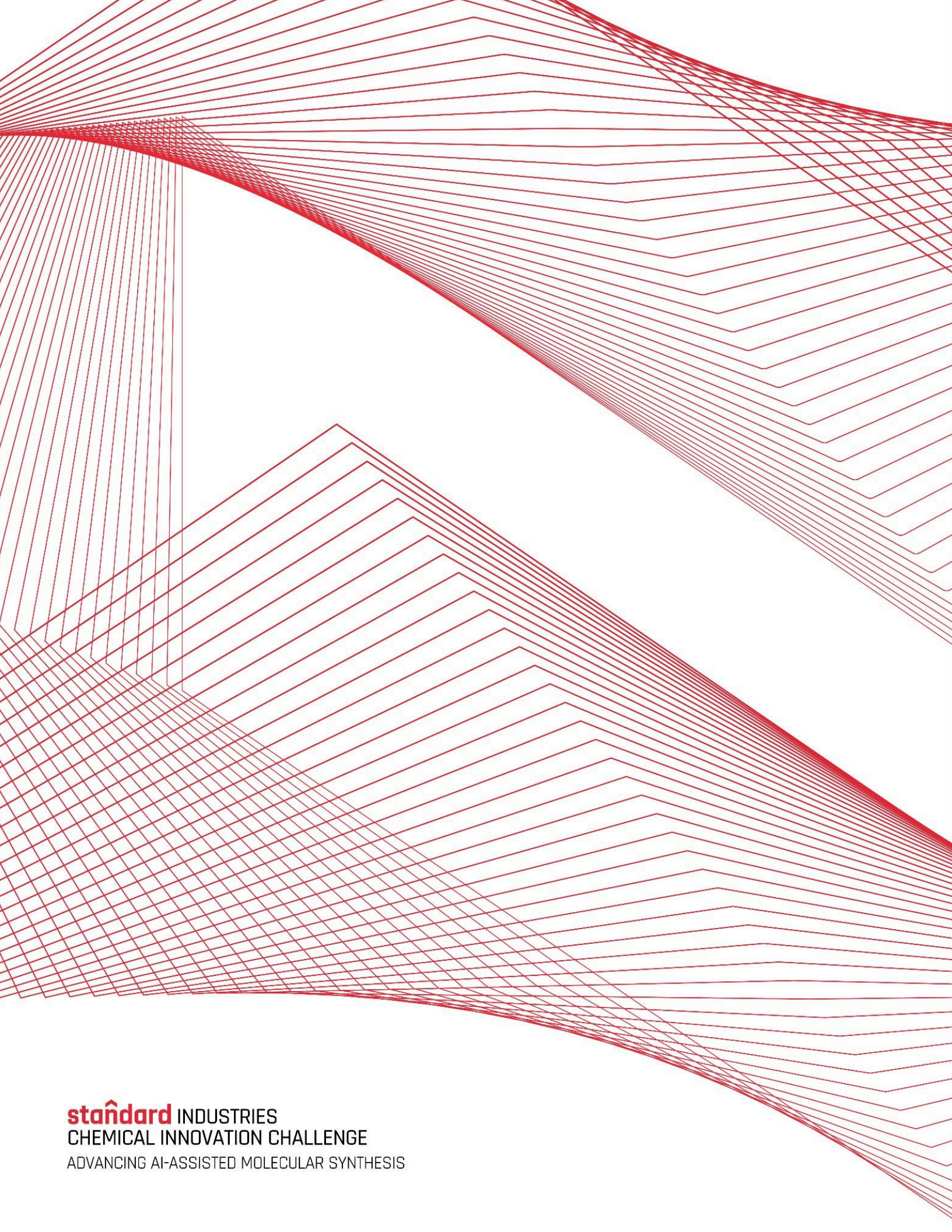

standard INDUSTRIES

CHEMICAL INNOVATION CHALLENGE

ADVANCING AI-ASSISTED MOLECULAR SYNTHESIS

Standard Industries Chemical Innovation Challenge (SI Challenge) Guidelines

Authors: Yaron Ben David, Robert Harding, Benjamin Gross, Shay Hershkovitz

-

Table of Contents

-

1. Background

1.1 Introduction: Standard Industries Chemical Innovation Challenge (the "SI Challenge" or "Challenge"

The SI Challenge is an ambitious initiative sponsored by Standard Industries Inc. and its affiliates, in particular W.R. Grace & Co. - Conn (collectively "Sponsor") aimed at transforming small-molecule synthesis through the integration of AI/ML-based tools. Its primary goal is to stimulate research in creating retrosynthesis tools imbued with chemical intuition. This involves leveraging reported chemical reactions and reactivity patterns to predict altogether new transformations and chemically-reasonable disconnects that have yet to be experimentally validated.

By encouraging participants (a participant in whatever form is hereafter referred to as a "Participant") to develop models tailored to molecular challenges with widespread relevance, the SI Challenge aims to promote a new generation of researchers integrating AI into their work. This evolution has the potential to reshape experimental synthetic chemistry, affecting areas like drug discovery, material science, environmental sustainability, and other essential scientific domains.

The SI Challenge aspires to nurture a vibrant community of innovators, highlighting their capacity to meet this ambitious target. Welcoming participants from around the globe, the SI Challenge is designed to harness diverse innovative capabilities. Notably, the Challenge embraces a variety of solutions, allowing participants to approach the goal from different perspectives.

1.2. Current solutions and their shortcomings

The focus of the SI Challenge can be stated simply, but has proven remarkably hard to do thus far: create a retrosynthetic tool that can propose novel chemistry and is able to display chemical intuition for the construction of synthetic routes to complex organic molecules. The organic molecules the challenge focuses on are defined as:

- Organic compounds containing carbon and one or more N,O, H, S, P, F, B, Cl, and Br
- Molecular weights less than 1000 Daltons.
- Containing one or more carbon-carbon bond

-

