八一八Anthropic和有效利他主义（EA）3.Anthropic核心圈 番外篇

原创 面包人小模型 一个人读书写字 2026年11月5日 08:15 上海

搜集了不少素材，正文没收录的整理成一篇。

01 PART Holden是哈佛2003届的学生

HARVARD REUNION

这一届学生分别在5年和10年后举行了校友重聚的活动，holden两次都作为代表发言。发言风格属于Harvard Lampoon 式-比较没品和嘴贱的，内容大部分不适合发表，只引用一段比较有代表性的2008年校友重聚活动演讲的内容。

https://www.harvard03.com/holdens-2008-speech

> *调查里还有一个很有趣的发现：**金融圈的人和非营利组织的人简直是宿敌，命中注定要互相毁灭。***
> *如果把各种职业大类放在一起比较，你会发现，金融从业者在某项特征上排第一的地方，非营利从业者几乎总是排最后，反过来也一样。比如，非营利组织的人最担心全球变暖，而金融圈的人最不担心。*
> *金融从业者和其他人差不多，都挺容易怀念大学生活；但非营利组织的人在这一项排最后。最后，再来看真正检验人格的一道题：如果要在“当公立学校老师”和“靠犯罪赚八位数”之间选，非营利组织的人最愿意去当老师。至于金融圈——我知道这一定会让大家非常震惊——**他们最有可能选择钱和犯罪。***
> *不过，在这一切表面差异下面，这两个群体还是有一个共同点。*
> *如果看“给自己的道德品质打最高分”的比例，金融圈排倒数第一，非营利组织的人排倒数第二。*
> *哎呀。*
> ***金融人、公益人，你们有没有想过，也许你们俩其实没那么不一样？要不干脆联手统治世界算了？***

03年的这段演讲是有视频的，但是视频里实际内容和文字稿是不一样的。并没有上面这段，所以就没有收录到正文里了。



02 PART Holden自导自演事件

METAFILTER

metafilter的高楼还有更典的：

Holden 当场想给 MetaFilter 捐钱“补偿”。 他被抓后自己在 thread 里说，自己钱不多、不是“Cristal set”，但愿意从自己口袋给 MetaFilter捐款，表示不是为了换取停止批评，只是想补偿违规。MetaFilter 用户立刻把这个理解成“你这是想拿钱把事摆平吗？”导致帖子越吵越长了

他还拿“sleep deprivation”解释判断失误。 这直接触发了 thread 里最著名的一轮暴打，有个用户写了超长 rant，大意是“我也没睡，你别拿缺觉当借口”。

03 PART Amodei 兄妹都是犹太人

THE AMODEIS



Alex Kantrowitz的节目 ，2025年7月31日播出的Anthropic CEO Dario Amodei: AI's Potential, OpenAI Rivalry, GenAI Business, Doomerism

母亲 **Elena Engel 是 Jewish American**，父亲 Riccardo Amodei 是意大利裔。

Jewish Insider 2026 年的人物稿，标题/正文直接把 **Dario 和 Daniela 两个人都称为 Jewish**，并再次写了他们母亲是 Jewish

04 PART Avital Balwit

AVITAL BALWIT

Dario唯一的幕僚长；Leopold的老婆



Avital 当时在 University of Virginia 双主修：**Political and Social Thought和Cognitive Science**

她拿到 2021 年 Rhodes Scholarship，计划去 Oxford 读 **BPhil in Philosophy**。但她对“去 Oxford 干什么”的描述非常具体，不是泛泛地说学哲学：

> 研究 **existential risk**，并与 Oxford 的 **Future of Humanity Institute（FHI）** 和 **Global Priorities Institute（GPI）** 一起工作。

她自述要去Oxford的理由：**Toby Ord**、**Nick Bostrom**、**William MacAskill**、**Hilary Greaves** 都在那里，而且 Oxford 集中了这个当时仍然很小的研究领域里最多的一批重要研究者。她希望参加他们的 seminar、做 RA、认识同领域学生。

2022 年 11 月 11 日：FTX Future Fund 团队集体辞职。联名声明里明确有 **Avital Balwit**、**Nick Beckstead**、**Leopold Aschenbrenner**、**Ketan Ramakrishnan**。

2023 年 6 月之前：Avital 已经是 **Anthropic Communications Lead**。Asterisk 2023 年 6 月刊的作者简介直接这么写，而且她正文也说 *“I happen to work at Anthropic”*。

