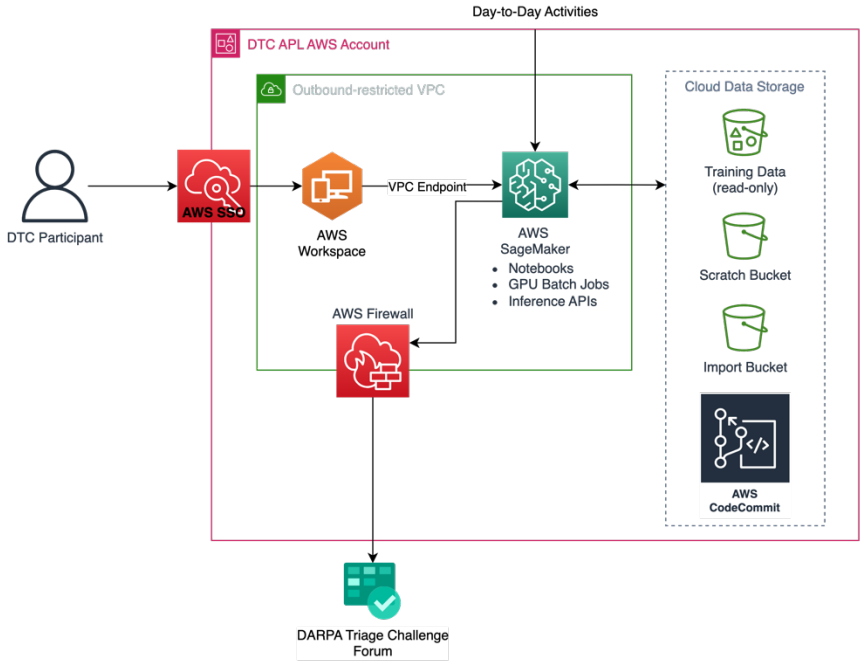

Figure 2 - DTC Participant AWS Architecture. Participants will perform all computations pertaining to sensitive data within the provided AWS ecosystem. Participants will be able to access the RITMO dataset stored in an S3 bucket, as well as instantiate computing resources (e.g., AWS EC2, AWS SageMaker, etc.) within their Workspace. Each team will be provided with a Scratch Bucket to share data or files among team members.

Teams would deploy AWS SageMaker instances to use GPU resources. Table 6 shows the available resources, their compute specifications, the maximum number of running instances per team, and the price per hour.

Instance Type	Compute Specifications				Purpose	Limit (per team)	Price per Hour ($)
	CPU	RAM (GB)	GPU	GPU Memory			
ml.t3.medium	2	4	0	0	General Purpose	15	0.05
ml.m5.large	2	8	0	0	General Purpose	15	0.115
ml.m5.2xlarge	8	32	0	0	General Purpose	15	0.461
ml.c5.xlarge	4	8	0	0	Compute Optimized	15	0.204
ml.c5.2xlarge	8	16	0	0	Compute Optimized	15	0.408
ml.c5.9xlarge	36	72	0	0	Compute Optimized	6	1.836
ml.r5.8xlarge	32	256	0	0	Memory Optimized	6	2.419
ml.r5.16xlarge	64	512	0	0	Memory Optimized	6	4.838
ml.p3.2xlarge	8	61	1	16	GPU - General	6	3.825
ml.p3.8xlarge	32	244	4	64	GPU - General	6	14.688
ml.g4dn.4xlarge	16	64	1	16	GPU - General	12	1.505
ml.g4dn.8xlarge	32	128	1	16	GPU - Training	6	2.72
ml.g5.4xlarge	16	64	1	24	GPU - Inference	1	2.03

Table 6 - AWS processor options.

Teams will be allocated a fixed budget at the beginning of each phase to spend on storage and computing resources. It will be the team's responsibility to manage their budget expenditures and resource consumption throughout the competition. All teams will be provided a dashboard to maintain and track their resources. Any remaining funds (up to 25%) at the end of a phase may be rolled into the subsequent phase. The Government's obligation for AWS funding under the DARPA Triage Challenge is subject to the availability of appropriated funds.

While working within Workspaces, participants are allowed to download data from the internet, allowing them to browse external websites and download open-source packages. However, participants are restricted from uploading data to ensure data does not leak outside the AWS ecosystem. Furthermore, if teams would like to use bespoke (privately developed) tools sitting outside AWS, they may package them and upload them to the Import Bucket, where they can retrieve them inside their SageMaker console.

