BOLD LAUNCH FESTIVAL · OXFORD · 3 AUGUST 2026

Frontier Research Outside the Big Labs

Zhengyao Jiang Cofounder & CEO, Weco AI

BACKGROUND

What we built in 3 years outside

- 2024

  AIDE — autoresearch agent

  4× the medal rate of the next-best agent, in OpenAI's MLE-Bench paper

- APR 2026

  Aiden — top contributor in OpenAI's hiring challenge

  7 records — the next best contributor: **3**

- JUL 2026

  AIDE² — effective recursive self-improvement

  8 days to build a better agent than we tuned for **2 years**

THE WHOLE TALK IN ONE PICTURE

The trade

More choice

your problem

your direction

what counts as progress

More extreme reward

closer to binary

mostly nothing,

occasionally a lot

THE FREEDOM SIDE

You choose your direction, and how you're graded

You choose what to work on.

You design how progress is measured.

Ours: public benchmarks, and whether people actually use our code.

THE PRICE

Outside, the reward is extreme

Inside — there is a floor.

Outside — nothing averages you.

The best work gets adopted by everyone. Useless work, by almost nobody.

Extreme success and extreme failure, interleaving.

THE FLIP SIDE

But you get credit assigned once it worked

When it works, the credit points at you.

Nothing to share it with. Nothing to dilute it.

This is what compounds.

METHODOLOGY — ONE STORY, N = 1

Our play, in one picture

01

Enter early

02

Accumulate

03

Cash the window

04

Compound the trust

In one sentence: trade early roots for credibility.

STEP 1

Enter early

2023: agents doing ML research — before the field had a name.

“Too early to productize” is what kept it available.

Cheap enough for two PhD students.

STEP 2

Persist through the silent years

No product. No feedback. Keep working.

What you are building is inventory — **results, code, understanding.**

STEP 3

Cash the window

The window only pays for what you already have.

Move immediately. Ship things people can use.

Code they can run. Results they can check. A story simple enough to travel.

STEP 4

Trust compounds into your own brand

MLE-Bench
top-scoring agent

“Autoresearch”
the field gets a name

Aiden
top contributor, OpenAI's hiring challenge

AIDE²
we opened this window ourselves

Each window costs less and pays more.

THE LATEST WINDOW

AIDE²: the first effective RSI

100
consecutive self-rewrites

8 days
unattended

9 / 10
changes rejected by a held-out gate

2 years
of hand-tuning, beaten

And it generalizes — the gains hold on benchmarks it was never selected on.

THE WARNING

This worked for me — run your own eval

1 What is your reward signal — and who chose it?

2 Where does your credit accumulate?

3 Can you get through the silent years?

You become what your reward signal measures.

Choose your signal.

Thank you · Zhengyao Jiang · Weco AI