Since first proposed by Corey in 1977, multiple retrosynthetic frameworks and approaches have been developed. In the last two decades, computer-assisted retrosynthetic tools have been developed, both for commercial use and through academic efforts.[1] While remarkable progress has been achieved, all approaches that we are aware of to date are limited in one or more of the following ways:[2]

- Trained on inadequate or inaccurate datasets.
- Poor prioritization and/or tree algorithm, leading to multiple similar syntheses, synthetic "dead ends", or highly inefficient/impractical routes.
- Biased toward reactions that are commonly used in the training sets.
- Requires Ph.D. chemists to build complex reaction "rules" to identify suitable compatibility of functional groups -or- proposes reactions where certain functional groups are known to be incompatible.
- Forward and retro-synthesis looks one step ahead and one step behind.
- Rules-based or hand-coded that make it difficult to train with new data.
- Allow reaction pathways that are difficult to scale (>100Kg) or have sustainability disadvantages.

Besides these challenges, to the best of our knowledge, none of these methods can accurately forecast unexplored organic chemistry. This "predicted" chemistry should have a plausible chance of success based on literature precedent, even if it has not been directly described in publications.

---

[1] Corey, E. J., & Cheng, X.-M. (1989). The logic of chemical synthesis. John Wiley & Sons.

[2] Shen, Y., Borowski, J. E., Hardy, M. A., Sarpong, R., Doyle, A. G., & Cernak, T. (2021). Automation and computer-assisted planning for chemical synthesis. Nature Reviews Methods Primers, 1(23); Skoraczyński, G., Kitlas, M., Miasojedow, B., et al. (2023). Critical assessment of synthetic accessibility scores in computer-assisted synthesis planning. J Cheminform, 15(6).

-

1.3 Why this Challenge is needed

The goal of the SI Challenge is to develop, test, and advance an approach for retrosynthesis with durable chemical intuition using artificial intelligence/machine learning. If successful, this solution would be transformative for organic chemists, resulting in access to novel chemical space, more efficient syntheses, enhanced scalability, and enhanced sustainability through the reduction in the use of expensive or toxic reagents. More importantly, it could serve to construct a more effective tool for what has been largely an empirical endeavor for over two centuries.

