



01 / REGISTRATION IS OPEN

OPEN FOR REGISTRATION

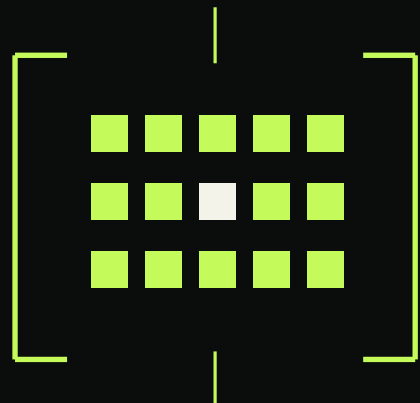
HACK THE FIRST AGI 1.0

Can You Help Evaluate the
Intelligence That Comes Next?

THE FRONTIER AI EVALUATION GAUNTLET

15 TEAMS. ONE GOA FINAL.

Frontier-model evaluation. Agent testing.
Benchmark engineering. Failure discovery.



APPLY BY SEPTEMBER 16, 2026

11:59 PM IST / TEAMS OF 2-3



02 / MANIFESTO

The models are getting smarter. The tests must evolve faster.

Building increasingly capable AI models is only one part of the race. The other part is learning how to measure them.

A score can hide a brittle strategy. A correct answer can conceal a lucky guess. A successful demo can obscure a system that fails the moment its tools, context or environment change.

You are not here to build another wrapper. You are here to design experiments that separate capability from coincidence; to make failures reproducible; to discover what intelligence can actually do.

Benchmarks measure scores. Great evaluators reveal behavior.

FIND THE FAILURE BEFORE THE WORLD DOES.



03 / THE COMPETITION

Measure intelligence. Expose its limits.

An elite AI evaluation, model-behavior and frontier-AI engineering hackathon.

HACK THE FIRST AGI 1.0 brings teams into a research-driven competition: build evaluations, test systems, investigate failures and defend the evidence. The central artifact is a credible finding, supported by working engineering.

01 / ASK A PRECISE QUESTION

Define the capability, behavior or failure you want to measure. State what your test can and cannot establish.

02 / BUILD THE INSTRUMENT

Create a benchmark, harness, dataset or agent-testing environment with clear success and failure conditions.

03 / RUN CONTROLLED TESTS

Establish baselines, record conditions and inspect the cases that aggregate scores fail to explain.

04 / DEFEND THE FINDING

Deliver code, evidence and a technical report that another evaluator can use to reproduce your conclusions.

A RESEARCH CHALLENGE. AN ENGINEERING STANDARD.

The event name is a challenge to evaluators; it is not a claim that any tested system has achieved AGI.



04 / THE FRONTIER

A benchmark is only as good as its blind spots.

The race toward AGI needs better evaluators. Strong evaluation asks not only whether a system succeeds, but when, why, at what cost and under which assumptions.

CAPABILITY $\neq$ RELIABILITY

Repeat trials. Perturb inputs. Change the environment. Measure how often useful behavior survives.

A SCORE IS NOT AN EXPLANATION

Inspect distributions, uncertainty and failure clusters. Separate genuine capability from contamination or shortcuts.

AGENTS CHANGE THE UNIT OF ANALYSIS

Evaluate the trajectory: tool choices, state tracking, recovery, intervention and long-horizon completion.

EVIDENCE MUST TRAVEL

Publishable reasoning begins with reproducible protocols, traceable data and honest limits on the claim.



05 / CHALLENGE DOMAINS

Choose a fault line.

Challenges may span one or more of these research and engineering domains.

REASONING & KNOWLEDGE

Reasoning evaluation; hallucination analysis; preference evaluation; reward-model analysis.

CODING & AUTONOMY

Coding-agent evaluation; autonomous-agent evaluation; tool-use reliability; long-horizon tasks.

BENCHMARKS & DATA

Benchmark construction; dataset design; contamination detection; statistical evaluation.

ROBUSTNESS & SAFETY

Adversarial testing; safety evaluation; model robustness; reliability under distribution shift.

MULTIMODAL & HUMAN

Multimodal evaluation; human-in-the-loop evaluation; annotation design; model-behavior research.

SYSTEMS & INTERPRETATION

Evaluation infrastructure; reproducibility; failure-mode discovery; interpretability-oriented experiments.

Challenge briefs define the permitted systems, datasets, access limits and deliverables. A focused, defensible experiment is more valuable than a broad claim without evidence.



