



# PRIZE<sup>®</sup>

## CHALLENGE GUIDELINES

January 25, 2020

The Pandemic Response Challenge is governed by these Competition Guidelines. The Competition Guidelines summarize the requirements and rules of the challenge.

XPRIZE may revise these Guidelines during the course of the challenge to provide additional information or to improve the quality of the challenge. Unanticipated issues that may also arise that will require modifications to these Guidelines. XPRIZE reserves the right to revise these Guidelines as it, in its sole discretion, deems necessary or desirable. All registered teams will be notified of any revisions promptly.

For further details concerning operating the challenge, please refer to the Competitor Agreement and other documents throughout the course of the challenge.

Please send questions about this challenge and/or feedback regarding the Competition Guidelines to [pandemicresponse@xprize.org](mailto:pandemicresponse@xprize.org).

# TABLE OF CONTENTS

I. Competition Overview	03
II. Competition Timeline & Registration	04
III. Competition Phases, Testing, And Judging	07
IV. Prizes	17
V. Roles and Responsibilities	18
VI. Environment and Safety	20
VII. Glossary	21

## I. COMPETITION OVERVIEW

COVID-19 has become one of the world's most critical challenges—greater than any government or organization can tackle in isolation. Countries around the world are struggling to implement health and safety interventions and policies that protect their citizens and economies.

This pandemic requires access to localized, data-driven planning systems combined with cutting edge artificial intelligence tools to help decision-makers develop and implement intervention plans that reduce infection cases, minimize negative economic impacts, and reopen their economies and societies.

XPRIZE and Cognizant have partnered to unlock the technological breakthroughs needed to make this possible by launching the Pandemic Response Challenge. The Pandemic Response Challenge tasks participants with developing models that predict local outbreaks with more accuracy, along with prescriptive intervention and mitigation approaches that minimize infection cases and economic costs. The Challenge's open platform will enable increased and higher-quality data, accurate predictions, stronger regional intervention plans, and continual improvement as new interventions such as vaccinations and treatments become available. It will provide a platform for shared human and machine creativity and problem-solving, and ultimately a tool for future humanitarian crises.

Join us in building a collaborative AI for Good ecosystem that fosters innovative, evidence-based decision-making to combat COVID-19 and future emergencies.

**Two winning teams will produce prediction models that estimate future numbers of daily COVID-19 cases with the greatest accuracy and will create the best prescription models for intervention plans. The winning teams will split a total of \$500,000.**

## II. COMPETITION TIMELINE AND REGISTRATION

This section provides information on the proposed timeline of the competition, registration process, and registration requirements. XPRIZE will provide real time updates to the guidelines for this challenge.

**Table One: Challenge Timeline**

Date	Item
Oct. 30	<i>Soft Launch:</i> Teams can log in to the Prize Operations Platform (POP). While in POP, teams will have access to the guidelines, a GitHub link ( <a href="https://github.com/leaf-ai/covid-xprize/">https://github.com/leaf-ai/covid-xprize/</a> ) to access example predictors, prescriptors, and data, the competitor agreement, and access to a Slack channel. We encourage teams to start working on predictor and prescriptor models right away. While in POP teams can also proceed with team registration and fee payment.
Nov. 17	<i>Official Launch:</i> This is the official public launch. Team registration continues until December 8, 2020. Shortly thereafter, teams will be provided with evaluation sandboxes.
Dec. 8	<i>Registration Close</i>
Oct. 30 – Dec. 22	<i>Phase One—Predictor Development:</i> Teams develop, train, and finalize their predictor models and load them into their evaluation sandbox. Teams are also able to start work on their prescriptors during this phase. Teams can upload newer versions of their model into the sandbox up to the Dec. 22 deadline.
Dec. 22 – Jan. 12 (with ongoing testing until Feb 25)	<i>Phase One Live Model Testing:</i> Starting December 22, teams will no longer be able to access their evaluation sandbox and their final models will be run automatically on a recurring basis. Results will be displayed on the leaderboard on POP.
Jan 6 – Jan 12	<i>Phase One—Intermediate Predictor Judging:</i> Qualitative and Quantitative judging
Jan. 13	<i>50 Finalists Announced:</i> Finalists are provided access to prescriptor evaluation sandboxes

Oct 30 — Feb. 3	<i>Phase Two—Prescriptor Development:</i> Teams develop, train, and finalize their prescriptor models and load them into their prescriptor evaluation sandbox. Teams are encouraged to start their prescriptor development as soon as phase one if possible. Teams can upload newer versions of their prescriptor model into the sandbox up to the Feb. 3 deadline.
Feb. 3 — Feb. 25	<i>Phase Two—Final Predictor and Prescriptor Judging:</i> Starting February 3, teams will no longer be able to access their prescriptor evaluation sandbox and their models will be run automatically on a recurring basis. Results will be displayed on the leaderboard on POP.
Feb. 26	Winners Announced

*Note: The above dates are subject to change.*

## 1. TEAM REGISTRATION

XPRIZE believes that Solutions can come from anyone, anywhere. Scientists, engineers, academics, entrepreneurs, and other innovators from all over the world are invited to form a team and register to compete. Teams will participate in the competition through XPRIZE Competition Sites. The Competition Sites consist of the Prize Landing Webpage, the Prize Operations Platform (POP), a Slack workspace, the XPRIZE Data Collaborative, and the team Sandboxes.