9.5.2 Preliminary Event LSIs

As part of the RITMO dataset, treatments and related clinical actions have been identified and grouped into LSI categories based on shared injury patterns and treatment paradigms. Each team in the Data competition is tasked with predicting the occurrence of these LSI categories, not the specific treatment or clinical action. For brevity, the LSI categories will be referred to as LSIs for the remainder of this document. LSIs with timestamps will be provided alongside the training and test datasets.

The preliminary list of LSIs includes: airway & respiration, bleeding control, blood products, cardiovascular procedures, chest decompression, crystalloid products, neurological products & procedures, RSI sedation medications, vascular access & monitoring, vasoactive & cardioactive medications, limb salvage, and damage control procedures. These LSIs may change in future years to reflect new data sources or changing priorities of the challenge. Details on the LSIs can be found in the data dictionary provided alongside the dataset.

9.5.3 Prediction Task

Submitted solutions will be evaluated as if they are performing an online prediction task, ingesting incremental continuous and discrete data within sequential windows. Given each window, the task is to predict the set of LSIs needed in the future relative to the observed data. Predictions are compared to the ground-truth LSIs occurring at future timepoints in the case. See Section 9.6 for details on the prediction task scoring criteria.

9.5.4 Preliminary Event Run Duration

The event run duration is defined as the cumulative processing time for solutions to produce prediction responses across all test cases. To ensure timely evaluation and efficient solutions, DARPA will set an evaluation run duration limit prior to each event as part of the Data Competition ICD. In Workshop 1 this time limit is 24 hours. In Competition 1 this time limit is 48 hours.

9.6 Data - Scoring Criteria

In the Secondary Triage (Data) Challenge, teams are evaluated based on sensitivity, specificity, and speed in predicting future LSIs from physiological signals and contextual health information. Results for the Data Competition will be announced at the prize ceremony on the last day of the competition event.

Submitted solutions in the data competition will be evaluated for accurate and early prediction of the need for LSIs using a set of held-out test cases. To simulate online prediction, solutions will be tested at multiple evaluation timepoints along each case. At each evaluation timepoint, solutions will be provided with incremental data from the past and any predicted LSIs will be compared against the ground truth set of LSIs present in the future within the same case, where past and future are defined relative to the evaluation timepoint. Details on the input data format and responses with LSI predictions are specified in the Data Competition ICD.

The Event Score is based on the performance of each team relative to the other teams. Within each case, teams will be ranked according to the accuracy and timeliness of their LSI predictions. Top-ranking teams will then receive a point for out-performing the other teams on the same case. The sum of points across cases in the test set will determine the winning team.

9.6.1 Definitions

All solutions will be evaluated at a predetermined set of *evaluation timepoints* within each *case*, where a case encompasses the available pre- and in-hospital data and timestamped LSIs from a single hospital admission. At each evaluation timepoint, timestamped data within the *observation window* will be provided to submitted solutions for the prediction task, where the observation window begins with the start of case and ends at the evaluation timepoint. The *prediction target* is the unique set of ground-truth LSIs within the *prediction window*, the time window beginning 15 minutes after the evaluation timepoint and extending to the end of care or four hours after hospital admission, whichever comes first.

The gap between observation and prediction windows represents the minimum lead-time necessary for actionable clinical predictions that would provide value for medical resource allocation and planning. Any LSIs occurring outside of the prediction window (i.e., within the observation window or the intervening gap) are excluded from the prediction target. See Figure 4 for an illustration of the evaluation timepoint and surrounding windows, and how the prediction target is derived from the underlying LSI instances.

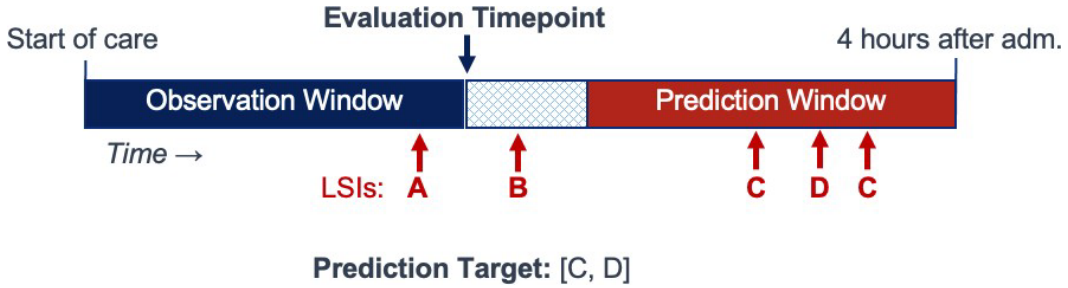

Figure 3 - Illustration of evaluation at a single timepoint. Solutions generate predictions using data occurring within the observation window preceding the evaluation timepoint. The prediction target is the set of timestamped LSIs (C and D) present in the prediction window, which begins after the evaluation timepoint and a predefined gap. LSIs occurring before the prediction window, whether within the observation window (A) or the intervening gap (B), are not included in the prediction target. Note that the prediction target is the set of unique LSIs present in the prediction window, not the frequency or order of LSIs.

