Data Competition Rules

Challenge Event 3

Version 6

August 24, 2026

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 This document is subject to change and may be superseded by later versions.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 Competition) 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.

Data Competition – DARPA-funded and self-funded			
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)
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	3/17/2025

Workshop - Month 4 *Lessons-learned session*	Virtual	1 day	Spring 2025
Challenge 2 - Month 12 *Evaluations / runs*	Virtual	TBD	8/30/2025
Challenge 2 - Month 12 *Awards /lessons-learned session*	Hybrid	1 day	10/04/2025
Year 3			
Workshop Part 1 - Month 4 *Evaluations / Runs*	Virtual	7 day	3/27/2026
Workshop Part 2 - Month 6 *Evaluations with new scoring*	Virtual	1 day	6/15/2026
Workshop - Month 4 *Lessons-learned session*	Virtual	1 day	7/8/2026
Final Challenge - Month 11 *Preliminary rounds*	Virtual	TBD	9/30/2026
Final Challenge - Month 11 *Finalists only - Runs and Awards*	In person	1 day	Fall 2026

Data Competition – DARPA-funded and self-funded			
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)
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	3/17/2025
Workshop - Month 4 *Lessons-learned session*	Virtual	1 day	Spring 2025
Challenge 2 - Month 12 *Evaluations / runs*	Virtual	TBD	8/30/2025
Challenge 2 - Month 12 *Awards /lessons-learned session*	Hybrid	1 day	10/04/2025
Year 3			
Workshop Part 1 - Month 4 *Evaluations / Runs*	Virtual	7 day	3/27/2026
Workshop Part 2 - Month 6 *Evaluations with new scoring*	Virtual	1 day	6/15/2026
Workshop - Month 4 *Lessons-learned session*	Virtual	1 day	7/8/2026
Final Challenge - Month 11 *Preliminary rounds*	Virtual	TBD	9/30/2026
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 and Data Competitions at each of the Challenge Events (Table 2).

Prizes

Systems Competition		Data Competition
AUTONOMOUS PLATFORMS Most lives saved on human-machine teaming lane	TECHNICAL CAPABILITY Overall performance on component testing gates	PREDICTION OF LIFE-SAVING INTERVENTIONS Overall performance on first look, continuous triage and resource allocation
Grand Prize $1,500,000†	1st Place $300,000*	Grand Prize $1,000,000†
Runner-up $500,000†	Runner-up $150,000*	2nd Place $300,000†
	Gate best † • Find and locate • Rapid triage • Trauma assessment • Accurate vitals $12,500 Each	ELIGIBILITY † All teams *Self-funded teams only

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 in the Data Competition are eligible for prizes in all of the Challenge Events.

Data Competition Prizes and Funding: All teams (DARPA-funded and Self-Funded) are eligible for the Grand prize in Phase 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 competitions (Primary Triage Systems tracks 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 - 12/2/2024
Phase 3	11/1/2025 - 6/30/2026

Table3. – 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 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 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	3/17/2025

Challenge 2	5/28/2025 – 6/30/2025	**Data** 8/30/25 - Submission 10/4/2025 - Awards
Workshop 3	12/5/2025 – 6/15/2026	Part 1: 3/27/2026 Part 2: 6/15/2026
Challenge 3	**Systems** 7/28/2026 – 8/30/2026 **Data** 6/28/2026 – 7/30/2026	**Systems** Window between: 11/1/26 and 1/15/26 **Data** 10/10/26 - Submission November 2026 - Awards

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 DARPA for use during DTC (henceforth dataset(s)). DARPA’s mission requirement and intent are to safeguard privacy and civil liberties and the 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 datasets, nor attempt to download or share the 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 DARPA. 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 training data that will be provided ahead of challenge events and used for evaluation. DARPA-provided datasets are comprised of data from two complementary academic hospital systems. Year 3 data will contain thousands of pre- and in-hospital cases. Data modalities include the following: 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. In addition to standard of care sensors, the dataset also contains ventilator data, pupillometry, and iSTAT portable blood analysis. Case count and data sources have increased 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).

9.3.2 Data Quality

DARPA-provided 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 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 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	*7/30/2025-8/30/2025*	*10/4/2025*
Challenge 3	*9/30/2026 – 10/10/2026*	*November 2026*

Table 5 - Submission window for the Datal Competition teams.

9.4.2 Submission scoring

Each qualified team must submit a solution for each of the tasks described below. Team solutions will be evaluated under multiple prediction tasks using held-out test data. Submitted solutions will be scored separately for each prediction task based on ground-truth LSI events. The Event Score for each team is a composition of performance metrics tailored to each prediction task. Details on scoring procedure and performance metrics are in Section 9.6 of this document.