-

2. Challenge Overview

The SI Challenge is intended as a three-phase competition designed to revolutionize small-molecule synthesis using AI/ML-based tools. Its main objective is to foster research in developing retrosynthesis tools enriched with chemical intuition (See the SI Challenge Official Rules, Terms and Conditions ("Challenge T&Cs") for governing terms and conditions in addition to these SI Challenge Guidelines). Ideally, this core capability will be combined with other elements of a fully integrated retrosynthesis tool, i.e. multi-step analysis, route suggestion, reaction confidence estimation, a full graphical user interface (GUI), and the ability to prioritize the relative effectiveness of different proposed chemical pathways based on input criteria such as starting material cost and availability, reaction condition constraints, and avoidance of user defined reagents. Although only the first three elements of a fully integrated retrosynthetic tool listed above are required for this Challenge, the approach taken should be compatible with further programming and algorithmic changes to allow prioritization, user interface improvement, and user customization mechanisms subsequent to the end of the Challenge.

Additionally, generalized chemical intuition is a difficult goal, and we do not anticipate the initial solution will have perfect sensitivity and specificity. A probability or confidence score should be assigned to each reaction step based on the training data and algorithm design.

2.1 Challenge Description

The Challenge will include the following phases:

I. White Paper Phase - Participants will be asked to submit a technical proposal, or "White Paper", detailing the following:

A. Overall approach to chemical intuition in a retrosynthesis pathway, including limitations of the approach that might impact generalizability.

B. Proposed initial training dataset.

C. Description of feasibility, defining AI/ML approach, chemical and reaction descriptors, and descriptions of limitations of feasibility.

-

As set out in these Guidelines and the Challenge T&Cs, Participants officially enter the Challenge by submitting their White Paper through a HeroX (HeroX.com) internet site dedicated to the Challenge. The White Paper, and any follow-on submissions throughout any phase of the Challenge, are individually and collectively Participant's "Submission". Participants acknowledge, by submitting their White Paper, that any information submitted during the White Paper Phase is not considered confidential information.

The White Paper, and all Submissions provided by Participants during the Challenge, will be evaluated by a panel of judges, comprising an interdisciplinary team of subject-matter experts in competition-related fields selected at the sole discretion of the Challenge Sponsor (the "Judge(s)" and, collectively, the "Judge's Panel"). Individual Judges may change, at Sponsor's sole discretion, if circumstances requiring substitution of an individual Judge should arise (e.g., due to unavailability or conflict).

II. Semifinalist Phase - Up to 10 White Papers from the White Paper Phase will be selected by the Judge's Panel as best meeting the SI Challenge criteria and the associated Participant's ("Semifinalists") will advance to the Semifinalist Phase. In recognition of their solution's potential and for their continued participation in the Challenge, these Semifinalists will be awarded Ten Thousand U.S. Dollars ($10,000) each. To participate in the Semifinalist phase, Participants must sign a 2-Way NDA with Sponsor (See representative template in the Challenge T&C's) prior to detailed sharing about their strategy and results. During this phase:

A. Semifinalists will advance their approach with a set of ten (10) or more molecules provided by Sponsor ("Test Molecules"). Semifinalists will advance their approach to: i) identify key transformations for the total synthesis of the Test Molecules; and ii) propose novel approaches to those transformations or the broader total synthesis.

B. Semifinalists will construct a rudimentary interface so that real time evaluation of Test Molecules can be evaluated.

C. A confidence score will be defined by each Semifinalist and included as an output of any reaction step.

D. Semifinalists can evaluate additional molecules of their choice to better demonstrate their solution's effectiveness, though this is not mandatory.

-

E. To reduce the influence of training data, the Challenge Sponsor may provide each Semifinalist with a training set. Each Semifinalist may augment this data with additional datasets.

Communication with the Challenge Sponsor for clarification is encouraged, however if any material clarification arises, Sponsor may, at its discretion, inform other Semifinalists of such clarification without details of the context or approach of the Semifinalist raising such clarification.

