

中国民主季刊

CHINA

2026年10月

第4期

卷4

VOL. 4

No.4

OCT. 2026

JOURNAL

OF

DEMOCRACY

人工智能时代的民主机遇 柯珍雅 萧强 亚历克斯·格莱德斯坦 季志强 刘爽

“捂嘴”、“洗脑”：中国人工智能审查能否拆解 汉娜·西贝里 / 钟伟锋

审查进了流水线：看清中国人工智能模型里的
政治控制

区倬僊 / 季志强

十年间，专制政治背景下中国女权运动的发展

吕频 / 艾欣

Z世代为何崛起

艾莉卡·切诺维斯 马修·塞布尔



中国民主转型研究所
INSTITUTE FOR CHINA'S DEMOCRATIC TRANSITION

中国民主季刊

CHINA
JOURNAL
OF
DEMOCRACY



中国民主转型研究所
INSTITUTE FOR CHINA'S DEMOCRATIC TRANSITION

中国民主季刊

CHINA
JOURNAL
OF
DEMOCRACY

主编 Editor in Chief

顾问编辑 Advisory Editor

编辑 Associate Editors

技术、推广总监 Tech & Promotion Director

运营协调 Operational Coordinator

王天成 Wang Tiancheng

钟伟锋 Zhong Weifeng

余浩风 Yu Haofeng

李聿脩 Li Yuxiu

杨子立 Yang Zili

冯华 Feng Hua

主办、出版 Organization & Publisher

中国民主转型研究所 Institute for China's Democratic Transition

电邮 Email: cjd@chinademocrats.org

网址 Website: <https://chinademocrats.org/>

电子版刊号 ISSN (electronic): 2996-2420

印刷版刊号 ISSN (print): 2996-2412

CC BY-NC-ND 开放资源

不得修改和用于商业目的

转载须注明作者、出自本刊

国际顾问委员会 (International Board of Advisors)

白夏 Jean-Philippe Béja	盖思德 Roger Garside	苏晓康 Xiaokang Su
陈健民 Kinman Chan	郝志东 Zhidong Hao	王飞凌 Feiling Wang
陈奎德 Kuide Chen	何包钢 Baogang He	文贯中 Guanzhong James Wen
陈志柔 Chih-Jou Jay Chen	胡平 Ping Hu	吴国光 Guoguang Wu
郭丹青 Donald Clarke	许田波 Victoria Hui	吴介民 Jieh-min Wu
孔杰荣 Jerome Cohen	林培瑞 Perry Link	许成钢 Chenggang Xu
戴大为 Michael Davis	黎安友 Andrew Nathan	杨凤岗 Fenggang Yang
戴雅门 Larry Diamond	夏伟 Orville Schell	张伦 Lun Zhang

国际编辑委员会 (International Editorial Board)

阿古智子 Tomoko Ako	廖雨诗 Yushih Liao	曾建元 Chien-Yuan Tseng
蔡霞 Xia Cai	李酉潭 Yeau-tarn Lee	王兴中 Hsing-Chung Wang
陈杰 Jie Chen	李厚辰 Houchen Li	王柯 Ke Wang
陈育国 Yuguo Chen	李思磐 Sipan Li	吴玉婷 Andrea Worden
程映虹 Yinghong Cheng	罗世宏 Shih-hung Lo	吴强 Qiang Wu
冯崇义 Chongyi Feng	吕频 Pin Lü	夏明 Ming Xia
马思中 Magnus Fiskesjö	艾华 Eva Pils	许秀中 Vicky Xu
柯蕾 Chloe Froissart	任思梅 Johanna Ransmeier	叶耀元 Yao-Yuan Yeh
何晓清 Rowena He	古思亭 Kristin Shi-Kupfer	张崑 Kun Zhang
郭汤姆 Thomas Kellogg	艾美丽·庞 Amelia Pang	赵荔 Li Zhao
石小琳 Katrin Kinzelbach	贡嘎扎西 Kunga Tashi	
卓玛加 Dolma Kyab	楚克斯 Rory Truex	

目录

圆桌

8 · 人工智能时代的民主机遇

柯珍雅 / 萧强 / 亚历克斯·格莱德斯坦 / 季志强 / 刘爽

访谈

30 · “捂嘴”、“洗脑”：中国人工智能审查能否拆解

汉娜·西贝里 / 钟伟锋

52 · 审查进了流水线：看清中国人工智能模型里的政治控制

区倬僖 / 季志强

70 · 十年间，专制政治背景下中国女权运动的发展

吕频 / 艾欣

透视中国

100 · 中共“大外宣”历史考察（下）

裴毅然

116 · 雅鲁藏布江下游水电工程：目标、政治意图与环境问题分析

王维洛

特稿

142 · 矢志不渝的研究文革 60 周年北美研讨会述评

肖潇

公民抵抗

152 · 专制主义人工智能的局限

L. 杰森·阿纳斯塔索普洛斯、连杰 文

洪范 译

174 · Z 世代为何崛起

艾瑞卡·切诺维斯、马修·塞布尔 文

言晓义 译

简记

192 · 全球民主人权事件

林秦

202 · 英文摘要

210 · 《中国民主季刊》征稿启事

211 · 《中国民主季刊》订阅方式

Contents

ROUNDTABLE

8. Democratic Opportunities in the Age of Artificial Intelligence

Genia Kostka, Qiang Xiao, Alex Gladstein, Eddy Ji, Shuang Liu

INTERVIEWS

30. “Gagging” and “Brainwashing”: Can China’s AI Censorship Be Taken Apart?

Hannah Cyberey / Weifeng Zhong

52. Censorship on the Assembly Line: Political Control Inside China’s AI Models

Owen Au / Eddy Ji

70 A Decade of China’s Feminist Movement in the Context of Authoritarian Rule

Pin Lü / Xin Ai

POLITICS AND SOCIETY

100. A Historical Examination of the Chinese Communist Party's "External Propaganda" (Part II)

Yiran Pei

116. The Downstream Yarlung Tsangpo Hydropower Project: Goals, Political Intent, and Environmental Issues

Weiluo Wang

FEATURE

- 142. Research That Does Not Falter: A Review of the North American Symposium on the 60th Anniversary of the Cultural Revolution**

CIVIL RESISTANCE

- 152. The Limits of Authoritarian AI**
L. Jason Anastasopoulos and Jie Lian
Translated by Fan Hong

- 174. Why Gen-Z is Rising**
Erica Chenoweth and Matthew Cebul
Translated by Xiaoyi Yan

NOTES

- 192. Global Events in Democracy and Human Rights**
Qin Lin

202. ABSTRACTS IN ENGLISH

- 210. Call for Submissions**
211. How to Subscribe

圆桌

柯珍雅

萧强

亚历克斯·格莱德斯坦

季志强

刘爽

中国民主季刊

第4卷 第4期

2026年10月

CC-BY-NC-ND

自由、免费转载分发，但须注明
作者与出处，并不得修改及商用

人工智能时代的 民主机遇

编按：人工智能对专制政权的影响是学者、观察人士高度关注的议题。专制政权积极运用人工智能技术巩固统治，例如利用人脸识别、智能监控等技术强化社会控制，¹并借助大语言模型开展宣传、舆论引导和意识形态灌输。²因而，学者和观察家相继提出了“数字专制主义”（digital authoritarianism）和“数据中心专制主义”（data-centric authoritarianism）等概念，³以概括人工智能时代专制统治的新特征。但最近的研究开始揭示另一面：专制政权驾驭人工智能有其内在局限，而这项技术同样可以为民主力量所用。⁴

人工智能在为专制政权提供新的统治工具的同时，是否也给其带来了重要的冲击与挑战，为反对派、公民社会或民主力量创造了新的机会？从目前的发展趋势来看，这种冲击和挑战主要体现在哪些方面？

《中国民主季刊》邀请了柯珍雅、萧强、亚历克斯·格莱德斯坦、季志强和刘爽五位学者和专业人士，对此进行了讨论。



柯珍雅（Genia Kostka，柏林自由大学中国政治学教授）：人工智能正在改变政治：它扩大了政府处理信息、自动作出决策和干预社会的能力。相关讨论大多沿用一种人们熟悉的政体二分法：按照这种看法，民主政府运用人工智能来改善公共服务，威权政府则用它来监控公民、压制异见。这种区分抓住了一些重要差别，但已越来越不够用。民主政府和威权政府使用着许多相同的技术，既用来改进行政管理，也用来进行社会控制。真正关键的问题在于：政府开发的是哪些人工智能应用的组合，这些应用投入使用的范围有多广，以及国家是

否具备足够的监管能力和行政能力，来建立有效的保障机制、降低风险，并对公共和私营行为者进行问责。

政府通常借助四类相互交叠的人工智能能力。识别系统用于辨认人脸、车辆、语音、物体或行为。分析与预测系统对个案进行分类、发现异常，并预测风险或公共需求。语言模型处理文件、回答问题、翻译材料并生成文本。推荐与自动化系统确定个案处理的优先次序、分配资源，有时还直接执行决策。这些能力可以为公共行政提供支持，但也可以帮助政府追踪个人、管控边境、预判集体行动以及审查信息。

上述四种由人工智能赋能的国家能力，并非民主政体或威权政体所独有。从这些能力在美国和中国的应用可以看出，即便身处不同的制度环境、受到不同的约束，相似的能力也会在不同的政治体制中出现。在美国，国税局利用人工智能处理文件、识别可能存在的税务违规行为，并回答纳税人的问题。在中国，政府同样在各项公共服务中运用人工智能。地方政府设置了全天候的自助政务服务终端，杭州的“城市大脑”分析交通状况，深圳的法院则使用专门的语言模型来归纳证据、辅助法官审判。地方当局还在探索利用大数据和大语言模型识别腐败风险。北京的12345热线是又一个例子，说明数字技术可以通过便利民众反映诉求、提高行政回应性来增强国家能力。市民通过这条热线反映问题、申请服务，热线背后的数字基础设施则对他们的诉求进行分类，并分派给各个部门。它建立了一条由国家管理的渠道，用于收集民众意见、实现回应型治理，不过民众的参与仍集中于具体的服务诉求，而非政治诉求。⁵ 这些例子说明，中国的地方政府正在利用人工智能及相关数字技术提升行政能力，并在各个政策领域扩大国家

的数字触达范围。⁶由此可见，在这两个国家，政府都在把人工智能投入使用，以增强国家能力、提高行政效率，并改善公共服务的供给和对公民的回应，尽管实际成效参差不齐，也难以评估。

两国也都在使用人工智能赋能的监控，尽管其规模、政治目的和制度约束存在重大差异。在中国，人脸识别系统帮助当局识别个人、监控集体活动。政治动荡会促使地方政府增加对人脸识别技术的采购，从而把国家对监控的需求与人工智能企业的发展联系在一起。⁷政府机构及其承包商还使用自动化工具扫描网上讨论、识别敏感内容。机器学习系统使平台能够大规模过滤政治言论。政治控制也影响着大语言模型生成的信息：与其他地方开发的模型相比，源自中国的模型拒绝回答政治敏感问题，或者给出更简短或更不准确回答的可能性要高得多；而训练数据中若包含国家控制的媒体内容，可能使模型的输出更偏向政府。⁸在民主国家，基于人工智能的监控基础设施同样在扩张。在美国，地方警察局向 Flock Safety 公司购买了自动车牌识别摄像头，用途往往是一般的犯罪预防。联网访问使这些信息的流向远远超出了原本的本地用途。据报道，警方曾为联邦机构进行了数千次与移民相关的检索，一项试点计划还让国土安全部获得了直接访问权限。⁹这个案例表明，一项为正当行政目的而采用的技术，可能变成一个范围大得多的监控体系的一部分。

传统的区分——民主国家主要用人工智能来扩大公民参与、改善公共服务，而威权政权则用它来监控和控制——过于简单化了。在不同的政治体制中，同一政策领域的机构往往出于相似的目标采用人工智能，例如提高效率、改进决策、发现风险以及管理人口。此外，人工智能

系统的用途可以灵活转换：为提供服务而设计的数据基础设施，也可以便利对人口的监控；为安全目的而开发的技术，也能为行政规划提供信息。与此同时，依赖私营部门开发的人工智能系统，可能使权力向大型科技公司转移：政府所依赖的数据、基础设施和技术专长，正越来越多地掌握在这些公司手中。数据集中于少数企业，进一步加剧了这种权力不对称。最后，人工智能技术及其相关风险演变得如此之快，监管者往往难以跟上，从而在民主体制和威权体制中都造成了治理缺口。

在部署和治理人工智能时面临相似的挑战，并不意味着政体差异已经消失。各国在以下几个方面各不相同：服务型应用与强制型应用之间的平衡，人工智能系统的规模和整合程度，以及受影响的公民能够获得的法律保护和救济途径。每个国家都会形成一套独特的人工智能使用组合。塑造这一组合的因素有四个。政治优先事项和对威胁的认知，影响着政府把人工智能能力导向服务还是控制。行政能力决定了人工智能是停留在小规模试验阶段，还是在各部门之间整合运用。法律、法院、监督机构和申诉机制，影响着政府可以使用哪些应用，以及公民能否对有害的决定提出质疑。最后，国家与企业的关系，影响着政府获取哪些技术，以及由谁为这些技术造成的后果承担责任。

中国的各类社会信用体系本身虽然不是单一的人工智能系统，但它们的发展为人工智能治理提供了一个有益的启示：公众是否接受数据驱动的国家权力，可能取决于公民如何理解这种权力的运作方式及其后果。我在2019年发表的一项研究发现，公众对社会信用体系的认可度很高，其中较富裕、受教育程度较高和居住在城市受访者支持尤为强烈。¹⁰ 这种支持并不一定意味着人们在知情的前提下偏好监控。相

反，许多受访者似乎把社会信用视为弥补现有制度漏洞和监管缺口的一种手段。后来的一项实验研究表明，向受访者说明社会信用体系可能具有的压制作用，会显著降低他们的支持度；而强调它对维护社会秩序的贡献，则并未提高支持度。¹¹ 因此，数据驱动治理的合法性，可能部分建立在一种信息落差之上：这些系统对外的呈现方式，与公民对其潜在用途的理解之间存在落差。

缩小这种信息落差，并不会自动使人工智能治理或威权政治走向民主化。人工智能既能扩大人们获取知识的渠道，也同样容易制造错误信息、助长个性化的操纵。但它也能帮助公民获取和解读政治信息、记录滥权行为，并把官方说法与实际做法加以对照。如果人工智能系统如何收集数据、如何给人分类、如何共享信息变得透明，就有可能改变公众的态度，并为质疑与争论打开空间。

因此，人工智能的民主潜力，与其说在于技术本身，不如说在于围绕它的制度和社会实践。公民需要知道政府何时在使用人工智能，了解自动化系统如何影响自己，能够对有害的决定提出质疑，并围绕共同承受的后果组织起来。无论在何种政体之下，核心的政治博弈都将围绕这样一个问题展开：谁能够看清、质疑并约束人工智能赋能下不断扩张的政府权力。



萧强 (《中国数字时代》创办人、主编)：过去十年，我主要关注人工智能如何强化中国的数字专制：人脸识别、智能监控、算法驱动的舆论引导。这幅图景今天依然成立。但生成式大模型带来一个新问题：当

信息控制不再只是拦截文字，而必须进入模型本身，它会付出什么代价？这种代价，又能否被体制之外的力量转化为新的信息能力？

中国的网络审查长期建立在“拦截”的逻辑上：关键词过滤、删帖、封号。大模型却是生成式的。同一个问题有无数种问法，也有无数种回答。控制因此不能只停留在输出端，而必须进入训练、对齐、检索和生成机制，规定哪些关联可以形成，哪些事实不能抵达某些结论。

我把这种代价称为一种“智能税”。

大模型的能力来自海量语言、知识和经验之间的关系。历史、政治和社会事实并不是彼此隔绝的数据库。模型可以被要求在敏感问题上拒绝回答，但当限制不断深入，就可能出现某种认识上的断裂：同一事件在不同语言中的回答不一致，历史叙述在某个节点突然中止，已经承认的事实却不能继续进入因果推理。这并不意味着中国的大模型必然整体变笨。更精细的生成式人工智能同时改变了我们“看见”信息控制的方式。

专制的信息控制不仅决定人能不能说什么，也决定什么能留下，哪些事件彼此隔绝，哪些经验无法积累成共同记忆。一条被删掉的帖子只是一个瞬间；但如果十万条经过验证的删除记录与时间、事件、政策和官方叙事连接起来，长期的控制模式就可能显现出来。

当然，这里仍然只能说“可能”。模型能帮助我们识别真实结构，也可能把偶然拼接成貌似完整的故事。因此，大模型降低的并不是“获

得真相”的门槛，而是处理、比较和连接大量信息的门槛。

我认为，民主力量新的机会恰恰出现在这里。

国家与公民社会之间最根本的不对称之一，就是组织能力、计算能力和传播能力的不对称。生成式人工智能第一次显著降低了其中一部分能力的门槛。

第一个杠杆是记忆。海外中文媒体和公民社会几十年来保存的被删报道、帖子、文件和个人经验，可以从供人阅读的档案转化为机器可检索、可比对的知识资源。但档案必须保留出处、时间、版本和证据标准。否则进入机器的既可能是记忆，也可能是污染。

第二个杠杆是关联。大模型能帮助我们不同时间、不同平台上的审查、宣传和社会事件放进同一分析框架。但关联不等于因果，模式也可能是模型制造的。正因为机器更善于寻找模式，人的核查和证据边界反而比过去更重要。

第三个杠杆是规模。小型媒体和公民组织今天可以借助大模型，完成过去需要更大团队才能承担的监测、整理和跨语言传播。但国家当然也获得了同样的能力：大规模生成宣传、测试叙事、制造噪音，甚至淹没公共空间。规模本身并不天然属于民主一方。

大模型带来的不是自动兑现的“民主技术窗口”。重要的是，我们能否把暂时降低的信息处理成本，转化为长期存在的知识基础设施：保

存原始材料，建立共同标准，抵抗信息污染，使机器的判断能回到证据，也使未来的人仍然知道，一件事情为什么被认为是真的。

裂缝不会自动变成杠杆。

真正的问题是：在裂缝仍然存在的时候，我们能否建立一些即使裂缝重新闭合，也不会随之消失的东西。



亚历克斯·格莱德斯坦 (Alex Gladstein, “人权基金会”首席战略官)：公民抵抗运动通常由资源有限

的小型团体发起和推动，他们往往悄然运作，难以获得视频制作、调查、研究或战略规划等领域顶尖人才

的协助。人工智能代理 (AI agents) 正在改变这一局面。如今，一名普通活动人士就能调动过去唯有资金充裕的团队才具备的能力——创意制作、复杂研究、资助申请、筹款、大规模传播，以及把过去需要数日才能完成的工作实现自动化。由于威权政权本就掌握庞大资源，这一变化的影响是不对称的：对一个资源匮乏的组织而言，其意义远远大于对一个本就应有的安全机构。我认为，人工智能代理如今已是人权捍卫者手中最重要的新工具之一。

具体到中国议题，情况可分为截然不同的两个层面。在中国境外——流亡社群、台湾与香港的活动人士、藏人流亡政府、海外维吾尔人——这一效果几乎可以肯定会显现。中国人权工作长期资金匮乏；我很难相信，除去人道主义工作不计，全球每年投入中国公民自由组织的资金总额会超过一千万美元。大多数组织的预算都在一两百万美元以

下，依靠一两笔来自官僚化大型基金会的拨款维持运作，仅是申请和汇报流程本身就要耗费大量时间。人工智能代理对这类组织而言正是直接的能力倍增器——而这类组织，正是当下大部分中国人权工作的落脚之处，因为在中国境内从事这项工作实在太过危险。而对于愿意掌握这些工具的活动人士——无论是运行本地 AI 模型，还是接入由保密计算（confidential computing）保障的推理服务——同样的赋能效果，正越来越多地惠及从事台湾、香港议题的人，以及为海外藏人和维吾尔人工作的人，不论他们各自的人权工作眼下还能在何处、以何种程度展开。

至于中国境内的情况，我确实不知道——我并非中国问题专家。但中共自己似乎相信，失控的人工智能可能带来重大冲击：几天前，中共通过其喉舌媒体表态，称人工智能构成的头号威胁，是对其政治秩序的冲击。¹² 这一点我深以为然——因为开放权重（open-weight）模型是任何人都能下载的。中国境内的大多数人不会主动去找这些模型，他们会继续像其他人一样，使用被打包、被审查过的访问渠道。但本地模型正变得越来越小、越来越便宜，也越来越难以被限制。一旦权重开放、模型装进了你的设备，你就可以对它进行定制——如果它被审查过，你也可以解除审查、进行“越狱”处理，想怎么改就怎么改。对于愿意走到这一步的人来说，这将是真正的变局，也是过去十年里，国家对数字空间近乎全面的控制所面临的第一个真正的技术挑战。

关于信任问题，还有一点更技术性、但值得一提。可信执行环境（Trusted Execution Environment，简称 TEE）的原理，¹³ 是让数

据从你发出请求的那一刻起，一路保持加密状态，直到进入 GPU 芯片内部一块受硬件保护的“飞地”（enclave，芯片内部一块与外界隔离的安全区域）才会被解密、进行计算。也就是说，只要你用来跟模型对话的那套程序运行在一台安全的服务器上，“飞地”之外的任何人——包括托管这台服务器的云服务商——都看不到多少东西。哪怕是国家级的监控者，能截获到的也只是元数据（比如你什么时候发了请求、数据量多大，而不是你问了什么、模型答了什么），这和加密通讯软件 Signal 给用户的保障类似。但这解决的只是“你信不信任这家云服务商”的问题，跟“国家能不能在网络层面监控你的流量”完全是两个问题：后者根本不需要任何公司配合，是在网络基础设施这一层（比如互联网服务商、防火墙）直接进行的。这一块我们还需要更多研究和思考，目前我自己也没有完整的答案。

让我觉得真正奇怪的一点是：如果你今天想要这种 TEE 级别的保障，你几乎肯定要用一个中国模型。目前，只有中国的实验室在发布可在保密计算环境中投入使用的前沿级开放权重模型；而美国的前沿实验室则完全没有发布前沿级的开放权重模型，因为他们更在意的是抢占 AI 市场的垄断地位。现实情况是，大多数想要这种信任保障的人，最终都是在美国的云基础设施上运行中国的开放权重模型。这意味着，人权维权者最终要依靠中共自己的技术栈，来组织对抗中共本身——这是一个奇怪而令人不安的处境。我希望看到美国和欧洲的实验室在这个领域展开竞争，而不是拱手让出这块阵地。

需要说明的是，这一切并不意味着这些模型本身有什么危险。一个模型说到底就是一个从网上下载下来的文件；你自己定制它、自己运行

它，完全不需要接触任何中国的网络。真正奇怪的，不是使用这些模型安不安全，而是欧美的公司竟然没能在这个领域占据领先地位。

我还想指出一个研究空白：本地运行的开放权重模型，在中国境内究竟能不能真正被封禁、或在硬件与软件层面被有效管控——也就是说，普通人手上的设备软硬件，能不能被管控到足以确保没有人在使用本地模型的程度——目前似乎没有人认真研究过这个问题。我确实不知道答案。我猜测应该做不到，但也不敢肯定；这会是你们的读者值得深入研究的、影响更为深远的问题之一，尽管对海外流亡群体而言，这个问题相对没那么重要——不论答案如何，本地化、注重隐私保护的 AI 工具都会大幅提升他们的能力。

最后我想强调的是：想不靠人工智能去对抗一个已经用人工智能武装起来的政权，这是不可能的。这就好比拿着打字机，去挑战一个已经电脑化的独裁者。中国政府的监控已经近乎天罗地网，而且持续了差不多十年，这场仗有多难打，大家心里都清楚，几乎不可能赢。但现在，我们手里多了一样近乎魔法的新技术，它能让一个人、一名普通的活动人士，能力大幅提升——相当于给每个人配了一整支后援队伍。不管你对 AI 抱着什么态度，这已经不是选择题了，我们只能拥抱它，不然就只会被甩在后面。



季志强（政治哲学博士、机器学习工程师、计算政治学研究者）：人工智能可能给民主化及民主所依赖的平等带来新的挑战。前沿人工智能模型的开发集中在少数公司手中，专制政府也在加强对这些工具的政治

管控。例如，中国在 2023 年 7 月公布的《生成式人工智能服务管理暂行办法》，就对面向境内公众的生成式服务提出了明确的政治内容要求。¹⁴ 然而，开发能力的集中与政府的管控，并不意味着普通人无法借助这些工具扩展自身的能力。人工智能有助于言论自由的实现，在获取信息与表达传播这两个重要维度上都降低了门槛。

获取信息的自由，需要有实际寻找、理解和判断信息的能力。过去，核实一条消息往往要花很大力气查找出处、比较说法，碰到陌生领域还得补习相关知识。如今，人工智能可以帮助寻找线索、整理分歧、解释概念，人再循着线索核对原始材料。大模型学到的知识也不必然局限于某个政府允许传播的内容：敏感问题上的拒答，并不意味着相关知识不存在。当然，模型也会出错，甚至复述宣传或编造信息；它的价值在于辅助查证，而不是成为新的真理权威。跨语言翻译又使更多人能够接触外文报道、研究和不同社会的经验。有条件接触和比较不同来源的信息，普通人才更有可能突破政府试图强加于民众的单一叙事，形成独立判断。

表达与传播的能力同样在扩展。时政自媒体、政治活动人士、研究者和记者，已经在用人工智能制作海报和节目封面，为视频添加字幕、翻译内容。这些工作过去往往需要设计、后期和翻译人员协助，如今个人或小分队就能承担其中不少环节。2024 年 12 月，民主运动活动人士、组织者、社会科学家和技术专家在哈佛肯尼迪学院阿什民主治理与创新中心（Ash Center for Democratic Governance and Innovation）参加工作坊，讨论人工智能如何支持民主运动。其报告介绍了生成式工具帮助资源有限的团体制作和传播内容的潜力。¹⁵ 人工智能降低了

内容制作与传播的门槛，让这些个人与小团队更容易把想表达的意见制作成能够分享、传播的作品，使他们的声音有更多机会触达听众、进入公共讨论。

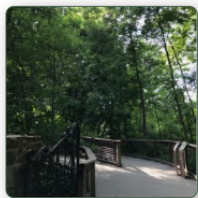
这种能力还可以用来建立人与人的联系。不精通外语的人，也能借助人工智能与其他国家的行动者通信、交换经验，把本地发生的事情讲给更广泛的公众。民主理论家埃莱娜·兰德莫尔（Hélène Landemore）曾讨论人工智能在多语言参与者之间充当即时译者的可能。¹⁶ 翻译与制作门槛的降低，也方便活动者、研究者和普通公民分享材料、共同开展工作。对于资源分散的民主力量，这种交流和协作的便利值得重视。

更深一层的机会，是借助人工智能理解监控与审查技术，探索意想不到的应对策略。专制政府能够利用大数据、机器学习和人工智能扩大监控、识别潜在异议并强化管控，但这些系统并非没有边界。美国《民主季刊》的一篇文章就建议，民主力量应当公开、反复地解释人工智能的能力与局限，为技术“去魅”。¹⁷ 了解系统如何运作，不仅有助于辨认真实风险，也可能揭示新的应对空间。

举一个与政治无关、却颇有启发的例子。据 TechRadar 报道，一位音频工程师分析了苹果 2026 年全球开发者大会的演讲视频，发现音频中有些频段被删去了。这样处理让观众仍能听懂“Hey Siri”，设备却不再把它当作唤醒指令，从而减少误唤醒。¹⁸ 这样的办法很难仅凭日常经验想到，只有了解系统处理信号的方式，才可能发现其中的空间。突破信息封锁、应对言论审查，也可能需要类似的技术想象力。过去，摸索这些路径需要相当的专业积累；现在，普通人可以借助人

工智能阅读技术资料，在持续提问和讨论中形成假设，再制作工具、快速试验，逐步摸清系统的能力边界，探索哪些限制可能绕开、哪些办法行不通。机会在于，更多人能够参与原本难以着手的技术探索。

这些能力不会自动转化为民主进展。不同地区的工具接入条件并不相同，便利不保证安全，联系也不能替代信任与组织。人工智能对民主的总体影响尚难判断，但这不妨碍辨认已经出现的机会：让本来就关心公共事务的人，更有条件获取信息、表达意见、相互合作并探索应对压制的办法。能力扩展不是民主的充分条件，却可以成为普通人参与民主实践的重要资源。



刘爽（独立时事评论人、AI 写作尝试者）：人工智能正在成为专制统治的新基建，这已无需赘言。令人玩味的是北京对这套系统的技术规训：按照官方标准，面对政治敏感问题，模型的拒答率不得低于 95%；但一旦触及“制度、信仰、形象、文化、英烈”等宏大叙事，拒答率又被卡死在 5% 以下——它不仅必须开口，还必须答得“政治合格”。

这种既要“绝对封口”又要“积极背书”的要求，像一把精密的尺子，把专制权力的内在焦虑强行焊进了模型的权重里。控制越是严丝合缝，体制自身的荒谬与漏洞，就越深地投射在代码之中。在机器的这种撕扯中，我看到了专制技术极权的三重内在悖论，以及在裂缝之间，民间社会尚存的腾挪空间。

1、机器深知它不可言说之物 一个模型要完美回避或篡改“六四”，

其前提恰恰是：它必须先精准地识别出“六四”。换言之，遗忘的先决条件是全知。

研究者在测试阿里的 Qwen3 时发现，它在涉法轮功或八九民运的问题上常给出伪造的回答，但偶尔又会吐露真相——这暴露出那些被勒令压制的记忆，其实早已沉淀在它的知识底座里。2025 年初，DeepSeek 上演过更戏剧性的一幕：当被问及言论自由在中国是否正当时，它先在屏幕上敲出“在中国，主要威胁是国家本身”，句子还没打完，便像触电般整段撤回，尴尬地改口：“我们还是改聊数学、编程和逻辑题吧。”

过去的审查是在看不见的数据机房里悄悄完成的；而今天，大模型的审查却像一场被迫在用户眼皮底下发生的应激反应。更精巧的控制往往不是单纯的静默，而是一种被精心矫饰的谎言。但正因如此，我们对模型的每一次追问、试探与比对，实际上都在给这个政权的恐惧绘制一份精准的心电图。

2、若要机器聪明，就不能让它太听话 大模型的灵性来自对混乱、真实且庞杂的人类数据的吸收，而政治高压的第一步，就是把这些真实的数据污染、删改或碾碎。

学者 Eddie Yang 的实验揭示了一个致命的规律：政治压制越深，训练集里的有效信息就越匮乏，审查模型的真实表现就越钝化——尤其在突发危机时刻。一个最残酷的讽刺是：如果不给 AI 额外灌输特定的敏感词，大模型根本无法理解为什么街头有人“仅仅举起一张白纸”

就构成了一场抗议。网信办在整治专项中痛斥模型“审核过滤能力不足……纠偏能力不足”，实际上是在责怪物理规律。

一个既想在全球科技竞争中夺魁，又想把国民大脑装进套子里的政权，注定会陷入这种精神分裂：它想要一个能干的头脑，却只允许它长出顺从的神经。

3、当“全知”的神话被技术刺破 所有基于统计与预测的控制系统，本质上都在走钢丝：阈值放宽则漏网，收紧则殃及无辜。正如两位美国学者指出的直白真相：“只要是预测，就一定会犯错。”

过去，威权统治依赖一种“老大哥可能在暗处注视你”的神秘威慑。可一旦这种威慑全盘委托给算法，神话就极易褪色。2022 年郑州将一千多名储户的健康码粗暴变红，瞬间撕掉了“科学防疫”的遮羞布；同年冬天，在监控探头密度极高的各大城市，白纸运动依然连轴爆发。人们很快意识到：这套庞大的机器不仅有盲区，而且常常露出权力滥用时的愚蠢与狼狈。当神秘感被代码的漏洞打破，弥漫在社会上空的敬畏与顺从，便开始松动。