Preliminary or partial results may be released to teams before the final ranking for review, however 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 processing time exceeded time limits, as described in Section 9.5.4.
- **Patient Expiration**: The patient expired or the case otherwise ended before another termination criterion was met.
- **Non-Compliant Submission:** The submission does not adhere to the ICD or otherwise fails at runtime.

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 24 hours of receiving data logs. 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 – Final Event Scenarios

9.5.1 Event Competition Environments

All computing operations and analysis involving the 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).

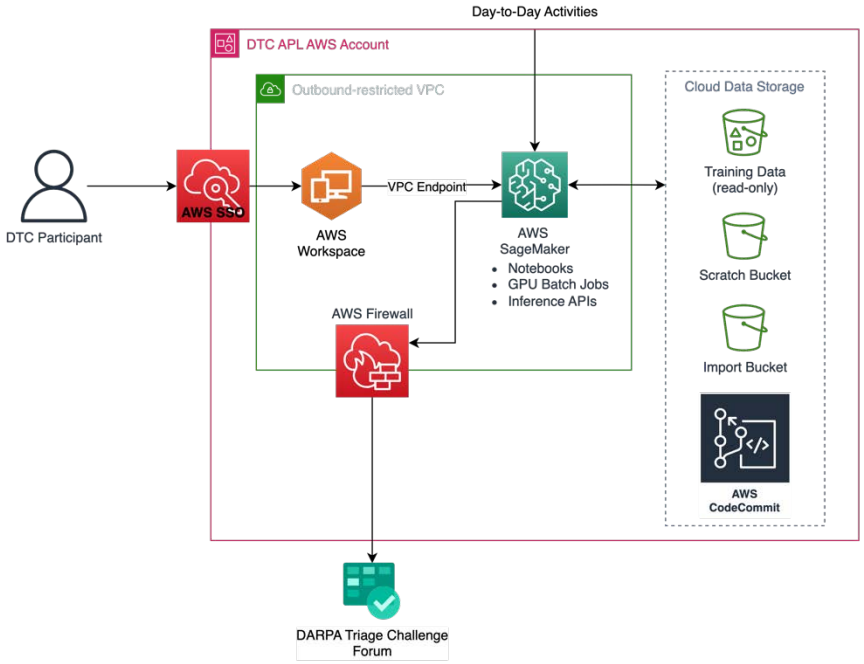

Figure 2 - DTC Participant AWS Architecture. Participants will perform all computations pertaining to sensitive data within the provided AWS ecosystem. Participants will be able to access the dataset stored in an S3 bucket, as well as instantiate computing resources (e.g., AWS EC2, AWS SageMaker, etc.) 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 some of the instance types available, their compute specifications, and the price per hour. The **ml.g4dn.4xlarge** instance type will be used for evaluation. For complete list of instance types and pricing available on AWS SageMaker, visit https://aws.amazon.com/sagemaker/ai/pricing/.

Instance Type	Compute Specifications				Purpose	Price per Hour ($)
	CPU	RAM (GB)	GPU	GPU Memory		
ml.t3.medium	2	4	0	0	General Purpose	0.05
ml.m5.large	2	8	0	0	General Purpose	0.115
ml.m5.2xlarge	8	32	0	0	General Purpose	0.461
ml.c5.xlarge	4	8	0	0	Compute Optimized	0.204
ml.c5.2xlarge	8	16	0	0	Compute Optimized	0.408
ml.c5.9xlarge	36	72	0	0	Compute Optimized	1.836
ml.r5.8xlarge	32	256	0	0	Memory Optimized	2.419

ml.r5.16xlarge	64	512	0	0	Memory Optimized	4.838
ml.p3.2xlarge	8	61	1	16	GPU - General	3.825
ml.p3.8xlarge	32	244	4	64	GPU - General	14.688
ml.g4dn.4xlarge	16	64	1	16	GPU - General	1.505
ml.g4dn.8xlarge	32	128	1	16	GPU - Training	2.72
ml.g5.4xlarge	16	64	1	24	GPU - Inference	2.03

*Table 6 - AWS processor options. **Bold** indicates instance type used for evaluation (ml.g4dn.4xlarge).*

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, access to external websites is limited and participants are restricted from transferring data to ensure data does not leak outside the AWS ecosystem. 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 LSI Events

As part of the dataset, treatments and related clinical actions will be identified and grouped based on shared injury patterns and treatment paradigms. Each team in the Data competition is tasked with predicting the occurrence of these LSI events, where LSI event refers to an instance of particular clinical actions on the patient. 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 list of LSIs for Phase 3 include: Invasive Airway, Blood Products, Chest Decompression, Surgical, and Vaso/Cardioactive Medications and Any LSI. Details on the specific LSIs can be found in the data dictionaries provided with each data release. New datasets in Phase 3 will use the updated LSI list, and the Phase 1 and 2 datasets will be re-released for use in Phase 3 (see the DTC Community Forum for dataset release announcements).