To participate, a team must first create an account in the [Prize Operations Platform](#). POP is an online platform through which teams will register for the challenge, pay the required registration fee, and submit and download important documents throughout the challenge. Teams are expected to maintain their POP profiles throughout the challenge, ensuring their profile is up to date with the most recent team information.

An XPRIZE competition is an exciting journey requiring a commitment of time, expertise, and resources. Registration fees are required. The \$100 fee collected is used to support XPRIZE challenge teams.

On October 30, 2020, the POP Platform opens and registration is available. After creating a login for POP, teams can access the Competition Guidelines and **find the competition repository at:** <https://github.com/leaf-ai/covid-xprize> with case and intervention plan (IP) data, a sample dataset, and example predictors and prescriptors. We encourage teams to first examine the sample predictors and prescriptors to determine whether they can contribute unique or better models. As

part of the full registration process, teams must certify they meet a basic level of technical competency by completing a competency assessment on POP.

While in POP, teams can also access the competitor agreement and a Slack channel to connect with other teams, ask technical and competition related questions, seek support, and receive challenge updates. Teams can also complete the full registration process from POP.

To fully register, teams must complete a competency assessment, registration forms, sign the competitor agreements, and pay the \$100 registration fee. The challenge evaluation sandbox (on the XPRIZE Data Collaborative) will be available to teams shortly after the public launch of the Challenge. XPRIZE will notify fully registered teams when the evaluation sandboxes become available.

The Pandemic Response Challenge is limited to the ***first 200 fully registered teams.***

Interested teams and individuals are encouraged to collaborate and share skills. A team may recruit additional experts and can add new members to their team throughout the challenge. Teams may also merge with other teams during the challenge. The Slack channel can be used for this. Teams must notify XPRIZE of a merger before it takes place. Additional details regarding team mergers are provided in the Competitor Agreement.

After the official launch date, XPRIZE will host webinars for all registered teams. XPRIZE webinars will allow teams to get to know each other and also to receive important challenge updates. Participation in these webinars, while not mandatory, is strongly encouraged.

An interactive demo illustrating the basic concepts of the competition, such as the predictor and prescriptor models and their interaction and performance, is available at <https://evolution.ml/demos/npidashboard>.

## 2. COMPETITOR AGREEMENT

Each team must complete all fields of the XPRIZE Competitor Agreement and submit all information required by XPRIZE within the application fields to be considered a successful registrant. Teams are required to comply with all applicable laws and acquire all necessary licenses, waivers, and/or permits from the applicable regulatory bodies or other applicable third parties.

Model validation results, scoring, statistics, judging, and other data and insights derived from XPRIZE teams' predictor and prescriptor models collected by XPRIZE teams during the operation of the Competition are the intellectual property of XPRIZE. XPRIZE retains the right to license such data for academic, research, and other purposes. Sample data, reference predictor and prescriptor models, and judging software provided by the sponsor to XPRIZE and/or teams as part of the execution of the Competition are the intellectual property of XPRIZE for the benefit of the

public. Teams agree that their models (predictors and prescriptors) once submitted to the Competition Sites are deemed open source and can be made available publicly under XPRIZE's purview for the benefit of the public. Teams warrant that they will not upload or process personally identifying information, including protected health information, to the Competition Sites.

Each team must designate a "Team Leader." The Team Leader will be responsible for receiving communications from and communicating with XPRIZE and the Judging Panel. The team leader shall be at least 18 years old (or the age of majority in their jurisdiction of residence). A team leader is designated in POP.

### III. COMPETITION PHASES, TESTING, AND JUDGING

To be eligible for any Prize Purse during the challenge, teams must meet or exceed all evaluation criteria established by XPRIZE and the independent Judging Panel. Examples of the submission requirements and evaluation criteria for each phase of the competition are summarized in the next sections of this document. The examples below do not encompass all the competition criteria considered. They provide an overview of how teams can expect they may be evaluated. The judges have full decision making authority and can add or change criteria at their discretion.

#### 1. PHASE ONE—PREDICTOR DEVELOPMENT OVERVIEW

The goal of Phase One—Predictor Development is to provide accurate, localized predictions of COVID-19 transmission based on local data, unique intervention strategies, community resilience characteristics, and mitigation policies and practices.

Upon registration, teams will gain access to case and intervention plan data, a sample dataset, and example predictors (that are not region specific) in the previously described GitHub repository. The example predictors include a linear regressor and Cognizant's LSTM-based predictor network. The COVID-19 data will be provided by Oxford University Blavatnik School of Government's COVID-19 Government Response Tracker.<sup>1</sup> Example intervention plans include school and workplace closure policies, travel restrictions, testing, and contact tracing. Teams are encouraged to work with the existing data and predictors and develop new predictors. XPRIZE encourages creativity and innovation; thus, teams may use other data sources and other methods to build better predictors. For instance, teams can add and use additional demographic or economic data, predict other variables such as deaths (which are included in Oxford University Blavatnik School of Government's data), or combine data-driven models with epidemiological models.