05 PART Searching for God in Silicon Valley –the free plus

FAITH



我成长于旧时代的边缘。我生来就是犹太人，父母给我取名为 Avital，就是为了让我永远不要忘记自己的犹太身份。但在我的童年时期，四处寻找教堂的，恰恰是我的母亲——家里有犹太血统的那一方。她发现，相比犹太会堂，基督教教堂更加友善，也更加便宜。其他犹太人会因为她离开社群太久而对她指指点点，而基督徒却没有这样的期待和要求。我便跟着她一起去教堂。我喜欢教堂里的彩色玻璃窗，喜欢唱圣歌，也喜欢那种试图触及某种比自身更加伟大、超越自我之物的感觉。但我始终无法真正跨越信仰的那道门槛。我一直把这归因于自己并非在宗教环境中长大——我觉得，一个人离开信仰之后再寻找归途，要比从未离开过困难得多。七岁时，我主动提出，希望在当时我们参加礼拜的圣公会教堂接受洗礼。吸引我的是那种敬畏感与共同体归属感，还有一个即使在当时就已经萌生的想法：接受洗礼是一种宣示自己严肃对待道德的方式。后来，那位牧师在一次布道中抄袭了别人的讲道词。我母亲指出这件事后，他的反应非常糟糕，于是我们便不再去那座教堂了。尽管如此，我仍然为自己的犹太血统感到自豪，也觉得自己对延续这个规模不大、且日渐式微的民族负有某种责任。这种牵绊让我始终无法沿着任何一条宗教道路走得太远，也让我一直处于某种信仰的十字路口。

节选

06 PART Holden 当然也是犹太人

HOLDEN KARNOFSKY

Holden 1981 年的出生公告直接写的是“son of Rabbi and Mrs Keith Karnofsky”



VOLUME LX, NUMBER 28 THURSDAY, SEPTEMBER 15,

New Jewish URI Chaplain

Rabbi Keith M. Karnofsky has been named chaplain to the Jewish students at the University of Rhode Island.

A native of Buffalo, New York, Rabbi Karnofsky earned an undergraduate degree in English from Columbia College in New York City. He holds graduate degrees in religion and psychology from Temple University, and he has completed all course requirements for a PhD in religion and psychology.

He was ordained in 1975 by the Reconstructionist Rabbinical College in Philadelphia, Pennsylvania. Most recently he was rabbi of the Jewish Reconstructionist Society of Brooklyn, New York.

He has also been the guest rabbi at Congregation Beth Israel in Media, Pennsylvania, and a counselor to Jewish inmates at the State Correctional Institution in Graterford, Pennsylvania. In addition, he was a teaching assistant in the department of religion at Temple University.

Rabbi Karnofsky is a certified FAA flight instructor, and for the past 18 months

RABBI KEITH M. KARNOFSKY

has served as first officer with Pilgrim Airlines of New London, Connecticut.

He and his wife, Andra, are residents of Pawcatuck, Connecticut.

Holden的父亲——Keith M. Karnofsky是一名是 rabbi。他年轻时本身就是 **宗教 + 激进社会行动**的混合体。1968 年 Columbia 学运期间，他参与抗议、被捕，当时正在 Hebrew Union College 学做 rabbi；*TIME* 采访他的原话：

> Activism is a Jewish thing.

Holden的父亲所在的 Reconstructionist Judaism 非常明确地认为：**“Jews are God’s chosen people”**这个概念应该从神学和礼仪里删除。

原因正是它容易隐含“我们比别人特殊/优越”。Reconstructionist prayer book 甚至主动把传统的“who has chosen us from among all peoples”改掉。

替换的概念是：Vocation

> *我们不是因为上帝偏爱我们而特殊；但我们作为一个共同体，*
> ***选择承担一种特殊的伦理使命和责任。***

不是 God chooses us，而是**我们选择承担这个 vocation。**

但是Holden 自己有没有谈过父亲、Jewish upbringing、vocation、religion 对自己道德观形成的影响。甚至他还把自己的中间名字删掉了。

Geremiah/Jeremiah是一个常见的带宗教色彩的名字。但是**Jeremiah 并不是一个“看到就必然是犹太人”的名字**。它来自 Hebrew Bible，在基督教世界也极其常见。

07 PART 我们和EA是盖棉被纯聊天的关系

EA AND ANTHROPIC

2025年3月28日，《WIRED》记者 Steven Levy 对 Anthropic 的长篇专访

Daniela Amodei：

> *我不是有效利他主义方面的专家。我不认同这个标签。我的印象是，这个词已经有点过时了。*

她的英文原话包含 ”I’m not the expert on effective altruism” 和 ”a bit of an outdated term”。