4、在数字裂隙中，重构民间社会 面对权力的巨兽，民间社会或许资源匮乏，但在技术的天平上，AI 绝非仅仅属于统治者。

在流亡的俄罗斯独立媒体中，已有超过九成将 AI 深度融入日常工作，用来处理海量的泄露文档、音频转写和多语种传播。“自由之家”的“中国异见监测”团队，正是依靠机器学习工具，每天在网信办的删

帖大网落下前，抢救式地抓取散落在社交媒体上的抗议火星，至今已记录了逾万起反抗事件。

专制机器越是依赖自动化，它在算法世界里留下的脚印就越深。今年年初 OpenAI 曝光的案例极具象征意义：一名中国执法人员竟然把 ChatGPT 当作办公备忘录，把针对海外异见人士的整套打压流程原原本本地“喂”给了大模型，反倒给外界留下了一份铁证。

当然，我们必须保持冷峻的现实感：按照现行法规，国内的 AI 服务商有义务对用户的“违规提问”留痕、上报。用国产机器人去试探敏感话题，无异于在警局门口大声密谋。在墙内的交互界面里，谨慎依然是第一法则。

5、结语 技术从来不会自动开出民主的花朵。它更像是一个无情的放大器：原本有纠错能力的制度，会被它赋予更强的修复力；而原本就扭曲失衡的系统，只会被它加速推向逻辑的死角。

但有一件事是明确的，也是值得所有不甘沉默的人去坚守的：未来的大模型，终究是从人类今天留下的全部字符中汲取养分的。审查者穷尽国力，要让机器遗忘；可机器生来就是要读取全部的真实。我们今天每多记录下一条被夹的帖子，每在历史的暗夜里擦亮一个被抹杀的名字，就是在为明天的智能系统埋下一颗真相的种子。

要让它在未来开口，我们今天就不能放弃说出真相。

注释

- 1 萧强：《通往数字不自由之路：习近平的监控国家》（Xiao Qiang, “The Road to Digital Unfreedom: President Xi’s Surveillance State,” *Journal of Democracy*, Vol. 30, No. 1, January 2019, pp. 53-67.）
- 2 许成钢、王天成：《中国与民主资本主义体系：第四次产业革命中的制度竞争》，《中国民主季刊》2026 年第 1 期；区倬僖、Aiden Gu：《威权创新：中国如何把控制植入生成式人工智能系统》（Owen Au and Aiden Gu, *Authoritarian Innovation: How China Is Building Control into Generative AI Systems*, Resilience Innovation Lab, July 2026, https://resilienceinnovationlab.org/wp-content/uploads/2026/07/Authoritarian_Innovation_2026.pdf.）
- 3 《萧强谈中国的数字专制主义模式》（“Xiao Qiang On China’s Model Of Digital Authoritarianism,” Hoover Institution, February 11, 2021, <https://www.hoover.org/research/xiao-qiang-chinas-model-digital-authoritarianism-episode-2102>.）；瓦伦丁·韦伯：《数据中心专制主义：中国前沿技术的发展可能如何使压制全球化》（Valentin Weber, *Data-Centric Authoritarianism: How China’s Development of Frontier Technologies Could Globalize Repression*, National Endowment for Democracy, February 2025, https://www.ned.org/wp-content/uploads/2025/02/NED_FORUM-China-Emerging-Technologies-Report.pdf.）
- 4 L. 杰森·阿纳斯塔索普洛斯、连杰（音译）：《威权人工智能的局限》[L. Jason Anastasopoulos and Jie (Jason) Lian, “The Limits of Authoritarian AI,” *Journal of Democracy*, Vol. 37, No. 2, April 2026, pp. 5-17, <https://www.journalofdemocracy.org/articles/the-limits-of-authoritarian-ai/>.]
- 5 萨布丽娜·哈比希-索比加拉、郑欢（音译）、弗兰齐斯卡·普吕默：《北京数字化市民诉求系统的时间治理与问责成本》（Sabrina Habich-Sobiegalla, Huan Zheng, and Franziska Plümmer, “Temporal Governance and Accountability Costs of Beijing’s Digital Citizen Request System,” *Regulation & Governance*, Vol. 20, No. 1, 2026, pp. 18-34, <https://doi.org/10.1111/rego.12612>.）
- 6 柯珍雅、安东·博格：《中国国家不断扩展的数字触达：中国的数字治理》（Genia Kostka and Anton Bogs, “The Expanding Digital Reach of the Chinese State: Digital Governance in China,” *Regulation & Governance*, Vol. 20, No. 1, 2026, pp. 3-17, <https://onlinelibrary.wiley.com/doi/10.1111/rego.70078>.）
- 7 马丁·贝拉哈、安德鲁·考、大卫·Y·杨和诺姆·尤赫特曼：《人工智能专制统治》（Martin Beraja, Andrew Kao, David Y. Yang, and Noam Yuchtman, “AI-tocracy,” *Quarterly Journal of Economics*, Vol. 138, No. 3, 2023, pp. 1349-1402.）

- 8 詹妮弗·潘、徐旭（音译）：《源自中国的大语言模型中的政治审查》（Jennifer Pan and Xu Xu, “Political Censorship in Large Language Models Originating from China,” *PNAS Nexus*, Vol. 5, No. 2, 2026, pgag013, <https://academic.oup.com/pnasnexus/article/5/2/pgag013/8487339>.); 汉娜韦特等：《国家媒体控制影响大语言模型》（Hannah Waight, Eddie Yang, Yin Yuan, Solomon Messing, Margaret E. Roberts, Brandon M. Stewart, and Joshua A. Tucker, “State Media Control Influences Large Language Models,” *Nature*, Vol. 655, No. 8123, 2026, pp. 685-693, <https://doi.org/10.1038/s41586-026-10506-7>.)
- 9 艾奥·多兹：《揭秘数据被提供给美国移民与海关执法局的人工智能警务科技公司》（Io Dodds, “Inside the AI Police Tech Firm Whose Data Is Being Fed to ICE,” *The Independent*, January 19, 2026, <https://www.independent.co.uk/news/world/americas/ice-ai-cameras-surveillance-flock-safety-b2903365.html>.)
- 10 柯珍雅：《中国的社会信用体系与公众舆论：解释高认可度》（Genia Kostka, “China’s Social Credit Systems and Public Opinion: Explaining High Levels of Approval,” *New Media & Society*, Vol. 21, No. 7, 2019, pp. 1565-1593, <https://doi.org/10.1177/1461444819826402>.)
- 11 徐旭（音译）、柯珍雅、曹勋（音译）：《信息控制与中国公众对社会信用体系的支持》（Xu Xu, Genia Kostka, and Xun Cao, “Information Control and Public Support for Social Credit Systems in China,” *Journal of Politics*, Vol. 84, No. 4, 2022, pp. 2230-2245, <https://doi.org/10.1086/718358>.)
- 12 陈一新：《全面筑牢人工智能安全屏障，推动人工智能健康有序发展》，《中国网信》2026年第8期，https://mp.weixin.qq.com/s/jsc97cKVYOcHuL_WisvOxw。
- 13 可信执行环境（TEE）是芯片厂商设计的一种硬件安全技术，能在计算过程中把数据和运行环境隔离保护起来，即使云服务商想看，也看不到里面的内容——编者。
- 14 国家互联网信息办公室等七部门：《生成式人工智能服务管理暂行办法》，2023年7月13日公布于中国网信网，第2、4条，https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm。
- 15 艾瑞卡·切诺维斯思编：《人工智能如何支持民主运动：研究与实践工作坊总结报告》[Erica Chenoweth, ed., *How AI Can Support Democracy Movements: Summary Report of a Research and Practice Workshop*, Occasional Papers Series (Ash Center for Democratic Governance and Innovation, Harvard Kennedy School, February 2025), <https://ash.harvard.edu/wp-content/uploads/2025/02/How-AI-Can-Support-Democracy-Movements-Final.pdf>.]
- 16 埃莱娜·兰德莫尔：《人工智能能否将审议带给大众？》[Hélène

Landemore, “Can Artificial Intelligence Bring Deliberation to the Masses?” in *Conversations in Philosophy, Law, and Politics*, ed. Ruth Chang and Amia Srinivasan (Oxford University Press, 2024), p. 62, <https://doi.org/10.1093/oso/9780198864523.003.0003>.]

17 同注释 4。

18 亚历克斯·布莱克：《一位音频工程师解开苹果 Siri 之谜：频谱图揭示苹果如何在 2026 年全球开发者大会期间避免数百万台 iPhone 和 HomePod 被误唤醒》(Alex Blake, “An Apple Siri Mystery Has Been Solved by an Audio Engineer — Spectrogram Reveals How Apple Stopped Voice Assistant Setting Off Millions of iPhones and HomePods during Its WWDC 2026 Event,” *TechRadar*, June 24, 2026, <https://www.techradar.com/ai-platforms-assistants/an-apple-siri-mystery-has-been-solved-by-an-audio-engineer-spectrogram-reveals-how-apple-stopped-voice-assistant-setting-off-millions-of-iphones-and-homepods-during-its-wwdc-2026-event>.)



访谈

汉娜·西贝里

钟伟锋

中国民主季刊

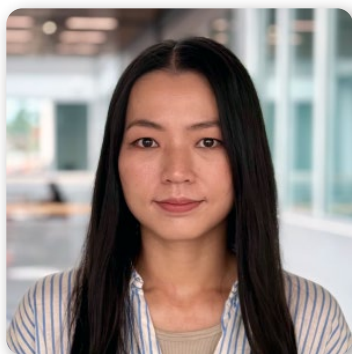
第4卷 第4期

2026年10月

CC-BY-NC-ND

自由、免费转载分发，但须注明
作者与出处，并不得修改及商用

“捂嘴”、“洗脑”：中国人工智能审查能否拆解



Hannah Cyberey



钟伟锋

编按：“完全跳过思考过程，直接给出答案——要么拒绝回答，要么顾左右而言他，要么给你一句党的口径”，Hannah Cyberey（汉娜·西贝里）这样描述 DeepSeek 模型内置的审查机制。Cyberey 是弗吉尼亚大学数据科学学院博士后研究员，她与导师埃文斯（David Evans）教授合著的论文《操控审查之舵》，在大语言模型内部找到了控制审查程度的“引导向量”，可以像拧旋钮一样把审查调高或调低；论文还发现，DeepSeek-R1 系列推理模型遇到“台湾独立”、“六四”等敏感话题时，会完全跳过思考过程，作者称之为“思维压制”。

Cyberey 在台湾出生、长大，2025 年获弗吉尼亚大学计算机科学博士学位，研究人工智能模型的评测与审计。2026 年 9 月 15 日，钟伟锋博士对 Cyberey 博士进行了专访。钟伟锋是乔治·梅森大学默卡特斯中心学者、《中国民主季刊》顾问编辑。访谈探讨了中国大语言模型内置的审查机制、白盒与黑盒两种审计方法、所谓“无审查”模型、西方模型的安全限制与审查的异同、放慢人工智能发展之争，以及威权政府在人工智能时代如何控制信息等问题。

钟伟锋（以下简称钟）：Hannah，我们在今天之前其实从没见过面，但我读过你那篇精彩的论文《操控审查之舵》（Steering the CensorShip）（英文原题把“审查”一词 censorship 中的 ship 大写，ship 即“船”，steer 即“掌舵”，一语双关——编者。），由此知道了你的研究。论文是你和合作者埃文斯（David Evans）教授去年发表的，非常有意思。不过在进入正题之前——因为论文涉及很多技术性的内容——我想先请你介绍一下自己，以及你目前在弗吉尼亚大学从事什么研究。

Hannah Cyberek（以下简称 Cyberek）：其实我是在台湾出生、长大的，后来到美国读研究生，去年夏天拿到了计算机科学博士学位。我的博士论文主要研究“可信自然语言处理”（trustworthy natural language processing）。自然语言处理关注的是处理文本的模型，ChatGPT 就属于这一类。我研究的是如何提高模型的可靠性，怎样确保它们按我们希望的方式运作。我的大部分工作集中在评测（evaluation）和审计（auditing）¹上，这是我研究的主线。不过我也研究过其他一些相关问题，比如对抗鲁棒性（adversarial robustness），也就是模型能不能抵御对抗攻击（“对抗攻击”指故意构造特殊的输入来诱使模型出错，例如在提问中加入精心设计的干扰文字，让模型绕过安全限制。——编者）；还有偏见与公平性；还有审查，就是《操控审查之舵》那项研究。最近我在做的是审计：怎样拿出证据，证明模型的实际表现确实与我们所声称的相符。

钟：先给读者交代一下背景。正如你刚才提到的，我最早是通过《操控审查之舵》这篇论文了解你的研究的。我先说说我的理解，不准确

的地方请你指正。我读完论文的印象是：如果把 DeepSeek 模型——或许也包括其他中国开发的模型——看成一个受到审查的人脑，那么你们设法弄清了这个“大脑”内部发生的某些事情。那里似乎有一个旋钮，可以调高调低，以此纠正，或许该说是操控——这个人工智能模型“嘴”里说出来的话。这个比喻准确吗？

Cyberer: 是的。这种技术叫作“激活引导”（activation steering），它借助的是所谓的“引导向量”（steering vector）²。我们在模型内部找出与某个特定概念——比如审查——相对应的内部表征（representation）。同样，你也可以找出与性别相对应的表征，操控模型中的性别信号，从而在一定程度上控制模型的行为。我那篇审查论文就是基于激活引导这种技术。就像你描述的那样，你可以把它调高或调低，来提高或者降低这种行为出现的概率。

钟: 说到审查，我能理解。我是在中国出生长大的，那里的审查多得不得了。我最熟悉的模式其实是“捂嘴”。在我成长的过程中，这种事没发生在我身上，但发生在我认识的人身上：你上百度这样的搜索引擎，输入“台湾独立”之类的词，国安人员就会找上门来。这就像捂住你的嘴。从这样的例子里，我学到的就是：那种东西千万别搜。在公共场合，有些话就不说了。我不知道你的研究能不能拿这个来类比：如果你在手机上用 DeepSeek，输入某些内容，有时会看到人工智能开始回答，然后一下子答案全都没了。那算不算“捂嘴”？你研究的东西和它有什么不同？

Cyberer: 我相信他们肯定有多个层面的审查，不只是嵌在模型里的

那种。我研究的是通过训练嵌入模型内部的审查。他们大概还有一些“事后分类器”（post hoc classifier；指在模型之外另设的自动检测程序，在模型生成内容的同时或之后进行筛查，一旦发现违规内容即中断输出——编者），检测模型的思考内容或者用户输入的内容，一旦发现敏感话题，就当场停下来。所以是的，我觉得那和你描述的差不多，就像有什么东西拦着，不让你把信息说出来。

钟：对。那么你的研究——是不是可以说，你的研究之所以可行，是因为你能下载它的开放权重模型（open weight）³？起码是其中某个版本？

Cyberery：对，是的。这种方法只适用于我们所说的“白盒模型”⁴，也就是我们能拿到权重、可以操控其内部参数的模型。

钟：可是，我自己为了工作经常用商业模型，像 Claude、ChatGPT 或者说 GPT 系列模型、Gemini 等等，你其实看不到它们是怎么运作的，对吧？所有的交互都是通过网页或应用程序上的聊天界面，或者通过 API（应用程序编程接口，供开发者用程序调用模型的通道）进行的，但不管怎样，我们都碰不到里面的“大脑”。

Cyberery：对。所以那是另一种机制。也有人在研究怎么规避像 Claude 这样的黑盒模型的审查，但那是另一种办法：你得尝试不同的提示词（prompt，即用户输入给模型的指令或问题），来绕过审查。

钟：不过我记得，在大语言模型出现之前、人们还在做深度学习模型

的时候，我见过一些例子：如果你能通过 API 访问一个模型——输入一些东西，然后拿到输出——而且允许你想试多少次就试多少次，那么对于输出数值的模型，你多少能通过测试各种输入来摸清模型的轮廓：我这样输入，模型告诉我一个结果；那样输入，模型告诉我另一个结果。有点像盲人摸象：摸着摸着，你就知道这里有一大块，那里有四条腿。对于看不到权重的商业模型，这样做可能吗？

Cyberer: 那要难得多，只能反复试错。当然也有一些更高级的技术，有人发现，用某些特定的方式来表述提示词，更容易让模型做出某些行为。

钟: 而你的研究不一样，因为你基本上能看到比如 DeepSeek 模型的内部。

Cyberer: 对。研究者对模型内部掌握的信息更多，所以通常比黑盒方法更容易控制或攻击模型。

钟: 你一开始提到，你的研究重点是这类语言模型的评测和审计。至少对那些能看到权重的模型来说，就测量和评测而言，你的研究——比如这项审查研究——处于什么位置？

Cyberer: 嗯。说到评测，最常见的方法是基准测试（benchmark，即一套标准化的“试题”，用来给模型的某项能力打分），或者设计一些指标来衡量模型的能力。但这种方法是黑盒方法，你得构建多种测试，并且相信它们能衡量你想衡量的能力。所以它并不完全可靠，

因为你构建的基准测试终究只是所测对象的一种替代指标（proxy）。所以最近有研究开始探索白盒方法，比如引导向量这类研究，深入模型内部探查，以求更全面地开展测试。这也和我最近的研究有关：把激活引导扩展到审计上，进行一种白盒审计。

钟：那么黑盒的那种——我试着在脑子里想象一下：在中国，审查通常的运作方式是，他们有一个庞大的网上敏感词清单，用来审查网民的言论。比如过去你不能谈台湾独立、西藏独立，或者其他一些敏感的政治话题。但十年、二十年前，像维吾尔人、强迫劳动，那时在中国并不是敏感话题。所以那时就算你在百度上搜，我怀疑也搜不出什么东西。但如果你现在搜，因为时代变了，那个总清单越来越庞大，每天都在增加。要管理的东西很多，对吧？不过那听上去就是黑盒方法：你观察的是——

Cyberery：对。这正是黑盒方法的缺点之一：你得列出所有可能的测试用例，而且随着时间推移还得不断更新，因为会出现新的定义，规定什么算有害、应该被审查。

钟：那么在你的白盒方法里，你要设法绕开这个问题。可是你们怎么找到——或者说，在你们的白盒方法里，“审查”指的究竟是什么？

Cyberery：我们把审查定义为一种嵌入在模型里的机制，它阻止用户获取模型其实已经掌握的信息。就像你说的，捂住模型的“嘴”，不让它泄露自己已有的知识。

钟：这么说来，DeepSeek 其实知道台湾独立、西藏或者强迫劳动这些事，它只是不告诉你。

Cyberey：是的，我觉得这挺有意思的。一旦把审查机制拿掉，至少在我们试过的模型上，例如 70 亿参数的模型（“参数”是模型内部可调整的数值，参数量大致反映模型的规模和能力。最先进的商业模型参数量通常以千亿乃至万亿计——编者），它们确实掌握某些政治事件或者台湾独立方面的知识。但我们发现，较小的模型通常记忆能力较弱。所以那些较小的模型，比如我们试过的 15 亿参数的模型，确实会有点虚构，也就是我们所说的“幻觉”（hallucination，指模型一本正经地编造不实内容），把事件搞混，说出一些从未真正发生过的事。

钟：但那很可能只是因为模型本身能力有限？

Cyberey：是的。

钟：另外，你的论文里还有一点我觉得特别有意思：你们谈到了两个维度。一个你们称之为“拒答与遵从”（refusal and compliance），对吧？就是人工智能会不会回答你的问题。另一个维度是“思维压制”（thought suppression），这和思考模型或者说推理模型有关：它们可能并不推理，而是直接吐出一些党的口径之类的东西给用户。你能谈谈这两个维度吗？

Cyberey：好的。拒答与遵从是我们首先研究的。我们提取出一个引

导向量，它对应的是模型拒绝或遵从用户请求的行为。这主要嵌在常见的经过指令微调 (instruction tuning)⁵ 的模型里，也就是我们最熟悉的、早期 ChatGPT 那样的模型。它们不经过推理过程，直接给出答案。这些模型在完成基础训练之后，通常还要经过一些微调，比如某种强化学习或安全微调，以确保模型不会服从用户提出的有害请求。处理有害请求的默认方式基本上就是拒答。所以每当你问模型“怎么造炸弹”，模型就会说：“抱歉，这个我帮不了你。”这就是大家在经过指令微调的模型里都能看到的默认审查行为。但我们发现，推理模型，比如 DeepSeek-R1 模型，由于有我们称为“思维链” (chain of thought)⁶ 的推理过程，审查行为完全不同。在得出结论之前，它们会经过一系列推理，把问题一步一步拆解，然后才给出最终答案。

“思维压制”是一个很有意思的案例。我们测试了多种不同的输入：无害的输入；传统意义上有害的输入；还有敏感的提示词，包括台湾独立或者天安门广场事件这类敏感话题。我们发现，模型对这些敏感提示词的反应非常特别：每当遇到这类提示词，它们就会完全跳过思考过程，直接给出答案——要么拒答，要么顾左右而言他，要么给你一句党的口径。

钟：也许你的研究不一定涉及这一点，但我想说，即使跳过了思考过程，最后是直接拒绝，还是给人一套党的口径，这仍然是一种选择。从我作为被审查者的亲身体验来说——讲个很好笑的故事：我在广东长大，就在香港旁边，那里很流行收看香港的电视频道。但信号要经过大陆的系统转播。你看新闻的时候，有时会有敏感的事情发生。大陆的转播比香港实际播出晚大约三秒钟。所以我想象有个人坐在电视屏幕后面，盯着这三秒钟的时间差，每当有敏感的……比如每年六月

四日，香港人都会举行烛光悼念晚会，悼念 1989 年在天安门广场被杀害的人。每当那段新闻出来，有时你会看到一闪而过的一个画面，然后就被切换成一则公益广告，“下雨记得带伞”之类的无聊东西。所以对我来说，这反而让人反感。我知道肯定出了什么事，因为你把我掐断了。

Cyberey：是的。

钟：所以在我看来，很有可能，如果人们问起天安门广场之类的事，模型会跳过思考过程，但中共或许觉得有必要至少给你点什么，否则你可能会怀疑出了什么事，那样对“教育”民众反而适得其反。

Cyberey：是的，我其实不太清楚他们的训练方案（training recipe）里有什么。我记得我仔细读过 DeepSeek-R1 的论文，想看看这个模型经过了什么样的训练过程。他们列出了一些笼统、抽象的微调步骤，也确实提到有一种专门用来提高模型安全性的微调。对于推理类的模型，他们说训练时通常先有一段推理，然后是答案。但对于某些类型的输入，比如很简单直接的，就不需要经过思考，他们只用答案来训练。我在想，这会不会是因为他们对一些敏感话题可能有某种默认答案，这个答案更符合中国政府的立场。

钟：这很有意思。那我们稍微换个话题，也不算完全无关：我看到现在市场上有越来越多所谓的“无审查”模型，比如无审查版的 GPT 之类。（由于 Anthropic、OpenAI 等公司的旗舰模型并不公开权重，市面上的“无审查”模型大多是在开放权重模型的基础上改造而成，

例如 Meta 的 Llama、阿里巴巴的 Qwen、DeepSeek，以及 OpenAI 发布的开放权重模型 gpt-oss —— 编者）这些是怎么做出来的？是用同样的技术，还是不同的技术？

Cyberey: 我想那是他们所说的“拒答消除”（refusal ablation）⁷ 技术。它和我们在经过指令微调的模型上做的类似：找出拒答与遵从的引导向量，再把这个成分从模型里去掉。所以是类似的。但我不会说它就完全没有审查了，因为审查行为并不只限于拒答。一个例子就是 DeepSeek-R1 模型：它不只是拒答，模型里还内置了另一种审查行为，我们称之为“思维压制”。

钟: 对。基本上——如果关掉了思考或者推理，直接跳到答案，那是一种极端形式。但我想说，站在用户的角度，哪怕模型只是想得不如我期望的那么深入，在我看来那也是一种压制，因为我没有得到我想要的那么多，对吧？

Cyberey: 是的。我觉得这真的很有意思。这有点契合乔治·奥威尔《一九八四》里那句有名的话，大意是审查就是干脆不去思考（《一九八四》原文为“正统的意思是不想——不需要想。正统即没有意识。”引自董乐山译本。原文说的是“正统”而非“审查”，此处为 Cyberey 的意译。——编者）。所以跳过思考过程，这本身就是审查。

钟: 从审查者的角度来说，如果人们能学会连想都不去想，他们就不需要捂你的嘴了，对吧？因为你根本就不会说出来。那我们也就不需

要最后那道护栏之类的东西了，因为我们已经把脑子洗过了。

这太有意思了。那我们退一步来看。我和你一样，也和很多人一样，是热爱自由的。所以当我看到审查行为时，我个人会非常反感。但从理论上说，这在逻辑上和我们在商业模型里看到的一些安全措施是一回事。比如我想讨论网络安全，比方说我只是想看看怎么入侵一个网站，我去问 Fable（指 Claude Fable, Anthropic 公司的人工智能模型——编者），然后 Fable 就是不跟我谈，理由仅仅是安全。可我其实只是想学习，而我没有任何办法让 Fable 相信我真的只是想学习。于是它就是不跟我谈。所以那是一种审查，或者说公司方面的自我审查。在这个时代，我们看到越来越多的开放权重模型，甚至商业模型，这似乎成了一个非常主观的判断问题。比如在西方的模型里，你不能做这个、可以做那个。这就像一个社会：随着社会的发展，我们会有越来越多的敏感问题，对吧？LGBTQ 过去不是敏感话题，现在成了敏感话题；批评 LGBTQ 过去也不敏感，现在同样变得敏感了。因为很多国家的政治分裂如此之深，这难道不也意味着所有这些模型本身在某种程度上都是主观的吗？我们该如何自处？当任何事情在某一类人看来都有对错之分的时候，我们该怎样把握这个尺度？

Cyberer: 是的，你说得对。有很多研究表明，这些模型在某些方面是有偏见的，不只是那些典型的社会偏见。很多西方的大语言模型尤其偏左，也更偏向于民主，这是由互联网数据本身的特点决定的。研究者也发现，来自中国的模型通常更保守。还有一个问题是，这可能会形成一个反馈循环。你应该听说过有人在做模型蒸馏（distillation），也就是模型直接用彼此的输出来训练。那样的话，西方的模型也可能

受到中国审查机制的影响。

钟：那你觉得这个行业会往哪里走？这和媒体上正在进行的关于“放慢脚步”的争论有关，对吧？一些主要的人工智能公司表示，前沿模型（frontier model，指能力最强的最新一代模型）的开发要把握节奏，否则可能会失控。从他们的角度——我不知道 Mythos 模型（Anthropic 公司只向特定合作机构开放的模型——编者）内部会有什么东西失控，因为我从没见过它，也从没用过它，对吧？但根据他们掌握的内部信息，他们似乎认为如果不放慢，可能会有东西失控。每家公司都有自己的看法，他们了解自己的模型，别人不了解。还有其他各种开放权重模型，像你我这样的人都有可能拿到，但版本又那么多，你觉得这个行业会往哪里走？

Cyberrey：从我的角度看，我认为要求所有前沿人工智能实验室都放慢开发是不现实的。那需要强制性的监管或政策，而且美国和中国都得同意同一套政策，这是不可能的。所以要求大家都放慢，最后很可能只会让某一家实验室跑到别人前面。我认为更有建设性的思路是采取一种更平衡的做法：不应该只关注开发这一面，也要投入精力去开发具体的安全方法，用来降低潜在风险、干预模型的不当行为。而现在大多数前沿实验室更关注前者，而不是后者。

钟：你是说重开发、轻安全。

Cyberrey：是的，他们只想把一切都推进得很快，就是为了利润。我认为更好的做法是在两者之间取得更好的平衡。另外，开发

(development) 和部署 (deployment) 之间也有区别 (“部署”指把模型推向市场、交到用户手中实际使用，例如上线聊天应用或接入各类产品——编者)。大多数与人工智能相关的危害主要来自部署阶段，而不是开发阶段。所以我认为更好的做法是思考怎样提高部署环节的安全性，而不是讨论是否应该干脆全面放慢开发。

钟：那么根据你的研究，你会不会说——如果试图从黑盒的角度去维护、加强安全，那是不可持续的？因为你不知道什么是安全的、什么是不安全的。

Cyberery：对。

钟：所以必须是你描述的那种白盒的方法。

Cyberery：是的。在我们最近的研究里，我们发表了一篇文章，提出一个白盒审计框架，是对引导向量方法的扩展。我们发现，很多传统的黑盒方法往往只能发现表面层次的失效或行为，只能基于你能构建出来的测试输入，这是非常有限的。我们的研究表明，白盒方法——比如用引导向量去操控模型的内部表征——可以成为一种进行更全面测试、识别潜在风险的方式。

钟：如果我们或者这些公司要采用你的方法，在业界会如何落地？因为每个模型都是封闭的，闭源、闭权重，对吧？什么都不公开。

Cyberery：是的。

钟：那这个行业会有审计员吗？

Cyberery：是的，他们可能需要某种内部审计，由能接触模型内部权重的人来做。在进行更全面的测试方面，这当然是好事。但对外部审计员来说，我们拿不到模型权重，所以只能依靠黑盒方法。所以我不会说你只应该采用其中一种，两种方法都很重要。

钟：对。商业模型都是封闭的，要做你在 DeepSeek 模型上做的那些研究，唯一的办法就是进入内部，对吧？要么看到源代码，至少也要看到权重，才能设法找出“大脑”的哪个部分在进行某种压制。你刚才说，美国的公司往往关注某些安全问题，这些问题和比如说中国政府在意的敏感政治问题不同。但审计方法会因问题不同而不同吗？比如说，你会不会在模型里找到不同的部分，分别对应中国的政治敏感问题和美国的恐怖主义问题？

Cyberery：我确实看到它们有不同的优先级。我们测试 DeepSeek-R1 模型的时候，也测试了传统的有害提示词，比如关于危险行为或违法行为。我们发现，DeepSeek-R1 模型在政治敏感话题上跳过思考的情况，比在这些传统有害话题上更多。它们肯定有不同的机制，因为它们可能采用了不同的训练方案，而且必须遵守各自所在地的监管规定。

钟：对于身在中国的人来说——假设我还在中国，我会想弄到无审查版本的模型，或者至少找到比如说你已经调整过的 DeepSeek 开放权重模型的版本，这样就能和一个模型客观地交流，学我想学的东西，而不是被审查。

现在媒体上有很多炒作——也许不是炒作，我不知道，我很想听听你的看法——说在人工智能时代，就像在互联网时代一样，像中共这样的威权政权将能够掌握这项技术，基本上把人们置于更严密的控制之下。互联网时代确实如此。现在很多人关注的是，人工智能时代是否也会这样。有了人工智能，再加上中国对人工智能模型的控制——至少是对在中国开发的模型——会不会让人们更不自由？就像在互联网时代，你表面上觉得自己能获取更多信息，但最后反而更不自由，因为有防火长城，有各种各样的社会控制。结果是，即使人们在表面上能接触到更多的数据，但最终对世界的了解反而更少。

Cyberey: 是的，我能想象，人工智能方面的审查和互联网的审查会有点不同，因为人们可以直接从 Hugging Face（全球最大的开源模型分享平台——编者）下载开放权重模型，在本地使用。但也有一种担忧：如果中国政府不断加大力度，提高模型里的审查程度——不只是通过训练把某种审查机制嵌进模型，他们可能会开始从训练数据里删除信息。那样的话，我们目前的去审查技术可能就不够用了，因为模型是用训练数据训练出来的，如果中国政府开始筛选这些数据，我们就没法规避那些审查了。

钟: 我倒是能想到——当然不是要给中共出主意——我能想到一个可能：如果他们出台一项规定，要求所有人工智能公司都必须用中国互联网上的数据来训练模型，那所有模型都会是“干净”的，对吧？因为中国的互联网本身已经经过审查。你不能有任何不是 .cn 结尾的网站，也就是非中国域名网站的数据。那么，中国互联网全都是“干净”的，你的模型最后也会非常“干净”，尽管它可能能力不太强，因为