9.6.2 Online prediction

Solutions will be tested at multiple evaluation timepoints along each case. At each timepoint, solutions will be provided with incremental data segments building upon previous data provided at earlier

timepoints, and solutions will be given the opportunity to predict the set of LSIs present in the prediction window relative to the current evaluation timepoint. See Figure 5 for an illustration of a series of evaluation timepoints, data segments, and prediction windows for a single case. At test time, solutions will be provided with the timestamp of the evaluation timepoint, timestamped data, and a case identifier. Note that all timestamps will be expressed as time elapsed relative to the beginning of the case. Evaluation timepoints will be tested in temporal order. After predictions from the final evaluation timepoint within a case have been collected, the sequence of predicted LSIs will then be scored against the sequence of ground truth LSIs over the case. See the Data Competition ICD for details on the spacing of evaluation timepoints and data format expected at evaluation.

To incentivize efficient processing and ensure time-bounds on the evaluation, the run-time of solutions will be recorded for each prediction. There will be a predetermined maximum processing time, equivalent to the data segment duration, after which the prediction is assumed to be the empty set and the evaluation will continue to the next evaluation timepoint or the next case, whichever is appropriate. The processing delay of correct predictions in each case will factor into the scoring, as described in the next section.

For each case, teams will be ranked in descending order according to a time-sensitive similarity metric comparing LSI predictions to ground truth LSIs over time. Teams with the same similarity metric will then be ranked further in descending order according to a timeliness metric. The similarity metric and timeliness metric are described in the following sections.

After ranking by similarity metric and by cumulative prediction lead-time, the top *k* teams receive a point towards their event score. The event score is the sum of points over all test cases (see Section 9.4.2). The winning team has the highest event score, or equivalently, the highest number of cases where they out-performed *N-k* other teams, where *N* is the total number of teams. The number of top-ranking teams that receive a point within each case, *k*, will be set ahead of each event. For the Phase 1 Competition Event, *k=3*.

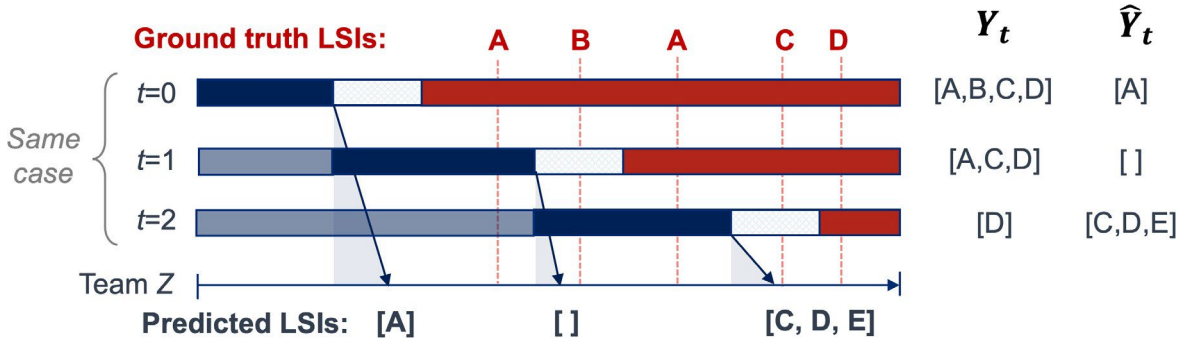

Figure 4 - Example of online prediction over a single case. There are three evaluation timepoints at which LSIs may be predicted, indicated by index t (left) and colored bars similar to Figure 4. At each timepoint, teams are provided with incremental data building on previous data provided at earlier timepoints (navy bars). Without loss of generality, LSIs are abstracted to the symbols {A, B, C, D, E}. Y_t is the set of ground truth LSIs in the prediction window at evaluation timepoint t, and $\hat{Y}_t$ is the LSI predictions at evaluation timepoint t. Arrows indicate processing delay at each evaluation timepoint. Prediction score is based on the sequence of predictions $\hat{Y}_t$ and the processing delay for correct predictions across the entire case.