9.5.3 Prediction Tasks

Submitted solutions will be evaluated on three separate predictions tasks using continuous (e.g., physiological signals) and/or discrete (e.g., health record) data segments from pre- and in-hospital data to predict LSI events in the future relative to input data:

(1) **First Look.** Prediction of existence and timing of LSI events in the future based on initial pre-hospital data.

(2) **Continuous Alert.** Prediction of existence of LSI within near-term horizon based on streaming in-hospital data.

(3) **Resource Allocation.** Assignment of medical resources (based on LSI events) within a simulated mass casualty event with multiple casualties based on initial pre-hospital data.

For Workshop 3 (WS3) Part 1, teams will submit solutions toward the first two tasks. Solutions for all tasks, including Task 3, will be required at WS3 Part 2. See Section 9.6 for details on the prediction tasks, input data, and scoring procedure.

9.5.4 Time Limits

The evaluation 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 duration limit prior to each event. The time limit for Challenge Event 3 will be based on number of cases processed per hour, and it will be applied to each task individually: 200 cases per hour for First Look, 10 cases per hour for Continuous Alert, and 2 scenarios per hour for Resource Allocation. If evaluation progress is slower than the minimum rate above, the evaluation of that task will end early and move on to the next task.

9.6 Data - Scoring Criteria

Teams will be evaluated based on accurate and timely prediction of future need for LSIs using physiological signals and contextual health information. Submitted solutions will be evaluated using a set of held-out test data. Results for the Data Competition will be announced at the prize ceremony on the last day of the competition event.

Solutions will be evaluated on three distinct tasks: First Look, Continuous Alert, and Resource Allocation. Each task will be scored separately and combined to form a single score for the challenge event. The remainder of this section contains overall definitions, scoring approach for each task, and the overall event score used to determine prizes.

9.6.1 Definitions

The datasets for training and evaluation are comprised of individual *cases*, where each case includes available pre- and in-hospital physiological signals and electronic health record (EHR) information from a single hospital admission. At evaluation time, solutions are provided with *input data* including one or more of the following: timestamped physiological waveforms and/or trends, timestamped interventions received, injury information, mental status, hospital admission time, and basic demographics. The *prediction target* is the information being predicted by solutions. The *prediction horizon* is the time window in the future in which the model predicts LSI events. Predictions provided as *confidence scores* take values between 0 and 1. The input data, prediction target, and prediction horizon differ by task, as described in subsections below.

9.6.2 Task 1: First Look

In this task, solutions use initial pre-hospital data to predict the existence and timing of LSI events received during pre-hospital care and up to four hours after hospital admission. See Figure 3 for an illustration of this task.

Figure 3. Illustration of Task 1: First Look . Based on initial pre-hospital data, confidence scores indicate existence of LSI in 15-minute time-bins up to four hours after Hospital Admission. Input data will include one or more of the following: EHR data, Casualty Report, prior LSIs, hospital admission time, and physiological signals. Output is confidence scores between 0 and 1 for each LSI and each time-bin; confidence scores are shown above for a single LSI only for clarity (red bars). In this example, the highest score (*) is used to calculate the LSI-specific Existence Score, and the rank of the time-bin containing the ground truth LSI (2) is used to calculate the LSI-specific Temporal Localization Score. An additional output confidence score indicates existence of any LSI received during the case, to be used to evaluate benchmark performance (not shown).

9.6.2.1 Input data and multiple runs

To examine the role of physiological signals and EHR data on the prediction of LSI events, solutions will be evaluated under multiple runs within the First Look task. Runs will differ by input data provided during the evaluation.

Input data has been divided into the following categories:

- Basic EHR: demographics, injury, GCS, initial EHR vitals, instantaneous labs (i.e., iSTAT)
- Casualty Report: triage category, injury, and mental status akin to the Systems Competition
- Hospital Admission time: the time-elapsed in the case when hospital admission occurs.
- Prior LSIs: Interventions received prior to the pre-hospital data, including LSIs
- Initial LSIs: Interventions received within the first 5 minutes of pre-hospital data, including LSIs
- Initial Vitals: First 5 minutes of continuous waveforms and trends from pre-hospital data

The following run schedule lists input data available at test time for each run:

Run 1. Casualty Report, Prior LSIs, Hospital Admission time.
Run 2. Casualty Report, Prior LSIs, Hospital Admission time, Initial LSIs, Initial Vitals.
Run 3. Basic EHR, Prior LSIs, Hospital Admission time, Initial LSIs, Initial Vitals.