Amanda Askell：

> *我确实在这里见过一些有效利他主义者，但这并不是公司的一种组织性主题。*

她的原话结尾是 ”it’s not a theme of the organization or anything”。

08 PART Anthropic历史LTBT名单

LTBT

TBT 历任人员完整名单

姓名	加入时间	离开时间	身份背景
Neil Buddy Shah	2023-06	仍在任	CHAI CEO、前 GiveWell 董事总经理；LTBT 主席
Jason Matheny	2023-09前	2023-12	RAND CEO，CSET 创始主任
Paul Christiano	2023-09前	2024-04	ARC 创始人、AI Alignment 研究者
Kanika Bahl	2023-09前	2026-01	Evidence Action CEO
Zach Robinson	2023-09前	2026-01	Effective Ventures US 前临时 CEO
Richard Fontaine	2025-04	仍在任	CNAS CEO、前三方委员会北美执行主任
Mariano-Florentino Cuéllar（Tino）	2026-01	2026-08-04	Carnegie 前总裁、前加州最高法院法官
Ben Bernanke	2026-06	仍在任	前美联储主席、Brookings 杰出研究员
Christopher Stubbs	2026-10	仍在任	哈佛物理学教授、JASON 成员

基本都八卦过了，没八卦过的就只有三个（Paul在下一篇）

- **Mariano-Florentino Cuéllar（Tino）**
  2026 年 1 月：以 Carnegie Endowment 总裁、前加州最高法院法官的身份加入 LTBT。
  2026 年 8 月 4 日：辞去 LTBT 职务，直接加入 Anthropic，成为首任 Chief Global Affairs Officer（首席全球事务官），负责政府关系和国际政策。



Holden Karnofsky 在2024年4月29日正式离开 Open Philanthropy，加入**卡内基国际和平基金会（Carnegie Endowment for International Peace，简称 CEIP）**，担任 Visiting Scholar（访问学者）。

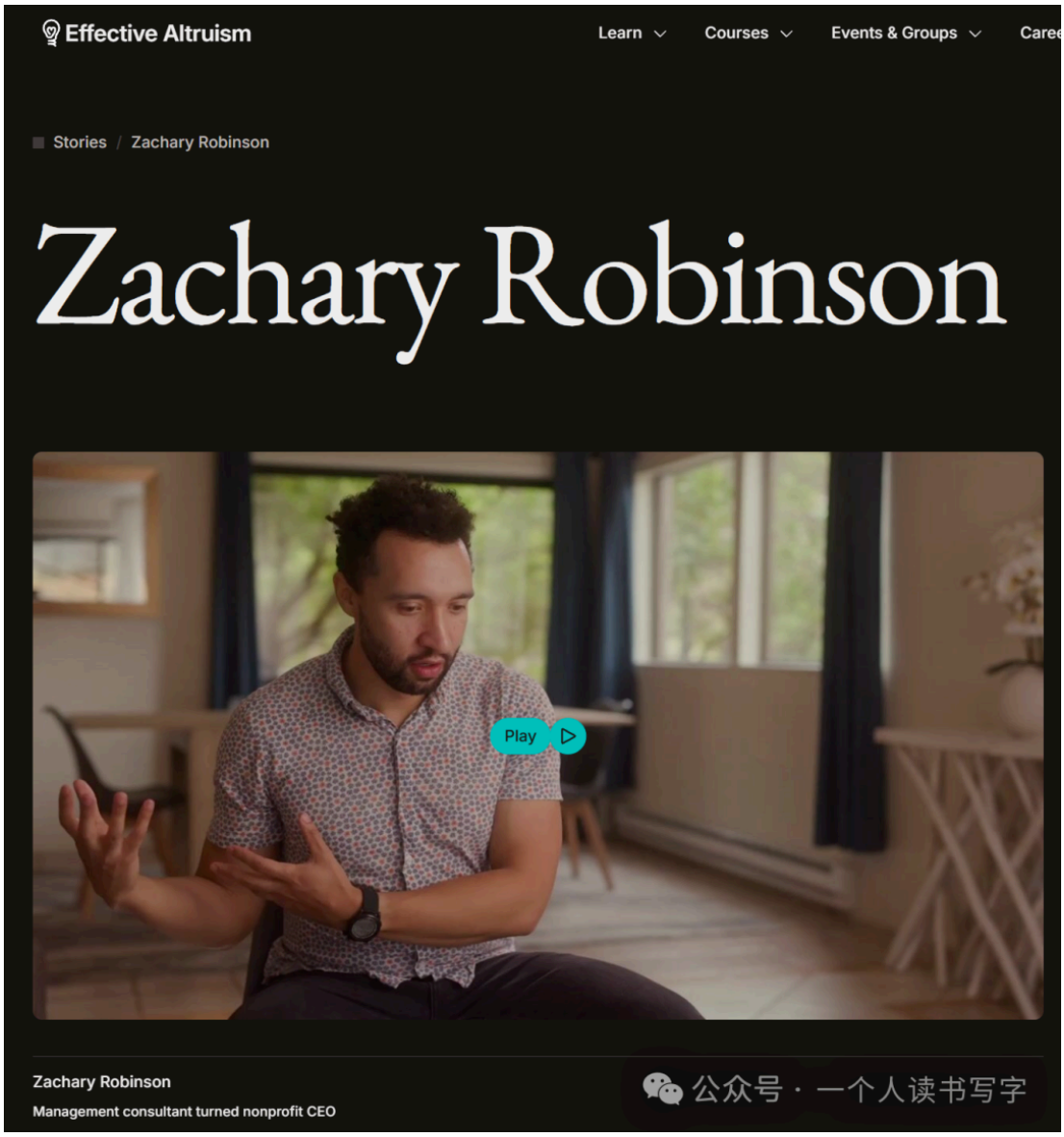
Holden 在离职声明里说自己希望将 AI risk 的讨论扩展到原本的社群之外，构建一个更大的 big tent。