At the end of the Semifinalist Phase, each Semifinalist will present its progress to the Judges Panel. Sponsor will provide several additional Test Molecules to Semifinalists at this time to demonstrate their solution in real-time. The Judges Panel will then select up to three (3) Semifinalists ("Finalists") judged as having the solutions that best meet the Challenge criteria to continue to the Finalist Phase. These three Finalists will each receive an award of One Hundred Thousand U.S. Dollars ($100,000) in recognition of their outstanding work, for their continued participation in the Challenge, and to support further development of their AI/ML models.

III. Finalist Phase - To participate in the Finalist Phase, Finalists must acknowledge and accept additional terms detailing the desired relationship between a Finalist selected as the Challenge Winner (the Participant with the solution judged to best meet the overall Challenge evaluation criteria) and Sponsor (These terms are set out in the Challenge T&Cs and are binding on all Finalists accepting the Finalist award and become effective on the Challenge Winner when the Winner accepts the Grand Award set out below. In the Finalist phase:

A. Sponsor will provide each Finalist with five or more Challenge Molecules. Finalists will have an opportunity to further refine their approach and solution with these five Challenge Molecules.

B. Finalists will present their solution's output for each of the five Challenge Molecules, defining the key transformations and the proposed synthetic routes.

C. Using appropriate intermediates/starting materials, Sponsor or a third party designated by Sponsor may attempt the key predicted transformation(s).

-

D. Based on the Challenge evaluation criteria, including real world results of the predictive model, the Judging Panel will select a Challenge Winner from among the Finalists to receive the Grand Award.

E. The Challenge Winner will adhere to the milestones set by the Challenge Sponsor and will provide updates on their progress.

2.2 Award Purse

The total award purse for this challenge amounts to One Million Four Hundred Thousand U.S. Dollars ($1,400,000). Sponsor expects that these awards will not only motivate Participants to produce groundbreaking work, but also provide financial support to the awarded Participants in advancing their research and contributing to developments in chemical synthesis and AI/ML:

- **Grand Award**: The Challenge Winner (Finalist with the AI/ML model that best fulfills the Challenge goal and evaluation Criteria) will be awarded One Million U.S. Dollars ($1,000,000), subject to the terms set out in these Guidelines. This substantial award highlights Sponsor's dedication to encouraging innovations in the field of organic synthesis and AI/ML, and it underscores the transformative potential of the winning solution.

- **Finalist Award**: Up to three (3) Semifinalists will be awarded One Hundred Thousand U.S. Dollars ($100,000) each,subject to the terms set out in these Guidelines.

- **White Paper Award**: Up to ten (10) Participants with White Papers judged to best meet the Judging criteria for advancing to the Semifinalist Phase (Semifinalists) will be awarded Ten Thousand U.S. Dollars ($10,000) each, subject to the terms set out in these Guidelines.

To collect a White Paper Award or a Finalist Award, awarded Participants must continue to the next phase of the contest.

-

2.3 Winning Participant Solution Requirements

The Challenge Winner will provide an AI/ML-driven research platform designed to identify multiple options for synthesizing a target molecule. The AI/ML platform will assign probabilities of success or confidence levels to each pathway based on how close it is to literature precedent. The AI/ML platform will propose pathways that include reaction steps that are not explicitly listed in the literature but are derived by "chemical intuition" or analogy to known molecular transformations. Preferably, the AI/ML platform will allow the user to prioritize pathways by excluding some types of reactions and show preference to other types of reactions. The AI/ML platform will provide a JSON output to allow streamlined judging. Preferably, the AI/ML platform will allow the user to exclude some types of expensive reagents and prioritize starting molecules of lower cost. Preferably, the AI/ML platform will allow the user to prioritize pathways based on the types of equipment necessary to perform the reaction steps in the pathway.

2.4 Eligibility for Challenge Participation:

Commercial Entities:

- Must be a company with no more than 30 employees at the time of entering the Challenge.
- The company should not have raised funds beyond a Series A round as of the time of entering the Challenge.

Academic Teams:

- Must be affiliated with an accredited academic institution at the time of entering the Challenge.