Note that Run 1 does not contain any physiological signals. All timestamps will be expressed as time elapsed relative to start of pre-hospital vitals data. More details about data fields provided during each run can be found in the Data Competition ICD.

Each run is evaluated and scored independently, and run scores are averaged into a single score for the First Look task, as described in Section 9.6.2.4.

9.6.2.2 Prediction target and horizon

There are two prediction outputs for the First Look task:

(1) A single confidence score representing existence of *any* LSI after input data and up to four hours after hospital admission.

(2) For each LSI, a time-series with confidence scores representing existence of an LSI event within 15-minute time bins into the future after input data and up to four hours after hospital admission.

The first prediction output is used as a benchmark to establish minimum performance differentiating between patients who do and do not receive an LSI within the prediction horizon up to four hours after hospital admission. The second prediction output is used to score LSI event prediction performance in the First Look task.

9.6.2.3 Benchmark

Let:

- $s_{i,any}$ be the confidence score that case i receives any LSI from the target list (see Section 9.5.2)
- y_i be the presence of any LSI after the input data and within the prediction horizon
- $AP(x_i, y_i)_{i=1}^{N}$ be the average precision between score x_i and label y_i for N cases indexed by i

The metric for determining minimum benchmark performance for First Look is the average precision of predicting existence of *any* LSI across cases in the test dataset:

$$\text{Benchmark Score} = AP\left(s_{i,any}, y_i\right)_{i=1}^{N}$$

If the minimum benchmark is not reached within a given run, the run receives a score of 0. Minimum benchmarks for each run will be based on prevalence of cases that received at any LSI within the evaluation dataset.

9.6.2.4 Scoring

The First Look task uses a composite score to separately evaluate predicting the existence and the timing of future LSI events:

$$S_r = \omega_E \cdot S_{E,r} + \omega_T \cdot S_{T,r}$$

where S_r is the score for run r, and $S_{E,r}$ and $S_{T,r}$ are the Existence Score and Timing Localization Score for run r, respectively. For Phase 3 Challenge Event, weights ω_E and ω_T will be 0.5.

The First Look task score is the average across R runs:

$$\text{First Look Score} = \frac{1}{R}\sum_{r} S_r$$

The Existence Score and Temporal Localization Score for each run are calculated from the time-series of predictions of LSI events, and they each take values between 0 and 1. In the following subsections, the run subscript is removed for clarity.

9.6.2.4.1 Existence Score

The Existence Score evaluates solutions' ability to predict which LSIs the patient receives within the prediction horizon at any time.

Let:

- $k \in \{1, 2, \ldots, K\}$ be an LSI type within the set of K LSI types
- $i \in \{1,2, \ldots, N\}$ be a case within the set of N cases in the test dataset
- $t \in \{1,2, \ldots, T_i\}$ be a time-bin within the set of T_i time-bins in case i
- $s_{i,k,t}$ be the confidence score that LSI k occurs in case i within time-bin t
- $y_{i,k}$ be 1 if at least one occurrence of LSI k occurs in case i, otherwise 0
- $AP(x_i, y_i)_{i=1}^{N}$ be the average precision between score x_i and label y_i for N cases indexed by i
- π_k be the prevalence of LSI k in the evaluation dataset

For cases $i = 1..N$ and LSIs $k = 1..K$, the maximum confidence score across time-bins is used as the prediction that LSI k occurs at least once in case i:

$$s_{i,k} = \max_t s_{i,k,t}$$

The average precision is then calculated for LSI k over N cases:

$$AP_k = AP\left(s_{i,k}, y_{i,k}\right)_{i=1}^{N}$$

The normalized average precision for LSI k accounts for differences in LSI prevalence:

$$\widehat{AP}_k = \frac{AP_k - \pi_k}{1 - \pi_k}$$

The Existence Score S_E is then the average of normalized average precision with zero-floor across LSI types:

$$S_E = \frac{1}{K} \sum_k \max\left(0, \widehat{AP}_k\right)$$

9.6.2.4.2 Temporal Localization Score

The Temporal Localization Score evaluates solutions' ability to predict when an LSIs occurs within a case, given that the LSI occurs at least once.

Let:

- $k \in \{1, 2, \ldots, K\}$ be an LSI type within the set of K LSI types
- $i \in \{1,2, \ldots, N_k\}$ be a case within the set of N_k cases in the test dataset that contain at least one occurrence of LSI k after the input data
- $t \in \{1,2, \ldots, T_i\}$ be a time-bin within the set of T_i time-bins in case i
- $j \in \{1,2, \ldots, M_{i,k}\}$ be an LSI event in case i of type k within the set of $M_{i,k}$ total occurrences
- $y_{i,k,t}$ be 1 if at least one occurrence of LSI k occurs within time-bin t in case i , otherwise 0
- $s_{i,k,t}$ be the confidence score that LSI k occurs in case i within time-bin t

For cases $i = 1..N$ and LSIs $k = 1..K$, rank the confidence score for all time-bins in case i for LSI k.

