MACHINE LEARNING MODEL ATTRIBUTION CHALLENGE

OFFICIAL RULES

1. SPONSOR

These Official Rules ("Rules") govern the operation of the Machine Learning Security Evasion Competition Contest ("Contest"). PlainText Group is the Contest sponsor ("Sponsor"). Partnering Organizations HuggingFace, Lincoln Network, MITRE, and Robust Intelligence are "Contest Parties".

2. DEFINITIONS

In these Rules, "we", "our", and "us" refer to Sponsor and "you" and "yourself" refers to a Contest participant, or the parent/legal guardian of any Contest entrant who has not reached the age of majority to contractually obligate themselves in their legal place of residence. By entering you (your parent/legal guardian if you are not the age of majority in your legal place of residence) agree to be bound by these Rules.

3. ENTRY PERIOD

The Contest starts on August 12, 2022 at 12:00am GMT-12 (Anywhere on Earth, or AoE) and ends on September 16, 2022 at 11:59pm GMT+12 (AoE), the period known as the "Contest Period". Solutions must be submitted to https://mlmac.io/submit-solution (“Solution Website”) during the Contest Period.

4. ELIGIBILITY

To enter, you must be an IT pro enthusiast or developer, 18 years of age or older. If you are 18 years of age or older but have not reached the age of majority in your legal place of residence, then you must have consent of a parent/legal guardian.

Employees and directors of Sponsor and Contest Parties, affiliates, advertising agencies, and Contest Parties are not eligible, nor are persons involved in the execution or administration of this promotion, or the family members of each above (parents, children, siblings, spouse/domestic partners, or individuals residing in the same household). Void in Cuba, Iran, North Korea, Russia, Sudan, Syria, Region of Crimea, and where prohibited.

5. HOW TO ENTER

The object of the Contest is to identify the underlying base model of a set of fine-tuned generative models by interacting with the fine-tuned models. Performance will be measured by the judging criteria listed below.

Follow these steps to create and submit an entry into the Contest:

1. Visit the https://mlmac.io/ ("Contest Website") and follow instructions register. This includes accepting these Terms of Use to gain access to your user ID / API key and the competition API ("API").
2. Use the API to interact with anonymous fine-tuned models, and compare to a list of potential base models.
   a. Use as few total API as possible, as the number of API calls is one of the ranking criteria for your solution.
3. When you have identified the underlying base model for each of the anonymous fine-tuned models, submit your answers, together with your user ID and email address to the Solution Website.

Entries must be received within the Entry Period to be eligible.

The entry limit is one unique entry per person per Contest.

Any attempt by any you to obtain more than the stated number of entries by using multiple/different accounts, email addresses, identities, registrations, logins, or any other methods will void your entries and you may be disqualified. Use of any automated system to submit fraudulent entries is prohibited.

We are not responsible for excess, lost, late, or incomplete entries. If disputed, entries will be deemed submitted by the "authorized account holder" of the email address, social media account, or other method used to enter. The "authorized account holder" is the natural person assigned to an email address by an internet or online service provider, or other organization responsible for assigning email addresses.

6. ELIGIBLE ENTRY

To be eligible, an entry must meet the following content/technical requirements:

- Your entry must be your own original work; **and**
- Your entry cannot have been selected as a winner in any other contest; **and**
- You must have obtained any and all consents, approvals, or licenses required for you to submit your entry; **and**

- To the extent that entry requires the submission of user-generated content such as software, photos, videos, music, artwork, essays, etc., entrants warrant that their entry is their original work, has not been copied from others without permission or apparent rights, and does not violate the privacy, intellectual property rights, or other rights of any other person or entity;; **and**
- Your entry may NOT contain, as determined by us in our sole and absolute discretion, any content that is obscene or offensive, violent, defamatory, disparaging or illegal, or that promotes alcohol, illegal drugs, tobacco or a particular political agenda, or that communicates messages that may reflect negatively on the goodwill of Sponsor or Partnering Organizations.

7. USE OF YOUR ENTRY

We are not claiming ownership rights to your entry. However, by submitting an entry, you grant us an irrevocable, royalty-free, worldwide right and license to use, review, assess, test and otherwise analyze your entry and all its content in connection with this Contest and use your entry in any media whatsoever now known or later invented for any non-commercial or commercial purpose, including, but not limited to, the marketing, sale or promotion of Sponsor or Contest Parties' products or services or those of any other Contest Parties, without further permission from you. You will not receive any compensation or credit for use of your entry, other than what is described in these Official Rules

By entering you acknowledge that we may have developed or commissioned materials similar or identical to your entry and you waive any claims resulting from any similarities to your entry. Further you understand that we will not restrict work assignments of representatives who have had access to your entry and you agree that use of information in our representativesâ€™ unaided memories in the development or deployment of our products or services does not create liability for us under this agreement or copyright or trade secret law.

Your entry may be posted on a public website. We are not responsible for any unauthorized use of your entry by visitors to this website. We are not obligated to use your entry for any purpose, even if it has been selected as a winning entry.