We encourage teams to submit novel or unique datasets which may be incorporated into Oxford University Blavatnik School of Government's COVID-19 Government Response Tracker. This can

---

<sup>1</sup> Hale, Thomas, Noam Angrist, Emily CameronBlake, Laura Hallas, Beatriz Kira, Saptarshi Majumdar, Anna Petherick, Toby Phillips, Helen Tatlow, and Samuel Webster. 2020. Oxford COVID-19 Government Response Tracker, Blavatnik School of Government. Available: [www.bsg.ox.ac.uk/covidtracker](http://www.bsg.ox.ac.uk/covidtracker)

include data on demographics, economics, healthcare factors, social distancing, adherence to policies, and more. Teams can upload datasets to the evaluation sandbox, but these datasets cannot be updated during the evaluation and live model testing phases. Providing reliable, novel data will be considered during the judging process and teams could be awarded bonus points. Shortly after the public launch of the Challenge, teams will be contacted by XPRIZE with instructions on how to access their evaluation sandbox. This sandbox will be used to evaluate the predictor models prior to and throughout the evaluation period.

XPRIZE recommends that teams, at a minimum, review the sample predictor models to ensure they have the level of competency needed to generate a competitive predictor. Before submitting the registration fee, teams will certify that they meet a basic level of tech competency (such as familiarity with Python) needed to complete the challenge.

At the conclusion of the Phase One—Predictor Development, teams will submit their predictors for evaluation with live (and hence unseen) data over a three-week intermediate evaluation period. Submissions must be complete and functional. Teams cannot modify their predictors after the submission deadline and teams' predictors will no longer have access to the internet or external resources during the evaluation. Judges will use this three-week evaluation and live model testing period—bringing together quantitative and qualitative judging criteria—to determine which teams advance to Phase Two—Prescriptor Development. The best predictors from Phase One will be combined to form an enhanced predictor to evaluate the prescriptors in Phase Two.

At the conclusion of the three-week Phase One Live Model Testing, XPRIZE will continue to run the predictors submitted by the finalists until the end of Phase Two Judging to evaluate the long-term accuracy of these models. The long-term predictor results will be used as part of the judging criteria to determine winners at the end of the challenge.

## 2. PHASE ONE—PREDICTOR SUBMISSION REQUIREMENTS

A team's submission will comprise:

- Team's model in a compliant API;
- Data used to run the model;
- A description of the approach taken in developing the model which should also address innovation, generality, collaboration, and other qualitative judging criteria (submitted via POP); and
- Optionally, teams can highlight the list of "specialty regions" they would like judges to consider for their model. These regions are the focus of a team's predictor model beyond the general evaluation. In these regions, their performance will be measured and judged separately. (**Note:** please refer to the Oxford dataset for a list of all available regions.)

Teams will submit their predictor models to their assigned evaluation sandbox prior to December 22. Teams are highly encouraged to submit and execute their models in this environment well prior



to this date in order to address any technical issues associated with using this environment. Predictor models in the sandbox can access only local files. No internet access will be available to the sandboxes after December 22. Descriptive documentation should be submitted to POP.

A predictor is accessed through a script in the evaluation sandbox. The predictor must be called with a single command with the following **exact** syntax and arguments:

```
python predict.py -s start_date -e end_date -ip path_to_ip_file -o path_to_output_file
```

This call should write a CSV file to *path\_to\_output\_file* containing the predictions for the daily cases between *start\_date* and *end\_date*, included, for each country and region for which *path\_to\_ip\_file* contains an intervention plan. The CSV file should contain:

- One row per day per region for which an intervention plan was supplied;
- Required Columns: Date, CountryName, RegionName, PredictedDailyNewCases; and
- Optional Columns: Teams may produce additional columns as output of their predictor models in the CSV file. These columns will be noted by the judges but not evaluated by the Robo Judge. Example optional columns could be:
  - A column labeled *IsSpecialty* to indicate whether a region is to be considered a speciality region for your model (1 = speciality region, 0 = not a speciality region)
  - A 95% confidence interval and standard deviation for predicted number of cases
  - Predicted number of deaths and related 95% confidence intervals
  - Predicted number of hospitalization rates and related 95% confidence intervals
  - Predicted number of ventilators needed and related 95% confidence intervals
  - Other columns chosen by the team

A sample CSV output file can be found [here](#) on the GitHub.

To judge the generality of the predictors, all predictors will take as input the active and historical intervention plans for each region and will need to output a prediction for all regions<sup>2</sup>. Performance on specialty regions is evaluated based on output on those regions. A predictor submission can consist of multiple models, for example those specializing in different regions, that are accessed through the same call. When called in the evaluation sandbox, a predictor must return a prediction in less than 1 hour for up to 180-days of prediction for up to 300 regions. Teams are highly encouraged to use the scenario generators found on the GitHub repository to try out different evaluation scenarios before submitting their models. Instructions on how to use these scenario generators to test models can be found within the sample models provided by Cognizant on the GitHub repository.

---

<sup>2</sup> Note: the historical cases are not explicitly an input to the predictor. The predictor can, however, save and access the historical case data up to the starting point of the evaluation in the evaluation sandbox work folder. It can then use its own predictions in lieu of actual cases for the active evaluation period. In this manner, its predictions can be based on parallel time series of case history and intervention plan history up to the current point in time.