For LSI events of type k in case i, define a tolerance window $W_{i,k,j}$ for each event $j = 1..M_{i,k}$. The tolerance window will always include the time-bin containing the LSI event timestamp and may include

time-bins before and after to account for timestamp ground truth uncertainty. LSI events with overlapping tolerance windows are merged into a single event for scoring. Tolerance windows will include the preceding time bin for all LSI events. For Blood Product LSI events, tolerance windows will correspond to pre-hospital and one-hour intervals in-hospital (first hour, second hour, etc.) to account for the temporal granularity of Blood Products ground truth.

The rank of LSI event j is the rank of the highest-scoring bin within the time-tolerance window for the event:

$$r_{i,k,j} = \min_{t \in W_{i,k,j}} \text{rank}_{i,k}(t)$$

where $\text{rank}_{i,k}(t)$ is the rank of the confidence score for LSI k at time-bin t in case i, with rank 1 corresponding to the highest score among time-bins within the case.

The reciprocal rank (RR) for LSI k and case i is the average inverse rank of LSI events within the case:

$$RR_{i,k} = \frac{1}{M_{i,k}} \sum_{j} \frac{1}{r_{i,k,j}}$$

The mean reciprocal rank (MRR) is calculated across cases for LSI k:

$$MRR_k = \frac{1}{N_k} \sum_{i} RR_{i,k}$$

The normalized mean reciprocal rank adjusts based on a naïve baseline that has decreasing prediction scores over time (i.e., MRR_k^{prior} is the mean reciprocal of bin number when LSI *k* occurred):

$$\widehat{MRR}_k = \frac{MRR_k - MRR_k^{prior}}{1 - MRR_k^{prior}}$$

The Temporal Localization Score S_T is then the average of the mean reciprocal rank with zero-floor across LSIs:

$$S_T = \frac{1}{K} \sum_{k} \max\left(0, \widehat{MRR}_k\right)$$

9.6.3 Task 2: Continuous Alert

In this task, solutions use streaming in-hospital data to predict soon-to-occur LSI events within a sliding prediction horizon up to 4 hours after hospital admission. This task is similar in format to the Phase 2 Data Competition, with shorter data segments and a sliding, near-term prediction horizon. See Figure 4 for an illustration of this task.

Figure 4. Illustration of Task 2: Continuous Alert. From streaming data segments (in blue) predict presence of each LSI within near-term horizon (in red). Input data are sequential 30-second segments containing one or more of the following: EHR data, Casualty Report, prior interventions, and physiological signals. After each data segment is provided (rows in figure), model outputs confidence score between 0 and 1 for each LSI indicating existence within the subsequent 15-minute prediction horizon. Ground truth is illustrated by LSI event at top intersecting with prediction horizon. Note that boxes representing data segments and prediction horizon illustrate sequential ordering, but are not drawn to the same time-scale. For Blood Products, the prediction horizon is 60 min.

9.6.3.1 Input data

Input data for this task is provided in *data segments* containing timestamped health record and available physiological signals data within the segment start and stop timestamps. All timestamps will be expressed as time elapsed relative to start of in-hospital vitals.

Data segments include the following data types:

- Basic EHR: demographics, injury, GCS, initial EHR vitals, instantaneous labs (i.e., iSTAT) available at hospital admission
- Casualty Report: triage category, injury, and mental status derived from pre-hospital data (if available) akin to the Systems Competition
- Prior interventions: Interventions received prior to hospital admission, including LSIs
- Vitals segments: 30-second consecutive segments of available continuous waveforms and trends from in-hospital data
- Intervention segments: 30-second consecutive segments of interventions received from in-hospital EHR data, including LSIs, time-aligned with Vitals data segment

There is a single run for the Continuous Alert task. The input data is provided in sequential order for a given case, representing prior-to-admission information, at-admission information, and continuous monitoring after hospital admission:

Segment 1. Casualty Report.
Segment 2. Basic EHR, Prior interventions.
Segment 3. First Vitals data segment, first Interventions segment.
Segment 4. Second Vitals data segment, second Interventions segment.
Segment 5. etc.

Segments will not exceed 4 hours after hospital admission or the end of the case, whichever is earlier.

More details about data fields provided can be found in the Data Competition ICD.