Individual:

- Must be able to demonstrate the ability to execute on their proposed solution. Ability will be determined at the Sponsor's discretion at the point of submission.
- Individuals must be 18 years of age to enter the Contest.

-

General Eligibility Exclusions (the following are ineligible to participate):

- Individuals or entities located in a jurisdiction where participation in the Challenge is prohibited or restricted by law or regulation.

- Individuals or entities with a residence in or who are a national of Cuba, Iran, Syria, North Korea, Sudan or, as applicable, Russia, Crimea and covered regions of Ukraine; or an entity or individual who is subject to export controls, embargos, or sanctions of the United States.

- Individuals or entities currently engaged with or employed by Sponsor, or that have immediate family members currently engaged with or employed by Sponsor, or those working on a proof-of-concept project, and their immediate family members, spouse, or significant other. Exception to this exclusion may be made at Sponsor's discretion.

- Individuals and entities that work directly with Judges in an academic or professional context and their immediate family members, spouse or significant other.

- Individuals or entities that have a business relationship or sponsored research affiliation with the pharmaceutical industry or contract manufacturing and development organizations will be determined as eligible or ineligible at Sponsor's discretion based on the nature of the relationship to such organizations.

- Individuals or entities that have or are currently serving as consultants for Sponsor are ineligible to participate. Exceptions to this exclusion may be made by Sponsor where Sponsor believes the consultant does not possess information that would provide an unfair advantage over other Participants.

- Individuals or entities that do not have the right to grant any licenses or other rights transfers as may be required in the Challenge Guidelines and/or the Challenge T&Cs.

- All Submissions related to the contest must be in English. Submissions in any other language will not be considered.

Participants shall meet the above criteria, and all other criteria in the Challenge T&Cs to be considered eligible to participate in the contest.

-

2.5 Challenge Timeline

- White Paper Phase: Challenge Launch and White Paper Development (8 weeks)
- Semifinalist Phase - Proof of Concept Development with up to 10 Semifinalists (20 weeks)
- Finalist Phase - AI/ML Model Development and Testing with up to 3 Finalist Teams (17 weeks)
- Winner Selection and Award Ceremony - (1 week)

Sponsor reserves the right to extend any Challenge Phase if no Participant team meets the required phase criteria. Timeline above is representative and subject to change at Sponsor's discretion. Sponsor will notify Participants of any changes to the timeline, or any other aspect of the Challenge, via the Challenge HeroX website. It is the responsibility of each Participant to periodically reference the Challenge HeroX website to make themselves aware of any such changes.

-

3. Testing and Judging

Testing and judging are some of the most important and valuable parts of any skills-based Challenge. This is true in terms of Sponsor's ability to fairly and objectively declare a Challenge Winner, but also in terms of the value that rigorous testing and judging provides Participants in deploying the winning Participant's solution after the Challenge concludes. This section lays out how we will test and evaluate the SI Challenge - i.e., the Judging Panel selection, evaluation criteria, and Awards.

3.1 Judges

A diverse group of subject-matter experts will be important in helping Sponsor evaluate and judge the Participants' Solutions:

- Judges – are an interdisciplinary team, from academia or industry, who are subject-matter experts in related fields of the SI Challenge. Judges are responsible for evaluating Participants and their Solutions against the Challenge Criteria.
- The Judges that evaluate Participants' Solutions are defined as the Judging Panel.

To ensure that Participants' Solutions meet Evaluation Criteria, the Judging Panel will include representatives with expertise in at least the following:

- Synthetic Organic Chemistry
- Computer Assisted Synthesis
- Artificial Intelligence

Note that while individual Judges may have a sub-specialty, all of the Judges selected to be on the Judging Panel will have some expertise in computer assisted synthesis/reactions.

-

3.2 Evaluation Model and Governance

Integrity in Peer Review

If they have not already, Judges will be required to complete the following two trainings, recently required of all National Institutes of Health (NIH) reviewers:

- "Review Integrity" – raises awareness of actions that breach review integrity and provides tools to prevent and report them.
- "Bias Awareness and Mitigation"– raises awareness of potential biases in the peer review process and provides strategies to mitigate them.