他还特别提到了两个人：

- **Mariano-Florentino Cuéllar（Tino）**：当时的 Carnegie 总裁，曾任加州最高法院法官，并在多届美国总统政府中任职。Holden 认为他重视高级 AI 风险。
- Matt Sheehan：Carnegie 的中国 AI 研究专家。Holden 明确认为研究中国 AI 生态非常重要。

Openphil一直给Carnegie打钱，光是26年就打了$5,975,000.

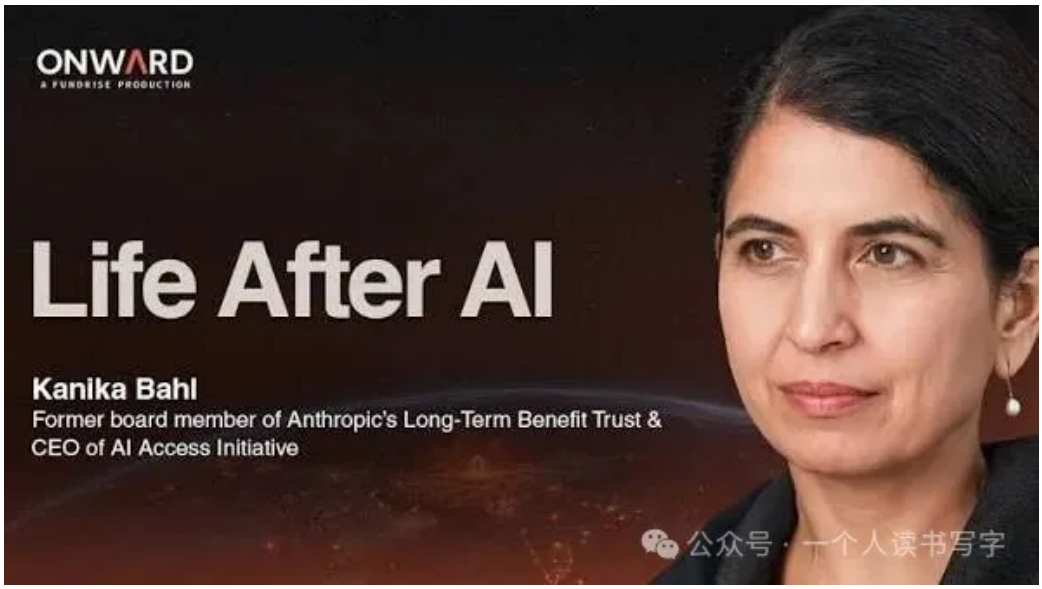
- Zachary Robinson 铁血EA人



时间	职位	具体职责
早期	Bain 顾问、Ivy Research Council 产品与战略主管	商业咨询和战略
至2023年初	Open Philanthropy Chief of Staff	管理传播、资助相关团队，制定机构政策，探索新的全球健康资助方向
2023年1月	Effective Ventures US 临时 CEO	接管该 EA 组织的运营
2023年4月	Effective Ventures US 董事	参与组织治理
2024年2月至今	Centre for Effective Altruism CEO	负责 CEA 及其 EA 社群建设业务

另外，Zach 在2023年初接手 Effective Ventures US，是在 FTX 崩盘之后。到2023年4月，EV US 董事会还包括 Nick Beckstead 和另一位 Open Phil 员工 Eli Rose。