9.6.3.2 Prediction target and horizon

Team solutions will produce LSI predictions as a set of confidence scores representing the existence of each LSI type in the 15-minute prediction horizon immediately following the vitals data in the input data segment. For the first two segments in the case (without vitals data), the prediction horizon includes LSIs that occur within the first 15 minutes in-hospital, including those received at-admission. For Blood Products only: the prediction horizon is 60 minutes (instead of 15 minutes) due to ground truth temporal resolution.

While solutions are expected to output predictions in response to every 30-second input segment, the evaluation only includes predictions in response to the first two segments (without vitals data) and predictions every 5 minutes after hospital admission to reduce skewed results due to overlapping prediction horizons. For Blood Products only: the evaluation includes the first two segments (without vitals data) and subsequent predictions every 60 minutes after hospital admission due to ground truth temporal resolution.

9.6.3.3 Scoring

The Continuous Alert task evaluates solutions' ability to predict occurrence of LSIs in the near-future from streaming data.

Let:

- $k \in \{1, 2, \dots, K\}$ be an LSI type within the set of K LSI types
- $i \in \{1,2, \dots, N\}$ be a case within the set of N cases in the test dataset
- $t \in \{1,2, \dots, T_i\}$ be prediction times within the set of T_i segments in case i
- $s_{i,k,t}$ be the confidence score that LSI k occurs in case i within the prediction horizon after segment t
- $y_{i,k,t}$ be 1 if at least one occurrence of LSI k occurs in case i within the prediction horizon after segment t, otherwise 0
- $AP((x_i, y_i) : i \in W)$ be the average precision between score x_i and label y_i for samples in set W
- $q \in \{1,2, \dots, Q\}$ be time strata for the evaluation. Preliminary time strata are: 0–30 minutes, 30–60 minutes, 1–2 hour, 2–4 hour. For Blood Products only: time strata are each hour after hospital admission.
- W_q be the set of predictions that occur within time strata q
- $\pi_{k,q}$ be the prevalence of LSI k within time strata q in the evaluation dataset

For cases $i = 1..N$ and LSIs $k = 1..K$, pool predictions within time strata and calculate average precision across pooled predictions:

$$AP_{k,q} = AP\big((s_{i,k,t}, y_{i,k,t}) : t \in W_q\big)$$

The normalized average precision for LSI k accounts for differences in LSI prevalence in time strata q:

$$\widehat{AP}_{k,q} = \frac{AP_{k,q} - \pi_{k,q}}{1 - \pi_{k,q}}$$

The task score is then the average of normalized average precision with zero-floor across time strata and across LSIs:

$$\text{Continuous Alert Score} = \frac{1}{K}\sum_{k}\frac{1}{Q}\sum_{q} \max(0, \widehat{AP}_{k,q})$$

9.6.4 Task 3: Resource Allocation

In this task, solutions assign available medical resources to casualties within a simulated mass casualty event. Solutions are presented with initial pre-hospital data from multiple cases simultaneously, along with available medical resources. Solutions must then determine which patients need which resources most, and allocate those resources in order to maximize the number of patients that receive the care they need. Solutions will be evaluated under different scenarios and within varying resource settings, as described in the next section. See Figure 5 for an illustration of this task.

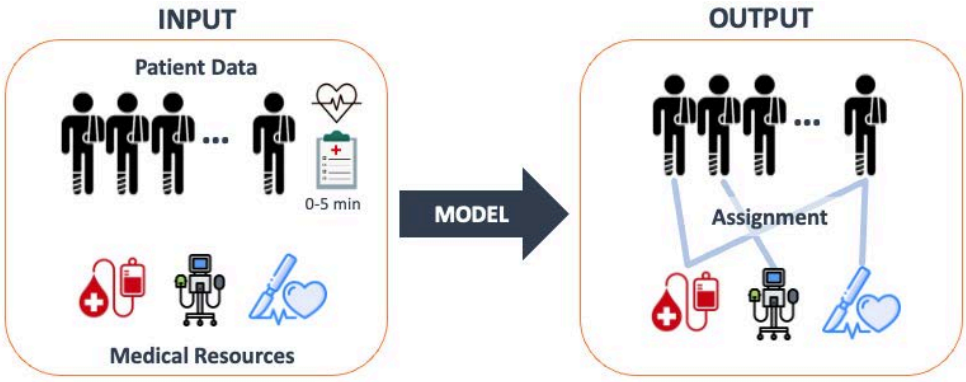

Figure 5. Illustration of Task 3: Resource Allocation. Based on a group of patients and available medical resources in a simulated mass casualty event, solutions assign resources to patients in order to maximize "Lives Saved" within the scenario.