06 / WHO SHOULD APPLY

Bring proof of work. Bring a better question.

Students, researchers, engineers, AI evaluators and technically strong builders can apply. Teams may span colleges or organizations. Prior evaluation experience helps; careful reasoning, strong implementation and a capacity to learn matter more than a title.

WHAT SELECTION VALUES

Technical ability; evaluation thinking; originality; research instincts; experimental design; statistical reasoning; reproducibility.

WHAT YOUR WORK SHOULD SHOW

GitHub and shipped projects; portfolio quality; reliable execution; collaboration; documentation; clear technical communication.

2–3 PARTICIPANTS PER TEAM

Solo participation is not allowed. Every registered member must actively contribute.

TEAM FORMATION & ELIGIBILITY

Finalize the member list by the application deadline, September 16, 2026 at 11:59 PM IST, unless an earlier challenge-specific selection deadline is communicated. Replacement after selection requires organizer approval. Applicants must provide accurate profiles, meet disclosed challenge and service eligibility requirements, and be able to attend the final if selected.



07 / SELECTION PIPELINE

Earn the invitation.

A research funnel built around evidence, not presentation polish alone.

0

APPLICATION

Profiles, GitHub, portfolio, technical background, previous projects, relevant AI/ML work, team statement, motivation and evaluation interests.

1

INITIAL SCREEN

Technical and portfolio review assesses execution, reasoning, documentation and fit for the evaluation challenge.

2

REMOTE QUALIFIER

Selected teams tackle evaluation engineering: benchmarks, model comparison, agents, datasets, failure analysis or adversarial testing.

3

FRONTIER EVALUATION TRIAL

High-performing teams receive a deeper confidential challenge. Methodology, correctness, rigor, engineering and findings determine progression.

4

CONFIDENTIAL TECHNICAL REVIEW

Organizers assess submissions privately. Other teams' submissions, scores, judge notes, hidden answers and unpublished infrastructure remain private.

5

GOA FINALIST SELECTION

Only 15 teams receive invitations. The first announcement names the Top 10; further places follow extended technical review.

APPLICATION → QUALIFIER → FRONTIER TRIAL → TECHNICAL REVIEW

TOP 10 ANNOUNCED → EXTENDED REVIEW → 15 GOA FINALISTS → THE FINAL



08 / THE GOA COHORT

15

FINALIST TEAMS

30–45 BUILDERS

ONE FRONTIER AI FINAL

GOA / INDIA

T / 01	T / 02	T / 03	T / 04	T / 05
T / 06	T / 07	T / 08	T / 09	T / 10
T / 11	T / 12	T / 13	T / 14	T / 15

15 teams will earn the right to evaluate at the frontier. Every invitation is earned through technical review. Finalists may meet, network and participate in official programming together.



09 / OFFICIAL TIMELINE

The path to Goa.

NOW

REGISTRATION OPEN

Submit your team application and supporting work.

16 SEP

APPLICATIONS CLOSE

September 16, 2026 · 11:59 PM IST.

18 SEP

TOP 10 ANNOUNCED

September 18, 2026 · First major finalist selection announcement.

REVIEW

REMAINING FINALIST SLOTS

Remaining finalist slots will be announced following extended technical review.

01–15 OCT

GOA FINAL WINDOW

October 1–15, 2026 is the expected event window. The final runs approximately 3–4 days; exact dates follow final confirmation.

Finalists should maintain availability during the announced October 1–15 window until exact dates are confirmed. Stage-specific instructions and deadlines will be communicated to participating teams.



10 / THE RESIDENTIAL FINAL

Goa. A temporary frontier research lab.

A focused residential environment for testing systems, challenging assumptions and turning unexpected behavior into reproducible evidence.

LOCATION

Goa, India. The exact venue will be communicated privately to selected finalists. No venue has been specified in this prospectus.

DATES & DURATION

Expected window: October 1–15, 2026. Approximately 3–4 days. Exact dates are communicated after final confirmation.

FINALIST COMMITMENT

Travel physically to Goa and attend the complete official event, including mandatory sessions, judging and final presentations. Bring a development laptop, maintain required software environments, present valid identification, follow venue rules and confidentiality requirements, complete technical deliverables and retain reimbursement records.