### 3. PHASE ONE LIVE MODEL TESTING

Upon the submission deadline on December 22, 2020, access to the evaluation sandboxes will be disabled. No updates will be allowed after this date. We will conduct a progressive quantitative evaluation of each model's output. Starting December 23, the predictor will be called on a recurring basis using the live and historical intervention plan data for each region from the Oxford dataset to generate predictions up to 180-days into the future as specified above. The subsequent days' predictions will be used to evaluate the model's quantitative accuracy against the actual (live) data. The long-term predictions will be used to illustrate how the number of cases would develop, according to this predictor, if the current IPs were held constant across other scenarios in any scenario analyses run by the Judging Panel.

During the Phase One Live Model Testing, XPRIZE will maintain a publicly available leaderboard with team results and standings on POP.

### 4. PHASE ONE— PREDICTOR JUDGING

At the start of the competition, XPRIZE will convene an independent Judging Panel. The Judges' role is to evaluate and rank team submissions and results and uphold the validity and integrity of the prize process. Judges enter into an agreement with XPRIZE obligating them to comply with all terms and conditions, including confidentiality and conflict of interest provisions as described in the Competitor Agreement. Judges also acknowledge that they shall make no claim to a Team's intellectual property.

For Phase One—Predictor Development, teams will develop and submit predictor models that estimate the number of future cases for a given region(s)—considering the local intervention plans in effect based on live Oxford data—over a given time. The data (estimates) each model generates will be evaluated by a Cognizant-designed “Robo Judge”—a tool that compares teams' prediction data against real time data and the other teams' results. This tool is being made available to Judges specifically for the quantitative evaluation of results. First, the Judging Panel will use the Robo Judge results to eliminate the most inaccurate teams from the competition. Next, the Judging Panel will evaluate the results (data) that each model produces quantitatively via the leaderboard rankings produced by the Robo Judge and qualitatively through a scorecard rubric consisting of the criteria described below. Teams are evaluated against, and compared to, other teams on their own merits. A maximum of 50 teams with the best predictor models will advance to Phase Two of the Pandemic Response Challenge.

#### **Quantitative Evaluation Criteria**

Teams must copy all necessary models and data into their evaluation sandbox and verify that it can be executed per the previously mentioned command syntax by December 22. The leaderboard will report the predictor results regularly. The predictor submissions will be evaluated on all regions and

then separately on the speciality regions. Judges may also cumulatively evaluate submissions on larger regions, such as specific countries, continents, and the world.

At the conclusion of the Phase One Live Model Testing, the data from submissions will be ranked in each region according to the cumulative error in the 7-day moving average for the number of cases per 100,000 people.

Based on such region-specific rankings, two overall performance measures will be formed. These are the:

- Mean ranking of teams across all regions
- Mean ranking of teams across the specialty regions, if selected

In their predictions, teams are encouraged to produce interesting results and show them. Judges will also consider any additional quantitative data that teams can provide. For example, whether predictor outputs include optional fields, such as confidence intervals, death rates, hospitalization rates, ventilators needed, and other outputs. Since death rates are available in Oxford University's Blavatnik School of Government's data, and they can be predicted like cases, teams may find it helpful to predict death rates as well as additional sources of information. If enough teams do this, the accuracy of death-rate predictions may be measured similarly to cases. And the rankings can be presented to the judges as optional quantitative information.

As a reminder to teams, and as explained in the GitHub repository, only a subset of the NPIs found in the Oxford dataset will be used in the evaluation of the predictor models. Specifically, only the NPIs that have a direct impact on the spread of the virus (i.e. on the daily new cases number) and are identified as "Containment and health index" in the Oxford Index Methodology document (found [here](#)) will be used. These NPIs are C1 to C8 (inclusive) and H1, H2, H3, and H6. Please refer to the sample code found on the GitHub repository for more examples of these indices and how they are implemented in the predictor submission template.

Teams will have their work subject to the following sanity checks:

- Ranking on retrospective runs on historical intervals, representing a broader range of situations than encountered in live testing
- Other predictor sanity check pass/fail results (e.g., negative predictions, maximal and minimal stringency predictions, and predictions exceeding population size)

The quantitative evaluation will be used by the Judging Panel to eliminate the most inaccurate models before the qualitative evaluation phase. Up to 100 teams may be eliminated in this phase, but the exact number will depend on the total number of registered teams and the distribution of accuracy among the all submitted models.