8. WINNER SELECTION AND NOTIFICATION

Pending confirmation of eligibility, potential prize winners will be selected by Sponsor or their Agent or a qualified judging panel from among all eligible entries received based on the following rank-ordered judging criteria:

(1) Highest number of correctly identified base models as determined by the solution submitted to the Solution Website
(2) Fewest total API queries through the API
(3) Earliest Submission Time

Potential prize winners will be determined within 20 business days following the Entry Period of the Attacker Challenge. The decisions of the judges are final and binding. If we do not receive a sufficient number of entries meeting the entry requirements, we may, at our discretion, select fewer winners than the number of Contest Prizes described below. If public vote determines winners, it is prohibited for any person to obtain votes by any fraudulent or inappropriate means, including offering prizes or other inducements in exchange for votes, automated programs or fraudulent identification. Sponsor will void any questionable votes.

However, in order to win a prize, your solution must be documented and published. Winners will be notified via the contact information provided during entry no more than 7 days following judging with prize claim instructions, including a request for a link to their published submission. Valid publication outlets include a document provided to Sponsor (for publication on a Sponsor or Contest Parties' website), a link to a publicly accessible internet website, such as https://github.com/ (e.g., for code), https://arxiv.org/ (e.g., for a whitepaper), or a blog post detailing the solution. Failure to publish within 15 days of notification will result in forfeiture of the prize.

If a selected winner cannot be contacted, is ineligible, fails to claim a prize or fails to return any forms, the selected winner will forfeit their prize and an alternate winner will be selected time allowing. If you are a potential winner and you are 18 or older but have not reached the age of majority in your legal place of residence, we may require your parent/legal guardian to sign all required forms on your behalf. Only three alternate winners will be selected, after which unclaimed prizes will remain unawarded.

9. PRIZES

Prize winners will be awarded the following prizes, which may be delivered, at Contest Sponsor's discretion, via credits of sufficient value.

One (1) "First Place Prize":

A cash equivalent prize valued at $3,000 USD .

One (1) "Honorable Mention Prize":

A cash equivalent prize valued at $1,500 USD .

Up to three (3) "Outstanding Student Submission":

Awarded to the three (or fewer, if there are an insufficient number of student submitters) highest ranking student submissions based on the evaluation criteria, consisting of free admission plus a $1.5K travel voucher to present a solution at CAMLIS 2022 (https://camlis.org) in October 2022. To be eligible, the submitter must be a current undergraduate or graduate degree-seeking student at an accredited university, and must submit their solution using the email address of their university (e.g., ending with .edu).

This "Outstanding Student Submission" award can be awarded in addition to a "First Place Prize" or "Honorable Mention Prize", or as a standalone award to an otherwise non-winning submission.

The Sponsor and Contest Parties reserve the right to validate whether the Contestant is a degree-seeking student at an accredited university.

We will only award up to two (2) prize(s) per person during the Contest Period. No more than the stated number of prizes will be awarded. No substitution, transfer, or assignment of prize permitted, except that Sponsor reserves the right to substitute a prize of equal or greater value in the event the offered prize is unavailable. Prizes are awarded AS IS with no warranty of any kind, either express or implied, including but not limited to, the implied warranties or merchantability, fitness for a particular purpose, or non-infringement. Prizes will be sent no later than 28 days after winner selection. Prize winners may be required to complete and return prize claim and / or tax forms (Forms) within the deadline stated in the winner notification. Taxes on the prize, if any, are the sole responsibility of the winner, who is advised to seek independent counsel regarding the tax implications of accepting a prize. By accepting a prize, you agree that Sponsor may use your entry, name, image and hometown online and in print, or in any other media, in connection with this Contest without additional payment or compensation to you, except where prohibited by law.

10. ODDS

The odds of winning are based on the number of eligible entries received.

11. GENERAL CONDITIONS AND RELEASE OF LIABILITY

The Contest Sponsor and Contest Parties shall have no liability whatsoever for any claims, losses, actions, damages, suits, or proceedings resulting from damage or

loss caused by malware to your computer or computer network, or damage or loss caused by malware to other computers or computer networks owing to your distributing of original or modified malicious software. To the extent allowed by law, by entering you agree to release and hold harmless Sponsor and its respective parents, partners, subsidiaries, affiliates, employees, and agents and Contest Parties from any and all liability or any injury, loss, or damage of any kind arising in connection with this Contest or any prize won.

All local laws apply. The decisions of Sponsor are final and binding.

We reserve the right to cancel, change, or suspend this Contest for any reason, including cheating, technology failure, catastrophe, war, or any other unforeseen or unexpected event that affects the integrity of this Contest, whether human or mechanical. If the integrity of the Contest cannot be restored, we may select winners from among all eligible entries received before we had to cancel, change or suspend the Contest.

If you attempt or we have strong reason to believe that you have compromised the integrity or the legitimate operation of this Contest by cheating, hacking, creating a bot or other automated program, or by committing fraud in any way, we may seek damages from you to the full extent of the law and you may be banned from participation in future Sponsor promotions.

12. USE OF YOUR ENTRY

Personal data you provide while entering this Contest will be used by Sponsor and Partnering Organizations and/or their agents for the administration and operation of this Contest.

13. GOVERNING LAW

This Contest will be governed by the laws of the State of New York, and you consent to the exclusive jurisdiction and venue of the courts of the State of New York for any disputes arising out of this Contest.