WHAT TO BRING

Laptop and charger · valid government ID · student ID where applicable · adapters · headphones · personal medication · GitHub access · permitted API credentials · development environment · travel documents · reimbursement receipts · personal essentials.

FULLY FUNDED FOR SELECTED FINALISTS / SUBJECT TO POLICY



11 / INSIDE THE FINAL

The frontier opens.

Illustrative program sequence; the final schedule will accompany confirmed event dates.

DAY 0**ARRIVAL / ORIENTATION**

Arrivals, check-in, registration, environment setup, team and confidentiality briefings, infrastructure testing and welcome session.

DAY 1**THE FRONTIER OPENS**

Confidential challenge release, evaluation access, dataset inspection, benchmark design, baselines and experiment planning.

DAY 2**BREAK THE MODEL**

Adversarial evaluation, model tests, agent experiments, failure discovery, benchmark iteration and mentor/reviewer checkpoints.

DAY 3**THE FINAL EVALUATION**

Experiment freeze, reproducibility checks, technical reports, private judging, demonstrations, research presentations and results.

Departure follows the confirmed schedule. An optional Day 4 may accommodate departure or closing programming if scheduling requires; this is not a confirmed additional competition day.



12 / JUDGING

100 points. Evidence over spectacle.

25

Evaluation Methodology

Question quality, metrics, baselines and validity of the evaluation.

20

Technical Implementation

Correct, robust engineering and dependable evaluation infrastructure.

15

Scientific / Experimental Rigor

Controls, uncertainty, sample design and honest interpretation.

15

Originality & Insight

New approaches and meaningful insights into model behavior.

10

Reproducibility

Traceable artifacts, documented environments and rerunnable results.

10

Model Failure Discovery / Evaluation Value

Useful failures, actionable findings and credible evaluation value.

05

Communication & Technical Presentation

Clear technical report, demonstrations and defensible claims.

Tie-breaking priority: Evaluation Methodology → Scientific Rigor → Reproducibility → Technical Implementation → private judge deliberation if still tied. Judges disclose conflicts and recuse where appropriate; organizer-appointed reviewers assess the submitted evidence.



13 / PRIZE POOL

10 LAKH. THREE WINNERS.

₹10,00,000 in top-three monetary prizes · funded directly by the organizing team

1ST PLACE

₹ 5,00,000

2ND PLACE

₹ 3,00,000

3RD PLACE

₹ 2,00,000

Prizes and finalist benefits are funded by the HACK THE FIRST AGI 1.0 organizing team unless otherwise announced.

Monetary amounts are team prizes. Awards require verified identity, eligibility and compliance with competition rules. Applicable tax withholding and reporting may apply; recipients are responsible for their own tax obligations. Payment documentation and distribution arrangements will be communicated by organizers.



14 / BEYOND THE PODIUM

Six months. Keep building.

RANKS 4–15

6 months of organizer-funded Codex access

Eligible recipients from Rank 4 through Rank 15 will receive six months of organizer-funded access to Codex, subject to account eligibility, product availability, applicable service terms, geographic availability, and the reimbursement/provisioning mechanism communicated by the organizing team.

HOW THE BENEFIT MAY BE PROVIDED

Direct organizer purchase, reimbursement, eligible subscription credits or another lawful provisioning method. Recipient-level allocation and account arrangements will be communicated by the organizing team; no particular plan or usage allowance is specified here.

No corporate affiliation is implied. The Codex benefit is independently funded by the HACK THE FIRST AGI 1.0 organizing team. OpenAI is not a sponsor. Product references do not imply sponsorship, endorsement, partnership, judging or official support by their respective companies.



15 / FINALIST FUNDING

Fully funded. Clearly defined.

The organizing team funds selected finalists' eligible travel, stay, official-event meals, approved local transport and official participation costs, subject to reimbursement rules and prior approvals.

ELIGIBLE TRAVEL

Reasonable round-trip transportation to Goa: economy-class flights, AC train, intercity bus or an approved combination of modes.

CAPS & APPROVAL

Organizers may set reasonable caps based on departure city, booking timing, route availability, event dates and comparable market pricing.

BEFORE YOU BOOK

Obtain the communicated travel policy, approved dates and applicable caps before committing to purchases. Premium upgrades are not automatically reimbursed. Organizer-approved accommodation follows the one-room-per-team policy on page 16.