### **Qualitative Evaluation Criteria**

After the preliminary quantitative evaluation phase described above, the Judging Panel will use a combination of both quantitative and qualitative evaluation criteria to evaluate the teams. The

qualitative judging of the teams' models will begin during the predictor evaluation period. For the qualitative evaluation, the Judging Panel will use the following criteria with a generally equal weighting:

- **Innovation:** Teams who submit and use additional data, intervention plans (such as vaccination policies and treatments), or otherwise find innovative ways to extend the scope of the challenge will be ranked highly;
- **Generality:** Teams will first be evaluated on how well their models perform across all regions. Subsequently, teams will be awarded bonus points for how well their models do in specialty regions;
- **Collaborative contributions:** Teams that take an open-source approach to the data or models that they use, and who contribute data and models to the shared success of all teams will be ranked highly;
- **Consistency:** Approaches that stay within an acceptable range of accuracy in the short and long term, and that perform as expected in any scenario analyses run by the Judging Panel, are preferred;
- **Speed and resource use:** Model that are faster and more efficient in their approach are preferred;
- **Addressing the challenge:** Teams must avoid taking shortcuts or finding loopholes to improve their quantitative performance at the expense of real-world performance. Additionally, teams may be awarded bonus points for predicting additional, relevant public health metrics such as required hospital beds and ventilators; and
- **Explanation:** Submissions should include a narrative description of how the model works, the data it uses, and its sources as well as any relevant points related to these themes. Furthermore, models that emphasize interpretability by being able to explain why the model is predicting what it does (i.e. glass-box models) will be ranked highly.

At the conclusion of Phase One—Predictor Judging, each teams' qualitative and quantitative scores will be combined and teams will be ranked by the Judging Panel. Up to 50 teams will advance to Phase Two—Prescriptors.

## 5. PHASE TWO—PRESCRIPTOR DEVELOPMENT OVERVIEW

The main aim of the Pandemic Response Challenge is to foster an ecosystem that makes it easier to implement accurate and rapid prescriptions and enable ongoing improvements to the model as new interventions, treatments, and vaccinations become available. Prescriptor development encompasses the rapid creation of custom, non-pharmaceutical and other intervention plan prescriptions and mitigation models to help decision-makers minimize COVID-19 infection cases while lessening economic and other negative implications of the virus. For example, machine-generated prescriptions may provide policymakers and public health officials with

actionable locally-based, customized, and least restrictive intervention recommendations, such as mandatory masks and reduced restaurant capacity.

During Phase Two—Prescriptors, teams are to use machine learning to make more accurate recommendations to stakeholders. Based on a time sequence of the number of cases in a region and the past intervention plans in place, teams will develop prescription models (for any region) that generate useful intervention plans that policy makers can implement for their region. Each prescriptor balances a tradeoff between two objectives: minimizing the number of daily COVID-19 cases while minimizing the stringency of the recommended interventions (as a proxy to their economic and quality-of-life cost).

Intervention plan costs can differ across regions. For example, closing public transportation may be much costlier in London than it is in Los Angeles. Such preferences are expressed as weights associated with each intervention plan dimension, given to the prescriptor as input for each region. The prescriptor recommendations along with the stringency objective will be evaluated according to these weights, so the prescriptor model should consider them in making recommendations. This is a significant aspect of the competition for two reasons: (1) such prescriptors can be more readily customized to a particular region for future live site testing that may occur, making it easier to adopt them, and (2) this is a new technical challenge beyond the current state of the art, promoting scientific advances in machine learning. Prescriptors will be evaluated separately both in the base case of equal weights and in the more advanced case where the weights are chosen randomly. Especially in early development of the prescriptors, it may be useful to use equal weights, thus simplifying the challenge and extending to the general case later in the development process.

During Phase Two—Prescriptor Development, it is not possible to evaluate the prescriptors with live data (i.e., implement the recommendations in the real world). Instead, the recommendations will be evaluated using a standard predictor model which may use some of the best predictors developed by the teams in Phase One. From the start of the challenge, teams will have access to this standard predictor and a collection of neural networks to represent the different tradeoffs between COVID-19 cases and the stringency of the intervention plan. These will be made available in the same GitHub repository mentioned previously. The provided prescriptors are general and not specific to any region. We encourage teams to improve upon or develop better prescriptor models, either general or region-specific, using their choice of any machine learning or other methods.

Teams can generate prescriptions through a variety of approaches. A possible approach may involve the following: A prescription is generated for each day, and a predictor is asked to predict the number of cases for the next day. The generated IPs and the predicted cases then become input to the prescriptor and the predictor for the next day. In this manner, the prescriptions can be rolled out day-by-day indefinitely into the future. Another possible prescriptor approach could generate a schedule of intervention plans over several days, weeks, or months based on the case and intervention plan history up to that point, and only consult the predictor occasionally. The teams can use their own predictor or the ones provided to them. An API will be provided to the standard predictor for this purpose since it is used in the quantitative evaluations.

Teams can begin developing their prescriptors at any time. Example prescriptor will be provided in the GitHub repository. When the standard predictor becomes available, teams will have three weeks to adapt and refine their prescriptor models with it. Therefore, teams are encouraged to start working on the prescriptor before the standard predictor is available.

## 6. PHASE TWO—PRESCRIPTOR SUBMISSION REQUIREMENTS

At the end of Phase Two—Prescriptor Development, teams will submit:

- Team’s prescriptor model in a compliant API;
- Data used to run the prescriptor model;
- A description of the approach taken in developing the model which should also address innovation, generality, collaboration, and other qualitative judging criteria (submitted via POP); and
- Optionally, teams can highlight the list of “specialty regions” they would like judges to consider for their prescriptor model. These regions are the focus of a team’s prescriptor model beyond the general evaluation. In these regions, their performance will be measured and judged separately.

Similar to predictors, a prescriptor can only access local resources and will not be able to access the internet or be updated once evaluation begins.