Guidelines for confidentiality and integrity of peer review will be based on the NIH guidelines. https://grants.nih.gov/grants/peer/guidelines_general/Confidentiality_CertificationsPR.pdf)

Blinding

After careful consideration and review of the literature, we believe blinding of Participants to Judges is not practical.

At the White Paper Phase, Participants will not know the names of the Judges on their panel, but the full list of Judges will be available. During the Semifinal and Final Phases, the names of Judges evaluating the Solution will be known to Participants, but the scoring by individual Judges will not be disclosed to participants.

-

3.3 Evaluation Criteria

Judges will evaluate Participants using multiple evaluation methods, including but not limited to: Screening, Benchmarking, Pass/Fail, and Scorecard. 'Screening' is criteria of minimum thresholds that must be met to qualify and are judged as 'pass/fail.' Benchmarking evaluates Participants against a defined 'baseline.' Pass/Fail identifies the minimum Criteria Participants must meet to advance. Finally, the scorecard Criteria ranks Participants alongside multiple Criteria. Points are accrued and totaled at the end of each phase and at the end of all phases to inform the results.

-

3.3.1 White Paper Phase

3.3.1.a Eligibility and Completeness

White Papers will be screened for eligibility and completeness on a rolling basis during the submission period by the Sponsor. In the event that prospective participants ("Prospective Participant") are not eligible, they will be notified. Incompleteness is defined as either a missing field on the White Paper form, or a field with insufficient detail. Prospective Participants will be notified if one or more fields are missing or if the Sponsor have determined one or more fields have insufficient detail. Prospective Participants will be given an opportunity to resubmit their White Paper. If Prospective Participants do not resubmit or if the Sponsor determines that there is still insufficient detail in one or more fields, the White Paper will not progress to Content Review by the Judges as described below.

3.3.1.b Content Review

After the White Paper submission period is over, White Papers that have complied with Eligibility and Completeness requirements will be evaluated by a Judging Panel of 6 or more Judges. At least three Judges will have primary expertise in AI/ML and/or molecular data mapping or modeling, and at least three Judges will have primary expertise in computer-assisted synthesis and/or computer-assisted reaction prediction. The objective of Judging Panel is to holistically assess the White Paper across the following Criteria:

Criteria	Description	Weight
Completeness	Does the approach, either using prior work or a novel approach, encompass a complete solution?	10%
Methodology	Is the approach methodologically sound? Ambitious proposals are strongly encouraged; however, the methodology must be plausible and computationally achievable.	30%

-

Team & Resources	Does the team have the expertise and computational resources to develop a solution within the competition timeframe? Has the team done prior work that de-risks the proposed approach?	20%
Impact	Is it likely that, if successful, the approach will have a meaningful impact on the field?	30%
Innovation	How does the approach differ or expand on what has been tried previously?	10%

-

3.3.3 Semifinalist Phase

In this Phase, Participants must produce a set of proposed retrosynthesis pathways for each of the 10 or more Sponsor-provided Test Molecules, ranked by a parameter indicating likelihood of success (the Confidence Estimate). As described earlier, one or more of these Test Molecules will be presented in real time with Judges present to validate the time-effectiveness and autonomy of the platform solution.

At the end of the Semifinalist Phase, Participants will be evaluated by a Judging Panel of 6 or more Judges. At least three Judges will have primary expertise in synthetic organic chemistry, and at least three will have primary expertise in computer-assisted synthesis and/or computer-assisted reaction prediction. The objective of the Judging Panel is to assess the Participants on the following criteria:

Criteria	Scoring	Evaluation Type	Weight
Breadth of solution	For what percentage of the Test Molecules did the solution provide pathways that were deemed achievable?	Quantitative	**30%**
Chemical Intuition	Were the transformations in the solution novel and meaningfully different from reactions existing in the literature?	Semi-quantitative assessment by the Judges	**25%**
Confidence estimation	Did the solution provide a confidence score for each reaction step? Was this confidence score meaningful?	Semi-quantitative assessment by the Judges	**25%**