- Kanika Bahl：



- 2005—2010年：Clinton Health Access Initiative 执行副总裁
- 2010—2015年：Results for Development 董事总经理
- 2017—2026年4月：Evidence Action CEO
- 2026年起：AI Access Initiative CEO

Givewell的 Relationship Disclosures页面节选：

Current staff and Board members
目前的员工和董事会成员

- **Clinton Health Access Initiative (CHAI) and Evidence Action.** GiveWell has recommended a number of grants to **CHAI** and **Evidence Action**. As of February 2025, Dr. Neil Buddy Shah (the CEO of CHAI) and **Kanika Bahl** (the CEO and President of Evidence Action) serve on Anthropic's **Long-Term Benefit Trust**. Daniela Amodei is the President of Anthropic and a member of Anthropic's Board of Directors; she is also a member of GiveWell's Board of Directors. As trustees, Buddy and Kanika have a role in selecting a subset of Anthropic's Board of Directors (not including Daniela). Daniela recuses herself from any decisions by GiveWell's Board related to Buddy, Kanika, and the organizations they lead.

  Clinton Health Access Initiative（CHAI）和 Evidence Action 组织。GiveWell 推荐了一些资助给 CHAI 和 Evidence Action 的项目。截至 2025 年 2 月，CHAI 的 CEO Neil Buddy Shah 和 Evidence Action 的 CEO 兼总裁 Kanika Bahl 都是 Anthropic 公司的长期利益信托成员。Daniela Amodei 是 Anthropic 公司的总裁，同时也是 Anthropic 公司董事委员会成员；她也是 GiveWell 公司董事委员会的成员。作为受托人，Buddy 和 Kanika 在选拔 Anthropic 公司董事委员会成员时有一定的影响力（不包括 Daniela）。Daniela 拒绝参与任何与 Buddy、Kanika 以及他们所领导的组织相关的 GiveWell 公司董事委员会的决策。

公众号 · 一个人读书写字

Kanika 2026年创立的新机构叫 AI Access Initiative（2AI）

顾问名单只有三个人



Dario Amodei

Anthropic CEO，2AI 顾问。



Kent Walker

Google / Alphabet 全球事务总裁，2AI 顾问，同时也在 Evidence Action 董事会。



Michael Kremer

2019年诺贝尔经济学奖得主，长期从事发展经济学与减贫研究。

公众号 · 一个人读书写字

这个跨界程度我不好评价。毕竟Dario是顾问，一切皆有可能。

09 PART Holden在S-1里只出现过2次

TENDER OFFERS

一次卖股票，
一次介绍他的薪酬水平和普通员工差不多（有点此地无银三百两的味道）。

2026年4月，Dario、Daniela 和 Holden 都参与出售了 Anthropic 股票，没有披露具体每个人卖了多少。

这一轮参与的员工加起来套现了$46.61亿，与G轮相同的估值基准

2月G轮和4月的老股交易：

	G轮融资2月	4月老股交易
交易对象	Anthropic新发行的优先股	老股东持有的A类普通股
资金流向	进入公司	进入出售股票的股东手中
金额	$302亿	$46.61亿
估值参考	投前$3,500亿	与G轮相同的估值基准
主要用途	公司融资	老股东变现、投资者取得股权

S-1里相关内容翻译：

股份要约收购（Tender Offers）

- 2026年4月

2026年4月，我们协助开展了一项股份要约收购。在该交易中，一个由股东组成的财团发起要约，以每股 25.9136美元 的现金价格，向我们的部分证券持有人收购公司A类普通股。

公司高管兼董事会成员 Dario Amodei 和 Daniela Amodei、公司高管 Krishna Rao，以及 Daniela Amodei 的配偶 Holden Karnofsky，均在此次要约收购中出售了所持有的公司A类普通股。

此次要约收购共计售出 179,883,540股 A类普通股，总收购金额为 4,661,437,298美元（约46.61亿美元）。

- **2025年6月**

2025年6月，我们发起了一项股份要约收购，以每股 5.61美元 的现金价格，向我们的部分证券持有人收购公司A类普通股。

公司高管兼董事会成员 Dario Amodei 和 Daniela Amodei 均在此次要约收购中出售了所持有的公司A类普通股。

此次要约收购共计售出 27,942,000股 A类普通股，总收购金额为 156,716,898美元（约1.57亿美元）。

喜欢作者

anthropic/EA · 目录

上一篇
八一八Anthropic和有效利他主义（EA）3.Anthropic核心圈

下一篇
Jensen和Dario的恩怨