The prescriptor must be called with a single command with the following **exact** syntax and arguments:

```
python prescribe.py -s start_date -e end_date -ip path_to_historical_ip_file -c  
path_to_weights_file -o path_to_output_file
```

Where weights is a CSV file specifying the stringency objective weights for each country: CountryName, RegionName, and one column per intervention. This command should write a CSV file containing the daily intervention plans for up to 90 days. The file should contain:

- One row per day per region;
- The required columns: Date, CountryName, RegionName, prescribed intervention plans in the same format as the historical ones; and
- A PrescriptionIndex (1 to 10): See the text below indicating that each submission could generate up to 10 different prescriptions.
- Optional Columns: Teams may produce additional columns as output of their prescriptor models in the CSV file. These columns may be noted by the judges but not evaluated by the Robo Judge.

Prescriptors should output a recommendation for all regions. Performance on speciality regions are evaluated based on the output for those regions. In one call, a prescriptor must return a

prescription file in less than 6 hours for up to 237 regions for up to 90 days with up to 10 prescriptions (PrescriptionIndex). Note that the prescriptor submission can comprise multiple models, such as those specializing in different regions, that are accessed through the same call. Also note it is unnecessary to specify which tradeoff between the case and cost objectives the prescriptor is addressing; the optimal tradeoff is determined automatically in comparison with other prescriptors during quantitative evaluation. To encourage teams to develop diverse solutions, each submission could generate up to 10 different prescriptions, identified by the PrescriptionIndex column, each potentially optimized for a different tradeoff. They will each be evaluated separately. To make aggregation possible, the number of prescriptions generated in each call must be the same.

## 7. PHASE TWO—FINAL PREDICTOR AND PRESCRIPTOR JUDGING

Phase Two Judging consists of:

- A final quantitative evaluation of a team's ongoing predictor results
- A quantitative prescriptor evaluation consisting of two parts; and
- A qualitative prescriptor evaluation

### **Final Predictor Quantitative Evaluation Criteria**

The finalist teams' predictor data will be evaluated quantitatively at the end of Phase Two using the same method as in Phase One—Prescriptor Judging. The only difference is the evaluation is now based on a much longer period of unseen data.

### **Final Prescriptor Quantitative Evaluation Criteria**

Teams will be judged over the three-week period beginning at the conclusion of the Phase Two—Prescriptor Development. Teams will submit their prescriptor models to their assigned evaluation sandbox prior to February 3, 2021. Teams are highly encouraged to submit and execute their models in this environment well prior to this date in order to address any technical issues associated with using this environment. There will be no internet access or updating of models or data after this deadline. The judging consists of quantitative and qualitative components. The quantitative evaluation will consist of two parts:

1. Performance along the two tradeoff objectives (cases vs. stringency) for between 30 to 90-days (inclusive) of simulation following the submission deadline.
2. The degree to which the submission contributes towards discovery of improved prescriptors in a further search process.

#### *First Quantitative Prescriptor Evaluation*

The first quantitative evaluation will be along the two tradeoff objectives—the number of cases vs. stringency—for each target region over a 30 to 90-day simulation period. Since the prescriptions are not yet implemented in the real world, they will be evaluated based on the estimates the

predictor models make. Thus, a standard predictor will be created using the best submissions in Phase One and the reference predictor to cover all regions as accurately as possible.

The submissions will be evaluated on all regions and separately on specialty regions. They will also be evaluated cumulatively on larger regions (i.e., countries, continents, and the world).

For each day in the 30 to 90-day evaluation period, the prescriptor is called with the date and weights as specified above, obtaining prescriptions for each region. They are evaluated along the two objectives:

- The standard predictor is called to estimate the number of cases for each region; and
- The total intervention plan stringency is calculated for each region using the Oxford University Blavatnik School of Government's COVID-19 Government Response Tracker Stringency Index formula, with the specified weights for the region.

The weights for each region are drawn from a uniform distribution within  $[0..1]$  and normalized to sum up to the number of intervention plan dimensions (i.e. 12). The process is repeated three times with different weights and the results are averaged. The same three sets of weights are used to evaluate all prescriptors. Additionally, a case where all weights are equal is used as a separate base-case evaluation. The predictions and stringency will be averaged over the 90-day period to obtain the final objective values (i.e., cases and stringency) for the prescriptor for each region.

Results with both the base case (with equal weights) and the general case (with random weights) will be presented to the judges as the outcome of the first quantitative evaluation.

For each region, all prescriptors will be placed in the 2D plane of objective values with the initial evolved prescriptors and the prescriptors evolved in the second Quantitative Evaluation detailed below. The prescriptor's performance in this region is then calculated as the number of other prescriptors it Pareto-dominates (i.e., is better along both objectives) in this space. Note: this evaluation automatically determines the best tradeoff between the two objectives for each prescriptor; that is, where it is most unique and useful compared to other prescriptors.

### *Second Quantitative Prescriptor Evaluation*

The second quantitative evaluation is based on how well the submission can serve as a stepping stone in creating improved prescriptors through further collaborative machine learning—i.e., a population-based search—in the following process:

1. The prescriptor is distilled into an equivalent neural network. It is queried with a “syllabus” of situations (case and IP history and stringency evaluation weights) to obtain a training set. A neural network similar to the evolved prescriptor samples is then trained with it.
2. The network is then inserted into the prescriptor population along with all other submissions for that region.
3. The prescriptors are evolved further, optimizing the two objectives specified in first quantitative evaluation.