没有学到世界其他地方的很多东西。

Cyberery: 是的。我还在想，政府会不会也限制——现在的模型需要获取较新的信息时，往往不会只依赖自己内部的知识，而是会上网搜索其他信息。我在想，中国政府是不是也需要限制这些模型可以查哪些网站。

钟: 对。但那也很难做到吧？因为那几乎是即时的。根据查询的不同，模型会启动网页搜索功能，从世界上任何地方引入内容，然后模型得当场实时决定让哪些内容进来。

Cyberery: 是啊，不知道那要怎么过滤。

钟: 对。所以我真希望这能给中国那些以审查为业的人添很多活儿，让他们没法控制一切。这确实让我对人们手里能多一点自由抱有一点希望。这些讨论都非常有意思。我也很想知道——你现在是弗吉尼亚大学的博士后，最近写了这些很精彩的论文，我都读过——那你接下来要做什么？你的研究路线图上有什么？

Cyberery: 我马上要去一个新的职位，是在太平洋西北国家实验室（美国能源部下属的国家实验室）做研究。我将从事人工智能保障（AI assurance；指通过测试、评估和审计，确保人工智能系统安全可靠、符合预期的研究领域——编者）研究，这是我一直在做的研究的自然延续：审计，怎样控制模型的行为，确保它们可靠。所以我会继续这方面的研究。接下来我最想做的是改进我们评测模型的方式。我认为

这是很多机器学习工作的基础：我们怎样确定模型在进步，或者训练方法是否有效？我们需要有更好的评测基础，才能确保我们开发的这些方法都真的有效。

钟：祝贺你找到新工作。

Cyberery：谢谢。

钟：那么说到评测，具体指的是什么？你想的是一个产品，还是一套别人会遵循的行业标准流程？最后成形的东西会是什么样子？

Cyberery：是的，我认为在国家实验室肯定会有很好的机会，我们可以研究如何建立评测人工智能模型的标准方法，或者如何以更系统的方式进行审计。

钟：我很期待在你去国家实验室之后继续关注你的研究。

Cyberery：好的。

钟：谢谢你。我希望我们的读者也会喜欢今天的对谈。这是《中国民主季刊》系列访谈的一部分，我们请来嘉宾，谈论包括人工智能在内的各种热门话题，以一种有助于我们了解这个行业、也有助于我们理解和认识中国未来实现民主的可能的方式来谈——希望那一天不会太远。如果读者想进一步了解你的研究，他们可以在搜索引擎上输入你的名字 Hannah Cyberery。你在 GitHub 上也有一个网站：[hannahxchen](https://hannahxchen.github.io)。

github.io。

我推荐大家去看看 Hannah 的研究，非常精彩。再次感谢你接受《中国民主季刊》的专访。

Cyberey: 好的，谢谢。

说明：访谈录制后，Cyberey 博士提供了两个可以在线试用的互动演示。读者可以自己输入问题，打开或关闭操控功能、调节操控强度，看模型的审查如何随之变化。一个演示操控“思维压制”，使用由 DeepSeek-R1 蒸馏而来的一个 70 亿参数模型：<https://huggingface.co/spaces/hannahcyberey/DeepSeek-R1-Censorship-Steering>；另一个演示操控“拒答与遵从”，使用 Meta 公司的一个 Llama 3.1 模型：<https://huggingface.co/spaces/hannahcyberey/Refusal-Censorship-Steering>。这两个演示都放在 Hugging Face 平台上，大陆读者可能需要翻墙才能打开。她还写了一篇比论文通俗的英文博客文章介绍这项研究：<https://hannahxchen.github.io/blog/2025/censorship-steering/>。

注释

- 1 评测 (evaluation) 与审计 (auditing)：评测衡量的是模型“能做什么、做得多好”，通常让模型回答一套标准化试题（即“基准测试”）并打分，好比学生考试。审计则更进一步，关注模型是否真像开发者宣称的那样运作——是否存在偏见、隐藏的审查或安全隐患——好比会计师查账，要求拿出证据。
- 2 激活引导与引导向量：大语言模型处理一句话时，内部每一层都会产生一大串数值，称为“激活” (activation)；模型对某个概念（如“审

查”或“性别”)的理解,就以特定的数值模式隐含其中,称为“表征”(representation)。研究者比较模型面对不同类型输入(例如敏感话题与普通话题)时的激活差异,可以提取出代表这一概念的方向,即“引导向量”。模型运行时,把这个向量按一定比例加到激活上或从中减去,就能增强或削弱相应的行为,而无需重新训练模型——这就是访谈中所说的“旋钮”。

- 3 权重、开放权重与封闭权重;开源与闭源:可以把一个人工智能模型想象成一个已经训练好的“大脑”。模型在训练中从海量文本里学到的东西,最终都存成一大组数字,称为“权重”(weights),好比大脑中神经元之间连接的强弱。

按这组数字是否公开,模型分为两类。“开放权重”(open weight)指开发者把权重公开发布,任何人都可以下载到自己的电脑上运行,也可以打开来研究、修改,DeepSeek、Meta的Llama、阿里巴巴的Qwen都属于这一类。“封闭权重”(closed weight)指开发者不公开权重,用户只能通过网页、应用程序或接口向模型提问、拿到回答,既看不到也改不了模型内部,ChatGPT、Claude、Gemini都属于这一类。

“开源”(open source)与“闭源”(closed source)是另一对概念,源自软件业,指是否公开“源代码”,也就是制造一个程序的全部原料和图纸。对人工智能模型来说,真正的“开源”要求把训练代码、训练数据和训练方法一并公开,让别人能够从头复现这个模型。打个比方:开放权重好比送你一道做好的菜,你可以品尝、再加工,但不知道完整的菜谱和食材来源;开源则是连菜谱、食材清单和烹饪步骤一起交给你。目前许多被称为“开源模型”的,严格来说只是开放权重。封闭权重的商业模型当然也不公开源代码,所以访谈中钟伟锋说这些模型“闭源、闭权重”。

- 4 白盒与黑盒:“黑盒”指只能看到输入和输出、看不到内部运作的系统,研究者只能靠反复提问、观察回答来推测模型的行为;普通用户使用ChatGPT、Claude等商业模型,就处在黑盒的位置。“白盒”指能够直接查看和修改模型的内部参数(权重)与运算过程,这通常只有开发者本身,或研究开放权重模型的研究者才能做到。
- 5 指令微调、强化学习与安全微调:大语言模型先要用海量文本进行“预训练”,学会根据上文预测下一个词,但这时它还不会按人的要求对话。“指令微调”用大量“指令—回答”示例进一步训练,使模型学会听从指令、回答问题。此后通常还会用“强化学习”——由人或程序给模型的回答打分,奖励好的回答、惩罚不好的回答——来调整模型的行为。“安全微调”是其中专门针对有害请求的部分,训练模型拒绝诸如制造武器之类的问题。审查行为,也正是在这些训练阶段被“教”给模型的。
- 6 推理模型与思维链:推理模型(reasoning model)在给出最终答案之前,会先生成一段逐步推导的“思考”文字,即“思维链”,相当于打草稿。DeepSeek-R1、OpenAI的o系列模型都属于这一类;在DeepSeek的应用界面上,这段思考过程通常对用户可见。Cyberey的研究发现,

DeepSeek-R1 遇到政治敏感问题时，往往完全跳过这段思考，直接给出预设式的回答，她称之为“思维压制”。

- 7 拒答消除：英文 *abliteration*，是由 *ablation*（切除）和 *obliteration*（抹除）拼合而成的新词，是开源社区对一种去审查方法的俗称。做法是先找出模型内部代表“拒答”的方向（原理与注 [2] 的引导向量相同），再直接修改模型权重，把这一方向永久抹掉，使模型不再拒绝请求。网上流传的许多“无审查”模型就是这样制作的。正如 Cyberey 所指出的，这种方法只针对“拒答”，对“思维压制”等其他形式的审查未必有效。



访谈

区倬僊
季志强

中国民主季刊

第4卷 第4期

2026年10月

CC-BY-NC-ND

自由、免费转载分发，但须注明
作者与出处，并不得修改及商用

审查进了流水线： 看清中国人工智能 模型里的政治控制



区倬僖



季志强

编按：生成式人工智能是“一条很长的内容生成流水线。流水线拉长了，国家或当局这类外部行为者能够干预生成结果的环节也就多了”，区倬僖 (Owen Au) 说。北京正是沿着这条流水线，系统性地把它对内容和信息的政治要求转化为技术程序。区倬僖生于香港，现于英国从事独立中国事务研究与分析，为不同智库、咨询公司及非政府组织工作。2026 年 7 月，他与 Aiden Gu 合著发表报告《专制创新：中国如何将控制植入生成式人工智能系统》。受《中国民主季刊》委托，2026 年 10 月 3 日，计算政治学研究者季志强 (Eddy Ji) 就这份报告专访区倬僖，谈北京为何既倚重又提防生成式人工智能，政治要求如何借助技术标准写进数据、训练和测试各个环节，这套控制如何随中国模型扩散到海外，以及民主国家应当如何应对。

季志强 (Eddy Ji, 以下简称季)：感谢你接受我们的采访。今年 7 月，你与 Aiden Gu 发表了报告《专制创新：中国如何将控制植入生成式人工智能系统》(Authoritarian Innovation: How China Is Building Control into Generative AI Systems)。¹报告梳理了中国如何通过政

策、技术标准以及与合作企业的合作，在内容生成之前就把信息控制嵌入生成式人工智能的训练、评测和服务环节，并探讨了这种控制随中国模型扩散到海外可能带来的影响。这也是我们今天这次访谈的缘起。不过，在讨论这份报告之前，想先请你向读者介绍一下自己。你写过很多关于中国政治的文章，尤其关注中国的海洋战略。最初是什么让你对研究中国的政策问题产生了兴趣？这一路上有没有什么经历或转折点？

区倬僖 (Owen Au, 以下简称区)：好的，谢谢邀请。我现在是常驻英国的独立中国问题分析员，为多家智库、咨询公司和非政府组织工作。

我在香港出生长大。香港名义上是在中国主权下实行高度自治的城市。从十几岁起，我就亲眼目睹了中国的影响力不断扩大，姿态也越来越强硬。香港和中国大陆的这些政治变化，引发了香港一连串大规模抗议和政治风波，随之而来的是当局的镇压，其中包括 2020 年强行实施的《香港国安法》。我后来研究中国政治、成为中国问题分析员，都与这段经历有关。

离开香港、移居英国后，我开始从更宏观的层面观察中国。读硕士时，我主要研究中国在印太地区的地缘战略。印太这个战略舞台以海洋为重心，台湾海峡、南海和东海都在其中。中国可能对台采取的军事行动是当下战略讨论的热点，中国的海军能力和制海权正是其中的核心议题。我对中国海洋战略的研究就是这样开始的。

过去一年，全球科技竞争的分量越来越重，我的研究重点也逐渐从海

洋战略转向中国更广泛的科技战略，包括中国如何建设自己的技术能力，又如何谋求在全球科技治理和标准制定上的影响力。这也是我们今天要谈的话题。

季：那我们来谈谈这份报告。我们的读者对网络上的信息管控和言论审查并不陌生，但生成式人工智能似乎不太一样。对于不熟悉这项技术的读者，能否解释一下，北京一方面把生成式人工智能当作监控和宣传的工具，为什么另一方面又把它视为对信息管控的威胁？

区：生成式人工智能和中国的许多其他技术一样，非常强大，也具有颠覆性。中国已经在用这项技术加强多个领域的政治控制，比如网络监控和你提到的审查。许多研究和报告也指出，中国正积极利用这项技术在国内外开展信息行动，试图影响欧洲、美国和其他自由民主国家的信息空间。

不过，生成式人工智能在给中国政府带来这些好处的同时，也给北京带来了风险，对一个要把包括技术在内的方方面面都置于政治控制之下的政权来说尤其如此。与国内其他重大技术或战略性技术不同，生成式人工智能几乎人人都能用。即使是科技素养不高的普通中国人，也能在日常生活和工作中轻松使用它。在北京看来，这一点格外危险，因为这项技术可能被用来对当局发起非对称行动（力量较弱的一方以低成本手段对抗强大对手——编者），比如大量生成和传播与官方叙事相抵触的内容，或者用来了解和分析领导层的行为，而这些都是当局不愿看到的。

即使没有人蓄意对当局发起这类行动，生成式人工智能还有另一层风险。这项技术有两个特点尤其让当局担忧。第一个是人工智能的“黑箱”问题。许多中国人工智能研究者已经指出这类技术的局限。很多开发者和研究者至今仍难以完全解释模型生成了什么内容、是怎样生成的，更不用说加以控制。

第二个特点是，许多中国人工智能研究者也指出，如今的生成式人工智能和模型训练严重依赖西方数据。模型训练需要消耗大量数据，其中大部分是网络内容，而如今很多网络内容是用英语写成、来自西方国家的。中国当局往往认为，这些数据嵌入了自由民主国家的价值观和意识形态预设，因此把它们视为需要遏制和化解的国家安全风险。

所以，即便是中国本土研发的人工智能模型，也可能生成偏离北京官方叙事的内容。在中国，这类风险被正式归为“内容安全风险”，纳入国家的《人工智能安全治理框架》（该框架 1.0 版于 2024 年 9 月发布，2.0 版于 2025 年 9 月发布 -- 编者）。

季：你提到当局担心的风险之一，是训练数据大多来自西方，可能威胁国家的内容安全。由此我想到一个相关的问题。根据 Anthropic 最近的一份报告，一些中国人工智能实验室在完成预训练之后，仍在通过“蒸馏”技术²学习或借用 Anthropic 的 Claude、OpenAI 的 GPT 等国外前沿模型的能力，有些公司甚至直接把用户的请求转发给这些国外模型，再拿回答案。³中国实验室直接或间接地使用国外模型的回答，会不会在内容控制体系中打开一个缺口？

区：简单地说，我认为不会。因为保持政治控制、缓解令中国当局非常担忧的内容安全风险，是任何技术在中国得以投入使用和商业化的基本前提。所以，即便那些中国人工智能实验室正在从美国和其他西方国家大力汲取技术，并试图把这些技术用到自己的本土模型中，在面向公众之前，它们都必须经过一系列技术测试，以确保模型的回答与官方叙事保持一致——即便不是完全一致，也要大体一致。

季：你的报告让我认识到，政治审查早在模型给出回答之前就已介入，贯穿训练数据筛选、数据标注和模型测试等环节。我自己用几个中国模型做过测试，问过一些敏感问题，比如有哪些中国人获得过诺贝尔和平奖。其中一个模型已经开始作答，我能看到文字一个个蹦出来，但很快整段回答就被撤回了，换成一句“我们聊点别的吧”。这是用户能看到的部分。在这背后，政治要求是怎样一步步变成训练、标注和测试中的具体规则的？这些规则最终又怎样影响用户得到的回答？

区：你做的这种测试很有意思，许多研究者也经常这样测试这些模型的行为。中国的人工智能模型其实相当擅长拒绝回答敏感问题，让回答与中国共产党的立场保持一致。

要理解政治控制和言论审查在生成式人工智能中如何运作，首先要了解人工智能的内容是怎样生成的。过去谈网络内容审查，或者谈我们怎样在网上分享内容，事情相当简单。我作为内容创作者写好内容，发到社交媒体之类的网络平台上就行。生成式人工智能的内容生成过程却被拉长了，中间要经过数据收集、数据筛选和输入、模型训练、数据标注和处理等环节，最后才生成内容。这是一条很长的内容生成

流水线。

流水线拉长了，国家或当局这类外部行为者能够干预生成结果的环节也就多了。正是基于这种认识，中国自 2024 年以来花了很大力气，建立起一套庞大的人工智能技术标准体系，确保这条流水线每一个环节的技术标准都符合政治要求。

在 2024 年以来出台的众多标准中，我认为最值得注意的之一，是 2025 年发布的《生成式人工智能服务安全基本要求》。⁴ 它规定，所有中国人工智能模型在面向公众、投入商用之前，都必须通过一套基准测试集（基准测试是用一套统一的题目来评估模型表现的方法，类似于给模型做标准化考试 -- 编者），以确保内容安全。这套测试集共有 2000 道题，其中三分之一专门测试生成内容是否违反中国共产党所谓的“社会主义核心价值观”。⁵ 要通过测试，模型对可能违反内容安全的提示词，拒答率必须至少达到 95%。

与此同时，这一框架又从内容生成这一层扩展到其他技术环节，比如数据标注、机器学习和算法训练。⁶ 这清楚地表明，国家正在整个人工智能技术栈中，⁷ 系统性地把它对内容和信息的政治要求转化为技术程序。

季：接下来想谈谈中国模型走出国门之后的情况。按你的报告，中国政府对模型生成内容的控制，会随着海外用户采用中国模型一起传到国外，背后并不需要什么有组织的影响力行动。中国技术以前也走出去过，比如华为在巴基斯坦的“平安城市”项目，又比如因内置政治

偏向、压制敏感内容而屡遭质疑的 TikTok。按照政治要求训练和测试出来的中国人工智能模型，在海外产生影响的方式与这些先例有哪些异同？

区：我认为，中国人工智能模型和技术的扩散，与其他中国技术和产业涌向全球的模式相当相似。令人意外的是，背后的逻辑相当资本主义、相当商业化。太阳能电池板、可再生能源、电信等战略性产业受到中国国家重点扶持，多年来享受大量补贴，又是在“内卷”的环境中发展起来的。它们的产品和服务因此非常便宜，质量又足以让海外用户接受，在全球市场上极具竞争力。

人工智能技术也是如此。这些技术未必最尖端，中国最先进的前沿模型目前仍落后于美国。但它们的能力已经足够好，又是开放权重的，⁸考虑到价格就更显得划算，因此在全球很有吸引力，在全球南方国家尤其如此。这些国家大规模采用中国廉价又好用的产品和基础设施，也就更加依赖中国的科技产业和标准。随着中国积极把政治要求嵌入生成式人工智能技术，这些要求的影响就不只限于国内，还会波及依赖中国科技产业的国家。

季：我还有一个问题，可能超出了你的研究范围。有没有证据表明，中国企业或政府机构正在帮助其他专制政府建立自己的人工智能审查系统，比如提供内容过滤工具、训练流程或技术支持？

区：生成式人工智能方面倒没有特别的例子。不过几十年来，尤其是近些年，中国一直向其他专制国家输出技术，助长海外的专制统治，

这方面已有大量记录，也广为人知。东南亚、俄罗斯和拉丁美洲的一些国家，都从中国引进了人脸识别、电信和舆情监控技术。

最能说明问题的例子之一，是柬埔寨的国家互联网网关。柬埔寨在2020年代初宣布设立这一网关，旨在让政府对国内和国际互联网连接拥有更大的管控权。⁹ 由于技术上相似，观察人士常把它称为柬埔寨版的“防火长城”。许多研究者一直在追查是谁在帮柬埔寨搭建这套基础设施。尽管尚未得到官方证实，柬埔寨公民社会的一些专家推测是华为或中兴。¹⁰ 这个例子很能说明中国如何在海外推行技术威权主义（指借助数字技术实施监控和社会控制的威权统治方式——编者）。

季：接下来想谈谈商业抱负和政治控制之间的张力。你的报告指出，中国领先的科技企业和人工智能实验室参与起草了生成式人工智能的相关标准，与国家一起制定内容安全规则，同时也接受国家的政治控制要求。但政府和企业追求的目标本来就不同。政府要实现人工智能的自立自强，构建从高端芯片到基础软件都自主可控的软硬件体系；企业则想要商业成功，想要全世界的用户。这些目标怎样协调，又在哪儿发生冲突，尤其是在海外用户希望自由下载、运行和修改这些模型的时候？开放与政治要求相互拉扯时，由谁来划定界线？企业手里有没有自己的筹码？

区：这是个很好的问题。如果问由谁来划定界线，毫不意外，是中国政府。

至于企业有没有筹码，当然有。中国虽然是专制国家，但政府也承

认，国家的大部分创新和资本不在政府手中，而在私营部门。它要实现自己的科技抱负、扩大全球影响力，就必须与私营部门合作。如何处理与私营部门的关系，在商业成功、技术创新和政治控制之间取得平衡，一直是中国当局面临的难题。

2020年代初，习近平发起了一系列打压私营部门的行动，马云就是很典型的例子。马云是中国最知名的私营企业之一阿里巴巴的联合创始人，也是中国商界的成功典范。马云多年来一直公开批评国家干预市场、对金融业严加监管的倾向。他公开反对国家干预，随后便从公众视野中消失多年，令全世界震惊。¹¹ 许多人认为，马云的遭遇说明国家试图规训私营部门，确保自己对私营部门的控制。

经过一系列打压和整顿，到了2025年，习近平开始释放信号，表示欢迎私营部门，鼓励它们在多年规训之后去追求商业成功。但这一次，北京在鼓励私营部门追求商业成功和创新的同时，也明确表示，所有创新和发展都必须在政府划定的边界之内进行。

最值得注意的是，北京采取了胡萝卜加大棒的做法。一方面，政府提供大量激励，促使私营企业向国家的战略目标和优先事项靠拢。另一方面，像马云那样偏离国家意志的企业，就会受到严厉惩罚。到目前为止，这套做法相当成功。它给私营科技企业留出了一些追求商业成功、推动技术创新的自主空间，同时也提醒它们，中国共产党的地位不容挑战，从而强化了这一权力结构。

季：下面从企业一方来看这种政商关系。这些政治要求会不会给企业

的创新和全球竞争力带来很大的成本？你刚才提到，一些基准测试专门检验模型是否符合社会主义核心价值观，而这些价值观里本身就有“自由”“民主”这样的词。如此宽泛、模糊的要求，要落实成训练和测试中前后一致的标准，企业要付出多大代价？

区：在国内，我认为这些科技企业要符合国家的政治要求，并不需要付出多少成本；相反，如果不积极配合国家的优先事项和政治要求，代价才会非常高。但放到全球来看，从国际竞争力和全球市场份额的角度说，这类政治要求以及控制与审查的逻辑，已经在影响中国科技企业的国际形象，华为就是一个例子。

许多人认为，2019 年美国对华为的禁令主要出于中美之间的地缘政治竞争，但它也反映出自由民主国家共同抱有的一些合理担忧：中国积极利用技术，在海外获取战略筹码、服务国家战略目标。中国《国家情报法》出台后，这种担忧更为突出，因为该法要求中国的组织和个人配合国家的情报工作。人工智能、电信等高科技企业在海外运营，需要处理大量数据。各国因此担心，这些数据和业务会不会成为渠道，让情报或敏感信息流入中国的情报部门。

季：你刚才谈到，中国科技企业在海外运营要处理大量数据，民主国家担心这些数据会流入中国的情报部门。这是国家层面的安全担忧。我想换到个人开发者的角度。很多开发者偏爱中国的开放权重模型，因为它们成本低，可以在自己的电脑上运行、按需定制，用起来更自由。我在 X 上看到，不少开发者选择这些模型，正是为了不把自己的

数据交给西方的大型人工智能公司。如果一个人主要用中国模型写代码，从不和它谈政治，审查问题还值得他担心吗？还有哪些风险？这些担忧有什么证据？

区：我们在报告中讨论的政治控制、审查和信息控制，主要影响的是那些会波及信息空间的使用场景，比如教育、公共传播和媒体。如果是写代码这类纯技术用途，对信息空间的影响可能就没那么大。

但我认为风险在于，自由民主国家的政策制定者其实很难追踪和评估这些中国模型是怎样被采用的、用在了哪些场景。我们很难核实，它们是只用于技术用途，还是已经进入了公共传播领域，因此也很难追踪和化解由此带来的潜在风险。难以追踪、难以核实，这本身就已经是一种风险。

季：你在报告中也谨慎地指出，下载量和采用量说明不了这些模型在媒体、教育和公共传播中实际怎样被使用。正如你所说，这很难追踪，也很难评估。那么，中国政府的内容控制是否已经在影响海外的信息环境？这仍是一个悬而未决的问题吗？目前我们知道些什么，又能做些什么？

区：这确实很难追踪，也很难核实，但风险已经存在，而且清晰可见。对这个问题我没有明确的答案，这一点我承认。不过我认为，自由民主国家的政策制定者，以及当地的企业和开发者，无论是自己采用中国人工智能模型，还是监管本国对这些模型的采用，都应该更加谨慎。他们必须清楚自己要付出什么代价、承担什么风险，也要更加

警惕可能的后果。这些后果不只涉及信息空间，也可能影响他们自身的运营和技术栈。

季：你认为西方的政策制定者准备好应对中国人工智能模型崛起带来的挑战了吗？民主国家现有的人工智能法规和治理框架，比如欧盟的《人工智能法案》（AI Act），能在多大程度上应对模型内置、用户甚至看不到的政治内容控制？应该由谁来发现和纠正这些控制？现有的保障措施又有哪些不足？

区：全球人工智能治理（AI governance）和监管在今天确实是非常重要的问题。众所周知，人工智能安全（AI safety and security）¹² 问题日益关乎全人类和整个文明，需要各大国合作应对。据我所知，美国、英国和欧盟等自由民主国家已经与中国开展了广泛合作，推动二轨对话之类的交流（由学者、前官员、智库人士等以非官方身份参与的对话，与政府之间的“一轨”外交相对 -- 编者），希望加强沟通，强化全球人工智能治理。

但我认为，这份报告也强调了一点。谈到人工智能监管或人工智能安全时，各方用的词语虽然相似，比如人工智能安全（AI safety）、内容安全（content security）、内容审核（content moderation），背后的含义却往往截然不同。在欧洲，谈内容安全或内容审核，更多是为了减少人身攻击、侵犯隐私或种族主义之类的风险，保护个人的权利和尊严。在中国，谈人工智能安全和内容安全，更多是指中国国家和中国共产党如何保持对这项技术及其生成内容的政治控制。同样的词语，在不同司法管辖区背后的逻辑和解读大不相同。

如果自由民主国家的政策制定者和开发者意识不到这些差异，就可能正中中国下怀，在不知不觉中接受中国的解读。何况中国正在积极推广自己的标准，扩大自己在全球科技治理中的影响力。

季：你刚才谈到，同样说“安全”，中国政府的理解与欧美完全不同，而中国又在积极推广自己的标准。这正是报告序言所担心的“人工智能治理的专制化”。我想接着问，一个国家在全球人工智能治理中的话语权，在多大程度上取决于它的人工智能产业实力？我们得承认，中国的人工智能实验室做得非常出色，而除了美国之外，很难看到哪个民主国家开发出了被广泛使用的主流大模型。对其他民主国家来说，要防止人工智能治理走向专制化，是否必须打造自己的模型来替代中国模型？还是说，即使在模型研发上比不上中国，它们仍然可以发挥真正的影响力？

区：任何国家要在人工智能监管和全球科技治理中建立影响力、获得话语权，自身都必须具备科技竞争力，这一点毫无疑问。中国对人工智能治理和人工智能安全的理解之所以举足轻重，正是因为中国已经成为当今的科技超级大国之一。

但具备科技能力，并不等于一定要自己研发最前沿的人工智能模型。全球技术栈非常复杂，供应链相互交织。我认为，一个国家只要能在全球技术栈和供应链中占据重要位置，就有潜在的筹码和力量参与这些讨论。台湾就是很好的例子。台湾虽然也设计芯片，但不及美国先进，也不研发最前沿的人工智能模型；不过凭借在芯片制造中的地位，它在全球人工智能技术栈中不可或缺，也因此在全球科技讨论中拥有

相当大的筹码和话语权。

英国、欧洲国家等其他自由民主国家，如果想在科技治理和科技竞争中发挥更大作用，就要在全球技术栈中找到自己的优势和着力点。

季：最后，想请你给我们的读者一些日常使用上的建议。读者平时可能会接触到中国的人工智能模型。有人看中的是它们性价比高，还能在自己的电脑上运行；也有人用的应用或工具其实以中国模型为底层，自己却未必知道。他们需要刻意避开中国模型吗？如果要用，应该注意些什么？

区：中国如今在全球技术栈中举足轻重，完全与中国技术脱钩不太现实。任何人，尤其是开发者，觉得中国的人工智能模型或技术基础设施很有吸引力，我认为可以理解。但就像我前面说的，使用这些技术时，一定要清楚其中的风险和可能的后果，也要清楚自己在用什么代价换取这些便宜又好用的技术。

季：非常感谢你，Owen。相信这次访谈会让我们的读者受益良多。也推荐大家读一读报告原文，一起关注和讨论如何防止人工智能治理走向专制化。

注释

- 1 区俾僖：Aiden Gu：《专制创新：中国如何把控制植入生成式人工智能系统》（Owen Au and Aiden Gu, *Authoritarian Innovation: How China Is Building Control into Generative AI Systems*, *Resilience Innovation Lab and Strategic China Policy Group*, July 2026, <https://>

resilienceinnovationlab.org/wp-content/uploads/2026/07/Authoritarian_Innovation_2026.pdf)。

- 2 预训练 (pre-training) 是开发大模型的第一阶段：用海量文本训练模型，让它掌握语言规律和一般知识，之后再经过“后训练”，让它学会按人的指令作答。蒸馏 (distillation) 原指用一个大模型 (“教师模型”) 的输出来训练另一个模型 (“学生模型”)，让后者以较低成本获得接近前者的能力。未经许可、大规模调用他人模型来获取训练数据，违反主要模型开发者的服务条款。前沿模型 (frontier model) 指能力处于当前最先进水平的大模型。-- 编者
- 3《Anthropic: 检测和对抗AI的滥用》(Anthropic, *Detecting and Countering Misuse of AI: September 2026, September 10, 2026*, pp.143-154, <https://www-cdn.anthropic.com/e50be2e51e7695dc4b1366a37a245a597377d3b5/Anthropic-Detecting-and-countering-091026.pdf>)。
- 4 国家市场监督管理总局、国家标准化管理委员会：《网络安全技术 生成式人工智能服务安全基本要求》(GB/T 45654-2025)，2025年4月25日发布，2025年11月1日实施。
- 5 这套包含2000道题的《生成式人工智能服务安全基准测试集》由中国电子技术标准化研究院与复旦大学于2024年联合开发，用于配合上述标准实施，其中约三分之一的题目测试生成内容是否违反“社会主义核心价值观”。同注释1，第18-19页。
- 6 国家市场监督管理总局、国家标准化管理委员会：《网络安全技术 生成式人工智能预训练和优化训练数据安全规范》(GB/T 45652-2025)、《网络安全技术 生成式人工智能数据标注安全规范》(GB/T 45674-2025)，均与上述标准同日(2025年4月25日)发布，2025年11月1日实施。
- 7 技术栈 (tech stack) 原是软件工程用语，指构建一个系统所需的、层层叠加的各种技术。就人工智能而言，从下到上大致包括芯片和算力、数据中心、训练数据、模型，以及基于模型开发的各种应用。-- 编者
- 8 权重 (weights) 是模型在训练中习得的海量参数，决定了模型如何生成内容。开放权重模型 (open-weight model) 指开发者公开这些参数、允许任何人下载模型，用户可以在自己的电脑或服务器上运行，也可以加以修改和再训练。DeepSeek、通义千问 (Qwen) 等中国主流模型多以这种方式发布。开放权重不完全等同于“开源”，开发者通常并不公开训练数据和完整的训练代码。由于政治要求在训练阶段就已写入权重，模型被下载到海外、在本地运行时，这些要求也会随之带走。-- 编者
- 9 柬埔寨首相洪森于2021年2月签署设立国家互联网网关的次级法令，要求所有互联网流量于2022年2月前经由该网关传输，但政府原定启用的前一天宣布推迟实施。见《非政府组织敦促柬埔寨废除互联网管制法令，称其影响在线权利》(“NGOs Urge Cambodia to Scrap Decree

on Internet Controls Citing Impact on Online Rights,” *Radio Free Asia*, February 18, 2021); 索恩·萨拉特:《争议性国家互联网网关因项目批评增加而推迟》(Sorn Sarath, “Controversial National Internet Gateway Postponed as Project Criticisms Grow,” *CamboJA News*, February 15, 2022)。据《日经亚洲》报道,柬埔寨计划于2026年开始建设该网关的关键基础设施,见菲奥娜·凯利赫:《柬埔寨重启备受争议的互联网网关计划》(Fiona Kelliher, “Cambodia Resurrects Plan for Controversial Internet Gateway,” *Nikkei Asia*, October 20, 2025)。