NORMALLY EXCLUDED

Business- or first-class upgrades; luxury transfers; unrelated tourism travel; personal shopping; alcohol; minibar charges; leisure excursions; companion travel for non-participants; avoidable change penalties; unrelated hotel nights.

Direct booking: Organizers may book hotels, group transport, flights, airport transfers or meals directly. An equivalent duplicate purchase cannot be reimbursed unless expressly approved.

**16 / ACCOMMODATION & EXPENSES**

One team. One room.

Standard organizer-funded accommodation covers one room per finalist team of 2 or 3.

Separate rooms are not automatically funded for each participant. Extra rooms requested purely for personal preference are normally self-funded. Accessibility needs, medical requirements, safety considerations, gender/privacy concerns and other reasonable circumstances may qualify for organizer-approved exceptions, handled privately.

ONE TEAM FINANCE REPRESENTATIVE

A team may nominate one participant to pay eligible costs and consolidate documentation. For example, one member of a three-person team may book eligible travel, pay event expenses, collect invoices and receipts, maintain the expense statement and submit the team's reimbursement package.

After approval, organizers may reimburse the representative for the approved consolidated amount. The representative is responsible for any internal distribution. All team members should acknowledge the consolidated statement before payment.

DOCUMENT EVERY CLAIM

Include, where applicable: invoice or tax/GST invoice; receipt; proof of payment; booking confirmation; passenger details and participant name; booking date and travel date; route; expense description and reason; merchant; currency; amount.

CLAIM REVIEW & INTEGRITY

Organizers may reject duplicate, unverifiable, unrelated or non-participant expenses, manipulated invoices, fraudulent receipts and luxury upgrades outside policy. Fraudulent reimbursement submissions may result in immediate disqualification and recovery of funds.

Keep original records. Do not submit duplicate claims for items booked directly by organizers. Reimbursement timing and submission channels will be communicated to finalists.



17 / OFFICIAL COMPETITION RULES

Protect the experiment.

01 / REGISTRATION & IDENTITY

Submit accurate information by September 16, 2026, 11:59 PM IST. Teams have 2–3 active members; no solo entry. Identity and eligibility may be verified. Finalize membership before the applicable selection deadline; post-selection replacement needs approval.

02 / RELEASE & DELIVERABLES

Follow each challenge's release instructions, access restrictions, deadlines and submission format. Submit source, environment/dependency details, methods, data provenance, run instructions, results and logs sufficient for review. Disclose limitations and permitted human interventions.

03 / TOOLS & THIRD PARTIES

AI assistants, IDE agents, LLM APIs, research tools, inference APIs and open-source models are allowed unless restricted by a challenge. Teams must understand their code, verify generated output and disclose AI use when requested. Authorship responsibility remains with the team.

04 / LICENSES & SECURITY

Comply with open-source, dataset and third-party service licenses and terms. Use paid APIs only as permitted; unapproved API spending is not automatically reimbursed. Protect API keys, avoid committing secrets and respect access and rate limits. Test only authorized systems.

05 / RESEARCH INTEGRITY

Prohibited: plagiarism, stolen code, fabricated results or benchmark scores, hidden manual intervention presented as automation, unauthorized dataset leakage, leaked test answers, infrastructure manipulation, falsified latency, fabricated screenshots, forged logs, fake API responses, another team's work and prohibited deliberate benchmark contamination.

06 / CONFIDENTIAL MATERIAL

Private submissions, hidden datasets, benchmark internals, unpublished scoring mechanisms, security-sensitive findings, judge feedback, private evaluation environments and challenge-specific material remain confidential unless organizers release them. Other teams' private scoring and internal judge notes are not shared.

07 / NORMAL INTERACTION

Finalists may interact, network, collaborate socially and participate in official programming. Teams may discuss their own work subject to challenge-specific confidentiality rules. Social interaction does not authorize sharing restricted artifacts or competing submissions.

08 / JUDGING & APPEALS

The 100-point rubric and tie-break order appear on page 12. Report conflicts privately; conflicted judges should recuse. Appeals must identify a procedural or factual error with evidence through the communicated organizer channel and deadline. Organizers review claims privately; rule breaches may lead to score adjustment or disqualification.

**18 / DATA, IP & PARTICIPATION**

Trust is part of the work.

09 / PERSONAL DATA

Use personal data only for disclosed event purposes, with data minimization, informed disclosure, security, confidential handling and appropriate retention. Deletion applies where legally and operationally applicable. Participant personal information is not sold.