-

Parameter selection	Does the solution provide a mechanism to prioritize or deprioritize pathways based on parameters such as starting material, excluded reagents/solvents, or other parameters preferred by the user? If not, is there a proposal for how to incorporate this mechanism?	**Quantitative**	**10%**
Real-time evaluation	For the real-time demonstration, did the solution provide a plausible set of prioritized reaction pathways within a reasonable time period?	**Quantitative**	**10%**

-

3.3.4 Finalist Phase

In this Phase, Participants must produce a set of proposed retrosynthesis pathways for each of the 5 or more Sponsor-provided Challenge Molecules, ranked by a parameter indicating likelihood of success (the Confidence Estimate). Each Participant will identify reaction steps that are produced by the chemical intuition capability of the solution (i.e., show that each such reaction step is not found in the public literature, but is plausible based on the public literature). Sponsor will choose one or more of these reaction steps to validate in the lab.

At the end of the Finalist Phase, Participants will be evaluated by a Judging Panel of 6 or more Judges. At least three Judges will have primary expertise in synthetic organic chemistry, and at least three Judges will have primary expertise in computer-assisted synthesis and/or computer-assisted reaction prediction. The objective of the Judging Panel is to assess the Participants on the following criteria:

Criteria	Scoring	Evaluation Type	Weight
Breadth of solution	For what percentage of the 5+ challenge molecules did the solution provide pathways that were deemed achievable?	**Quantitative**	**25%**
Novelty	Were the transformations in the solution novel and meaningfully different from the reactions demonstrated in the literature?	**Semi-quantitative assessment by the Judges**	**25%**
Confidence estimation	Did the solution provide a confidence score for each reaction step? Was this confidence score meaningful?	**Semi-quantitative assessment by the Judges**	**10%**

-

Prioritization Mechanism	Does the solution provide a mechanism to prioritize or deprioritize pathways based on parameters such as starting material, excluded reagents/solvents or other parameters preferred by the user? If not, is there a proposal for how to incorporate this mechanism?	**Quantitative**	**20%**
Real-world validation	Could key transformations be performed in the lab with reasonable yield?	**Quantitative**	**20%**

-

4. Appendix

4.1 Technical Details:

4.1.1 Allowed Reaction Classes:

Almost all known chemical reaction steps are allowed as part of the Challenge Evaluation Criteria. Photoredox, electrochemistry, and biocatalytic or enzymatic reaction steps are not the focus of this Challenge. These reaction types are allowed as valid steps in a multi-step retrosynthesis pathway if literature references or the training data contain identical reactants and products. Steps that use photoredox, electrochemistry or enzymatic reactions derived by chemical intuition are not considered valid. Purification steps are not considered for this Challenge.

4.1.2 Output Format

The participants in the Semifinalist Phase and the Finalist Phase will be provided with a JSON format to transmit their proposed multi-step reaction pathways to the Judges. The format will include (for each step):

- Reactants
- Products
- Conditions (e.g. Temperature, pH, catalyst). Estimation of reaction conditions is preferred in the SemiFinalist Stage, but not required.
- Closest Chemical Similarity Index (Tanimoto) to chemicals in training data
- Predicted Confidence Score of this transformation
- Closest literature precedent, if known. Closest literature precedent is preferred but not required as it makes Participant's results more explainable and easier to validate.
- Scalability Index (as described below.).

-

4.1.3 Training Data

Sponsor may provide each Semifinalist and Finalist team access to training data to provide all participants in these phases with access to adequate starting data. Participants are allowed to add additional training data, or substitute a different database of data from their own sources. The purpose of requesting the closest chemical similarity index score of the chemicals in the training data to the reactants and products is to allow the Judges to separate the effects of algorithm and training data.

4.1.4 GUI

The Graphical User Interface included in the Participant's AI/ML platform is not a feature that will be evaluated by the Judges. The focus of the Challengejudging is Participant's solution of an algorithm for successful retrosynthesis prediction.