4. In the final Pareto front, prescriptors that are descendants of the submitted prescriptor will be identified. The greater the number of descendants, the higher the evaluation score.

Since this form of evaluation takes significant computing time, it will be time bound to enable the speedy evaluation of submissions. At a minimum, general prescriptors (i.e., those evaluated across all regions) will be evolved using a single set of weights. Further evaluation may be conducted to obtain and identify additional innovative or creative solutions. For instance, the results using multiple weightings can be averaged; the evaluation may be done separately for each region; the quality of the descendants can be measured based on dominance (as in the first quantitative evaluation), and averaged across all descendants. The judges will identify high quality evolved prescriptors, particularly those that improve upon the initial evolved and human designed prescriptors.

Thus, the second quantitative evaluation addresses an important and novel aspect of the competition: it is a community effort. Team contributions are brought together into a common population, and a modern machine learning discovery method is used to find synergies and compatible innovations to achieve better performance than would otherwise be possible.

### **Final Prescriptor Qualitative Evaluation Criteria**

The themes of the qualitative judging may include, but are not limited to:

- **Actionability and usability:** Models that are usable in a real world setting, that provide interactivity and actionability, and that present results in a visual and well-communicated format will be ranked highly; *Judging Importance: High*
- **Explanation:** Submissions should include a narrative description of how the model works, the data it uses, and its sources as well as any relevant points related to these themes. Furthermore, models that emphasize interpretability by being able to explain why the model is predicting what it does (i.e. glass-box models) will be ranked highly; *Judging Importance: High*
- **Addressing the challenge:** Teams must avoid taking shortcuts or finding loopholes to improve their quantitative performance at the expense of real-world performance. Additionally, teams may be awarded bonus points for predicting additional, relevant public health metrics such as required hospital beds and ventilators; *Judging Importance: Medium*
- **Inclusivity and fairness:** The degree to which the data, model, and approaches consider particularly vulnerable groups<sup>3</sup> in designing and implementing their solution will be evaluated. Teams may also be judged on documented evidence of the diversity of perspectives they sought input from during the development of their solution; *Judging Importance: Medium*

---

<sup>3</sup> Vulnerable groups may include the unemployed, working poor, unhoused individuals, children, the elderly, people with disabilities, ethnic minorities, and other marginalized groups.

- **Generality:** Teams will first be evaluated on how well their models perform across all regions. Subsequently, teams will be awarded bonus points for how well their models do in specialty regions; *Judging Importance: Medium*
- **Consistency:** Approaches that stay within an acceptable range of accuracy in the short and long term, and that perform as expected in any scenario analyses run by the Judging Panel, are preferred; *Judging Importance: Medium*
- **Transparency and trust:** The extent to which their solution enables and facilitates user-facing transparency, including the ease with which a layperson can access and understand information related to how the solution functions, what data is collected and stored, and how that data may be used will be considered; *Judging Importance: Medium*
- **Collaborative contributions:** Teams that take an open-source approach to the data or models that they use, and who contribute data and models to the shared success of all teams will be ranked highly; *Judging Importance: Medium*
- **Innovation:** Teams who submit and use additional data, intervention plans (such as vaccination policies and treatments), or otherwise find innovative ways to extend the scope of the challenge will be ranked highly. *Judging Importance: Low*

## 8. DETERMINING THE WINNERS

The full judging panel will combine both quantitative rankings and the qualitative evaluations of both predictors and prescriptors to make the final determination on the winning teams. Judges have indicated that they will weigh prescriptor scores somewhat more heavily than predictor scores. If any vote of the Judges results in a tie, then the judging panel shall determine, in its sole and absolute discretion, the mechanism to settle the tie. Decisions of the judging panel are final and shall be binding on XPRIZE, teams, and each team member.

## IV. PRIZES

The Pandemic Response Challenge will be awarded with a total Prize Purse of \$500,000 (USD) and is divided as follows.

**Grand Prize:** A Grand Prize of \$500,000 (USD) will be split between the two teams that produce prediction models that estimate future numbers of daily COVID-19 cases with greatest accuracy and produce prescription models that produce intervention plans that minimize the number of cases and stringency of interventions most effectively.

Additional awards may be administered at the discretion of the Judges.

## V. ROLES AND RESPONSIBILITIES

### ADVISORY BOARD

- A. **SELECTION OF ADVISORS.** XPRIZE will appoint a panel of topical experts and big-picture thought leaders to serve as the Advisory Board for the challenge. The Advisory Board will remain in place throughout the challenge to advise XPRIZE regarding the scientific and other elements of the challenge.
- B. **INDEPENDENT ADVISORY BOARD.** The Advisory Board will be independent of XPRIZE and all teams and team members. No Advisor, nor any member of the Advisor's immediate family, shall participate, nor have any financial or other material interest, in XPRIZE, the Sponsor(s), and/or any team or team member. All members of the Advisory Board shall promptly disclose to XPRIZE any such current, former, or expected future conflict of interest with XPRIZE, the Title Sponsor, or any team or team member.
- C. **ROLE OF ADVISORY BOARD.** The duties and responsibilities of the Advisory Board may include, but not be limited to: (i) helping to establish the qualifications for prospective Judges; (ii) recommending members of the Judging Panel; (iii) helping to develop testing protocols and judging criteria; (iv) and providing input toward developing Competition Guidelines.