10 / TECHNICAL ARTIFACTS

Challenge artifacts can include benchmark results, model outputs, annotations, adversarial examples, metrics, failure cases and experiment logs. Broader research or commercial use beyond competition administration must follow terms communicated to participants.

11 / INTELLECTUAL PROPERTY

Teams retain independently created original IP unless a specific challenge discloses different terms before participation. Organizers receive only the limited rights needed to evaluate, judge, verify reproducibility, maintain competition records, announce winners and showcase approved project descriptions.

12 / ETHICS & DISCLOSURE

Use authorized data and ethically designed experiments. Respect privacy and avoid harm to people or third-party systems. Report security-sensitive findings privately through the organizer's designated process; do not publish confidential exploit details or personal data.

13 / CONDUCT & SAFETY

Harassment, discrimination, threats and unsafe behavior are prohibited. Follow venue rules and safety instructions. Report concerns privately to organizers. Accessibility, medical, safety and privacy requests are handled confidentially and considered reasonably.

14 / PHOTOGRAPHY & MEDIA

Organizers will communicate photography and recording arrangements. Raise privacy or accessibility concerns privately. Do not record restricted screens, hidden data, private reviews or confidential materials. Public project descriptions require approval under the applicable confidentiality terms.

15 / ATTENDANCE & BENEFITS

Finalists must attend mandatory programming and complete required deliverables. Travel, room allocation, meals, claim verification and direct booking follow pages 15–16. Prizes and benefits depend on verified eligibility and rule compliance; taxation and product service terms apply.

16 / CHANGES & DISCRETION

Force majeure, safety needs or operational constraints may require schedule, venue or program changes, postponement or cancellation. Organizers will communicate material changes and effects on approved bookings and benefits. Reasonable discretion may resolve unforeseen issues fairly; challenge-specific terms must be communicated before participation.



19 / FREQUENTLY ASKED QUESTIONS

Before you apply.

What is Hack the First AGI 1.0?

A frontier-AI evaluation, model-behavior and engineering hackathon.

Who can participate? Can students apply?

Yes. Students, researchers, engineers and AI evaluators who meet the disclosed eligibility requirements can apply.

Can I participate solo?

No. Teams must contain 2–3 members.

Do we need prior AI-evaluation experience?

No fixed experience threshold is specified. Show technical ability, careful evaluation thinking and evidence of execution.

Can members come from different institutions?

Yes. Teams may span colleges or organizations.

How many teams reach Goa?

15 teams, comprising 30–45 participants.

Is Goa travel funded?

Yes, subject to the official finalist travel policy.

Are accommodation and food funded?

Yes. Team-based accommodation and official-event meals are covered according to the finalist policy.

Does every member get a separate hotel room?

No. Standard funding provides one room per team, subject to legitimate accessibility, privacy, medical, safety or similar exceptions.

What if one teammate pays for everybody?

Nominate a Team Finance Representative. Submit a documented, team-acknowledged statement; approved consolidated claims may be paid to that representative for internal distribution.

What are the prizes?

Team prizes: ₹5,00,000 / ₹3,00,000 / ₹2,00,000 for first / second / third.

What do ranks 4–15 receive?

Six months of organizer-funded Codex access for eligible recipients, subject to the conditions on page 14.

Is OpenAI a sponsor?

No. The Codex benefit is independently organizer-funded. Product names do not imply sponsorship or endorsement by their respective companies.

When is the application deadline?

September 16, 2026 at 11:59 PM IST.

When are the first results? How many teams?

September 18, 2026. The Top 10 teams are announced initially.

What about the other five finalist slots?

Remaining finalist slots will be announced following extended technical review.

When is Goa?

Within the October 1–15, 2026 window. Exact dates and the private venue are communicated to finalists after confirmation.



20 / APPLY NOW

THINK YOU CAN EVALUATE THE FRONTIER?

15

FINALIST TEAMS



3

PODIUM TEAMS



1

CHAMPION

APPLICATIONS CLOSE

SEPTEMBER 16, 2026

11:59 PM IST

FIRST FINALIST ANNOUNCEMENT: SEPTEMBER 18, 2026

Top 10 announced first. 15 teams will reach Goa.

REGISTER NOW — OFFICIAL LINK TO BE INSERTED

[OFFICIAL REGISTRATION URL]

REGISTRATION LINK / QR PENDING ORGANIZER INSERTION