4.1.5 Confidence Score

The Confidence Score for each reaction step is intended to be estimated by the algorithm as a measure of how close the proposed reaction step is to known transformations in commercial and public data sets. Sponsor expects a multi-step retrosynthetic route to contain some steps that will have high confidence scores and some steps that will have lower confidence scores (particularly for the steps that require significant chemical intuition). The method for estimating these confidence scores is not prescribed, and Participants are encouraged to develop/propose their own approach. In the Finalist Phase some experimental validation is planned, and Sponsor intends to focus on the steps with lower confidence scores.

4.1.6 Viable Starting Reactants

The multi-step pathways proposed by the Participant's solution should start with starting materials that have a commercial source with a reasonable price point (e.g., at or about $50/g). The Aldrich catalog is an acceptable source of this information.

-

4.1.7 Use of Open Source Software

Participants are allowed to build their program on top of open source software, with the following limitations:

- "**Software**" means any and all (i) software, computer programs, systems, platforms, applications (including mobile apps), application programming interfaces, firmware, middleware, microcode, routines, compilers, assemblers, and software implementations of algorithms or models (including for machine learning, deep learning, and other artificial intelligence technologies), in each case whether in **source** code, object code, or other form; (ii) databases, data files, libraries, and data compilations; (iii) screens, user interfaces, report formats, templates, menus, buttons, and icons; (iv) development and design tools, diagrams, descriptions, protocols, flow charts, and other work product used to design, plan, organize, and develop any of the foregoing; and (v) documentation, including specifications, files, scripts, developer notes, comments, annotations, user documentation (including user instructions, guides, and manuals), and training materials, relating to any of the foregoing.

- "**Open Source Software**" means any Software that is distributed as "free software," "open source software," or pursuant to any license identified as an "open source license" by the Open Source Initiative (www.opensource.org/licenses) or other license that substantially conforms to the Open Source Definition (opensource.org/osd) [(including the GNU General Public License (GPL), GNU Lesser General Public License (LGPL), GNU Affero General Public License (AGPL), MIT License (MIT), Apache License, Artistic License, and BSD Licenses)].

- Participant agrees not to use, and has not used, any Open Source Software in any submission at any phase of the Contest in a manner that does, will, or would reasonably be expected to require the: (A) disclosure or distribution of any Software that is part of participant's submission, any derivative thereof, or or any other proprietary Software in source code form; (B) license or other provision of any Software that is part of participant's submission, any derivative thereof, or or any other proprietary Software on a royalty-free basis; or (C) grant of any patent license, non-assertion covenant, or other rights under any participant or Sponsor Intellectual Property or rights to modify, make derivative works based on, decompile, disassemble, or reverse engineer any Software that is part of participant's submission, any derivative thereof, or or any other proprietary Software.

-

4.1.8 Multiple Synthetic Routes Desired

Participants are encouraged to provide up to 5 alternative synthetic routes for a given Target Molecule or Challenge Molecule. A synthetic chemist prefers options.

4.1.9 Estimated Scalability Index

The target of the SI Challenge is to develop a solution to estimate retrosynthetic routes that are practical, scalable (to >100kg), and sustainable. To estimate these features, we define a Scalability Index from 1-10 based on the type of reaction and a limited set of reaction conditions. Each reaction step has a Scalability Index, and the Index of the entire pathway is the lowest Scalability index of the steps contained within it. A set of reaction types and their Scalability Index is listed below. Participants may ask for a Scalability Score for additional reaction types. The Sponsors may opt to refine this index during the course of the competition.

Reaction Scalability Table

Reaction	Relative Scalability Score 0-10
Acetylation	10
Acyl Chloride esterification	8
Aldol Condensation	2

-

Alkene epoxidation	4
Alkyne Bromination	2
Azide-Alkyne Cycloaddition	0
Carbonyl Reduction	6
Carboxylic Acid Esterification	10
Carboxylic Acid to Acyl Chloride	10
Aryl Chloride Coupling	8
Heck	10
Negishi transformation	10
Nitro reduction	6
Olefin Metathesis	8

-

Reductive Amination	10
Sulfide Oxidation	8
Suzuki Coupling	10
Wittig Reaction	4

-