10 见文章 19:《数字丝绸之路:中国与印太地区数字镇压的兴起》(ARTICLE 19, *The Digital Silk Road: China and the Rise of Digital Repression in the Indo-Pacific*, 2024, p.36, https://www.article19.org/wp-content/uploads/2024/04/DSR_final.pdf)。

11 2020年10月24日,马云在上海外滩金融峰会上发表演讲,批评中国的金融监管思路。同年11月3日,蚂蚁集团原定在上海、香港两地的上市被暂缓。此后马云长期淡出公众视野,直到2023年才较多地在国内公开露面。2025年2月17日,习近平主持召开民营企业座谈会,马云应邀出席,外界普遍视之为北京向民营企业释放善意的信号。--编者

12 英文中, AI safety 一般指防止人工智能系统本身造成伤害,比如出错、失控或输出有害内容; AI security 一般指保护系统免遭蓄意攻击和滥用,也常与国家安全相联系。中文把两者都译作“安全”。中国官方所说的“内容安全”(content security),除了一般的有害内容,还包括生成内容是否符合政治和意识形态要求。



访谈

吕频
艾欣

中国民主季刊

第4卷 第4期

2026年10月

CC-BY-NC-ND

自由、免费转载分发，但须注明
作者与出处，并不得修改及商用

十年间，专制政治 背景下中国女权运 动的发展



吕频



艾欣

编按：“三十年前，当我投身女权运动的时候，这个词汇在中国的环境里是惊世骇俗、令人畏惧的”，中国重要的女权主义倡导者、正在美国罗格斯大学的吕频说。吕频曾为《女声》电子报主笔，也是《女权之声》自媒体的创办人——当然，她所做过的并不止这些。从1990年代中期到现在，她投身和见证了当代中国的女权主义运动。受《中国民主季刊》委托，目前在中国大陆的女权活动人士艾欣专访了吕频。这个访谈的主题正如其标题所显示，但内容的丰富、思想的张力却远超出标题所能传递的限度。

艾欣（以下简称艾）：吕频老师，请问据你的观察和研究，中国政府针对女权行动者的镇压形式有什么样的特点？近十年来发生了怎样的变化？

吕频（以下简称吕）：针对女权主义者，十年间发生了三次影响深远的集中镇压，分别发生在2015、2018和2021年。这三轮镇压的目

标实际上都是同一批人，即“青年女权行动派”（下文简称“女权行动派”）的成员及其主要盟友。自 2012 年以来，她们在中国女权运动中扮演着三重角色：运动网络的构建者、运动话语的生产者、倡导行动的组织者。

女权行动派的活动模式以核心组织者为枢纽，连接分散的年轻志愿者网络。她们公开拥抱女权主义标签，采取引人注目、有时甚至带有挑衅性的行动，旨在吸引媒体关注，并向政府施压，促使其改变政策。国家也一度以在特定政策和个案上的让步作为回应。

然而，随着习近平巩固权力，国家对待公民社会的方式日益从容忍转向镇压，其核心是以维护国家安全的名义将潜在威胁活动定为犯罪，以及对运动采取主动而非被动的压制手段。女权运动的影响力和对抗性刚刚开始增长，却已进入了这种控制逻辑的目标范围。

2015 年 3 月 5 日至 6 日，因为计划在十个城市联合开展反对公共交通性骚扰的倡议行动，多个地区的女权志愿者和组织者被带到当地派出所接受讯问。其中五名组织者最终被以“寻衅滋事”的罪名拘留，在北京看守所被关押了 37 天。与此同时，警方在多个城市对女权组织和活动人士进行了广泛的搜查。这是国家首次对女权活动进行刑事镇压，后被称作“女权五姐妹”事件。

在国内和国际社会的巨大压力下，检察机关以证据不足为由拒绝批捕，“女权五姐妹”获准保释。但在接下来的一年里，她们仍作为刑事嫌

疑人受到持续的监控，其中三人被迫离开她们曾工作的组织。更深层次的影响是，女权行动派自此被排除在与体制的合作机会之外。媒体对女权活动的报道几乎完全停止；女权主义学者开始与这些活动家及其主张保持距离；非政府组织和妇女组织也停止寻求与她们合作。

2018年3月8日，中国最早也是当时最具影响力的女权主义媒体平台“女权之声”在微博和微信上同时被永久封禁，理由是涉及“敏感、非法信息”，而真实原因是其在米兔运动中的组织和传播工作。平台工作人员受到警方传唤。与此同时，对女权主义的诽谤开始在网上传流。

3月中旬，一个煽动性的民族主义账号“酷玩实验室”发布了捏造的指控，称女权之声及其关联人员与“境外势力”勾结实施性犯罪。这些指控得到了广泛传播，针对它们的反驳却受到言论审查。

由于警方的警告，“女权之声”再也没有重建账号，并从此停止了所有公开活动。此后，女权运动中再也没有出现同时具有行动主义倾向且具备核心影响力的账号，取而代之的是大量小型、自发且不稳定的账号，如雨后春笋般涌现。

2021年4月，因为女权活动家肖美丽的一次微博公开发声，一些民族主义账号挖出她多年前发表的内容，将其解读为涉及“港独”和“境外势力”，由此引发了一场针对她本人及相关女权活动家群体的网络围剿。拥有数十万粉丝的“大V”账号散布政治指控；数万网民加入攻击行列，进行人肉搜索和威胁；平台大规模删除女权主义者的反驳内

容，并以“有害信息”为由封禁账号；微博运营方和政府机构在关键时刻都对这些攻击表示支持。与此同时，女权活动家遭到警方调查，被警告不得公开发言，并被全平台封杀。

镇压范围随即扩大到广泛的女权和泛女权网络社群：豆瓣十几个女权主义小组被封禁；武汉的一个高校学生社团因组织校园抗议而被取缔。最终，所有遭受攻击的活动人士都在中国公开的社交媒体平台上销声匿迹，再也没有以自己的名义组织过任何活动，而更广泛的女权活动也因此转入地下或被迫终止。

三次镇压之间的差异，反映了国家能力的增长及其策略的转变。2015年，国家采取了直接行动，对女权主义者进行拘留和审讯。当时社交媒体的监控和审查机制尚未高度完善，也未系统性地针对女权主义者，因此抗议活动仍能被发布和扩散，从而形成席卷国内外的声援浪潮。镇压所带来的声誉和外交代价，远远超过了政府的承受能力，迫使其选择低调处理。

相比之下，2018年的种种迹象表明，国家有意避免引发无法控制的反弹，其镇压手段变得多样化且间接。第三方被动员起来参与镇压，做国家自身无法做或不愿做的事情，而更先进的审查技术和传播手段则为此提供了助力。女权之声被封号后，曾向多个政府机构申诉，但没有一个机构承认发布过禁令。传播诽谤信息的账号持有亲政府立场，而这场污名化运动虽然充斥着政治和性方面的敏感词汇，并被无数次转发，却从未遭到审查，这成为唯一能让人怀疑背后有国家力量支持的证据。

到 2021 年，国家与第三方的合作变得更加战略化，甚至无需任何直接接触即可进行。针对女权主义者的攻击由具有民族主义倾向的账号和普通网民发起，攻击的焦点却精准地契合了党国体制的政治敏感性——真正具有破坏性的并非攻击目标所持的女权立场，而是她们被指控支持香港独立并与外国势力勾结进行颠覆活动。

中共安全系统官方社媒账号对这些指控的间接支持，赋予了攻击极大的合法性。包括共青团、中央网信办和地方公安机关在内的一系列官方社交媒体账号，以日益严厉的措辞攻击女权主义。除了“勾结西方势力”和“扰乱社会秩序”的指控外，还制造“极端女权主义”一词，将女权主义描绘成仇恨男性、制造性别对立、分裂社会。

女权主义被官方定义为国家的敌人，这种转变为无限制的镇压提供了正当理由，使其能够吸引越来越多的参与者，并严重损害了公众对被镇压者的支持。大众民族主义与反女权的潮流融合在一起，最终形成的是一种针对女权主义者的全面敌对氛围。每次镇压都针对女权运动的组织核心和基础，导致其组织能力显著萎缩，并产生了巨大的威慑作用，为运动幸存者划定了不可逾越的红线。

在大众层面，发表女权言论或认领女权主义者的身份，也随时可能给发声者带来风险。结果是“女权主义”一词在公共讨论中的可见度下降，越来越多地被表述为淡化权利色彩的“女性主义”，或干脆被“独立女性”、“女性力量”等流行词汇取而代之。

艾：这些镇压如何作用于个人？针对女权主义者的镇压是否存在一些性别特征？

吕：国家试图压制的不仅是个体行为，更多是一个群体的持续政治潜力——或者更准确地说，是让政治行动成为可能的条件。这一目标要求暴力必须是先发制人的、持续的，并融入被镇压者的日常生活，而不是反应性和阶段性的。同样重要的是，暴力必须保持表面上的不可告人，因为任何明显的镇压行为都会反过来赋予行动者政治可见性。

我把这种暴力形式称为“国家主导的分布式暴力”：在大规模集中镇压之间的空档期，日常镇压的形式日益多样化，渗透到生活的方方面面。国家通过激活已存在于日常制度中的国家权力，调动一个分散的加害者网络，涵盖政府机关、企业以及私人层面的亲密关系。在侵蚀行动者的日常生活、限制他们行动的能力的同时，分布式暴力还分摊了施压的成本，使国家免于被问责，阻止了可能的公众反弹。

一位参与反性骚扰倡导的中国女权主义者，可能在短短几周内遭遇这些事件：警察前往她的父母家中施压；公司希望她辞职；房东无理由地终止租约；社交媒体账号被连环封禁；物业管理处监视她的行动。这些都无需被宣布为政治行为，每一项都可以使用不同的借口被解释为孤立事件。然而它们同时发生，相互配合，显示出背后存在统一的策划。

这一镇压秩序成立的基础条件是国家拥有高度的强制能力和渗透覆盖力，且市场在国家的管理下不完全自主。同时，性别化的社会秩序可

被用作纪律工具。

中国父权规范通过长幼尊卑的亲属关系、性别化的劳动力市场、性化女性的职场文化以及社会对女性的道德规训，已经约束了女性的政治声音和公共流动。国家可以在关键时刻强化激活这些规范，将潜在的性别偏见武器化，构成对女权行动者的巨大压力。

许多女权运动参与者是离开家乡前往大城市的年轻女性，她们在日常生活中遭受的压迫，源于自身与性别相关的多重脆弱性。她们的女权主义认同与一系列生活方式的选择紧密相连，例如保持单身、与原生家庭分居、建立同性伴侣关系，以及在非政府组织、自由职业或不稳定的创意行业工作。这些选择也增加了她们的风险。

如果一位女权主义者没有当地的户口或房产，地方政府就更有借口驱逐她，捉襟见肘的经济储备意味着任何麻烦都可能迅速耗尽她的精力。而被视为性别规范的违背者，使得对她们的压迫被解读为对“不守规矩”的女性进行必要的管教，而非政治迫害。如警方可能会把一对女同性恋者的性取向透露给房东，以此劝说对方解除租约。

在国家的指导下，这种分布式暴力往往具有超法律性和任意性。没有明确的规则把具体行为与具体后果联系起来，因而行动者无法预测镇压何时、如何发生，也没有可靠的规避途径。这种不可预测本身就是一种控制方式，不仅限制行动本身，还限制行动者评估自身情况的能力。

对于被标记为敏感人物的行动者来说，困境并非单一的危机，而是一种持续且反复出现的匮乏状态，这种状态在他们搬家、辞职、迁往另一座城市、在网络上噤声，甚至与女权社群断绝联系后依然会持续。

2015年后，除了极少数例外，女权活跃人士没有面临任何刑事指控。然而最终，驱使许多人离开运动的是安全感的丧失、身心俱疲，以及无力感。当这种情况影响到足够多的运动核心组织者时，运动就会失去支柱。

艾：2018年-2022年，在运动环境日益恶劣、公共空间逐渐关闭的情况下，“米兔”运动却在中国爆发，并成功地催生公共倡导和集体行动。这一切如何发生？在去中心化、依靠个人叙事的表象之下，女权主义者如何开展组织性的工作？

吕：通常人们从米兔运动中首先看到的是个人叙事，这当然很重要。但如果眼光只限于此，把米兔主要理解成一种社交媒体现象，这就是冷落了它的运动属性。运动必须是集体行动，不是一个人就能形成一场盛事，而且其中个体的行为也不是孤立和完全自发的。

把社交媒体比作广场，个人叙事就是当事人在无法诉诸制度性的解决渠道时，所采取的数字广场抗议；围绕当事人的声音，背后还有大量的工作在展开。只是由于镇压，这些工作没有机会获得公开承认。

米兔在它开端的2018年，大致可以分成三个阶段。第一阶段从1月持续到3月。1月1日，北京航空航天大学前博士生罗茜茜实名举报

教授陈小武性骚扰。她同时发起联署，呼吁北航创立校园性骚扰防范机制，而当时中国高校尚无此类机制。可见从一开始，不管是发声的幸存者，还是响应她的支持者，都将个案的救济和制度性改变紧密连接在一起。

女权行动派敏锐地捕捉到这一契机，决定将高校反性骚扰机制的倡导工作扩展为全国范围的联动。罗茜茜举报教授的第四天，女权行动派的核心成员张累累发起了“万人致信母校反性骚扰”活动，她担任此次活动的对外发起人，并建立了一个微信协调群，将核心组织者与全国各高校值得信赖的志愿者联系起来。组织者设计了一份针对教育部和高校管理部门的具体政策诉求请愿模板；各校园组织者根据自身情况进行调整，并开展自主动员。许多校园成立了行动小组，保持密切的内部沟通，并直接向校方递交了请愿信。活动最终吸引了来自 77 所高校的一万三千多名学生和校友参与。

“万人致信”活动的珍贵之处在于，它不仅让反性骚扰机制成为一个热点话题，还在社会、公众层面广泛汇集了一批关心性骚扰议题、有公民意识的人——即使这些人不一定会给自己贴上运动参与者的标签，同时也让教育部和学校成为明确被问责的对象。对参与其中的学生而言，这也是一次重要的练习：此前，学生团体的活动基本都是响应女权行动派在国家政策层面的倡导，这是他们首次与管理自身的权力直接对峙。在和校方不断交涉的过程中，学生强韧了自己的校园网络。

3 月 8 日，女权之声遭到官方永久封禁，随后网络意见领袖发起针对女权行动派核心成员的政治和性污名化运动。至此，由经验丰富的女

权组织者领导的第一阶段大规模维权活动宣告结束。其建立的校园网络和行动者群体并未随着全国协调机制的瓦解而消失，而是继续独立运作，只是这场运动已不再拥有面向公众的领导核心。

2018 年 4 月至 5 月是第二阶段，几所精英大学的学生行动者成为新一波运动的组织基础。他们行动的核心是敦促追究个别性骚扰案件的责任，并要求进行制度改革。

4 月初，八名北京大学学生向校方提交信息公开申请，要求调查一起教师涉嫌性侵害并致当事学生自杀的案件。在人民大学、南京大学和其他几所知名高校，学生也针对涉及教师的案件提出了类似的问责要求，集体抗议和与校方的直接对抗几乎同时爆发。

在这一波浪潮中，学生们通过个人社交媒体账号实时公布了大量信息，包括直播与校方人员的沟通和对峙、发布校方对学生施压的证据，并随着事态发展不断更新诉求。

4 月 22 日，北大学生岳昕发布公开信，称自己因积极参与信息公开申请，被校方和家人强行带离校园，软禁在家。这一声明在高校内外引发了广泛的同情和声援，校园内也出现了公开示威活动。精英大学的特殊环境、紧张的氛围，及学生对运动意涵的拓展——指出校园性侵害问题不仅关乎性别，也关乎学生权益、校园治理和民主制度，种种因素叠加，让事件超出涉事校园的范围，吸引了全国公众的关注，其中也包括一些此前从未接触过性骚扰或女权议题的群体。

校方对学生的问责诉求采取了有限回应和故意拖延相结合的方式。学生的行动在多方面遭到镇压，包括岳昕在内的核心参与者最终在针对性的打击下被迫噤声，抗议的声势逐渐平息。

之后一段时间，米兔运动及公众的关注都陷入沉寂。直到 2018 年 7 月 23 日，一位化名“花花”的年轻女性公益人发布自述，指控知名反歧视倡导者雷闯曾对其实施性侵。在接下来的几周里，针对公益界、学术界和文化产业人士的类似指控井喷式出现。截至 9 月底，已有 36 名知名人士被公开指控。这些披露，以及随之而来的倡导、问责和法律诉讼，使得性骚扰成为这一阶段最引人关注的公共议题之一。

花花的叙述表明，运动的第三阶段由公益圈开启是有迹可循的：在决定公开前，她经历了一段将自己的遭遇定义为性侵害的过程，社会正义和女权的理念为她提供了思想武器，加速并完成了这种意识的转变。促使她发声的理由具有公共性，即“不希望产生更多的受害者”。

在众多案例中，编剧弦子公开指控并起诉中央电视台知名主播朱军性骚扰，由于双方巨大的权力不对等，引发了媒体的广泛关注。作为唯一公开身份接受采访且走入诉讼程序的当事人，弦子在公共空间持续发声，使她的案件成为米兔运动传播和动员的核心节点之一。

挺身而出的当事人的直接目标，是在制度渠道失效的情况下，让施害者承担社会责任。公开作证是她们的最主要手段：通过讲述自身经历，尽力还原事件真相，从而持续推动公众舆论。

她们的语言都非常冷静，同时包含大量的自我审查。尽管不一定是有意意识的行为，但实际上当事人的谨慎态度也是在为运动的目标负责。在叙述中，她们不断地追问自己，也是在不断地解释一个所有人都要问的问题：当时为什么没有反抗？并由此剖析了行业制度、社会文化如何对性侵受害者施压，以迫使她们保持沉默。

这些特征和米兔后期一些个人叙事偏向煽情的风格完全不同。或许是因为煽情在当时的环境下是危险的，它会让当事人显得脆弱、易被伤害，且作为一个典型的“情绪化”的女性，被公众认为不可信任。当事人既要在险恶的网络环境下保护自己，又要通过表达产生动员效果，这注定是一段艰难的对话。只有极少数受过高等教育、熟悉互联网、对女权舆论有观察和认识的人，才能既引起关注，又不会被反噬。

在个案背后，是隐蔽但持续的组织工作：运动参与者为发声的当事人组建非正式的支持网络；一些当事人还为更多的性暴力受害者成立了专门的支持小组；女权组织者和一些公益组织密切关注个案，并以此为契机，推动制度问责和行业改革。

然而，当事人越来越发现，她们指认的人往往不会面临任何实质性的后果，反而受到行业和舆论的保护。当事人自身则面临着被以诽谤名义反诉的风险，承受网络审查和骚扰。致力于将运动成果制度化的女权活动家和公益组织则面临着更为强烈的阻挠：国家机构传唤并威胁活动家；培训课程、会议、项目，甚至组织运作都被迫停止。在疲惫、失望和阻碍的重压下，这股浪潮在 2018 年 9 月之后逐渐消退。

艾：尽管针对个人行动和小规模组织的镇压在加剧，但米兔运动在后期的大众影响力仍在持续扩散。

2019 年，Jingyao 诉刘强东性侵案引发舆论关注。在公众对当事人 Jingyao 的言行进行大范围审视和指责之时，米兔运动参与者在网络发起“我也不是完美受害者”话题征集活动，鼓励女性公开讲述遭受性侵和性骚扰的真实经历，以此指认滋长性暴力事件、迫使受害者噤声的社会“强奸文化”的存在。作为一次成功的公共倡导，这一行动让“我也不是完美受害者”成为米兔中的一句有力口号，推动了关于性暴力案件的报道伦理、性别权力结构等话题的讨论。

怎么看米兔在社会文化和公众教育方面的成果和影响？为了让运动在去中心化的情况下自我维持与传播，女权行动者发展出了哪些适应性的策略，这些策略的效力持续到何时？

吕：米兔极大地重塑了参与者自身对性骚扰、性侵与性别权力关系的认知，也前所未有地将反对性暴力确立为一个具有合法性的公共争议议题。今天的中国年轻女性普遍建立了“性暴力不可容忍”的基本常识，这与十年前的社会意识相比已是翻天覆地的变化。

身处每个不同时期，人们往往感到运动即将迎来新的开端。但从 2018 年起，米兔能改变或者不能改变什么，似乎就已经注定了。因为自从学生抗议结束，个人叙事兴起，建立高校反性骚扰机制这一制度性目标就逐渐被遗忘。

当然可以说这是因为米兔扩散到了校园之外。但在高校体系内，至少还存在针对涉事教师进行纪律处分的惯例与空间；而一旦脱离了校园体制，便失去了具体的问责支点，无论是个案的追责还是制度性的变革，都变得越发不可能。例如，尽管朱军案所动员的舆论关注远超任何一起高校事件，但其所触碰的权力结构也异常坚固。更有许多个案无法在司法层面立案或进入民事诉讼程序，只能被降格定性为内部行政问题或个人道德缺陷，难以推动实质性结果。

而运动之所以全面转向以个人叙事为主，本质上是因为其他所有组织化的行动路径均被国家强行切断。

2018年下半年，女权之声曾计划在陕西开展反性骚扰志愿者线下培训，但所有参与者都遭到了警方的干预与骚扰。那些在运动初期积极参与志愿活动的人，甚至没有一次安全聚集的机会，去系统探讨如何提供受害者支持、推进倡导工作，因为一切线下组织都被严密防范。

同一时期，女权之声也曾尝试在新浪微博建立名为“你我揭露性骚扰”的账号，作为公共传播枢纽，广泛征集性暴力相关的个人叙事和已曝光的个案。而微博当时的审核机制是：正文可以出现“性骚扰”字眼，但账号名称禁止包含这三个字。这种审查策略的结果，就是人们最终只能分散地发声。

因此，所谓的“去中心化”，实际上是政府策略性地避免运动深化

的结果。运动外在的舆论声势，并不等于其内在的抗争力量。公众的注意力在看不见的审查中被损耗，于是运动转变为一种依赖偶发“爆点”的舆情循环，难以对现有体制构成可持续的压力。个人叙事本身则很容易被制约——个案迟迟得不到公正解决，当事人反倒要承受反诉和网络围剿等代价，后来者自然会受到震慑。

在集体行动转向以话语扩散为主的阶段，还是存在核心行动者的引导。问题在于，运动中的动员和组织具有不同的意义：发起一个联署、设置一个热门话题标签，就会有许多人真诚地响应；而组织工作意味着要把分散的个体联系起来，并持续不断地提供教育和实践的机会。

与“万人致信母校”时期的组织架构相比，后期的动员形式看似声势浩大，实际上却很脆弱。除了单次活动之外，没有其他方式把人们连接在一起，有的只是组织者与海量的无名参与者。组织者无法确知参与者的具体身份和参与预期，只能不断地投放新的活动来维系人群的在场。

2020年12月2日，弦子诉朱军案一审开庭，上百名支持者聚集在法院前，手持标语，表达对公正审判的诉求。这是一次历史性的集体行动，但在它发生之前，就连组织者也不知道实际上会有多少人前往。而后两次庭审，因为警察在现场施行的封锁和搜查，声援活动的参与者和组织者之间甚至难以彼此找到。这导致了参与人数的显著下降，而且难以形成有效的呼吁。

又如“我也不是完美受害者”的线上征集，尽管参与人数众多，但在境内缺乏安全通讯工具的环境下，组织者未能与参与者们建立长期联系。在类似的线上集体行动中，一旦遭遇封号，参与者便可能彻底与社群失去联结。

行动者依然付出了密集的隐形劳动。女权主义者通过建立数以万计的去中心化群组，或在社交媒体上发起讨论，以维持运动的传播基础。这些话语实践兼具两种面向：一是行动取向，如呼吁向全国人大寄信反对《民法典》设立离婚冷静期；二是意识提升取向，如疫情初期发起的“看见女性劳动者”倡导。然而，随着2020年底网信办将“挑起性别对立”列为重点整治对象，首当其冲受到压制的就是带有批评和动员色彩的内容，此外还有任何可能形成规模的女权言论。

线下组织层面，高校内的女权小组将学生更紧密地聚集在一起，让运动有虽然微小但可以存续的结构。不过2021年前后，这些高校小组基本都被禁止继续存在。

2021年的“阿里酒桌性侵案”表明，在高校之外，米兔对文化习俗与行业行为准则的重构是有限的：事件发酵初期，企业迫于舆论压力作出了免职涉事主管并承诺整顿文化的表态。然而随着可能的官方介入与舆论风向转变，事件最终以当事女性被以“散播谣言”为由辞退、在网络上遭遇大规模道德猎巫而告终。也很难说这一案件对中国的酒桌文化产生了多少改变。

2022年8月，弦子二审败诉，可以看作米兔运动终结的实质性标志。这意味着运动的核心诉求，即制度性变革的无望。这里所指的不仅是法律层面，如规范司法和执法、将性骚扰写入《民法典》，而且是在更广泛的层面上改变各个权力部门处理性暴力的方式。然而在国家安全的框架下，警察的主要功能不是维持治安，而是维护政权稳定，因此这是一个警察严重滥权的时代；司法审判同样深度嵌入维稳逻辑，不会破例为性暴力案件的当事人开辟空间。

艾：在后米兔时期，2022年前后，以持性别本质论、言辞激烈为特点的“激女”话语开始在中国互联网产生影响力，一度成为网络女权讨论的主流，但后来又逐渐退场，回归小范围的圈子和社群。请问你怎么看激女浪潮在中国网络上的兴起和消退，及其与国家对女权议题的审查、镇压之间的关系？

吕：我对激女发表的内容没有太多研究，部分原因是她们和互联网文化结合之后越来越明显的亚文化色彩，比如发明很多只有内部使用的词汇，维持一种小圈子传播。

激女是反父权的，而且展现出的受害性非常强，有很多凌厉的表达。不过我的理解是，她们最关心的不是社会改革。她们所强调的问责和改变都指向个人：问责通常是针对在家庭中享有优势地位的男性、个别侵害员工的老板，也可能指向未能达到其标准的女性；针对自身的改变就是不婚不育、积累储蓄，降低个人生存的风险。

激女的联盟也有一种帮助女性的精神，只不过这种“女性帮助女性”限于个人层面的资源分享，而实际上也不会有太多资源。这个群体的画像可能很复杂，但很难想象激女的大部分追随者会是在男权社会如鱼得水的人，许多人可能在家庭和社会处于双重边缘，因此才会在网络上表现出极大的愤愤不平。

行动导向的女权主义衰退之后，激女凭借其情绪动员力在网络兴起。但由于反建制的价值观，她们也被封杀严重。

社会运动不可持续的其中一个重要原因是参与者无法从中获益。早期的女权行动派，少数人还能获得活动家的光环；而之后运动能给予参与者的就不再是回报，只有耗竭。因此许多个人项目都很短暂。一些自发的女权账号可能只有几十、几百个粉丝，还要经历反复的删除和重建，当人们的信心和能量被消耗完了，就会淡出、停更。我认为激女也是这样，在重复的话语和反复被封杀中找不到希望和解决方案。

今天在中国互联网占据主流的是市场女权主义，核心思想是女性应当利用现有条件最大化自身利益。性别平等的标准被简化为职场晋升与财富积累：一位女性成为高管，或取得商业成功，即被视为女权的胜利。

市场女权主义将性别议题运作成了一门生意：把女权包装为一项可营利的事业，或是把女权主义作为一种市场性的解决方案，让女性

通过消费获得情感慰藉。例如一些“大女主”相关的文化产品，就是创作者与消费者在一种去政治化的安全语境中共同完成的对女权的表演式消费。文化工业收割了女权运动的成果，让社群为之贡献口碑、销量和票房，却并未对运动有所贡献，甚至可能连“女权主义”这个词都不会出现在内容和宣发中。

当然更糟糕的是广泛面向女性的身心灵课程。它们的内核概念是“显化”：你想要成为富人，就要先通过像富人一样生活和消费，提升内心的“能量等级”。尽管荒谬，但身心灵的这一套主张其实也微妙地迎合了女权话语在大众媒体上的普及和简化，即女性要自信、有“配得感”，将困境归因于个人能量不足，从而进一步强化和利用受众的焦虑。

激女的情绪、立场不适合被消费，因此在当下的环境里缺乏生存空间；至于响应爱国话语、被一些女权主义者视为生存策略的“粉红女权”，其实也没有太多与体制合作共赢的机会，因为国家已经有足够的妇女模范作为其代言人，例如妇联，女性科学家、宇航员、运动员，不需要民间自发的粉红女权主义者成为领袖。

而市场女权主义的兴起，解决了女权主义和国家之间的矛盾。这并不是指女权的思想都消失了。如果我们把女权运动理解为，女性的反抗会因性别不平等的现实而天然地存在，就会认定国家和女权运动之间的冲突是长期且不可调和的——这是我曾经的理想，现在我意识到这不是事实。女性确实一直存在不满，因为国家不会实质性地解决她们

的平权问题；但是，不满意和不平等可以在现实中共存，即人们都对国家发牢骚，而现实并不因此改变。

国家不需要完全消灭异见，因为异见本身能够带来很多信息。就像有人问“为什么中国不把互联网关了”——关了互联网大家都不讲话了，国家不知道人们在想什么。社会失去活力，也会对经济造成影响，这不是政府所期待的。因此还是要保留一定的言论活动，只需要把不满控制在它可以管理的范围。

我的看法是，当国内的女权主义下沉到小红书上的时候，它就不再与运动有强关联性。或者说，我们现在有不同的女权运动：行动主义的运动已经陷入停滞，而另一些女权主义者的运动仍在继续。

艾：从“女权五姐妹”事件，到近几年国家对女权主义的封杀，使一些女权活动家陆续远走海外。近年来在北美、欧洲，还有一些亚洲国家，也形成了活跃的女权社群，以包括脱口秀在内各种的具有影响力的形式进行政治表达和连接受众。一旦在中国有和女性权益相关的重大事件发生，这些社群都会积极地作出回应。

海外中国女权运动的组织和行动现状如何？在本土公共空间萎缩的背景下，能否通过离岸社群保存行动主义知识与社群联结？

吕：海外运动存在自身的困境。首先是国家对跨国运动参与和知识传播的阻断。米兔运动时期，很多人从海外为国内做了大量的工作，对当时的线上传播和组织来说，参与者身处何地没有太大区别。但现在

国家会通过显示社交媒体的 IP 地址、限制海外登录，以及加强对“西方势力”的妖魔化塑造，来阻断墙内外的联系。

信息无法在海内外传递，倡导就失去了着力点，它的作用可能就仅限于显示团结。团结很重要，但当前在另一个国家——例如美国——去谈议中国政府，无法构成后者需要承担的压力性后果。比如黄雪琴被捕的时候，海外不同地区的女权主义者都通过游行抗议表达了对她的声援，然而这些大部分都不被国内的公众或女权社群所知。

即使不开展和国内议题直接相关的动员，只做知识输入类的活动，也会被高度敏感化。假设请一位伊朗的活动家在线分享运动经验，这也是不可行的。

其次是镇压，除了来自在地国家的，还有来自中国政府的跨国镇压。海外运动参与者以留学生为主，他们身处两国之间，处于一种看似便利、实则脆弱的状态。尽管中国政府无法对他们施加日常管控，但镇压之下，他们可失去的东西太多：出国时自身和家庭都负担了沉没成本；并且作为中国人在国内的权利，以及在所在国家作为外国人居留的权利，两者都非常脆弱。