9.6.4.1 Assumptions

Solutions will output an assignment of available medical resources to individual patients within the provided patient population. Medical resources needed by each patient are determined from historical patient data, where medical resources are mapped to LSIs. Because patient response to treatments other than what they received is unknowable, this task employs several assumptions for purposes of simulation:

1. Resources received by the patient in the historical dataset are treated as a proxy for resources needed for survival.
2. Patients are considered "saved" only if all required resources are assigned.
3. Patients assigned resources that they do not need are not adversely impacted; however, the over-assignment of resources consumes limited quantity that may otherwise benefit other patients.
4. Patients who expired during care in the dataset are treated as non-survivable within the simulation, assuming that care in a resource-constrained mass casualty setting would not exceed the care received in the original clinical setting.

These assumptions are reflected both in the ground truth patient data ("resources needed") and in the scoring approach, as described in the following sections.

9.6.4.2 Input data

Input data is comprised of both Patient Data and Medical Resources:

Patient Data is a list of individual case data representing multiple casualties needing care simultaneously within a simulated mass casualty event. Individual case data will be similar in

format to the input data used in the First Look Run 2 task (see Section 9.6.2.1) with the following data types: casualty report, previously received LSIs, and initial pre-hospital vitals signals and trends data.

Medical Resources is a list of resources available for allocation to patients within the scenario. Each Medical Resource is mapped to one or more LSI groups used in Tasks 1 and 2, as described in more detail in the next section. Generally, units represent the number of patients that can be supported locally within the resource setting. The resource types are:

- *Evacuation*. Movement of 1 patient to a high-resource setting without resource constraints where any and all resources are available. Patients assigned this resource do not consume any local resources below.
- *Hospital Bed*. Baseline staffing, monitoring, space, and infrastructure required to provide any definitive local medical care to 1 patient. Any patients treated locally who received at least one LSI will need at minimum this resource (see section 9.6.4.5 for details on resources needed).
- *Surgery*. Specific staffing, equipment, and space required for 1 patient to receive surgical care.
- *Blood*. Abstracted unit of blood products, defined as 1 unit of packed red blood cells (PRBC) or 1 unit of whole blood (WB). Unlike other resources, multiple units may be assigned to a single patient (see Section 9.6.4.4 for details on resource assignment and consumption).
- *Ventilator*. Specific equipment required for 1 patient to receive an invasive airway.

Details on specific data formats can be found in the Data Competition ICD.

9.6.4.3 Scenarios

To examine model performance across a variety of simulated mass casualty settings, solutions will be evaluated under multiple scenarios within the Resource Allocation task. Each scenario will constitute a different patient population (i.e., injury distribution) and resource setting (i.e., quantity of medical resources available).

While exact characteristics of patient populations and resource settings will not be provided prior to evaluation, the following storylines will drive design of evaluation scenarios: *Active Shooter* (biased toward penetrating injuries) and *Building Collapse* (biased toward blunt injuries). Scenarios for both storylines will be composed of 50–250 patients. Patients will not appear within the same scenario twice; however, the same patient may appear in multiple scenarios. Resource settings are categorized as *High*, *Medium*, and *Low*. See Table 6 for ranges of resources available within each resource setting.

Resource Type	High	Medium	Low
Hospital Bed	90-120	40-60	8-15
Surgery	10-15	5-10	0
Blood	80-200	40-100	0-5
Ventilator	25-50	10-20	0-5
Evacuation	15-30	5-10	0-5

Table 6. Resources settings (High, Medium, Low) with ranges for each resource type.

9.6.4.4 Model output and resource consumption

In this task, solutions assign available resources to patients within the provided patient population. Patients

are first assigned to one of three distinct groups:

1. Patients to be evacuated.
2. Patients to receive at least one medical resource locally.
3. Patients who do not receive any medical resources.

Patients in Group 1 consume one *Evacuation* resource and no other medical resources, patients in Group 2 consume one *Hospital Bed* resource, and patients in Group 3 do not consume any resources. Within Group 2, patients can then be assigned any of these additional resources: *Surgery*, *Ventilation*, *Blood-Low*, and *Blood-High*. *Surgery* and *Ventilation* each consume one resource of the same name per patient. *Blood-Low* and *Blood-High* represent low and high transfusion burden per patient; *Blood-Low* consumes 1 *Blood* unit and *Blood-High* consumes 3 *Blood* units. Patients may not be assigned multiple instances of the same resource.

9.6.4.5 Life-saving interventions and medical resources needed

Medical resources needed serves as the ground truth for scoring, and will be provided in future data release for Task 3. Resources assigned and resources needed determines patient outcome (survival or death) within the simulation. Medical resources needed is derived from the same life-saving interventions used in Tasks 1 and 2. Within the assumptions described above, medical resources needed indicate the minimal set of assigned resources required in order for the patient to survive. Patients who did not receive any LSIs do not need any medical resources.

The following table shows how the life-saving intervention received by a patient determine the medical resources needed within Task 3.