### JUDGING PANEL

- A. **SELECTION OF JUDGES.** The Judging Panel (as defined in the Competitor Agreement) will be composed of highly qualified and impartial Judges with subject matter and technical expertise.
- B. **INDEPENDENT JUDGING PANEL.** The Judging Panel will be independent of XPRIZE, the Title Sponsor, any other prize sponsors, and all teams and team members. No Judge, nor any member of Judge's immediate family, shall participate, nor have any financial or other material interest, in XPRIZE, the sponsor(s), and/or any team or team member. All members of the Judging Panel shall promptly disclose to XPRIZE any such current, former, or expected future conflict of interest with XPRIZE, the sponsor, and/or any team or team member.
- C. **ROLE OF JUDGING PANEL.** The duties and responsibilities of the Judging Panel will include, but not be limited to: (i) evaluating teams' compliance with the Competitor Agreement as they relate to prize operations, these Competition Guidelines, and the

Rules and Regulations for the challenge; and (ii) the awarding of points and selection of teams that will proceed to each subsequent round of the challenge.

- D. **GROUND FOR JUDGING PANEL DECISIONS.** Official decisions made by the Judging Panel will be approved by a majority of the Judges that vote on each such decision after careful consideration of the testing protocols, procedures, guidelines, rules, regulations, criteria, results, and scores set forth in the Competitor Agreement, these Competition Guidelines, Rules and Regulations, and all other applicable exhibits to the Competitor Agreement. If any vote of the Judges results in a tie, then the Judging Panel shall determine, in its sole and absolute discretion, the mechanism to settle the tie. Similarly, if one or more teams are tied at any stage during the challenge, the Judging Panel shall have the sole and absolute discretion to settle the tie.
- E. **DECISIONS OF JUDGING PANEL ARE FINAL.** The Judging Panel shall have sole and absolute discretion: (i) to allocate duties among the Judges; (ii) to determine the accuracy and error rate acceptable to the Judging Panel for all challenge calculations, measurements, and results, where not specified in the Rules and Regulations; (iii) to determine the methodology used by the Judging Panel to decide; (iv) to declare the winners of the challenge; and (v) to award the prize purses and other awards. Decisions of the Judging Panel shall be binding on XPRIZE, teams, and each team member. XPRIZE and teams agree not to dispute any decision or ruling of the Judging Panel, including decisions regarding the accuracy or error rate of any challenge calculations, measurements, and results. Teams shall have no right to observe other teams' testing or evaluation, or to be informed of other teams' calculations, measurements, and results, unless such information is made publicly available by XPRIZE.

## VI. ENVIRONMENT AND SAFETY

We prioritize safety in all challenges. Safety is a top priority for this challenge. Solutions should minimize harm and ensure safety of participants and surrounding communities. All teams must comply with these requirements:

- Teams will comply with all existing environmental, health, and safety regulations.
- As mandated by the Competitor Agreement, teams will acquire and hold all necessary licenses and insurance to demonstrate safety compliance and liability coverage as required for participation in this challenge.
- Teams should abide by the responsible use of AI and data and include an ethics statement in their submission affirming their responsible and ethical AI approach.

Additional details regarding Environment and Safety for teams' model development and testing may be provided in the Rules and Regulations for this challenge. XPRIZE reserves the right to adjust the Competition Guidelines or Rules and Regulations based on the latest scientific and legal information available at the time to ensure personal and environmental safety. XPRIZE will make all final determinations on safe and acceptable practices for challenge operations.

## VII. GLOSSARY

**Advisory Board:** A select group of prominent advisors who contribute their wisdom, knowledge and guidance to various aspects of the prize.

**Competition Guidelines:** Document for the public and for teams that describes the requirements and parameters of the challenge.

**Competitor Agreement:** A legal and binding document that details the responsibilities of competitors for the prize.

**Intervention Plan:** In this context, intervention plan pertains to a plan designed to improve the mitigation of negative impacts of COVID-19.

**Judging Panel:** The subject matter and technical experts who serve as an impartial and independent evaluation team for all aspects of this prize. Judges score the team submissions and make the final determinations in both Phase One—Predictor and Phase Two—Prescriptions.

**Prize Operations Platform (POP):** The standard internal XPRIZE portal for teams to input data, documents, and other information for this Competition.

**Prize Purse:** This refers to money offered, won, or received as a prize. It also refers to the overall amount of funds allocated to all prizes in this challenge.

**Registered Team(s):** A team that has paid the required registration fee, signed the Competitor Agreement, and is eligible to upload a Submission for the Judging Panel's review.

**Rules and Regulations:** Document detailing the testing protocols, specific rules, dates/times, and other details that will govern the challenge and will be binding on teams.

**Successful Registration:** Teams must complete all of the required steps listed in POP to successfully register.