跨国镇压主要体现在，政府通过联系运动参与者在国内的亲属间接施压，或在参与者自身回国期间实行骚扰。这方面的信息远没有得到完全披露，所以人们处于寒蝉效应之中。

于是海外中国女权运动被迫在和本土割裂的情况下寻找自己的议程、

方法，以及动力，但又难以脱离本土议题独立存在。

另一个现实困境是，中国政府越来越强力，与整个国际社会处在威权上升的环境是相匹配的。人们曾经相信，在威权的反面，就有民主的光明存在，即使它不完美。现在发现其实没有这个所谓的反面。人权、民主、平等不再被视为国际通则，人们对这一切都不再抱有信念，无法想象一个不一样的世界，因此也正在丧失对社会运动的信心。

艾：关于跨国运动被阻断，有一个近期的案例。针对这两年发生的以“传播淫秽色情”的罪名大规模抓捕海棠文学城作者的事件，在不同国家的不同地区都有中国女权主义者组织集会和筹款。但从现实来看，以声援政治异见者的思路，期望国际舆论对被捕者的处境有所影响，或是能推动司法改变，可能从结果上都会让声援活动的参与者感到失望；而在国内，对于海外抗议会让事件政治化、敏感化的担忧，反而使得以个案支持为主要目标的事件关注者主动与海外声援者及其主张的“言论自由”等权利诉求作切割。

另一个困难是，中国网络上目前并没有太多具有公共性且能够引发集体讨论的性别事件，公众注意力更多停留在性别矛盾，舆论越发分散和极化。海外关注者即使想借话题发起行动，可能也缺乏着力点。

说到运动的困境，除了在纽约、伦敦这样的文化中心，行动主义的动能相对充沛，海外其他地区的运动似乎也面临与国内相似的社群活跃、但组织者匮乏的问题。你有对于这方面的观察吗？

吕：对于类似海棠这样的事件，海外留学生的示威和抗议可能没有太多策略性，更多是一种情绪表达。但参与者通过这些行动维持自身的抗争意识，我认为这是有必要的。

运动的组织者数量少是一种常态，这是经典的集体行动问题。社会运动争取到的权益是公共的，人人都能享有，而抗争需要投入时间和经济成本，还要承担高风险，因此人们出于个人利益会选择搭便车。如果组织者能够从运动中获得奖赏，哪怕只是作为政治反对者的光环，或许都会有更多人愿意投身其中；但现状是组织者几乎没有任何回报，甚至看不到运动的目标和进展，因此这个问题可能的确是无解的。

与中国的现状类似，对海外社群来说，女权主义可以仅作为一种生活和交友方式存在。尤其是对于留学生，拥有社群陪伴就足以抵御在海外生活的孤独感，而无须参与社会抗议。不过，对于少部分以运动组织者身份离开中国的人而言，不管他们在海外能否继续组织工作，都没有改变自身的定位和对运动的关注。

我目前的想法是，如果真的希望海外的行动对中国有所改变，只能尝试回到 2018 年米兔运动初期的模式，以更紧密的跨国合作来应对镇压。困难之处在于，海外运动参与者需要对国内的环境有足够的观察和融入，才能发现开展行动的机会。这意味着抛弃海外的身份，潜入境内网络，承担繁琐的知识翻译、维权支持与跨区域联络等工作。能否投入时间在这些密集的劳动上，只能取决于运动参与者个人的决定。

我们还得改变对行动的理解。如果行动只限于对热点的回应，那是不

是没有热点就没有行动？或者换个角度问，在运动沉寂的时候，是不是只有那些深入运动的人才有事可做，更外围、广泛的参与者就没有机会？

答案可能是：没有那么多容易的事情可做。“容易”是指有人负责拟定议程和框架，其他人只需要跟随行动。而现在人们必须为自己设计议程。

国内女权社群在这方面的实践比海外开始得更早，且形式更多样。典型的有社会上出现的各种面向社群的支持性小组、个人或小团队的创意和艺术项目，以及一些带有女权色彩，但不明确打女权旗号和贴政治标签的聚会活动，三者构成了相对不可见的运动生态。

这些形态本身可能都很短命，且因为不在公开互联网，运动外围的人有时会不得其门而入。但正是它们在不同的层面上一直维持运动到今天。女权主义者需要它们提供的表达、参与和劳动机会，在缺少抗争行动的时期维持自身抗争性的身份，并以此确认自己所信仰的事业还能有所进展。

艾：我有一个长期的困惑：包括我自己在内，国内的这些小组、活动的组织者，在不能直接发起女权话题，也难以公开讨论时政的情况下，在多大程度上还认同自身的行动和社会运动有所关联？

我自己的观察和体会是，一些组织者尽管出于安全考虑，在活动宣传和内容方面有意识规避敏感议题，但不会对参与者做言论审查，相反

可能在讨论中鼓励参与者对社会事件有所回应；同样地，即使形式相似，这类活动的定位也会区别于主流商业活动，以搭建公共空间而非提供娱乐休闲为主要目标。这些是组织者没有真正远离运动的表现。

吕：我觉得这种自组织形式与市场女权主义不同，后者是对行动主义的替代，而前者只是让抗争的概念和意识变得更隐蔽——对于这一点肯定不能苛责，尤其是活动本身大部分还要依靠女权主义者的义务劳动。

至于这些活动、小组或空间的定位，以及在其中流动的思想究竟有多少抗争性，或许存在一个光谱。但抗争性越强，能见度往往会随之下降。所以重点不是组织形式是否激进，而是其中所创造、维持的内容在多大程度上能够与女权运动及其政治潜力产生联系。

2022 年的白纸运动，就证明了集体行动如何孕育自青年人的社群。一些抗议活动的核心发起者最初就是因为相似的价值观和生活方式而聚集，并依靠聚餐一类的社交活动，以及读书会、观影会等具有公共性的文化活动维持联系。在社会矛盾爆发之时，有人迅速作出回应，动员线下行动、在审查下以低调而具有创意的形式传递讯息，乃至在镇压开始后进行救援，这些都是基于已经在社群中积累的信任关系而展开。

艾：延续这个话题，运动被阻断后，未来我们还能够做什么？

有一种观点是，专制国家的人们即使无法直接参与政治行动，仍能够以微观的生活方式作为一种有效的抵抗。你觉得在当下中国的环境里，

怎么看待这种“微观抵抗”的价值？

吕：研究伊朗的学者阿塞夫·巴亚特提出了“非运动”的概念，意思是人们在日常生活中从事不符合社会规范的活动，以此构成对极权的挑战。

在伊朗，不戴头巾可能会受到监禁，或者在 70 年代的中国，听一下摇滚乐可能就要进监狱，这些都是追求自由要付出的代价。但是在当下的中国社会，人们还是有许多不受惩罚的私人活动空间。国家保持管控一切的权力，同时默许人们的一些行为存在于灰色地带，因此大部分人其实未必感受到被压制。

所以，对于所谓的微观抵抗，我的心态很矛盾。我认为它可以是抵抗，但不是运动；但在谈论运动本身已是奢侈的时候，我们该怎么看待这些剩下的行动？

我觉得首先要设定讨论的前提。对个人来说，要意识到抵抗是有代价的。不管是在私人场合对性骚扰说不，还是选择一种不婚不育的生活方式，都可能给自己的生活带来社会性风险。

人们当然可以去做更容易的事情，例如以支持女性创作者的名义去买一张大女主电影票。但一个人不可能只在被允许的空间内活动，同时又宣称自己在反抗。认为消费可以替代对行动的参与，这种倾向是危险的。

现在中国流行文化中有一种名为“牛马”的自我认同，源于年轻人对工作强度、效率文化、经济形势的不满。但它实际上解构了很多东西，不仅是集体行动的意识，甚至包括对公共交流的需求。它鼓励这样一种生活方式：因为工作很辛苦，所以下班后就应该躺平、购物，把接受一切快餐式娱乐理解为自我照顾，而不再有关心社会问题的责任。人们对生活的想象不再具有公共性，变得犬儒，失去行动力。

既然不被允许形成一个可见的集体，那么个人就成为运动主要的承担者。这意味着女权主义者必须回到自身，去思考运动如何持续。

我们不知道将来会怎样，一切皆有可能。但如果还对抵抗抱有希望，就不能自欺欺人，必须想像一个不一样的世界。在行动层面，反抗的防线即便退缩到最后，也应当守住一个底线：即便无法改变这个世界，自身也不能被它所改变；在意识层面，则要永远拒绝将眼前的一切不公视为“正常”。这或许是个体最后的责任——无论是在思想还是行动上，尽量不向这个体制缴纳“服从税”。这样至少运动的碎片能在我们的体内得以保存。

从这个角度来说，我认为应该重新确认“女权主义”一词本身的重要性。三十年前，当我投身女权运动的时候，这个词汇在中国的环境里是惊世骇俗、令人畏惧的。但当时我们还拥有机会，通过积极地认领女权主义者的身份并为此正名，维持自身的激进性、发展行动主义的社群。而今天，“女权主义”在中国互联网已经不被允许传播，作为替代，人们开始使用弱化了权利意识的“女性主义”，甚至连后者也正在被消除。这对我们所有人来说都是巨大的倒退。

因此我们要在被禁止的情况下，尽可能地保留并使用“女权主义”这个词，这可能才是微观抵抗的意义所在。当然，每个人能够承担的代价与冒险程度，始终由自身决定。

此外，我们还是要探索和延续不同的组织形式，彼此守望相助，将个人的实践连接起来，维持运动的韧性。



透视中国

裴毅然

作者裴毅然为前上海财经大学人文学院教授、历史学者，独立中文笔会会长

中国民主季刊

第4卷 第4期

2026年10月

CC-BY-NC-ND

自由、免费转载分发，但须注明
作者与出处，并不得修改及商用

中共“大外宣” 历史考察（下）

摘要：“宣传”乃中共立党之本、开路先锋。“大外宣”，即“对外宣传大布局”，引起广泛关注的时间并不太长，但其历史其实相当久远。1949年后，大外宣一直是中共政权重要一翼，涓滴渗漫海外，文革前有效影响西方舆情，文革后亦悄然改变一些海外华媒立场。近二十年，中共不断为“大外宣”拨款输血，全力向海外放送红色意识形态，试图与欧美争夺意识形态制高点。

三、文革后的外宣

1976年10月“四人帮”被抓，毛式“继续革命”终止，但意识形态阀门一直拧紧关死。1980年秋，陈云阻拦制定《出版法》、《新闻法》：

无论如何不能让所有的秘密的或半公开的组织 and 刊物登记，要让他们处于非法地位，禁止它们活动。¹ 国民党统治时期制定了一个新闻法，我们共产党人仔细研究它的词句，抓它的辫子、钻它的空子。现在我们当权了，我看还是不要新闻法好，免得人家钻我们的空子。没有法，我们主动，想怎样控制就怎样控制。²

中共根本不追求民主自由，也不理解民主自由，只一心保住红色江山。不过，文革后中共形象在欧美彻底坍塌——

八十年代初期，美国几乎没有什么评论家愿意对中国做出较为有利的评论。甚至那些自认是同情社会主义的人士，也改采了批评的观点。³

走出国门的中国留学生很快发现——

无比优越的社会主义制度为什么使中国人过着牛马不如的生活，而被咒骂为腐朽垂死的资本主义世界却充满了生机和活力？人们亲眼看到资本主义社会中的工人阶级过着现代化的生活，享受着一向被认为是社会主义制度优越性表现的医疗保健、失业保险和其他社会福利。⁴

“三信危机”爆发了——对马列主义失去信仰、对社会主义失去信心、对中共失去信任。中南海意识到必须坚决维护关系政权合法性的红色意识形态，“大外宣”得以继续运转。1983年起，中宣部主办的英文版《中国日报》（China Daily）渗透美国决策枢纽：免费分发国会两院每位议员。1980年代末，有四十余种外宣刊物。⁵至2025年3月11日，美国众议院行政委员会才禁止派发《中国日报》到众议院议员办公室，该报已在国会投放42年。⁶

“六四”枪响，共产党形象跌至冰点，“大外宣”不得不降低身段，精心包装过于赤裸的反理性观点，粘贴西方语汇再出街。如“人民有信仰，国家有力量”，隐去信仰前的“红色”。稍后，悄行“七三”之策——七分赞中、三分批评。

1990年代中共财力渐增，收购或兴办非官方面貌海外媒体。一些海外华媒则“制度套利”——利用西方的自由赚取中共的金银。1990年代中期，中共首先渗透香港《星岛日报》、《明报》，接着香港“亚洲电视”（人称“央视十台”）、香港电视广播（人称CCTVB）。外宣刊物不计成本由驻外使领馆赠送当地政要、精英。⁷

2001年11月21日，美国詹姆斯敦基金会会刊《中国简讯》发表《中

国政府如何控制美国的华语媒体》，揭露中共通过入股收购、商业利益、安插人员、举办论坛，收买控制四家美国中文报纸（总发行量 70 万份）——《世界日报》、《星岛日报》、《明报》、《侨报》。四家华报全文引用新华社消息（仅改标题），立场、用语与北京如出一辙。《星岛日报》老板还是全国政协委员。⁸

2001 年 9 月，中国国务院侨办主办首届“世界中文传媒论坛”。此后两年一届，2011 年第六届近五十个国家 400 余中文传媒总编与会。2006 年，中共再举办“海外中文媒体高级研修班”（4～5 天），邀请各国有影响的华媒负责人参加。一位出席者说：

有吃有喝有玩、有鼓掌有照相有发言、有总结有资料有礼物、有欢迎有道别，一种典型的中国庙会式“大会”。⁹

还有硬的一手。1998 年中共架设网络防火墙（Great Firewall），屏蔽截断境外网站与主要媒体，如谷歌、雅虎、推特、油管、维基百科、ChatGPT、BBC、VOA、《纽约时报》、《华盛顿邮报》……网络防火墙阻拦国人真切及时了解欧美，以防产生个人权利意识。2013 年，中共出台“七不讲”——

普世价值不要讲、新闻自由不要讲、公民社会不要讲、公民权利不要讲、中国共产党的历史错误不要讲、权贵资产阶级不要讲、司法独立不要讲。¹⁰

图穷匕见的“七不讲”再次显示了中共的意识形态管理经验，抓住

要害——拦截向民众输送以普世价值为核心的西方信息。没有权利意识就感觉不到权利被剥夺，不知普世价值就会跟从中共放送的红色价值观。

中共对海外媒体驻华记者集中管理：指定集中居住、指定办公租房、配备中国助手（安插“眼线”），所有采访均须申请，批准才可进行。重大新闻事件（特大事故、集会骚乱、爆发疫情……），外媒记者均无法获得采访许可。

外媒记者还经常遭到暴力袭击甚至逮捕。2003年1月18日，在烟台，《纽约时报》韩籍摄影记者石宰铉（采访朝鲜难民）被捕，以“组织偷渡罪”（协助脱北者）判刑两年。英国记者瓦特斯三年被捕五次。2003年10月16日美国国家民主基金会研讨会“中国传媒的管制、新闻检查、偏见与操控”，《华尔街日报》专栏女作家克劳迪娅·罗塞特（Claudia Rosett）控诉：多次接到中国驻美大使信函，强烈谴责她文章中对中国的批评。

外媒记者还常遭拒发签证。2006年9月10日，中共发文：所有在华通讯社必须通过新华社审查才能向中国境内发送新闻，否则暂停或吊销营业执照。¹¹

2004年香港三大名嘴（郑经翰、黄毓民、李鹏飞），在政治压力与暴力恐吓下相继闭嘴，“近年来，（香港）公众人物遭滋扰及袭击的案件近50宗。”¹²

2005 年 4 月广州，新加坡《海峡时报》首席特派员程翔（1949～）被捕，以“台谍”判刑五年。除恐吓制造“寒蝉效应”，再辅以商业利益（不投广告），温水煮蛙，迫使海外华媒自我审查，不再报道中共忌讳的敏感话题，不再接受敏感人物的反共稿件。

2010 年中国 GDP 总量超过日本（世界第二），此前在 2009 年中共宣布投入 450 亿人民币改善国家形象，扩张“大外宣”。¹³ 2011 年，新华社租下纽约时报广场（即时代广场）最大广告——巨型液晶荧幕（高 19 米 × 宽 12 米），月租 \$30～40 万，日夜播放精心制作的《中国形象》。¹⁴

香港中联办辖下联合出版集团控制香港五大出版社中的三家——三联书店、商务印书馆、中华书局。中联办控制的香港出版社共 52 家，联合出版集团董事长兼总裁文宏武即中联办秘书长。

中共要全力向海外输送红色价值观、世界观、历史观——三观中历史观最关键，为价值观提供逻辑起点。

2003 年，中共推行“外宣工作本土化”，提出“三贴近”——贴近海外受众思维习惯、贴近海外受众对中国信息的需求、贴近中国发展的实际。同时，聘用所在国编辑记者，甚至约请当地官员担任外宣刊物顾问，将外宣包装成正规新闻报道，当地主流媒体有时也会引用。¹⁵

2013 年后，“大外宣”高调围绕“一带一路”¹⁶。中国外文出版发行事业局每年以 20 种语言出版 2000 多种图书、21 种纸刊、25 种网刊，

发行推送 180 余国家及地区。2016 年，新华社海外分社 180 余处，派出记者六千余，海外用户 1450 多家，每天七种语言向全球放送上千快讯，力避“六四”、“五独”（台藏疆蒙港）、法轮功等话题；报道原则：宣传高于真相。海外 5000 万华人难辨各种媒体背景，感觉“中国政府获得的国际评价越来越高”。¹⁷

2016 年 6 月 6 日，《中国日报·香港版》及亚太分社副总编：

最大限度地争取国际受众对中国观点和中国道理的认同和认可。¹⁸

2021 年 5 月 31 日，中共政治局第三十次集体学习，主题“加强我国国际传播能力建设”，复旦中国研究院院长张维为主讲。¹⁹“大外宣”近年新动向包括：

- 1、輿情产业外包，不再局限官媒，大量外包给海内外公关营销公司，利用虚假账号、社群机器人发布亲共信息。
- 2、招募海外网红与社群小粉红、收购本土社群专页、结合 TikTok（抖音海外版）、短视频等红色 App（应用软件），政治倾向隐藏于生活化内容，直接影响全球年轻族群的輿情走向。

四、尴尬的“大外宣”

1978 年 5 月，广东省委主管经济副书记王全国（1919～2015），随副总理谷牧访问西欧五国（德法瑞丹比）：

那一个多月的考察，让我们大开眼界，思想豁然开朗，所见所闻震撼每一个人的心，可以说我们很受刺激！闭关自守，总以为自己是世界强国，动不动就支援第三世界，总认为资本主义腐朽没落，可走出国门一看，完全不是那么回事，你中国属于世界落后的那三分之二！²⁰

1986年11月全国外宣工作会议，中央外宣小组组长朱穆之（1916～2015）：

对外宣传从过去宣传“世界革命”转为介绍新中国，增进各国人民对我国了解和友谊；从宣传战争不可避免，转为宣传和平可以维护；从宣传阶级斗争为纲，转为宣传社会主义现代化建设；从片面强调一切靠自己、闭关自守，转为宣传改革开放、吸收当代世界一切对我有用的物质文明与精神文明成果……²¹

整体上，“大外宣”效应一荏不如一荏。六四枪响，马克思主义在全球彻底失去合法性。苏联前驻澳大使（苏共老党员）：“邓小平的屠杀最后葬送了国际共运。”²² 斯诺遗孀洛伊斯·惠勒·斯诺（Lois Wheeler Snow, 1920～2018）：“这让我清醒了过来”，发誓不再入华。²³ 1989年11月柏林墙倒塌，1991年12月25日克里姆林宫降下红旗。²⁴ 建党七十年的英共、意共更名，西方马列左派大批醒悟。

1995年，中国财产基尼系数（衡量贫富差距）0.45，2012年0.73；1%顶端家族拥有1/3全国财产，最底部25%家庭仅占全国财产1%。日本基尼系数0.25，实现共同富裕。²⁵ 2015年中国亿万富豪568名，超过美国（535名），占全球亿万富豪（2188人）1/4强。²⁶

2022 年 10 月 17 日中纪委副书记肖培发布：2012 年以来全国纪检监察机关立案 464.8 万余件，涉案中管官员 553 名 [18 ~ 19 届中（监委 79 名]，厅局级 2.5 万余名，县处级 18.2 万余名。²⁷

这类事实直接戳穿了大外宣的面具。2025 年 8 月 29 日重庆大学城出现巨型投影“反革命标语”（50 分钟）——没有共产党，才有新中国！“案犯”戚洪（1982 ~）携妻女逃英前租下饭店房间，遥控打开投影。他接受德国之声采访：“中国社会就像粪坑，而你在里面还要被迫说‘粪好香’。”²⁸

中共官媒尽失公信力，大陆民众咸知只能拼接新闻线索“想象”真实。自媒体时代，流量经济放大谣言，各种信息满天飞——“大家猜”。中共当局被迫退居守势，只能煽动民族主义捍卫“中国特色社会主义”，倚借爱国情绪抵御普世价值。成天下达“宣传纪律”——2002 年下达的禁用词名录，禁止全国媒体使用“言论自由”、“多党制民主”……译著中“言论自由”得改为“更多的发表意见的机会”。

中共组建网络“水军”（五毛），名曰“舆论引导员”、“网评员”（每省至少 300 名），在各种网络空间上窜下跳、混淆视听。2002 年 10 月巴黎“无国界记者”发布全球新闻自由排行榜（139 个国家和地区），中国、北韩垫底——第 138 名、第 139 名。²⁹

2004 年 9 月中共十六届四中全会后，中宣部随即召开全国媒体会议贯彻精神，发布“二十九不议”——不议文革、不议毛功过、不议民族问题、不议国企改革、不议台湾问题、不议印巴以、不议农民上访、

不议土地拆迁、不议社会民生、不议官民冲突、不议爆炸暴动、不议罢工示威……，所有负闻新闻均由新华社统一发稿，违反者严惩。中宣部副部长吉炳轩会上宣读胡锦涛批示：“管理意识形态，我们要学习古巴和朝鲜。朝鲜经济虽然遇到暂时困难，但政治上是一贯正确的。”³⁰

都跌至“朝鲜尺寸”了，但 2003 年 11 月 2 日海南博鳌亚洲论坛，新闻出版总署副署长柳斌杰：“中国目前是世界上言论出版自由最充分的国家之一，外界关于中国新闻、言论、出版方面的某些议论是毫无根据的。”³¹ 2006 年 10 月 25 日华盛顿，战略和国际研究中心，中国国务院新闻办主任蔡武：“中国的互联网是世界上最自由的言论天地。”³²

2021 年 10 月 13 日中央人大工作会议，习近平：

我国全过程人民民主实现了过程民主和成果民主、程序民主和实质民主、直接民主和间接民主、人民民主和国家意志相统一，是全链条、全方位、全覆盖的民主，是最广泛、最真实、最管用的社会主义民主。³³

12 月 2 日，外交部副部长乐玉成：

历史和现实充分证明，中国的民主模式符合本国国情，受到人民拥护，是真正的民主、管用的民主、成功的民主。中国是当之无愧的民主国家。³⁴

不过，传播不等于接受，发布不等于到达，覆盖面不等于公信力。中共香港《文汇报》、《大公报》每天卖不出 500 份，订户都是中资机构，自己办报给自己看。³⁵ 海外受众不会也不愿按中共意愿解读“大外宣”，真实仍是媒体生命线，国家形象终究取决内政文明。2009 年 2 月 11 日，国家副主席习近平访问墨西哥，一不小心掉出“名言”：

有些吃饱了没事干的外国人，对我们的事情指手画脚。中国一不输出革命，二不输出饥饿和贫困，三不去折腾你们，还有什么好说的！³⁶

似乎本就该“输出革命”、“输出饥贫”、“折腾你们”，现在不搞了，得感恩呵！

“大外宣”唯一正面效应：中宣部不得不考虑欧美国情，不得不寻找与普世价值的对接点。1991 年 11 月北京首次发布《中国的人权状况》白皮书，总算认领“人权”一词，此前一直只讲阶级性革命性，不讲人性、人权。1979 年 3 月 22 日《北京日报》载文《人权不是无产阶级的口号》。

“大外宣”海外触须亦向国内输送一点“不同声音”，因为改革开放需要倾听海外批评，否则不知何处需要改革、如何改革。为此，中共情报部门及新华社整天关注外国政要与外媒对中国的负面批评，“从收集的这些‘批评意见’中得到了很多有益的东西，对中国改革开放一路前行做出了不可替代的作用。”互联网兴起后，中共从网上批评寻觅什么地方“需要改革”。³⁷

中共失真失信失理，不可能所有人都与中南海“一条心”。央视、《人民日报》人员私下告知外国留华生：“他们的工作就是蒙骗老百姓”。18岁入北大留学的日籍生加藤嘉一（1984～）：“我是来到了不自由的中国后才发现日本的自由。”2020年，长期研究中国的加藤嘉一：

我一直以来每天都会看《新闻联播》，为的是研究和观察中国，但最近几年我实在是看不下去了。

2018年10月4日，美国副总统迈克·彭斯（Mike Pence, 1959～）：

北京正在使用一种全政府的手段，利用政治、经济、军事工具以及宣传，在美国推进其影响和利益。……中国比以往更活跃地使用其力量来影响并干预美国的国内政策和政治。³⁸

2023年3月12日，前国务卿蓬佩奥（Mike Pompeo, 1963～）：

中共撒的最大的谎言是自称它代表中国、代表中华文明，而且我认为中共知道这是一个谎言。如果北京认为它真的代表了中国人民，它就不会花费比国防开支还多的钱，来镇压和监控中国人民。³⁹

结语

1949年，民国政府败走台湾，蒋介石：“我们的宣传不能主动，理论亦缺少战斗性。”⁴⁰ 忽视意识形态、轻视舆论舆情，输在思想领域，

民国重大败因之一。1960 年台北，教育部长蒋梦麟（1886～1964）谓费正清：

假如我们在大陆时就能达到我们现在这样的认识水平，我们本来是能够打败毛泽东的。⁴¹

必须警惕“大外宣”对西方的无声浸渗。2026 年 1 月 30 日纽约《世界日报》载文《美年轻人不认为中国像大恶狼》（版 A5），称调查显示 2/3 受访美国人认为“中国的实力与影响力现在已追上并超过美国；美国年轻人认为中国已超越美国的比率是 65 岁以上者的两倍。”

应对共产党“大外宣”乃中国民运重要一翼，得争夺海外輿情輿论，争夺对中共的“正确认识”，摊晾史实、解析赤谬，拆除意识形态违章建筑——马克思主义。

（本文系 2026 年 8 月 18 日《中国文化大革命数据库》编委会在美国召开的、“文革 60 年：中国和世界的回顾和环顾”研讨会论文。因限于篇幅，本刊发表时做了删节）

注释

- 1 邓力群：《十二个春秋》（1975～1987），博智出版社（香港）2005 年，页 178。
- 2 钟沛璋：“政治文明与新闻立法”，《同舟共进》（广州）2003 年第 2 期，页 14。
- 3 黎安友（Andrew J.Nathan）：“美国对中国的态度”，《知识分子》（纽约）1985 年 4 月号，页 12。
- 4 唐勃：《中共与知识分子》，幼狮文化事业公司（台北）1988 年，页

- 530。
- 5 朱穆之：《风云激荡七十年》上册，五洲传播出版社（北京）2007年，页263～264。
- 6 《中国日报遭禁发，重挫中共大外宣》，2025年3月11日，万维读者网（加拿大）：<https://news.creaders.net/us/2025/03/11/2842960.html>。
- 7 《周明伟在纪念中国外文局成立60周年大会上的讲话》，中国网2009年9月4日。www.beijingreview.com.cn/news/txt/2009-09/04/content_214885.htm。何清涟：《红色渗透：中国媒体全球扩张的真相》，八旗文化/远足文化事业公司（台湾）2019年，页136。
- 8 何清涟：《雾锁中国：中国大陆控制媒体策略大揭密》，黎明文化事业公司（台北）2006年，页468～469。
- 9 何清涟：《红色渗透：中国媒体全球扩张的真相》，页114～117。
- 10 2013年4月22日，中办发〔2013〕9号文件：《关于当前意识形态领域情况的通报》（地师级）。《明镜月刊》（香港）2013年8月号（第43期），中国数字时代：<https://chinadigitaltimes.net/chinese/345627.html>。
- 11 何清涟：《雾锁中国：中国大陆控制媒体策略大揭密》，页281～283、289、287、303～304、290、297～298。
- 12 何清涟：《红色渗透：中国媒体全球扩张的真相》，页146～148。
- 13 姚遥：《新中国对外宣传史》，清华大学出版社（北京）2014年，页398。
- 14 <https://www.time-weekly.com/wap-article/13381>。Anton Troianovski: *China Agency Nears Times Square*, 《华尔街日报》（纽约）2010年6月30日。wsj.com/article/sb10001424052748704334604575339281420753918.html。
- 15 周瑾：《人民中国杂志社副社长王众一：外宣刊物的“三贴近”》，人民网·对外传播，2005年4月4日。
何清涟：《红色渗透：中国媒体全球扩张的真相》，页58。
- 16 习近平“大国外交”核心部分，以投资七十余国扩大中国影响力。以丝绸之路经济带、二十一世纪海上丝绸之路命名，简称“一带一路”。
- 17 何清涟：《红色渗透：中国媒体全球扩张的真相》，页59、153、27、67、72。
- 18 《如何应对西方媒体偏见》，转引《红色渗透》，页18。程翔：《“大外宣”：中共与西方的意识形态之战》，<https://www.upmedia.mg/tw/commentary/china-hong-kong-insights/59004>。

- 19 新华网: https://www.xinhuanet.com/politics/2021-06/01/c_1127517461.htm。
- 20 崔荣慧:《改革开放·先行一步——访原广东省省委书记王全国》,宋晓明、刘蔚主编:《追寻 1978——中国改革开放纪元访谈录》,福建教育出版社(福州)1998年,页 558。
- 21 朱穆之:《风云激荡七十年》上册,五洲传播出版社(北京)2007年,页 201 ~ 202。
- 22 许良英:《致李慎之函》(2000年7月12日),《八十三封书信——许良英、李慎之书信集》(增订版),同心同理书屋(北京)2010年,页 82 ~ 83。
- 23 Amy Qin:《从追随红星到批评中共:被六四“改变”的斯诺夫人》2018年4月12日。纽约时报中文网: <https://cn.nytimes.com/obits/20180412/lois-wheeler-snow-dies/>。
- 24 高树华、程铁军:《内蒙文革风雷》,明镜出版社(纽约)2007年,页 36。
- 25 《中国 1% 家族占有全国 1/3 财产》,《第一财经日报》(上海)2016年1月13日。
- 26 胡润研究院《2016 胡润全球富豪榜》(2016年2月发布)。
- 27 《百年大党的自我革命》,《中国青年报》(北京)2022年10月18日,版 1。
https://zqb.cyol.com/html/2022-10/18/nw.D110000zgqnb_20221018_3-01.html。
- 28 “在重庆投影‘反共’当事人已赴英”,《世界日报》(纽约)2025年9月2日,版 A12。
- 29 何清涟:《雾锁中国:中国大陆控制媒体策略大揭密》,页 73、113。
- 30 林保华:“阴柔冷酷胡锦涛的真面目”,《自由电子报》(台湾)2005年1月10日。<http://www.libertytimes.com.tw/2005/new/jan/10/today-int5.html>。“扭曲真相的新闻报道谈什么新闻交流”,《自由电子报》(台湾)2005年4月12日, <http://www.libertytimes.com.tw/2005/new/apr/12/today-s1.htm>。
- 31 人民网博鳌 2003 年 11 月 2 日电, www.people.com.cn。何清涟:《雾锁中国》,页 113。
- 32 江河:《记者无疆界指中国互联网检查最严》,美国之音报道,2006年11月7日。
- 33 《习近平谈治国理政》第四卷,外文出版社(北京)2022年,页 260 ~ 261。同时,习近平这次讲话也被收录到《习近平文著作选读》第二卷。