9.6.3 Similarity metric: Jaccard Index

The Jaccard Index (a.k.a., intersection-over-union) has been adapted as a time-sensitive similarity metric for online prediction (see Figure 6): it rewards for correctly predicted LSIs and penalizes for false positive, missing, or late predictions. Correct LSI predictions occur when at least one ground truth LSI matching the

prediction is within the prediction window for the evaluation timepoint; false positive LSI predictions occur when the predicted LSI does not occur in any prediction time window; missing LSI predictions occur when the ground truth LSI is never predicted; and late LSI predictions occur when a ground truth LSI is predicted only after it no longer falls within the prediction window.

The Jaccard Index is the ratio of total count of unique correct LSI predictions across timepoints to the count of unique LSIs in both predicted and ground truth LSIs across the case. This similarity metric is bounded between 0 and 1, where 1 indicates correct prediction of all LSIs without any false positives and 0 indicates that none of the ground truth LSIs were correctly predicted in time.

Implicit in this metric are the following consequences: (1) the order of LSIs predicted at time *t* does not impact the metric, (2) the quantity of LSIs predicted at time *t* does not impact the metric, (3) prediction of a particular LSI at time *t* is equivalent to predicting it at all subsequent timepoints within the case, and (4) prediction of a particular LSI at time *t* cannot be retracted at subsequent timepoints. These are all intentional consequences of the union over intersections at each timepoint *t* in the numerator.

For cases with no ground truth LSIs, the similarity metric collapses to 0 for solutions that falsely predict an LSIs at any time *t*, and is undefined for solutions that correctly do not predict any LSI at any time *t*=0,…,T. In this special case, any team that correctly predicts no LSIs over the entire case receives a point toward their event score.

$$J = \frac{\left|\bigcup_{t=0}^{T}\left[\hat{Y}_t \cap Y_t\right]\right|}{\left|\bigcup_{t=0}^{T}\left[\hat{Y}_t \cup Y_t\right]\right|}$$

Figure 5 - Time-sensitive similarity metric based on Jaccard Index (J), also known as Intersection-over-Union. The denominator is the count of unique LSIs across predictions ($\hat{Y}$) and ground truth (Y) over the entire case (t=0,…,T). The numerator is the count of unique correct LSI predictions at time t over the entire case.

9.6.4 Timeliness metric: Prediction Lead-Time

Figure 7 shows the prediction lead-time timeliness metric, measuring how early a given solution correctly predicted LSIs within a case. It is the sum of time intervals between the simulated prediction timestamp and the ground truth LSI timestamp, where the *prediction timestamp* is the evaluation timepoint plus the processing time of the solution, as recorded by the evaluation system at run-time. If the same LSI was predicted at multiple evaluation timepoints, the earliest correct prediction is used.

The prediction lead-time is accumulated across each correctly predicted LSI occurrence, thereby rewarding the solution with an earlier correct prediction of multiple LSI occurrences of the same type.

$$PLT = \sum_{\substack{\text{Earliest correct predictions} \\ \text{for each LSI type}}} \text{LSI timestamp} - \underbrace{(\text{Evaluation timepoint} + \text{processing delay})}_{\textit{Simulated prediction time}}$$

Figure 6 - Timeliness metric Prediction Lead Time (PLT). The time difference between the ground truth LSI timestamp and the simulated prediction time, accounting for processing delay, is summed across all correct predictions. For cases with correct prediction of multiple LSI instances from the same group, they count as separate elements in the sum. Using the example in Figure 5, there are three items in the sum: the time difference between both ground truth instances of [A] and the correctly predicted [A], and the time difference between the ground truth instance of [D] and the correctly predicted [D].

9.6.5 Final Ranking

For the Secondary Triage Data Competition, the final ranking will be determined based on each team's event score. In the event that multiple teams have an identical event score, those teams will be ranked in descending order according to the sum of prediction lead-times across all test cases.

10 Appendix 1 DTC Glossary

Chief Official – Program manager or higher DARPA authority for the DARPA Triage Challenge.

Systems Competition –Primary Triage Competition run with actors on a real course (Track A, B).

Virtual Competition – Primary Triage Competition run on a virtual platform (Track C).

Data Competition – Secondary Triage Competition (Track D, E).

Chief Judge – DARPA-designated individual with the sole and final authority to make any decisions related to the rules or scoring.

Judge – DARPA-designated individual with authority to make decisions related to rules, scoring, and safety, with decision-making authority only superseded by the Chief Judge.