Life-Saving Intervention (LSI)	Medical Resources
Invasive Airway	Hospital Bed, Ventilator
Blood Products (1-2 units PRBC and/or WB)	Hospital Bed, Blood-Low
Blood Products (3+ units PRBC and/or WB)	Hospital Bed, Blood-High
Chest Decompression	Hospital Bed
Surgical	Hospital Bed, Surgery
Vaso/Cardioactive Medications	Hospital Bed

Table 7. Mapping of LSIs from Tasks 1-2 to Medical Resources in Task 3. Blood Products LSI is unique in that it maps to two different medical resources (Blood-High and Blood-Low) depending on total volume of Packed Red Blood Cells (PRBC) and/or Whole Blood (WB) received.

Resources needed by patients who received multiple LSIs is the union of mapped resources in the table above: for example, a patient who received one Invasive Airway LSI and two Surgery LSIs needs the following medical resources: *Hospital Bed, Ventilator, Surgery*. Blood Products is a special case: for patients who received one or more Blood Product LSI(s), they either need the resource *Blood-Low* or *Blood-High* depending on the total units of Packed Red Blood Cells (PRBC) and Whole Blood (WB) received. The same patient cannot need both *Blood-Low* and *Blood-High* resources.

An exception to the above mapping based on Assumption 4 in Section 9.6.4.1: patients who expired during care in the dataset do not need any medical resources, regardless of whether they received any LSIs in the original clinical setting. These patients are treated as non-survivable within the simulation.

Medical resources needed by each patient will be provided in the next data release.

9.6.4.6 Scoring

The Resource Allocation task evaluates solutions' ability to save lives within a simulated mass casualty event by optimizing assignment of available medical resources to patients based on predicted need. The Lives Saved score is defined as the proportion of patients within the scenario that needed at least one medical resource and were assigned all medical resources needed.

There are two special cases for determining Lives Saved in terms of resources needed and consumed:

- A Patient selected for *Evacuation* receives any and all resources needed (i.e., it is a "wildcard" resource), and they count towards Lives Saved if they needed at least one medical resource.
- For a patient who needs the *Blood-Low* resource (1 unit consumed), assignment of *Blood-High* (3 units consumed) satisfies the blood resources needed by that patient for survival. However, the converse is not true.

Based on the assumptions described in Section 9.6.4.1, the following rules apply to the calculation of the Lives Saved score for this task:

1. Patients must be assigned all resources needed in order to contribute to the Lives Saved score, unless they are assigned the *Evacuation* resource (see next rule).
2. Patients who are assigned the *Evacuation* resource contribute to the Lives Saved score if and only if they needed at least one medical resource.
3. Patients who do not need any medical resources (i.e., they did not receive any LSIs or they expired during care) do not contribute to the Lives Saved score, regardless of resources assigned.
4. Resources assigned to patients who do not need them are still consumed from remaining resources, and they have no impact on the Lives Saved score.

Within a scenario, the ordering of assignments represents prioritization by the solution under resource constraints. Resources are consumed sequentially in the order assignments are provided. Resources assignments must not exceed the number of each resource available within the scenario. See the Data Competition ICD for more details on how the response format is structured and interpreted for scoring.

The Lives Saved Score for a given scenario is the ratio of number of "saved" lives to the number of "savable" lives:

$$L = \frac{E + S}{T}$$

where L is the Lives Saved Score, E is the number of evacuated patients who needed at least one medical resource, S is the number of remaining patients who needed at least one medical resource and were assigned all medical resources needed, and T is the maximum number of patients needing at least one medical resource who could be assigned all medical resources needed (i.e., the maximum possible "saveable" given the resource limitations). This score is calculated for each scenario, and the Resource Allocation score is then the average of scores across scenarios:

$$\text{Resource Allocation Score} = \frac{1}{R}\sum_{r} L_r$$

where R is the total number of scenarios and L_r is the Lives Saved score for scenario r.

9.6.5 Event Score

The Event Score is the weighted sum of task scores:

$$\text{Event Score} = \alpha \cdot \text{First Look Score} + \beta \cdot \text{Continuous Alert Score} + \gamma \cdot \text{Resource Allocation Score}$$

where α, β, γ are task-specific weights with the following values: $\alpha = 0.4, \beta = 0.4, \text{and } \gamma = 0.2$.

9.6.6 Final Ranking

The final ranking will be determined by decreasing Event Score. In the event of a tie, the team with the shortest evaluation run-time will be ranked higher.

9.6.7 Minimum Benchmarks to Win Prizes

In Phase 3 there is a minimum benchmark for winning prizes. Teams must achieve each of the below performance requirements:

- Exceed at least 75% of the Event Score of baseline algorithms provided by the DTC IV&V team.

10 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).

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.