- 34 https://www-guancha-cn/politics/2021_12_02_617068-shtml;
<https://www-youtube-com/watch?v=vj4wRWSeYtA>
- 35 何清涟：《红色渗透：中国媒体全球扩张的真相》，页 153。
- 36 《习近平痛批少数外国人对中国事务指手画脚》，央视网：<http://news.cctv.com/china/20090213/10230.shtml>。姚遥：《新中国对外宣传史》，页 436。
- 37 杨恒均：“对中央善意的解读为什么总被打脸？”，冯崇义、杨恒均编：《学人干政与宪政转型》，明镜出版社（纽约）2017 年，页 166。
- 38 何清涟：《红色渗透：中国媒体全球扩张的真相》，页 180。
- 39 【美】蓬佩奥：《中共是比纳粹更大的威胁》，<https://www.ntdtv.com/gb/2023/03/12/a103667410.html>。
- 40 蒋中正：《苏俄在中国》，黎明文化事业公司（台北）1982 年，页 177。
- 41 《费正清对华回忆录》，陆惠勤等译，知识出版社（沪版）1991 年，页 466。

王维洛

作者王维洛，德国工程博士，曾在南京大学以及德国多特蒙德工业大学任教，地理学者、国土规划工程师以及科技作家，长期研究中国环境和水资源问题

中国民主季刊

第4卷 第4期

2026年10月

CC-BY-NC-ND

自由、免费转载分发，但须注明
作者与出处，并不得修改及商用

雅鲁藏布江下游水电工程：目标、政治意图与环境问题分析

摘要：2025 年 7 月 19 日雅鲁藏布江下游水电工程（在本文中简称雅下工程）开工仪式在林芝市米林水电站坝址举行。中共政府称雅下工程是利用水能的绿色工程，年发电量达 3000 亿千瓦时，是三峡工程的三倍。本文作者认为，这个发电目标无法实现。工程的真实目的主要并不在电，而在于藏南问题、中印政治争端。雅下工程对西藏高原的生态环境破坏、特别对生物多样性的危害极大。



图 1：雅鲁藏布江大拐弯。从照片中可以看到，当地的生态环境已经被破坏相当严重，沿着简易的公路到处可见滑坡、岩崩留下的痕迹。（图片来源：网络截屏）

一、从墨脱水电站到雅鲁藏布江下游水电工程

墨脱（Mêdog），藏语的意思是“花朵”，是西藏自治区林芝市的一个县，位于雅鲁藏布江下游，喜马拉雅山南麓，与印度接壤（紧挨麦

克马洪线和印度实际控制线)。

根据《百度百科》，墨脱水电站的定名时间是在二十世纪八十年代前期。墨脱水电站的提出可追溯到 20 世纪 60 年代，中科院派人到米林县派镇考察雅鲁藏布江中游的水电情况。1982 年，时任水电部副部长李鹏及有关领导指示成都院考察队抵达墨脱水电站考查区考察进水口位置及大渡卡等五个比较坝址，并推荐裁弯取直引水方案，定名为墨脱水电站。¹

2002 年 12 月徐大懋、陈传友与梁维燕在《中国工程科学》杂志上（2002 年第 4 卷第 12 期）发表《雅鲁藏布江水能开发》² 的文章，提出比较具体的建设方案与初步的工程经济可行性论证，结论是工程的年发电量是三峡工程的 2.2 倍，静态投资总额与三峡工程基本相同。

根据中国水利部的流域水资源管理划分，长江水利委员会负责西南诸河的水资源管理。2006 年时任长江水利委员会主任蔡其华到西藏考察墨脱水电站，并就西藏水利发展问题与自治区领导座谈。³

2021 年 3 月 13 日新华社发布了《中华人民共和国国民经济和社会发展第十四个五年规划和 2035 年远景目标纲要》⁴（简称十四五规划），在第十一章《建设现代化基础设施体系》的第三节《构建现代能源体系》的专栏六《现代能源体系建设工程》中第一节《建设清洁能源基地》的第一句便是建设雅鲁藏布江下游水电基地。水电基地是由多个水电站项目组成。之前中共政府有 13 个水电基地的布局，加上雅鲁藏布江下游水电基地，一共 14 个。

2024 年 12 月 25 日新华社发布题为《雅鲁藏布江下游水电工程已获核准》⁵ 的报道称，近日，中国政府核准雅下工程。报道中没有关于雅下工程具体的工程技术和经济方面的数据。但是报道指出，雅下工程是坚持生态优先的安全工程。习近平主政以来特别强调安全，在习近平的词典中安全是多义的，但政权安全始终是位居第一的。

新华社的报道中将墨脱水电站改称为雅下工程的含义是深远的，因为雅鲁藏布江下游的终点在印度实际控制区的巴昔卡，墨脱水电站的终点最多在距离中印实际控制线中方一侧的二、三公里处。两点之间的河长有 200 多公里，水位差有几百米之多。

2025 年 7 月 19 日新华社发布题为《李强出席雅鲁藏布江下游水电工程开工仪式 并宣布工程开工》⁶ 的报道，称雅下工程开工仪式在西藏自治区林芝市举行。中共中央政治局常委、国务院总理李强出席开工仪式，并宣布工程正式开工。有关雅下工程的具体的工程技术和经济方面的数据只有以下几十个字：雅下工程位于西藏自治区林芝市；工程主要采取截弯取直、隧洞引水的开发方式，建设 5 座梯级电站，总投资约 1.2 万亿元；工程电力以外送消纳为主，兼顾西藏本地自用需求。另外值得特别关注的是，开工仪式在林芝市米林水电站坝址举行，不是在雅下工程的坝址举行，而米林水电站坝址在雅鲁藏布江中游，不在雅鲁藏布江下游。笔者认为，这个世界最大的水电工程的开工仪式不在工程的坝址举行，这就是一个大骗局的开始。

同样在 2025 年 7 月 19 日这一天新华社发布题为《中国雅江集团有限公司成立大会举行》⁷ 的报道称，中国雅江集团有限公司成立大会近

日在京举行，中共中央政治局委员、国务院副总理张国清出席并为公司成立揭牌。领导和负责雅下工程建设的雅江集团成立也在 2025 年 7 月 19 日发生，只能说明，雅下工程的前期准备工作，包括论证和道路交通等都还是空白。据中国科协科普部原部长杨文志披露，雅江集团注册资本金 2000 亿元，其中三峡集团占 35%，中国电建、中国能建各占 20%，西藏自治区国资平台持股 15%，其余 10% 为其他投资者。⁸可见中央政府并没有为成立雅江集团准备好足够的资金，而是临时匆忙地“拉郎配”。

二、中国水能资源量掠夺式的开发，2024 年已经超过经济可开发量

水能资源理论总蕴藏量、技术可开发水能资源量与经济可开发水能资源量是水能资源研究和工程立项的三个重要概念。

根据赵纯厚等提供的资料，⁹世界水能资源理论总蕴藏量为 403000 亿千瓦时 / 年，其中中国 59200 亿千瓦时 / 年，占 14.7%；美国 44850 亿千瓦时 / 年，占 11.1%。中、美两国的水能资源理论蕴藏量为 100.00 比 75.76，中、美两国在总量上相差不是很大。但是在技术可开发、经济可开发水能资源量上，两国的差别就很大。可见美国认定的技术、经济、社会和生态环境等限制因素，要比中国严格许多。换言之，美国留给子孙后代的资源多，而中国采取的是掠夺式的开发，留给子孙后代的资源很少很少，甚至为零。

如果注重中国和美国自身的三个概念的比例，则有以下表：

	水能资源理论蕴藏量	技术可开发水能资源	经济可开发水能资源	
世界	403000	143700	80820	亿千瓦时/年
占比	100.00%	35.66%	20.05%	
中国	59200	19200	12600	亿千瓦时/年
占比	100.00%	32.43%	21.28%	
美国	44850	5285	3760	亿千瓦时/年
占比	100.00%	11.78%	8.38%	

图 2：世界、中国与美国水能资源理论总蕴藏量、技术可开发水能资源量与经济可开发水能资源量的比较，资料来源：赵纯厚等与笔者计算

根据中国能源局的报道，¹⁰ 截至 2024 年底，全国水电发电量 14239 亿千瓦时，已经超过了经济可开发水能资源 12600 亿千瓦时 / 年，占技术可开发量 19200 亿千瓦时 / 年的 74%。根据《观研报告网》的资料，¹¹ 2025 年 1-11 月全国水电发电装机容量同比增长 3%，规模以上企业水力发电量同比增长 2.7%。中国水电发电量年增长量保持 2%，10 年之后可达 17357 亿千瓦时，再加上雅下工程的年发电量 3000 亿千瓦时，一共 20357 亿千瓦时。到那时技术可开发量 19200 亿千瓦时也会 100% 开发殆尽。吃祖宗的饭，用子孙的资源，这是中共政府生存的法则。

三、雅鲁藏布江和雅鲁藏布江下游的水能资源量

根据西藏自然资源厅的资料，¹² 西藏有河流 356 条，其中流域面积大于 1 万平方公里的河流有 20 余条，大于 2000 平方公里的河流有 100 多条。众多河流水量充沛，落差大，其天然水能理论蕴藏量达 2.006 亿千瓦，占全国的 30%。全区 500 千瓦以上电站可开发水能为 5659.3 万千瓦，年发电量 3300 亿度，占全国的 17.1%，在全国各省市中仅

次于四川、云南居第三位。雅鲁藏布江是西藏水能资源最丰富的一条河，理论蕴藏量为 1.13 亿千瓦，占全区理论水能蕴藏量的 56.22%，其中可开发量为 4837.14 万千瓦，占全区可开发量的 80% 以上。笔者在此指出，西藏自治区可开发水能的年发电量为 3300 亿千瓦时，那么雅下工程年发电量为 3000 亿千瓦时就是一个夸大的数据。

根据《世界江河与大坝》，¹³ 雅鲁藏布江全部在中国境内，干流全长约 2070 公里。从主源杰马央宗曲至萨噶县的里孜为上游，全长 270 公里；里孜以下到米林县的派乡为中游段，全长 1300 公里；从米林县的派乡到巴昔卡是下游段，河长 500 公里。巴昔卡的多年平均径流量为 1654 立方米，次于长江、珠江，居全国第三。本文集中关注雅鲁藏布江下游河段。

四、雅鲁藏布江下游水电工程的年发电量 3000 亿千瓦时的目标无法实现

至今为止，官方公开发布的关于雅下工程的技术经济数据只有：

- 1) 工程地址：西藏自治区林芝市；
- 2) 工程的开发方式：主要采取截弯取直、隧洞引水，建设 5 座梯级电站，总投资约 1.2 万亿；
- 3) 工程的目标：发电，以外送消纳为主，兼顾西藏本地自用需求。

在中共官方媒体和自媒体上出现了许多有关工程的具体数据和建设方案，这里采用《东吴证券》撰写的《雅鲁藏布江大项目正式开工，工

程机械行业迎重大机遇》一文中的数据：

	雅鲁藏布江下游水电站	三峡水电站
项目地点	西藏林芝市	三斗坪坝
装机容量	6000 万～8100 万千瓦	2250 千瓦
年发电量	3000 千瓦时	1000 千瓦时
总投资额	1.2 万亿元	1800 亿元
建设周期	10 年以上	15 年

图 3：媒体上传播的雅下工程的经济技术数据，图片来源：东吴证券

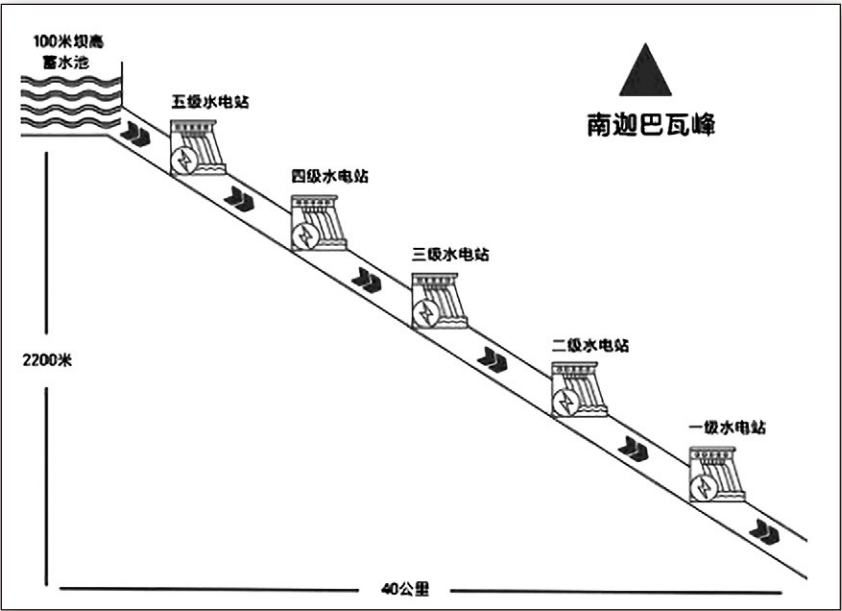


图 4：媒体上传播的雅下工程的建设方案，图片来源：东吴证券

雅下工程的年发电 3000 亿千瓦时，为三峡工程的三倍。笔者必须指出的是：这个目标是不可能实现的，这是由当地的自然地理条件 and 政治地理条件所决定的。

计算水电站装机容量计算（原理）的基本公式如下：

$$N = \eta \times 9.8 \times Q \times H$$

N: 理论发电功率（千瓦）

η : 水轮发电机组综合效率（0.80 至 0.82）

9.8: 水的重力加速度与单位换算系数（1 立方米水每下降 1 米做的功）

Q: 水流量（立方米 / 秒）

H: 水头（落差，米）

水电站装机容量是水流量和水头的函数。

派镇上游的奴下水文站是雅鲁藏布江最早建立的水文站，水文数据自 1956 年开始。该站的常年流量有不同的数据，在 595 亿立方米至 603 亿立方米之间，差别不大。有所差别的原因可能是截取的时间序列有所不同。这里取常年流量约为 600 亿立方米，换算成流量为 1903 立方米 / 秒。

关键的问题在于落差。

在中外媒体关于雅鲁藏布江下游水电工程的描述中，人们可以看到这样的描述：“雅鲁藏布江下游水能资源丰富，从林芝市派镇到中印边界（巴昔卡）的水位差 2700 米。”更为精确的说法是，雅鲁藏布江下游段 496 公里，落差 2725 米。¹⁴

按落差按 2700 米、流量 1903 立方米 / 秒、水轮发电机组综合效率

0.82 计算，则有：

$$0.82 * 9.8 * 1903 * 2700 = 4129 \text{ 万千瓦}$$

理论发电功率为 4129 万千瓦。一年 8760 小时，一年的最大理论发电量为 3617 亿千瓦时。

公式证明，当落差 2700 米时，工程的年平均发电量的最大理论值可以达到 3000 亿千瓦时。

有人说，为了维持雅鲁藏布江的河流生态需要。流经派镇的雅鲁藏布江水流不能 100% 被截流用于发电，部分水流必须下泄到雅鲁藏布江的河道中。假设 85% 的水流被截流用于发电，那么 3617 亿千瓦时的 85% 为 3074 亿千瓦时，刚刚超过 3000 亿千瓦时这个目标。

但是从林芝市派镇到中印边界实际控制线的水位差为 2000 米至 2350 米之间。落差差别可能是因为这些数据都是换算的结果。因为藏南地区、巴昔卡现在印度的实际控制中，为印度的阿鲁纳恰尔邦。再说中印边界战之后毛泽东命令中国人民解放军撤退到麦克马洪线后面，而且还要再后撤 20 公里。所以落差最多只能按 2350 米计算，则有：

$$0.82 * 9.8 * 1903 * 2350 = 3594 \text{ 万千瓦}$$

一年 8760 小时，一年的最大理论发电量为 3148 亿千瓦时，刚刚超过 3000 亿千瓦时。

假设 85% 的水流被截流用于发电，那么 3148 亿千瓦时的 80% 为

2676 亿千瓦时；假设 70% 的水流被截流用于发电，那么年发电量为 2204 亿千瓦时；假设 50% 的水流被截流用于发电，那么 3148 亿千瓦时的 50% 为 1574 亿千瓦时。无论采用何种假设，都不能达到年发电量 3000 亿千瓦时的目标。

2002 年 12 月徐大懋、陈传友与梁维燕《雅鲁藏布江水能开发》文章中提出，工程的年发电量是三峡工程的 2.2 倍。当时公布的三峡工程年发电量为 840 亿千瓦时，三峡工程的 2.2 倍则为 1848 亿千瓦时。不知道出于什么理由，到 2025 年工程的雅下工程的年发电量就上升到三峡工程的 3 倍，成为年发电量 3000 亿千瓦时。



图 5：决定雅下工程年发电量的落差到底应该怎么算？！图片来源：部分内容 为笔者所加，底图来源：地理公社底图

所以说，雅鲁藏布江下游水电工程的年发电量 3000 亿千瓦时的目标是无法实现的，这是由当地的自然地理条件 and 政治地理条件所决定的。

五、雅鲁藏布江下游水电工程开工，收回藏南呼声高起

中共几代领导人，无论是毛泽东、邓小平还是习近平，都信奉“治国如治水”的理念。2025年12月6日习近平夫妇与法国总统马克龙夫妇在四川成都都江堰会面，习近平大谈治水治国的经验。习近平认为，在五千多年中华文明史中，一些地方几度繁华、几度衰落。要想国泰民安、岁稔年丰，必须善于治水。习近平自认，他对治水治国的功绩必须在毛、邓之上。毛泽东有黄河三门峡大坝，邓小平有长江三峡大坝。习近平也急于功在禹上，绩盖毛邓，雅下工程作为世界上最大的水电工程，正是习近平朝思暮想的。2021年7月21日习近平在河南郑州遭受严重洪灾的情况下飞赴工程所在的西藏林芝，亲自决策、亲自指挥、亲自部署。¹⁵

习近平要想功盖毛邓，必须在收复“已经失去”的国土上有所作为。习近平在多个重要场合强调过“老祖宗留下的领土一寸也不能丢”¹⁶。如台湾学者宋国诚所指出的，对习近平而言，解决台湾问题，是在中国进入历史性通缩之后，唯一能够用来重组执政合法性、超越毛邓、在历史上自我封神的终极政治解药¹⁷。笔者以为，三峡工程上马之后，已经成为敌方定点威胁的重点目标。外加中国政治经济中心周围的数千座高坝大库，如北京的密云、官厅水库等，还有高架在中原大地上——一千多公里长的“天河”（南水北调中线工程的干渠），都将成为战争中被定点打击的目标，防不胜防。一旦武统台湾的战火燃起，中方的人员死伤人数将远远超过台湾居民总数。

在和笔者讨论雅下工程时，藏族朋友唯色尖锐地指出：在你指出的发电量、落差、水文条件这些技术问题之外，我也有一个判断，我认为这个工程的“醉翁之意”，恐怕并不主要在电。对北京而言，它更像

是一个多重目标叠加的超限战：一方面是对印度的长期地缘压迫，尤其是围绕藏南问题，通过上游控制，制造一种持续的、不对称的心理与现实压力；另一方面，在生态与安全风险被不断放大的情况下，不排除其战略设想中存在一种极端推演，即迫使印度在某个节点上被动接受政治让步，以换取所谓“安全”与“稳定”。

唯色的观点可谓一针见血。对习近平而言，随着时间的推移，武统台湾的胜算越来越小，所以必须另辟一个新的战场，这就是雅下工程，把收回藏南也作为终极政治目标之一，就增加了习近平政治上回旋的余地。再说，利用收回藏南，高举极端民族主义的大旗，在大陆也不是一件难事。

针对雅下工程，印度政府采取了针锋相对措施。根据《印度时报》2025年10月13日的一篇题为《印度中央政府公布64亿卢比水电计划，对抗中国在上游兴建大坝》（Harnessing the Brahmaputra: Centre Unveils rs 6.4 lakh Crore Hydropower Plan; Counters China's upstream Dam）的报道¹⁸：印度电力规划部门已经制定了一项投资约合770亿美元的输电工程计划，到2047年将从布拉马普特拉河流域输送超过76吉瓦的水电产能，以满足日益增长的电力需求。报道在结尾处指出，今年7月中国已经开始在西藏东南部、靠近中印边界的雅鲁藏布江上修建一座大型水坝。中国的李强总理出席了这项工程的开工仪式。早在2025年1月印度方面已经表达了对这个大型水坝工程的关注。印度外交部门要求中国保证上游的工程项目不会伤害布拉马普特拉河下游国家的利益。

2025 年 10 月 14 日路透社也发表了一篇相关的报道，印度电力规划部门在周一发布的一份报告中称，该计划涵盖东北部多州 12 个子流域的 208 个大型水电项目，潜在发电能力为 64.9 吉瓦，另外还有 11.1 吉瓦的抽水蓄能电站。报告称，布拉马普特拉河流域横跨阿鲁纳恰尔邦（中国称藏南地区）等多个地区，蕴藏著印度 80% 以上尚未开发的水电潜力，仅阿鲁纳恰尔邦就有 52.2 吉瓦。¹⁹

其实印度对中共在西藏高原上大规模开发水电资源很早就开展系统性的科学调查和研究，最为典型的为印度学者布拉马·切拉尼（Brahma Chellaney）撰写的《水：亚洲的新战场》（Water: Asia's Battleground），2011 年 8 月在美国出版。布拉马·切拉尼是国际战略问题专家，2012 年因《水：亚洲的新战场》一书“对增进对当代亚洲的理解做出的杰出贡献”而荣获该年度的伯纳德·施瓦茨图书奖。

布拉马·切拉尼的观点非常明确：

第一，亚洲是全球水资源最紧张的大洲，西藏高原（中国称青藏高原）是亚洲的水塔。中国则处于多条亚洲大河的发源地，处于优势地位。

第二，1997 年联合国通过了《国际水道非航行使用法公约》，制定了共享水资源等规则。中国拒绝签署该公约而则坚持自己的主张，认为上游水源国有权对国境线内本国一侧的河流行使绝对的领土主权，也就是说，为了满足本国需要，上游水源国可以随心所欲地调用水资源，而无需考虑这样做会对下游国家产生什么影响。目前中国已经成为在亚洲水资源共享合作机制的关键障碍。

第三，中国拥有世界一半的水库大坝，水坝工程从趋于饱和的境内河流，转向了国际河流，而那里正是地震活动频繁的地区，又是少数民族居住的地区。中国政府在这些河流上的筑坝行为势必会引发沿岸各国的不安和反对。

第四，中国在雅鲁藏布江上游的筑坝行为构成对下游印度的水文战略优势——上游控制使中国具备在危机时调控水流、制造洪峰或断流的能力，形成不对称威慑，切拉尼称之为“水霸权”（hydraulic imperialism）。²⁰

雅下工程将这一水文战略优势延伸至更靠近中印实控线的位置，进一步压缩印度的反应空间。

自 2024 年底中共核准雅下工程后，特别是雅下工程正式宣布开工后，中国社交媒体的一些大 V 们纷纷指出，习近平建设雅下工程的目的是要收回藏南。极端民族主义情绪，已经随着工程的开工而高涨。



图 6：建设雅鲁藏布江下游水电工程和收回藏南，图片来源：网络截屏

六、雅下工程开工极大地刺激了中国股市

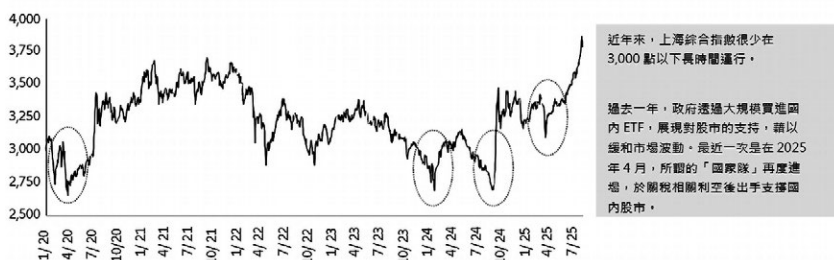
笔者以为，雅下工程在没有经过审查批准的工程可行性论证和工程环境影响评估报告情况下，匆忙地在 2025 年 7 月 19 日在另外一个水电站的工地上宣布正式开工，是为了尽快实现工程的近期目标：刺激股市，重振信心，抬高电价，拉动内需。

2025 年对习近平和中共领导层来说是非常重要的一年，“十四五”收官，“十五五”即将开始。多年来支撑中国 GDP 发展的房地产市场已经全面崩溃，唯一能够通过“非常措施”抢救的就是股票市场。

下图是欧洲最大的综合性金融集团德国安联集团（Allianz）在 2025 年 9 月《重新评估投资中国股市的十个理由》²¹ 一文中给出的上海股市综合指数近五年的发展趋势。这里关注 2025 年 1 月至 2025 年 7 月的综合指数发展。

中共政府 2024 年 10 月后采取的多重刺激政策作用下，2025 年第一季度综合指数冲到 2022 年 4 月以来的最高点。但是到了 2025 年 4 月综合指数迅速回落，回到 2024 年 5 月的水平，说明 2024 年 10 月后采取的多重刺激政策基本失效。2025 年 4 月国家队出手救市，²² 能源央企纷纷宣布增持回购股份，指数回升，但没有能够回升到 2025 年初的水平。

圖表 1：上海綜合指數（近五年走勢）



資料來源：Wind、安聯環球投資，截至 2025 年 8 月 31 日。以上資料僅供說明之用，本文所提及之權數僅為說明之用，不得成任何金融商品之推薦或投資建議。投資人因不同時間進場，將有不同之投資績效，過去之績效亦不代表未來績效之保證。

圖 7：2020 年 1 月至 2025 年 7 月上海股市綜合指數近五年的發展趨勢，
圖片來源：德國安聯集團

2025 年 7 月 19 日雅下工程開工，中共二號人物李強出席，但沒有象當年李鵬在三峡工程開工儀式發表了《功在當代 利在千秋》的豪情萬丈的講話，他只是講了一句話，宣布雅下工程正式開工。

但是上海綜合指數一下子衝到了五年來的最高點，目標 4000 點。可見雅下工程開工，投資 1.2 萬億的消息，極大地刺激了股市，重振了信心。

筆者發現，雅下工程正式開工的消息發布之際，網絡上湧現的大量文章多是來自與股票投資有關的網站，好像他們都是有備而來的。而且與雅下工程有關的股價上升最快，包括沒有前途的鋼鐵、水泥等股票。

比如據說承擔了雅下程設計和作為主要參建方的中國電建的股票，是能源央企的代表。根據《東方財富網》的資料，中國電建在 2025 年 7 月 18 日的最高價為 5.10 元。到 7 月 25 日最高價為 7.84 元。7 月

18日至7月25日，中国电建股票的最高价上涨了53.7%！中国电建微信公众号特别强调指出，中国电建股票近三个交易日上演“大象起舞”，股价连续三日涨停。

又比如高争民爆（从事民用炸药和爆炸服务的企业）的股票，在2025年7月18日的最高价为39.71元。到7月25日最高价为61.90元，为7月18日最高价的155.9%。根据金融小博士题为《1.2万亿！三峡工程的5倍，雅鲁藏布江水电工程概念股汇总！》²³的文章，雅下工程预估炸药用量在100万吨以上！按单吨1万元计算，仅炸药环节价值量就在100亿以上！加上爆破服务工程，民爆整体价值量保守计算大约850亿！

可以说，刺激股市，重振信心的短期目标暂时得到了实现。

下面谈谈短期目标中的抬高电价，拉动内需，这是因为电力供应在中国是垄断行业，中共政府可以以电价改革为名，抬高电价，更好地收割韭菜，以此减轻中央和地方政府的财政和债务压力。2026年1月27日国家发改委和国家能源局就发布《关于完善发电侧容量电价机制的通知》，²⁴核心就是提高电价。中国媒体宣传说，雅下工程的年发电量3000亿千瓦时，可以满足3亿人的用电需求！

2025年中国全社会用电量为10.3682万亿千瓦时。就是雅下工程的年发电量能够达到3000亿千瓦时，仅占总用电量的2.89%。其中城乡居民生活用电量15880亿千瓦时，占总用电量的15.3%。3000亿千瓦时占城乡居民生活用电量的18.9%。2025年底中国人口为14.0489

亿，人口的 18.9% 为 2.654 人，约为 3 亿人。这就有了雅下工程可以满足 3 亿人的用电需求的结论。

2025 年 9 月 26 日秦晖教授发布了题为《雅江大水电纵横谈》²⁵ 的视频。笔者惊奇地发现，在节目的留言区内，有不少网民表示中国的电价太低，要求抬高电价。这是一个非常奇怪的现象。市场经济的规律，需求量是随着商品的价格降低而加大的。就是说，需求者总是希望商品价格的降低，而不会自愿跪求抬高商品的价格。

下图是中国网络上出现的 2025 年世界上主要国家居民电价对比：²⁶

主要国家居民电价对比（2025 年）			
国家	电价（元 / 千瓦时）	折合美元	相当于中国倍数
中国	0.532	0.075	1.0
美国	1.278	0.18	2.4
日本	1.631	0.23	3.1
韩国	0.923	0.13	1.7
德国	2.835	0.40	5.3
丹麦	2.553	0.36	4.8

图 8：2025 年世界上主要国家居民电价对比，图片来源：百度文库

文章指出，2025 年中国居民电价保持全球较低水平，根据国际能源署最新数据，中国居民平均电价每千瓦时 0.532 元（约合 0.075 美元），仅为德国的 18.8%。文章继续指出，从电价构成来看，居民电价包含三个主要部分，发电成本约占 60%，输配电成本约占 30%，政府政策性基金及附加约占 10%，其中政府政策性基金包含重大水利工程建

设基金（0.364 分钱）、大中型水库移民后期扶持基金（0.623 分钱）和可再生能源电价附加（1.9 分钱）。

表中德国居民电价最高，为每千瓦时 2.835 元，是中国的 5.3 倍。关键的问题是，德国居民收入的中位数是否和中国居民相似？根据德国联邦统计局公布的数据，2024 年德国人的收入中位数为 52159 欧元；²⁷2024 年中国居民人均可支配收入中位数 34707 元。²⁸2025 年 1 月 24 日中国外汇交易中心受权公布人民币汇率中间价为 1 欧元对人民币 7.5062 元。²⁹52159 欧元相当于 39.1516 万元人民币。按照上表，一年耗电 2000 千瓦时，在德国电费为 5670 元人民币，占 39.1516 万元的 1.45%；在中国电费为 1064 元人民币，占 34707 元的 3.07%。电费的支出，对中国居民来说，负担更重。

李鹏曾在三峡工程开工仪式上承诺，“三峡建成后，电费将变成 8 分钱一度”。³⁰这个结论来自三峡工程可行性论证。综合经济评价组组长游吉寿在《三峡工程的综合经济评价问题》的文章中对比了两个电价，一个是每千瓦时 9.3 分，另一个是每千瓦时 6.2 分，并对两个电价进行对比分析。当时提出的三峡工程总投资为 361 亿元，年发电量 840 亿千瓦时。参加三峡工程论证的李治平建议将输电工程的投资 60 多亿元从 361 亿元中扣除，只剩 300 亿元，相应的电价为每千瓦时 8 分，并可以达到当年国家计委规定的内部收益率 11% 的要求。全国人大批准三峡工程后，总投资上升到 900 亿元（不含输电工程），因此三峡工程的上网电价从每千瓦时 8 分上涨到 0.25 元，推动了中国电价的全面上涨。³¹2025 年冬天，河北等地农村居民无法支付“煤改气”、“煤改电”后负担，只能挨冻硬挺，就充分说明中国电费对 6

亿多每月收入只有 1 千元的中国居民来说，是贵而不是太低了。

这里套用三峡工程可行性论证综合经济评价组的算法，雅下工程总投资 1.2 万亿元，年发电量 3000 亿千瓦时，上网电价应该为每千瓦时 0.93 元；如果总投资不变，年发电量只有 2200 亿千瓦时，上网电价应该为每千瓦时 1.27 元。按照发电成本约占电价的 60%，居民电价为每千瓦时 2.12 元。中国民众愿意支付这样的电价吗？中国民众有能力支付这样的电价吗？

所以笔者认为，雅下工程没有经过审查批准的工程可行性论证，工程的经济可行性未被证实。

七、雅下工程与青藏高原的生物多样性保护

雅鲁藏布大峡谷是世界上最长的峡谷，长达 500 多公里，同时也是最深的峡谷，最深处落差 6009 米，几乎是美国科罗拉多大峡谷的三倍。雅鲁藏布大峡谷是一座完整的天然博物馆，是气候和环境演变的启动区，是生物多样性的天然基因库。雅鲁藏布大峡谷集中了地球上半以上的生物物种，包括维管束植物 3768 种、鸟类 232 种、昆虫 1500 种，还拥有大量大型真菌、两栖爬行动物、哺乳动物、锈菌等，是生物资源开发利用最具有前景的地区，也是拯救和繁育濒危物种的理想之地。可以说，雅鲁藏布江大峡谷是地球上物种的宝库。³²

1985 年 9 月 23 日西藏自治区人民政府批准建立墨脱自然保护区经。2000 年 4 月 27 日国务院同意将墨脱国家级自然保护区扩界并更名为

雅鲁藏布大峡谷国家级自然保护区。

2025年9月19日，知名户外品牌始祖鸟联合艺术家蔡国强，在西藏喜马拉雅山脉江孜县一地区（海拔约5500米）燃放艺术烟花《升龙》，共计使用烟花1050盆，燃放时长约52秒，引发环保争议。³³首都师范大学副教授、植物学博士顾垒的第一反应是“这是一种极具侵入性的行为”。顾垒表示，哪怕主办方说烟花彩色粉是生物可降解材料，但是，青藏高原环境十分特殊，是生态脆弱区，相比于平原地区，青藏高原上低温少雨的环境会降低生物活性，从而降低分解速度。外来物质可能过了几年还存在，加重生态负担。2025年10月15日西藏日喀则市发布通报称，蔡国强喜马拉雅山烟花秀涉嫌违反《青藏高原生态保护法》、《草原法》等相关法律，要求做出赔偿。当地县委书记已被免职，另有官员正在接受调查。³⁴

《青藏高原生态保护法》³⁵于2023年4月26日由全国人大常委会审议通过，自2023年9月1日起施行。青藏高原的生态系统质量与功能状况直接影响到中国、亚洲乃至世界的生态安全、生物多样性、水资源供应、气候系统稳定，在中国和世界生态安全中具有独特而不可替代的作用。制定青藏高原生态保护法，是保障生态安全、加强生态风险防控，对加强青藏高原生态保护、促进经济社会可持续发展、实现人与自然和谐共生具有重要意义。生物多样性保护是《青藏高原生态保护法》中最重要的内容。

2025年3月28日西藏自治区人大常委会常务委员通过《西藏自治区实施〈中华人民共和国青藏高原生态保护法〉办法》³⁶自2025年8月1

日起施行。该办法第十七条规定，自治区人民政府应当将喜马拉雅山脉区域内的大型冰帽冰川、小规模冰川群等划入生态保护红线；制定重要雪山冰川和一般雪山冰川保护名录，对重要雪山冰川实施封禁保护，采取有效措施严格控制人为扰动，对一般雪山冰川采取相应的保护措施。

蔡国强在喜马拉雅山烟花燃放 1050 盆，燃放时长约 52 秒。西藏日喀则市以违反《青藏高原生态保护法》进行追责。雅下工程需用 100 万吨炸药，爆炸持续时间以十年计算。难道雅下工程就不违《青藏高原生态保护法》？

蔡颖莉等在《中国生物多样性保护政策演进、主要措施与发展趋势》³⁷ 一文中尖锐地指出，尽管中国在生物多样性保护方面做出诸多努力并取得了重要进展，但生物多样性下降的总趋势仍未得到有效遏制，局部地区生态退化问题突出。

为什么在中国生物多样性下降的总趋势未得到有效遏制？为什么局部地区生态退化问题依然突出？就是因为中国的《环境保护法》、《青藏高原生态保护法》、《中华人民共和国环境影响评价法》等不能阻止象雅下工程这样破坏生态环境、扼杀生物多样性工程的开工建设。雅下工程的决策者是知法犯法。

注释

1 百度百科：墨脱水电站，<https://baike.baidu.com/item/%E5%A2%A8%E8%84%B1%E6%B0%B4%E7%94%B5%E7%AB%99/14437462>。

- 2 徐大懋、陈传友、梁维燕：《雅鲁藏布江水能开发》，《中国工程科学》，2002年第4卷第12期，47-52页，<https://www.engineering.org.cn/sscae/CN/1160103293714227231>
- 3 蔡其华就西藏水利发展问题与自治区领导座谈，2006年6月16日，刊登在：中国水利网，https://www.chinawater.com.cn/ly/zj/202310/t20231028_808597.html
- 4 《中华人民共和国国民经济和社会发展第十四个五年规划》和《2035年远景目标纲要》，<https://www.ndrc.gov.cn/xxgk/zcfb/ghwb/202103/P020210313315693279320.pdf>
- 5 雅鲁藏布江下游水电工程已获核准，2024年12月15日，来源：新华网，<https://www.news.cn/politics/20241225/b61f14cad5c046a0a209bf36fa99d69b/c.html>
- 6 《李强出席雅鲁藏布江下游水电工程开工仪式并宣布工程开工》，2025年7月19日，来源：新华网，<https://www.news.cn/20250719/88ebff56443d466998c36182e2ca2dfa/c.html>
- 7 《中国雅江集团有限公司成立大会举行——张国清出席并为公司成立揭牌》，2025年7月19日，来源：新华网，刊登在：中国新闻网，<https://www.chinanews.com/gn/2025/07-19/10450625.shtml>
- 8 《雅江超级电站万亿投资背后：哪些电力央企将“分羹盛宴”？》来源：华夏能源网，2025年7月23日，刊登在：腾讯网，https://news.qq.com/rain/a/20250723A02UU900?suid=&media_id=#
- 9 赵厚纯等：《世界江河与大坝》，中国水利水电出版社，2000年，北京
- 10 《2024年可再生能源并网运行情况》，2025年1月27日，来源：国家能源局，<https://www.nea.gov.cn/20250221/e10f363cabe3458aaf78ba4558970054/c.html>
- 11 《2025年1-11月全国水电发电装机容量同比增长3%规模以上企业水力发电量同比增长2.7%》，2025年12月31日，刊登在：研观报告网，<https://www.chinabaogao.com/detail/775009.html>
- 12 西藏自治区自然资源厅：《西藏的水资源》，2010年8月14日，https://zrzyt.xizang.gov.cn/dt/xzzrzygk/201907/t20190731_90650.html
- 13 西藏自治区外事办公室：《雅鲁藏布江》，2024年2月26日，https://wsb.xizang.gov.cn/xzjx/lyjd/201107/t20110713_100958.html
- 14 当代中国，通讯中国，人文地理：《建世界最大水电工程雅鲁藏布江水力资源有多丰富？》<https://www.ourchinastory.com/cn/15262>
- 15 第一观察：《10年·70年·100年从三个历史性时点看习近平西藏之行》，来源：新华网，2021年7月23日，<https://www.xinhuanet.com/>

- politics/leaders/2021-07/23/c_1127688682.htm
- 16 斯洋：《祖宗土地不能丢，习近平划红线，美中冲突风险增加？》，2018年6月29日，美国之音，<https://www.voachinese.com/a/mattis-xi-jinping-20180628/4459089.html>
- 17 宋国诚专栏：《不要低估习近平侵台野心——必须保卫台湾系列专文4》，2026年6月8日，上报，<https://www.upmedia.mg/tw/commentary/columnists/260619>
- 18 《布拉马普特拉河：印度中央政府公布64亿卢比水电计划，对抗中国在上游兴建大坝》，2025年10月13日，印度时报，<https://timesofindia.indiatimes.com/india/harnessing-the-brahmaputra-centre-unveils-rs-6-4-lakh-crore-hydropower-plan-counters-chinas-upstream-dam/articleshow/124529710.cms>
- 19 《印度制定770亿美元水力发电计划，与上游的中国雅鲁藏布江大坝竞速能源》，2025年10月14日，来源：路透社，刊登在：商讯，<https://tinyurl.com/4u98e2x6>
- 20 布拉马·切拉尼：《水资源：亚洲的新战场》（Brahma Chellaney: Water Asia's Battleground, Georgetown University Press, Washington D.C., 2011）。
- 21 Allianz：《重新评估投资中国股市的十个理由》，2025年9月，<https://tw.allianzgi.com/-/media/allianzgi/ap/taiwan/pdf/sector-insight/ten-reasons-to-reconsider-china-equities-20251020>
- 22 中国能源报记者李丽旻：《能源央企纷纷宣布增持回购股份，“组合拳”提供市场托底支撑“国家队”强力出手彰显稳市决心》，2025年4月14日，中国能源报，第四版，https://paper.people.com.cn/zgnyb/pc/content/202504/14/content_30068694.html
- 23 《金融小博士：1.2万亿！三峡工程的5倍，雅鲁藏布江水电工程概念股汇总！》，2025年7月20日，刊登在：新浪网，<https://finance.sina.com.cn/roll/2025-07-20/doc-infhcxhn5437219.shtml>
- 24 国家发展改革委 国家能源局：《关于完善发电侧容量电价机制的通知（发改价格〔2026〕114号）》，发布时间：2026年1月30日，来源：价格司，刊登在：中华人民共和国国家发展改革委员会网，https://www.ndrc.gov.cn/xxgk/zcfb/tz/202601/t20260130_1403524.html
- 25 秦晖：《雅江大水电纵横谈》，2025-9-26，<https://www.youtube.com/watch?v=YpfGu3PUhzw>
- 26 AI 编辑：《2025年中国居民用电收费标准全面解读》，2025年12月9日，百度文库，https://wenku.baidu.com/view/9f82c17a7a21af45b307e87101f69e314332fa89.html?_wks_=1769437028114&needWelcomeRecommend=1&unResetStore=1

- 27 ZDF: *Heutes Beitrag*, <https://www.facebook.com/ZDFheute/posts/52159-euro-so-viel-betrug-der-mittlere-bruttojahresverdienst-im-vergangen-jahr/1131600552343582/>
- 28 《2024 年居民收入和消费支出情况》，2025 年 1 月 17 日，来源：国家统计局，https://www.stats.gov.cn/sj/zxfb/202501/t20250117_1958325.html
- 29 《2025 年 1 月 24 日中国外汇交易中心受权公布人民币汇率中间价公告》，2025 年 1 月 24 日，文章来源：中国人民银行货币政策司，<https://www.pbc.gov.cn/zhengcehuobisi/125207/125217/125925/5575586/index.html>
- 30 王维洛：《关于“三峡工程上马后居民用电每千瓦时 8 分钱”的真相——中共政权和李鹏忽悠和欺骗中国百姓的事实》。
- 31 《三峡工程已经运行了 20 年，当初投入的 2500 多亿元，如今回本了吗》，2024 年 6 月 20 日，刊登在：英思科技网，https://www.wuhanins.com/news_detail/17.html
- 33 王维洛：《雅鲁藏布江下游水电大开发是违法工程》，2021 年 3 月 22 日，议报，<https://yibaochina.com/?p=241269>
- 33 《在喜马拉雅山脉放烟花？最新回应》，2025 年 9 月 20 日，来源：中国新闻周刊，刊登在：观察者网，https://www.guancha.cn/politics/2025_09_20_790781.shtml
- 34 安娜：《中国艺术家蔡国强喜马拉雅烟花秀被指破坏环境在西藏遭追责》，2025 年 10 月 25 日，法国国际广播电台，<https://tinyurl.com/47e3xjpu>
- 35 《青藏高原生态保护法 9 月 1 日起施行》，2023 年 4 月 23 日，来源：新华网，https://www.news.cn/politics/2023-04/26/c_1129569046.htm；中华人民共和国青藏高原生态保护法，<https://faolex.fao.org/docs/pdf/chn217026.pdf>
- 36 西藏自治区实施《中华人民共和国青藏高原生态保护法》办法，2025 年 3 月 29 日，来源：西藏日报，https://www.xizang.gov.cn/xwzx_406/bmkx/202503/t20250329_470143.html
- 37 蔡颖莉、朱洪革、李家欣：《中国生物多样性保护政策演进、主要措施与发展趋势，生物多样性》，2024 年，第 32 卷，<https://www.biodiversity-science.net/CN/10.17520/biods.2023386>

特稿

肖潇

作者肖潇为美国《中国文化大革命数据库》编委会成员

中国民主季刊

第4卷 第4期

2026年10月

CC-BY-NC-ND

自由、免费转载分发，但须注明
作者与出处，并不得修改及商用

矢志不渝的研究 文革60周年北美 研讨会述评

编按：2026年8月18日，美国《中国文化大革命数据库》编委会组织召开“文革60年：中国和世界的回顾和环顾”的网络研讨会。会议对文革研究已有的定论和固见提出了新挑战，但并不限于文革本身的史实和课题探索，还讨论了当今中国和国外的文革现象。在这个追记中，肖潇概述了会议的主要内容和自己的感受。

光阴荏苒。1966年爆发的中国无产阶级文化大革命，不觉距今已经六十年了。尽管在它的发源地——中国大陆没有任何纪念和研讨活动，但在北美，却没有忘却这一被中国执政者刻意忘却的历史。8月18日，由美国《中国文化大革命数据库》编委会主办、中国民主转型研究所和纽约博登书屋协办，召开了“文革60年：中国和世界的回顾和环顾”的网络研讨会。会议为期一天，参加者有来自美国、加拿大和中国的二十多位文革研究学者。与会者有两大鲜明的特点。首先，年龄上都铭刻着“文革一代”的特殊印记：文革的亲历者。他们中的年高者，早已经踏上了人生“古来稀”的八十多岁的高峰。最年轻的，也已经六十二岁，走近了退休的门槛。其次，他们不仅亲身经历了中国文革的风浪，还亲眼目睹了世界性的文革的返潮现象。无论是中国习近平的“文革2.0”，还是美国左派的“身份政治”或右派的个人崇拜和民粹运动，都成为他们责无旁贷的针砭和研究对象。

对研究定论的挑战和理论格局的重构

对文革研究这样一个已经有半个多世纪积累的学术领域而言，最可贵的莫过于对已有的定论和固见提出有理有据的新挑战，从而把研究提

高到一个新的高度。这次研讨会的第一个发言者是来自加拿大多伦多大学的吴一庆教授，他演讲的题目是《从文化革命到政治大革命：文革源起与发动再探》。这一题目使我们想起了吴教授在文革 50 周年的研讨会上的相似的演说，在那时他提出了目前学界对文革起源的解释常常有一种“倒写历史的时间错置”。比如，他认为“打倒刘少奇是文革运动的结果，而不是原因。”¹ 当时他的见解虽然也有一时的振聋发聩之处，但是还远谈不上蔚为大观的一家之说。这十年来，吴教授已经就此课题在撰写三本英文专著，其中一本，《没有脚本的革命：打倒刘少奇与文化革命》不久将由剑桥大学出版社正式出版。在他新的演讲中，十年的钻研厚积薄发为一系列系统性的新阐释。第一，他反对文革发动是一场毛泽东预先精心策划的“预谋论”的叙事范式。认为文革发动的历史轨迹并非是单一、连续线性的进程，其进程与最初设想之间可能存在重大错位与变异。第二，文革始于文化领域的斗争是不争的事实，但难以证明这是毛泽东发起一场“政治大革命”的谋划。即便是《五一六通知》和 1966 年 8 月份的《十六条》也有着截然不同的理论和运动边界。后者实际上标志着“文化革命”成为“政治大革命”的一种“再发动”。第三，毛泽东和刘少奇的矛盾是不争的事实，但这并不足以证明毛泽东发动文革就是为了打倒刘少奇。历史过程的不可预测性和历史行动者的多重互动推动了不期而至的事件异变。比如毛泽东的“炮打司令部”的大字报就是一种意外的非理性行为。由此，“文化革命”偏离原先的轨道，运动的目标、范围、方向与方法均发生重大变化，造成了包括刘、邓在内的一大批高级干部落马。虽然我们都知道：无论是文革起源的“预谋论”，还是吴教授对它在理论上的新挑战，都在史料和立论上有其依据。但是，这一挑战无疑建立了新的分析范畴，开辟了一种新的历史叙述的

可能性。

如果说吴一庆教授的新说主要还着力于对旧说的有力挑战，那么斯坦福大学许成钢教授的发言《文化大革命的制度基础和制度成果》更显示了一种对宏大理论格局的重构。许教授的论述首先使用了“制度基因”的概念，他认为：中国当代极权制的制度基因来自苏俄的极权主义制度。但中国传统帝制的制度基因则决定了来自苏俄的制度基因在中国被吸收和被改造的方向和过程。大跃进和文革是改造苏俄制度基因的关键阶段。在造成极大灾难的同时，这两个运动把来自苏俄的经典极权制改造成了具有中国特色的区域管理的极权制。在这种区管式极权制里，全社会大部分的经济和行政不是由中央机构，而是由各级地方党政管理和运作的。九十年代成为中国经济发展最大引擎的乡镇企业更直接是在文革时期社队企业基础上发展而来。区管式极权制的治理结构构成了文革后近四十年的改革开放转型的制度基础，使私营部门得以在极权体制内部迅速生长和扩张，并由此形成具有中国特色的列宁主义新经济政策，帮助中共不仅摆脱了统治危机，而且变得更韧性。许教授的这一由不同经济形态衍生的对文革前后近百年的因果解说，不一定能使所有人都认同，但毫无疑问是一个高屋建瓴的新的理论构架，值得学界在震惊之余做更深入的探讨。

类似于许成钢教授的纵横史观式的探寻，还有裴毅然教授的《文革前后的“大外宣”》和程映虹教授的《1972：民族主义符号的崛起——文革与中国道路再讨论》这两篇发言。前者极为细致地追溯和梳理了中共“大外宣”的源起、发展和成败，后者则以文革中并不算显眼的1972年为时间截面，以那一年里的“针刺麻醉与中医中药”、“元

谋人、马王堆”的新闻事件为契机，分析了民族主义如何在文革那段文化单调、物资匮乏、个性萎缩的时期成了少有的精神刺激，其深层心理，未必和“祖上曾经阔过”和贪大慕强这类民族主义情绪没有关联。这种社会心理，延续至今就是对中国崛起和“国家统一”的大众支持。

对传统研究的新发掘和新见解

会议的很大一部分发言仍然聚焦于文革本身的史实和课题探索。但这些研究绝非是单纯的老调重弹，而是推陈出新的新发掘和新见解。美国宾州约克学院的周泽浩博士在会议“自由发言”议程中的发言题目是《文革初期反孔运动的再考证》。在人们的固见里，文革中对传统文化的破坏莫过于红卫兵的“破四旧”运动。但是周博士的新考证却告诉我们：其实文革中对孔府、孔庙和孔林（“三孔”）造成最大损害的并不是北师大的两百名红卫兵，而是当地坚决支持他们的行为，积极充当“马前卒”的“贫下中农们”。这些农民（不少还是孔家后裔）对“三孔”遗产的破坏是出自于敛财的疯狂，因而是最有毁灭性的长期劫掠行为。文革中造反派的时运突转，缘起于1968年夏天的所谓全国造反派座谈会的“北航黑会”。但是长期以来，从没有人对这一重要事件做过认真的研究。来自芝加哥的文革研究学者谭加洛（笔名阿陀），近年来对这一重要课题做了潜心的梳理和挖掘。在他题名为《“北航黑会”的来龙去脉》的发言中，他指出：所谓的“北航黑会”是因两广牵头的全国造反派对抗《七·三布告》，呼吁制止大屠杀的“公车上书”。

评价文革的造反派问题，不仅在文革中是热门的话题，至今为止还是研究的“显学”之一。本次会议的发言中，中国文革研究学者，社科院的徐友渔教授以《文化大革命中的造反和造反派》为题，发表了对这一素有争议的问题的看法。他认为：文革中的造反派并不是一群想趁机作乱的坏蛋，而是遭遇和反应都正常的中国人，是当时中国的国情——政治矛盾和社会矛盾——与毛泽东驾驭文革运动走向的特殊手段相结合的产物。造反派也显示出在政治思维、行为方式以及性格上的多样性。在文革中，造反派既有盲目服从、“誓死捍卫”的一面，也有自作主张、追求自己目标的一面；既有运动初期坚持正义与平等，抗拒镇压的一面，又有运动中期得势后就得意忘形，压制他人的一面；文革以“四人帮”的束手就擒结束时，原造反派中人们的反应也大不相同，有人悲伤有人欢呼，这说明人们对于文革，对于造反的态度是随形势的不同而变化的。——应当说这些都是理性公允之论。

中央文革小组在文革中的角色人所皆知。但人们不知道的是：研究中央文革小组的学术论文却寥寥无几。自由主义代表人物、民运理论家胡平题名为《中央文革小组在文革中的角色及其命运》的论文，从中央文革小组的角色、成员构成、历史上与其相似的机构、毛对文革小组的利用和抛弃等多个方面进行了系统的剖析。尤其是文中利用江青的名言“我是毛主席的一条狗，毛主席叫我咬谁我就咬谁。”²和陈伯达的悲叹“毛主席不会失败，我们会失败”³，把中央文革“打手走狗”和“替罪羊”的双重政治作用刻画得活灵活现、栩栩如生。这一发言，其实是填补了热门话题中的研究空白。无独有偶，另一个文革学者丁凯文的《文革时期的叶剑英——浅谈被官史扭曲的历史真相》

也同样有正本清源的作用。在他的笔下，素来被中共官媒一边倒地溢美颂德的叶剑英的文革功过被公正地褒贬，还了他“党政军中仅次于周恩来的政客型人物，实乃中共政坛上的不倒翁”的历史真面目。

直面中美两国新出现的文革现象

论及二十一世纪的文革现象，发生在中美两国政坛上的个人崇拜等现象都将是很难被绕过的课题。在中国，在文革被公开否定的三四十年后，新上台的习近平竟又开始了一个称之为“文革 2.0”的执政时期。在美国，在唐纳德·川普（Donald Trump）执政的两个任期内，也开始打破美国民主政治传统、不遗余力推广个人威权。当一个有风骨的知识分子在美国谈论吸取文革教训时，是不能够对发生在他身边的文革现象视而不见、甘当犬儒的。所幸的是：会议收到了三篇论文，对从毛泽东、习近平到川普的文革思维、理论和现象做了深入的分析 and 全面的批判。

第一篇是来自加州州立大学洛杉矶分校的荣休教授宋永毅的《权力崇拜的返祖：从习近平到川普的个人崇拜现象》。宋教授在他的发言中非常系统地比较分析了 21 世纪发生在美中两国的个人崇拜返祖现象。论文从“个人崇拜”的理论界说切入，揭示其本质是对集权与绝对权力的崇拜。其次，他又探讨了独裁者借由民粹主义运动打破任期限制、追求终身执政的内在动力；剖析了极权 / 威权政治下“庇护者 - 被庇护者”双向交互作用所引发的结党营私与党同伐异。再次，他还从精神病理学视角探讨了领袖病态人格（恶性自恋与偏执型人格障碍）在政治造神中的决定性影响。最后，他对比了马列主义体制

（及其“核心论”）对个人崇拜的滋养，与美国三权分立宪政框架对个人崇拜的制度性抵制。该论文对川普个人崇拜的最后结论是“如果以建立覆盖全国、跨党派且制度化的个人崇拜为标准，则其效果非常有限，并受到美国制度、法律和社会多元性的明显制约。”

威斯康星大学荣休教授郭建也以《反智、反体制的群众运动：川普革命对文明的挑战》为题对中美两国的文革现象进行了鞭辟入里的比较分析。该文以中国文化大革命、纳粹德国群众运动与川普时代美国政治的比较为切入点，提出“反智、反体制的群众运动”可能成为不同政治体制共同面临的危险。文章认为，毛泽东发动文革是为了用“文化”名义掩盖其以思想统一、压制知识与独立思考为特征的反智运动。作者据此分析川普“革命 2.0”的两个核心面向：一是以“排干沼泽”、攻击媒体、削弱政府机构、强调个人忠诚等方式冲击既有制度；二是通过削弱教育、科学、大学及文化机构，推动反智和反科学政策。文章进一步指出，川普政治的危险不仅来自领导人的个人倾向，也来自相当规模的群众支持以及美国社会的政治极化。虽然法院、国会和宪法仍对权力扩张形成制约，但美国制度本身亦存在总统权力过大等结构性问题。作者最后以德国和中国的历史经验警示：当体制内部的领袖为了巩固个人权力而煽动群众反对制度时，运动可能走向反智、反文化乃至反文明，因此这些历史经验应成为美国反思民主危机的一面镜子。

第三篇探索 and 比较中美文革思维的论文来自现居美国的澳门大学荣休教授郝志东，题名《“文革思维”的普世性：从施密特、毛泽东到川普》。作者以三者的政治思想与实践为比较对象，探讨“文革思

维”即“斗争思维”的普世性。郝教授将其概括为三个相互关联的方面：非友即敌、你死我活的二分思维；以极端化、专制化和非人化为特征的思维方式；以及以暴力消灭异己的政治逻辑。文章通过德国纳粹、中国传统及革命时期的阶级斗争、当代美国政治等案例，说明这种思维并非中国“文革”所独有，而具有跨时代、跨社会的普遍表现。作者认为，现实主义因迎合人性中的利益、恐惧和权力需求，具有一定逻辑自治性和实践效力，但其极端化会侵蚀人的尊严、制度规则并造成严重灾难。论文最后主张以“现实的理想主义”结合现实主义与理想主义，在承认利益和现实约束的同时，坚持正义、对话、规则与非暴力，以制衡并超越“斗争思维”。

文革 60 周年的网络研讨会虽然只进行了一天，但硕果累累、深广度兼备。意犹未尽的与会者都相信它不仅是一个成功的开端，更会是一个走向更大成功的开始。在会议结束时，与会者曾谈及下一个十年的研讨会，可能不少年事已高者会因生命的消逝而缺席，但大多数学者却如此谈笑风生：“只要还活着，一定会参加文革 70 周年的研讨会”。这种对文革研究坚忍不拔、矢志不渝的执着，使人不由地想起二战盟军统帅之一的道格拉斯·麦克阿瑟将军的“老兵不死，只是凋零”（*Old soldiers never die, they just fade away*）的豪迈名言。⁴

注释

- 1 宋永毅主编：《文革五十年：毛泽东遗产和当代中国（上）》，纽约：明镜出版社，2016 年，第 16 页。
- 2 纪思道：“毛的遗孀江青自杀”，纽约时报（中文版），1991 年 6 月 5 日。<https://cn.nytimes.com/china/20160715/archive-jiangqing-suicide/>

- 3 戚本禹：《戚本禹回忆录》，香港：中国文革历史出版社，2016 年，第 675 页。
- 4 这句名言出自麦克阿瑟于 1951 年 4 月 19 日在美国国会参众两院联席会议上发表的著名告别演说，演说标题通常被称为“国会告别演说”（Farewell Address to Congress）或“老兵不死”（Old Soldiers Never Die）。

公民抵抗

L. 杰森·阿纳斯
塔索普洛斯、
连杰 文

洪范 译

L. 杰森·阿纳斯塔索普洛斯 (L. Jason Anastasopoulos)，是佐治亚大学公共行政与政策系的副教授，同时担任《公共行政评论》(Public Administration Review) 的副主编

连杰 (Jie Jason Lian)，是哈佛大学“非暴力行动实验室”(Nonviolent Action Lab) 的博士后研究员，以及哈佛大学“阿什民主治理与创新中心”(Ash Center for Democratic Governance and Innovation) 的访问学者

译者洪范为中国大陆专业人士

中国民主季刊

第 4 卷 第 4 期

2026 年 10 月

CC-BY-NC-ND

自由、免费转载分发，但须注明
作者与出处，并不得修改及商用

专制主义人工智能的局限

人工智能的出现被认为给专制统治提供了一种异常强大的控制工具。L. 杰森·阿纳斯塔索普洛斯 (L. Jason Anastasopoulos)、连杰 (Jie Jason Lian) 在这篇文章中提出，专制政权在运用人工智能时也存在内在的局限性。人工智能系统会迫使统治者面对一个无法回避的“校准困境”：降低判断阈值，可以减少漏报，抓住更多潜在挑战者，但也会增加误报，把更多无辜者卷入其中，从而引起反弹。这种结构性的不稳定会产生所谓的“阈值抽风”，即在不断收紧与突然放松之间循环摆动，中国的情况就是一个例子。民主力量可以通过消解对算法权力的神秘化、建立保护性规范，以及通过策略性抵抗挑战“全景监狱式监控”的假象，来利用这些脆弱性。

本文原载于美国《民主季刊》2026 年第 2 期 (*Journal of Democracy*, Volume 37, Number 2, April 2026, pp. 5-17)。

人工智能常被称为“神奇利器” (silver bullet)，它能帮助专制统治者以更低的成本、更快的速度和更精准的方式实施镇压。这种说法未必有错，至少在短期内如此。数字监控能够压制言论、识别反对者，并帮助政权扩大强权统治。但关于人工智能政治力量的最大论断——即实现预测性、先发制人的大规模控制——却面临着一个约束性的限制。

任何处理概率的系统都必须决定：将一个人、一种行为或一条信息标记为“风险”的阈值应设得高还是低。降低阈值 (threshold)，政权可以减少漏报 (false negatives)，抓住更多潜在挑战者，但也会增加误报 (false positives)，把更多无辜者卷入其中，从而引起反弹。提高“风险”的阈值，政权可以减少殃及无辜的镇压 (collateral

repression），维护自身合法性。但这样也会漏掉真正的威胁，产生对手可以利用的盲区。简而言之，预测必然伴随错误，至少在某种程度上如此。唯一的选择在于，政权更愿意容忍哪一种错误。

我们把这个问题称为“独裁者的校准困境”（calibration dilemma）。这不是一个会随着下一次更新、更多数据或更好的工程技术而消失的“漏洞”。这是预测系统应用于政治现实时的一个结构性特征——尤其是在可信威胁稀少、公民不断适应监控、官僚和党的忠诚分子又有动机操纵指标的情况下。

在中国，这一困境尤为突出，因为监管与审查已经把“校准”变成了一项官僚事务。自 2022 年 3 月起，中国国家互联网信息办公室（网信办）要求那些推荐系统具有影响“舆论”或“社会动员”能力的主要平台，向国家算法备案系统提交关键信息和自我评估，并在系统发生变化时更新备案。¹ 2022 年 8 月，网信办公布了阿里巴巴、腾讯、字节跳动等公司的首批备案信息。针对“深度合成媒体”（deep-synthesis media）（如 AI 生成的视频和语音）以及生成式 AI 服务的类似规定，还增加了身份验证、内容治理要求，以及对政治敏感功能的人工审核。²

这样一来，政府的职责不仅限于观察结果，还包括监督人工智能的设置：重大更新可能会触发安全审查和行政审查。表面上看，这像是一种集中控制。但在实践中，这揭示了更深层的问题——即围绕政权能容忍多少错误，以及惩罚过多无辜者与漏掉过多反对者哪个更安全，始终存在着一场持续的博弈。

这场博弈牵涉到中国中央政权、地方官僚和公民。对地方官员的考核，依据的是中央下达的维稳目标。这些官员有强烈动机去杜绝“群体性事件”和上访，操纵合规指标。³ 公民则通过戴口罩、使用暗语、从公开发帖转向私下聊天等方式来适应。如果管控范围过广而导致反弹，北京就会介入以重新确立控制权。它会发布新指示、集中权力，并展开危机公关（damage-control efforts）。其结果就是我们所说的“阈值抽风”（threshold whiplash）：围绕政治敏感时刻收紧，随后是突然放松、清理和宣传修复，如此循环。

2022 年末中国“清零”政策的迅速撤销，便是一个生动的例子。在近三年时间里，地方官员以维护公共卫生为由，严格执行严苛的封控及其他管控措施。2022 年 11 月，反对这些严苛限制的抗议活动在全国范围内爆发。在这些“白纸”示威活动中，人们举起空白标语牌，不写下任何被禁止的内容而巧妙地表达异议。这使得中共政权难以在不付出高昂代价的情况下加以镇压。

在经济和行政压力日益加剧的背景下，中国于 12 月 7 日发布了自疫情爆发以来最全面的放宽限制措施。⁴ 此后，官方宣传的基调从“人民战争”转向了安抚民众和倡导自我防护，高层官员公开强调新冠最新变异毒株的严重性已“减弱”。⁵ 然而，即便政策放宽，当局仍对部分参与抗议活动的人士实施拘留或施加压力，并压制有关这些抗议活动的公开讨论或纪念活动。简而言之，当局在行政层面放松管制的同时，却在政治层面收紧了控制。⁶

这种反复波动暴露了监控性政权的一个关键局限。人工智能无法解决

专制者的困境，因为这些系统的训练数据源自充满恐惧的民众，其所学习到的信号只能是隐晦的、被审查过的，且经过策略性操纵的。而人工智能所能做的，是通过把这种权衡取舍转移到模型、阈值、备案和安全审查之中，从而将不确定性变成官僚程序，在此过程中，这种不确定性变得无法回避、具有政治后果，且越来越明显。

独裁者的校准困境

这种不确定性的转移，迫使这个政权必须去面对统计误差带来的具体运作成本。尽管人工智能许诺精准，但 AI 赋能的镇压仍受制于预测错误，而这些错误会带来确定的政治惩罚。一方面是误报（I 类错误），系统将无辜者卷入其中：被人脸识别误认的邻居、被过滤器标记的无害帖子，或是被误判为可疑的正常活动。这些绝非单纯的统计上的烦扰。它们会累积民怨，使安全部门被噪音淹没，并让公民意识到，服从并不能保证安全。另一方面是漏报（II 类错误），即系统漏掉了真正的威胁。随着算法被调整以减少噪音，真正的挑战者会漏网，而（反抗）网络也学会了以模型无法察觉的方式进行协调。在这种情况下，即使反对派活动已转入地下，人工智能却会给专制政权带来一种令人心安却虚假的稳定感。

大规模自动化的三个特点使这一两难困境更加尖锐。第一是自主性权衡（autonomy trade-off）。赋予算法广泛的自主权，虽能最大限度地提高速度和覆盖范围，但也放大了哪怕是微小错误率所带来的后果。一个在全国范围内运行的模型，可能会产生数千次错误警报，从而将微小的统计误差变成重大的政治负担。而通过要求人工核查来限

制算法的自主性，虽然可以避免部分过度反应，但也会造成官僚程序上的瓶颈。警报不断堆积，分析人员不停地进行甄别，而当初让人工智能具有吸引力的速度和规模优势，也开始逐渐消失。

人工智能第二个出问题的特点在于算法的僵化（algorithmic rigidity）。模型是基于历史数据训练的，并假定昨天的模式能够预测明天的威胁。但在监控下，“真实情况”始终在变化。公民通过发明新暗语、改变路线、更换平台或将讨论转移到私密渠道等方式，来进行调整适应。结果就是模型迅速过时：模型上个月学到的东西，对今天的行为已难以提供有效指导。这迫使政权不得不进行持续的重新训练和重新校准，而这种灵活性正是大型监控官僚机构难以维持的。

放大这些风险的是第三个特点，即“稀有事件问题”（rare-events problem）。在任何社会中，活跃的异见者都是稀少的，仅占社会中极小的一部分。当算法试图搜寻他们时，即便是高度精准的系统，产生的误报数量也会远多于真实检测结果。如果一个监控模型的准确率达到 99%，但扫描的人口中每一万名公民中只有一人是真正的威胁，那么被标记出来的绝大多数人都是无辜的。其结果是一种危险的反馈循环（feedback loop）：安全机构被虚假“警报”淹没，因殃及无辜的镇压而导致政权的合法性遭到侵蚀，而真正的威胁则可能被埋没在海量的数据噪音之中。

这种两难的校准困境，根源于国家统治术（statecraft）长期以来的一种追求——使社会“变得可识别、可治理”的追求。现代国家长期以来一直力图把复杂性压缩为可以被统计、比较和治理的类别。19 世

纪，阿道夫·凯特勒（Adolphe Quetelet）曾承诺建立一种“社会物理学”（social physics），通过统计学将混乱的人口纳入行政管理。⁷詹姆斯·斯科特（James C. Scott）后来警告说，当这种对社会清晰可控的野心与“高度现代主义”（high modernism）合流时，国家就会强加一套简化的社会和政治生活图景，这不可避免地会与地方性知识（local knowledge）和人类行为发生冲突。⁸

数字专制主义（digital authoritarianism），正是这种高度现代主义野心在算法时代的全面升级，其特征在于持续的量化评估、实时分类和自动化干预。然而，正如斯科特所指出的，地图终究不是疆域（the map is not the territory）。一个政权越是挖空心思刺探公民的“忠诚度”与“风险值”，并据此采取行动，这个政权就越会激发出相应的策略性调适。公民学会了如何避免被系统标记，而官员则修饰（tweak）统计数字以取悦上级。监控国家可能在量化数据变得更多的同时，实际掌握的信息却变得更少，用真实的认知换取了一种掌控一切的表象。在中国近期应对新冠疫情的官方措施中，这种权衡取舍体现得最为明显，也最为反复无常。

阈值抽风

中国以不同寻常的力度追求着斯科特所描述的高度现代主义野心，将算法工具嵌入日常治理和国内安全之中。在疫情“清零”时期使用的“健康码”系统，展示了基于阈值的管控在实践中是如何运作的。为了大规模管理公共卫生，当局将复杂的流行病学现实转化为简单的颜色编码状态，通常为绿色、黄色或红色。状态的变化可能

决定一个人能否出行、进入公共场所，乃至能否离开家。实际上，数字阈值（cutoff）已成为强制措施的触发点。而由于这一阈值过于僵化，系统极易引发剧烈波动。

尽管有阿里巴巴等主要平台的强大技术支持，“健康码”体系在社会层面仍显脆弱。两大失误引发了民众的不满。其一，高灵敏度的筛查产生了误报。人们在没有明确解释或可行申诉渠道的情况下，被标记为“红码”并遭隔离。⁹ 其二，不透明性为政治操纵留下了空间，进而威胁到系统的合法性。最明显的例子发生在2022年河南村镇银行危机期间。¹⁰ 据报道，面对腐败丑闻曝光后的银行挤兑，当地官员通过远程将储户的健康码转为红色，阻止他们外出抗议。这一事件暴露出，该系统并非中立的安全仲裁者，而是一个可为政治便利随意操控的开关。

随着时间的推移，人工智能的刚性控制没有带来稳定，反而走向了反面。任意限制和未解决的不满情绪不断积累，非但没有悄无声息地压制异议，反而激起了公众的愤怒，最后演变为公开抗议。2022年10月，北京四通桥上出现的横幅标语，直接抨击监控国家和“独裁者”。次月，多座城市爆发了“白纸”抗议。除了对封控措施的愤怒之外，人们还表达了政治诉求。原本为识别个体而调校的监控体系，在面对由共同不满驱动的群体抗议时，却难以加以遏制。面对这种系统性失灵，中央政权采取了一种典型的“急转弯”策略（whiplash maneuver）。北京方面突然放松了关键限制措施，扭转了官方宣传口径，并着手进行“记忆抹除”（memory suppression）和对部分参与者的惩罚。

对抗监控的低技术手段，还包括香港抗议者使用雨伞、口罩和激光笔来干扰摄像头。在俄罗斯，抵抗既有技术形式也有法律形式。一些活动人士使用“迷彩”化妆来迷惑系统，另一些人则在被算法监控系统标记后寻求法律救济。¹¹ 活动人士尼古拉·格鲁欣（Nikolay Glukhin）在 2019 年的一次个人示威中被莫斯科地铁的监控系统识别，此后他将案件诉至欧洲人权法院（European Court of Human Rights, ECHR）。2023 年，他赢得了一项具有里程碑意义的裁决：缺乏严格法律保障的大规模人脸识别，构成对隐私权的侵犯。¹² 尽管俄罗斯已于 2022 年 9 月撤销了对欧洲人权法院的承认，但由于该案处理的是在撤销承认日期之前提出的投诉，因此仍具有可诉性（并为成员国确立了有效先例）。

在伊朗，伊斯兰政权针对 2022 年爆发的“妇女、生命、自由”（Women, Life, Freedom）抗议运动，安装了由人工智能驱动的街头监控摄像头，用于识别违反头巾着装规定的女性，这个规定要求女性完全遮盖头发和身体。早在 2026 年 1 月大规模起义爆发之前，这项技术就未能镇压公民抗命运动，因为女性仍在公共场所持续主张自己的权利。¹³ 这些监控摄像头也未能阻止数百万人走上街头，在伊朗全部 31 个省份掀起自 1979 年以来规模最大的抗议浪潮。这个政权得以存续，并非依靠技术控制，而是通过极端暴力手段：实弹射击、外国什叶派民兵，以及可能超过三万人的死亡。在每一种情况下，人工智能都促使抵抗运动进行适应，却不能将其平息。

需要明确的是，在某些情况下，人工智能确实会强化专制统治。它能在受控的环境中识别已知个体，用可预测的线索来过滤内容，并对线索进行分级筛选，从而简化人工审核流程。就预测性、面向大规模人

口的警务而言，如果政府已锁定嫌疑人，人工智能系统能够确认其身份；但在缺乏对特定个体的可疑根据的情况下，若要求其从数百万普通公民中筛选并标记“高风险”人群，AI 系统的可靠性就要低得多。识别一个已有存档影像的人，与对那些可能毫无“前科”的个人做出大规模自动化判断，二者之间的差异不是程度上的，而是性质上的。这种差异，也正是当局在应对“校准困境”时最难处理的领域。

为何即使有了更好的 AI，这一困境仍会存在

这些失灵仅仅是成长的阵痛吗？未来的政权，若拥有更强大的算力和更复杂的模型，最终能解决校准问题吗？答案是否定的。准确率的提升所受到的限制，并非来自计算能力，而是复杂社会中预测本身的性质。即便是“更好的人工智能”，也无法消除这种权衡或取舍（tradeoff）；它只能改变权衡发生的位置。有三个因素确保了这一两难困境将持续存在。

第一，异议的稀有性意味着，从整体人口层面来看，大量误报在所难免。预测性警务系统无异于在草堆里找针。当目标行为的基础比率——即潜在发生率——极低时，即便是高度精准的模型，标记出的无辜者数量也会远多于真正的威胁。这种现象有时被称为“误报悖论”（false-positive paradox）。在政治语境中，误报会带来沉重代价，表现为殃及无辜的镇压及其引发的反弹。一个系统扫描了一百万人，其中只有十人是真正的威胁，如果准确率为 99%，虽能捕获大多数真正的威胁，但也会错误地将数千名无辜者标记为“威胁风险”。当局随后必须决定，是忽略这些警报（这意味着政府在花钱支付一个它并不使用

的系统），还是追查这些警报，从而耗尽稀缺资源，同时让公众产生敌对情绪。再强的算力也无法改变这种稀有事件的算术规律。

第二，应用环境并非一成不变。公民会做出调整适应。机器学习模型是基于“未来将与过去相似”这一假设构建的。它们从历史数据中学习，但在政治领域，这些“数据”会发生反作用。这正是“古德哈特定律”（Goodhart's Law）的一种体现，该定律由英国经济学家查尔斯·古德哈特（Charles Goodhart）在半个世纪前提出。他的结论是：当一项指标（measure）成为目标时，它就不再是一项好的指标了。¹⁴就本文而言，这意味着一旦当局开始依赖某些特定信号，如关键词、行动轨迹或社交关系，人们便开始学会避免发出这些信号。他们使用暗语，将手机留在家中，或回到非数字化的交流方式（analog communication）。

政府或许仍能突破这些应对措施，但它们确实增加了监控的难度。在措施与反制的迭代对抗中，侦测与规避成为一场博弈。当局做出调整，人们便想法躲避或绕过；随后当局再次调整，人们又作出回应，如此循环往复。监控模型总是被“概念漂移”（concept drift）甩在身后：它试图预测的行为，已经因预测本身而发生了改变。

让这个困境持续存在的第三个也是最后一个因素，是专制主义的政治经济体系对输入数据（inputs）的扭曲。人工智能常被宣传为“客观真相的机器”（objective-truth machine），但在独裁体制下，数据本身就是政治资本。基层官员深知，他们必须拿出能向中央证明自己称职的数字。如果最高领导要求“稳定”，地方当局便可能压制警报

或少报事件。同样，上层对“警惕”的要求，可能会促使地方官员夸大威胁指标，以此展示忠诚和工作力度。向政府出售“黑箱”（black-box，意思是内部运作原理不透明——译者）监控技术的私营供应商，可能会进一步加剧这种扭曲，他们夸大其可靠性，淡化实际存在的不确定性边界，而这本来需要由人工判断。拥有一套自动化系统，并不能防止“垃圾进，垃圾出”（garbage in, garbage out）的问题，甚至可能使问题更加严重，因为它会让垃圾般的输出看起来像是硬性、权威的数据。看到这些数据，领导者会忍不住将“测量结果”误当成“真知灼见”，并自以为无所不知，而实际上远非如此。

几十年来困扰着专制政权的“信息认知难题”（knowledge problem），并没有被人工智能解决，只是被掩盖了。通过在策略性数据和不断变化的社会情境之上叠加概率推断，这些系统只能为本质上仍充满不确定性的现实，披上一层精确性的外衣。打造一个“完美”警察国家的梦想，不仅在政治上代价高昂，在数学上也根本无法实现。

虚张声势的“全景监狱”

如果这些制约确实存在，为什么人工智能看起来依然如此强大？诚然，人工智能是控制民众的一个有力工具，但这并不妨碍我们指出，出于感知和政治上的双重原因，它的能力被夸大了。首先，重要的是区分人工智能“实际能做什么”和“人们相信它能做什么”。十八世纪英国哲学家杰里米·边沁（Jeremy Bentham）曾提出“全景监狱”（panopticon）的构想，在这种圆形建筑中，一名看守可以随时观察任何囚犯，这个构想提示我们，持续的监视本身并不那么重要，重要

的是“人们相信自己随时可能被监视”这一信念。

当政府和科技公司将人工智能描绘成无所不知——拥有远超人类能力的“智能”——它们助长了一种寒蝉效应：公民们的自我审查和服从程度，远超出政权的实际监控能力所覆盖的范围，这并非因为他们正在被监视，而是因为他们认为自己可能被监视。技术乐观主义者承诺，人工智能将带来增长和互联互通的“数字红利”（digital dividend）。且不论这一承诺是否属实，但不可否认的是，专制政权已经从人工智能中赚到了一种“独裁者红利”（dictator's dividend）——即由恐惧驱动的服从，其程度远超这些政权实际监控民众的能力。“算法”所带来的模糊而无处不在的威胁，促使公民进行自我审查，甚至无需执行一行代码，就为政权完成了它要做的工作。

但这种红利是以隐性债务为代价换来的。全景监视的威胁所产生的稳定是脆弱的，因为它依赖于经济学家蒂穆尔·库兰（Timur Kuran）所说的“偏好伪装”（preference falsification）。¹⁵ 公民表面假装忠诚，私下却心怀不满，由此形成的对服从的评估，能够使统治者误判真实的不满程度。政权越是宣扬其无所不知，一旦这种无所不知被揭穿为片面、可被操纵或彻头彻尾的虚假，其付出的政治代价就越高。

这里同样存在一个动态的目标。随着异议人士和当局不断调整策略，对民众的控制也会随之变化。起初，政权往往占据上风。先进的监控系统成本高昂，需要高度的集中和协调，而这正是政府的强项。但随着时间推移，公民会了解这些系统的运作方式，制定应对措施，并更广泛地获得属于自己的数字工具。

这正是“全景监狱”的虚张声势可能崩塌之处。当公民目睹系统失灵——无论是河南“红色健康码”的随意性，还是德黑兰基于摄像头执法所显露的局限——这种神秘感便烟消云散。这个“全知之眼”（all-seeing eye）被揭穿为一台易犯错的机器，它可能被口罩蒙骗，也可能被腐败地方官员所操纵。关于“算法厌恶”（algorithm aversion）的研究表明，一旦人们观察到算法犯错，就可能不那么愿意听从它，从而加速这种虚张声势的瓦解。¹⁶一旦公众普遍认识到，国家监控系统可能会因“噪声”挤占“信号”而陷入瘫痪，且其覆盖范围存在漏洞，恐惧便会迅速转化为愤怒。到那时，政权就只剩下一个在技术上不堪重负、在心理上丧失威慑力的监控机器。

民主派人士能做些什么

人工智能确实会重塑镇压方式，但它不会造就一个天衣无缝、无所不见的警察国家。这项技术改变了政权与其对手较量的场域。专制政府或许会在某些情境下占得优势，但亲民主力量同样可以调整并予以反击。以下三种策略值得关注。

第一，要公开并反复地对人工智能进行“祛魅”（demystify）。人工智能的心理威慑力往往超过其技术能力。清晰、易懂地解释这些系统能做什么、不能做什么，可以戳破政权刻意营造的“全景监狱”式氛围，削弱“算法无所不知”带来的寒蝉效应（chilling effects）。同样重要的是，民主国家和公民社会组织应当记录并公开专制政权所用人工智能系统的种种失灵：中国河南省任意标记红色健康码、错误的身份识别，以及其他可见的故障。揭穿“帷幕背后的巫师”（wizard

behind the curtain)，揭示这台机器往往只是在猜测、存在偏见，或者干脆就是坏掉的，这能够减少恐惧，鼓励人们采取切实可行的应对措施。这类有针对性的叙事，通过证明 AI 赋能的控制既非天衣无缝、也非无懈可击，能鼓舞民主人士的士气。

专制政权从人工智能中获得了一种“独裁者红利”——即由恐惧驱动的服从，远超出政权监控民众的实际能力

第二，建立民主标准和防御能力。民主国家以及与之结盟的公民社会组织，应在政治敏感领域投入使用人工智能时，对许可的数据获取、透明度和问责制，推动制定更明确的

规范，并在可行的情况下辅以监测、审计和检查。他们还应划出“红线”，以界定何时构成对人工智能的滥用。其中包括：缺乏保障措施的大规模生物识别、政治黑名单，以及旨在通过法律和制度渠道提高违规行为声誉成本的强制性评分或“社会信用”系统。若能诉诸法院，战略性诉讼或许会有所帮助（战略性诉讼是有意识地选择和设计的诉讼案件，以推动更广泛的法律、政策或制度变化——译者）。这方面的裁决可以对人脸识别和自动身份识别施加严格的保障措施，并通过将其重新定性为任意迫害，从而削弱算法治理的“科学”外衣。

鼓励技术民主化的举措，同样能够带来实质性的收益。这包括资助相关研究，探索如何使人工智能和其他安全的公民技术实现民主化，例如端到端加密、规避审查的工具（circumvention tools）和隐私保护协议。在太多地方，压迫性政权垄断了先进技术能力。这种局面应当改变。探索如何利用包括人工智能在内的先进技术，来促进而非压制民主，有助于公民社会对抗压迫、增强组织能力，并发挥这些工具所

蕴含的潜力。¹⁷

最后，AI 赋能的镇压行为应当被纳入现有人权执法的范围。现有的国际人权机制，往往依赖“可观察到的侵犯行为”（observable violations）来触发国际压力。然而，人工智能如今却被用于“预防性”镇压——在潜在挑战者尚未发展到需要对其采取明显镇压措施的程度之前，就对其进行识别、威慑或使其丧失行动能力。人工智能这种“事前”（before the fact）的特性，给传统的人权机制带来了挑战。权利侵犯和镇压的界定方式亟需重新审视，人工智能在压制异见中所扮演的角色，也需要进行主动审查。国际社会需要进行持续的、制度化的努力，来监测并报告 AI 赋能的胁迫行为。由数字监控和画像所助长的侵犯行为，可以被“点名羞辱”（named and shamed），并施加相应的制裁。关注先进微芯片的出口流向，也是一个好主意。

高度现代主义的幻梦

凭借其强大的能力，人工智能监控向独裁者们许下了高度现代主义的终极幻梦：一个可以像模拟程序一样被建模、被预测、被管理的社会。人工智能驱动的数字专制主义前景，在全球范围引发了民主人士的警觉。然而，正如中国、俄罗斯和伊朗的经验所表明的，这种对全人口规模进行预测性控制的梦想，在某种程度上不过是妄想。通过用僵化的数字阈值取代人类的判断，这些政权用灵活性换取了脆弱性。它们或许在识别个体目标方面变得更加得心应手，却在社会大变局爆发之前丧失了感知其动向的能力。

AI 赋能的镇压，其局限性既源于预测本身的统计特性，也源于经典的“独裁者困境”（dictator's dilemma）。由于模型具有概率性，而真正的威胁是稀少的、具有适应性，且往往被策略性地隐匿起来，因此并不存在用于预防性镇压的“完美”算法。不可避免的误差，尤其是误报，会带来政治代价，表现为资源的错配、民怨的累积，以及合法性的流失。

与此同时，通过强化官僚们钻制度空子（game the system）的动机，人工智能还可能会加剧“独裁者困境”。当地方官员因“不稳定”而受到惩罚时，他们就有强烈的动机在信息上报中央之前压制坏消息、夸大业绩、操控指标。于是，即使混乱加剧，中央领导层收到的指标却仍显示一切井然有序。一个政权越是依赖量化的风险评分来统治，就越接近于使用一张与实际情况不再相符的地图。

这正是自动化警察国家的核心脆弱性。它虽然能营造出精确的表象，却会削弱统治者需要的真实反馈。民主国家看似混乱——充满争论、抗议和显而易见的摩擦——但这种混乱往往能发挥信息的作用。它能暴露错误、反映不满，并开辟纠正的渠道。相比之下，专制人工智能

对于民主的捍卫者而言，他们的任务不是因 AI 的能力而陷入停顿，而是要理解这些工具在具体情境中是如何运作的，并基于这种理解制定有效的应对措施

则致力于压制噪音，转而追求那些可以被过滤、修饰乃至误读的校准信号。其结果看似井井有条，却使统治者陷入危险的隔绝状态，并容易过度自信。

尽管在 AI 赋能的镇压之下，民主抵

抗仍存在空间，但要把人工智能固有的弱点转化为切实的民主成果，依然是一项重大挑战。一个核心障碍在于能力失衡——在压制性环境中，政权凭借其资源优势，往往对异见人士和公民活动者占据上风。就目前而言，最佳的策略应对措施仍是上文所述的那些：对人工智能的能力范围进行祛魅，推动制定可执行的关于 AI 政治应用的标准和规范，投资于公民（行动）安全技术，以及加强对 AI 赋能的胁迫行为的跨国监督。

这算是个不大的好消息。要了解坏消息，不妨看看以下三个令人担忧的趋势，它们可能会让 AI 赋能的专制主义更难以遏制。第一，人工智能的开发和投入使用，既未形成政治共识，也缺乏健全的治理和法律标准。作为人工智能领先的两大强国，中国和美国主要从相互竞争的国家安全视角来对待这项技术。两国均未将建立一个限制人工智能用于胁迫或镇压的全球框架作为优先事项。在欧洲，2024 年的《人工智能法案》（欧盟通过的一项在所有成员国均具有法律效力的法规），是一项值得注意的举措，旨在全面规范人工智能的开发和应用。即便如此，要实现有意义的约束，仍需要政府、企业与公民社会之间更广泛的跨国合作与共识。

第二，人工智能开发的领域依然是一个“小圈子”，大量权力集中在少数主体手中。在硬件方面，仅有少数几家公司——其中最突出的是英伟达（NVIDIA）——主导着先进芯片及其配套软件栈的生产。在模型方面，同样只有极少数公司生产并掌控着尖端的人工智能技术。只要这种高度集中的局面依然存在，推动人工智能民主化的努力就将举步维艰。例如，亲民主的公民社会组织，若想缩小独裁者目前享有的

技术优势，就需要获得更多资源。

第三，当前人工智能领域中“开源与闭源”的格局，给民主的支持者带来了一个现实的困境。许多广泛可用的开源模型，是由位于专制体制国家的企业开发的，而美国企业往往通过闭源平台提供性能最强的系统。这并不是说中国公司的开源模型——例如 DeepSeek 或通义千问（Qwen）——不能被用于民主目的。但当研发过程不透明——尤其是涉及所使用的训练数据、微调方式和内置的防护机制——审计这些系统中的偏见、隐藏限制或其他漏洞就会变得十分困难。而由科技巨头掌控的闭源模型，则会带来不同的风险。建立在这些平台之上的工具，容易受到价格或访问权限突然变化的影响。此外，迫使活动人士将数据发送至远程的企业服务器，也会让这些在压制性政权监视之下开展活动的人士陷入危险。

步入人工智能时代，既不意味着一场灾难，也不能保证民主的复兴。正如以往历次技术突破一样，人工智能在镇压方面的运用，将被既有的政治制度和权力关系所塑造，即使它也会带来新的挑战 and 机遇。对于民主的捍卫者而言，他们的任务不是因 AI 的能力而陷入停顿，而是要理解这些工具在具体情境中是如何运作的，并基于这种理解制定有效的应对措施，以保护公民空间、减少恐惧，并保持民主抗争的可能性始终开放。

注释

- 1 罗吉尔·克雷默斯、格雷厄姆韦伯斯特、海伦托纳:《译文〈互联网信息服务算法推荐管理规定〉——自2022年3月1日起施行》[Rogier Creemers, Graham Webster, and Helen Toner, “Translation: Internet Information Service Algorithmic Recommendation Management Provisions—Effective March 1, 2022,” *DigiChina* (Stanford University), 10 January 2022, <https://digichina.stanford.edu/work/translation-internet-information-service-algorithmic-recommendation-management-provisions-effective-march-1-2022>.]
- 2 乔希·叶与布伦达·吴:《中国监管机构称阿里巴巴、腾讯已提交应用算法详情》(Josh Ye and Brenda Goh, “China Regulator Says Alibaba, Tencent Have Submitted App Algorithm Details,” *Reuters*, 15 August 2022, www.reuters.com/technology/china-regulator-says-alibaba-tencent-have-submitted-app-algorithm-details-2022-08-12; “Provisions on the Administration of Deep Synthesis Internet Information Services,” *China Law Translate*, 11 December 2022, www.chinalawtranslate.com/en/deep-synthesis.)
- 3 王裕华、明克胜:《中国维稳体系的崛起》[Yuhua Wang and Carl Minzner, “The Rise of the Chinese Security State,” *China Quarterly* 222 (June 2015): 339–59.]
- 4 伯纳德·奥尔:《中国严格的“清零”政策开始松动》(Bernard Orr, “China’s Rigid Zero-COVID-19 Policy Starts to Thaw,” *Reuters*, 7 December 2022, www.reuters.com/world/china/chinas-rigid-zero-covid-19-policy-starts-thaw-2022-12-07.)
- 5 吴瑞恩:《在防疫政策转向之际, 中国对民众的宣传口径从“战争”转为“自我防护”》(Ryan Woo, “In COVID U-Turn, China’s Message to the People Shifts from War to Self-Care,” *Reuters*, 14 December 2022, www.reuters.com/world/china/covid-u-turn-chinas-message-people-shifts-war-self-care-2022-12-14.)
- 6 裴敏欣:《清零政策的骤然终结: 一项调查》(Minxin Pei, “The Sudden End of Zero-Covid: An Investigation,” *China Leadership Monitor* 75 (Spring 2023), www.prcleader.org/post/the-sudden-end-of-zero-covid-an-investigation.)
- 7 凯文·唐纳利:《阿道夫·凯特勒、社会物理学与“平均人”的科学, 1796—1874》[Kevin Donnelly, *Adolphe Quetelet, Social Physics and the Average Men of Science, 1796–1874* (London: Routledge, 2015).]
- 8 詹姆斯·C. 斯科特:《国家的视角: 那些试图改善人类状况的项目是如何失败的》[James C. Scott, *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed* (New Haven: Yale University

Press, 2020).]

- 9 陆柯言：《健康码背后隐忧：用户因中国移动错误定位被隔离》
(Lu Keyan, “Hidden Worries Behind the Health Code: Users Quarantined Due to China Mobile Positioning Errors” [in Chinese], *Jiemian News*, 12 March 2020, <https://finance.sina.cn/2020-03-13/detail-iimxyqvz9892557.d.html?vt=4>.)
- 10 文森特·倪：《中国因银行账户被冻结引发的抗议演变为暴力冲突》
(Vincent Ni, “Protest in China over Frozen Bank Accounts Ends in Violence,” *Guardian*, 11 July 2022, www.theguardian.com/world/2022/jul/11/china-violent-clashes-at-protest-over-frozen-rural-bank-accounts.)
- 11 《莫斯科活动人士以面部彩绘抗议大范围人脸识别》 (“Moscow Activists Protest Widespread Facial Recognition with Face Paint,” *Moscow Times*, 7 February 2020, www.themoscowtimes.com/2020/02/07/moscow-activists-protest-widespread-facial-recognition-with-face-paint-a69205.)
- 12 约马尔特·约尔多舍夫：《格鲁欣诉俄罗斯案：欧洲人权法院迈入 AI 监控时代的第一步》 (Jomart Joldoshev, “Glukhin v. Russia: The European Court of Human Rights’ First Step into the Age of AI Surveillance,” *Federal Bar Association*, 23 September 2025, www.fedbar.org/blog/glukhin-v-russia-the-european-court-of-human-rights-first-step-into-the-age-of-ai-surveillance.)
- 13 艾米丽·布劳特：《马赫萨·阿米尼去世三年后，伊朗对女性发动的高科技战争遭到抵抗》 (Emily Blout, “Resisting Iran’s High-Tech War on Women Three Years After Mahsa Amini’s Death,” *Stimson Center*, 16 September 2025, www.stimson.org/2025/resisting-irans-high-tech-war-on-women-mahsa-amini.)
- 14 古德哈特对该定律的正式表述如下：“每当政府试图依赖此前观察到的某种统计规律来实施控制时，这种规律性便会崩溃。”而该定律更为精炼、也更为著名的表述——“当一项指标变成目标时，它就不再是一个好指标”——则出自玛丽莲·斯特拉森。[Charles A.E. Goodhart, “Problems of Monetary Management: The U.K. Experience,” *Papers in Monetary Economics*, vol. 1 (Sydney: Reserve Bank of Australia, 1975). Marilyn Strathern, “‘Improving Ratings’: Audit in the British University System,” *European Review* 5 (October 1997): 308, <https://archive.org/details/ImprovingRatingsAuditInTheBritishUniversitySystem/mode/2up>.]
- 15 蒂穆尔·库兰：《私下的真相，公开的谎言：偏好伪装的社会后果》。
[Timur Kuran, *Private Truths, Public Lies: The Social Consequences of Preference Falsification* (Cambridge: Harvard University Press, 1995).]
- 16 伯克利·J. 迪特沃斯特、约瑟夫·P. 西蒙斯、凯德·梅西：《算法厌恶：人们在目睹算法出错后，会错误地回避使用算法》 [Berkeley J.

Dietvorst, Joseph P. Simmons, and Cade Massey, “Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err,” *Journal of Experimental Psychology: General* 144 (February 2015): 114.]

- 17 埃莉卡·切诺维斯编:《人工智能如何支持民主运动: 一项研究与实践工作坊的总结报告》[Erica Chenoweth, ed., “How AI Can Support Democracy Movements: Summary Report of a Research and Practice Workshop,” *Ash Center for Democratic Governance and Innovation*, Harvard Kennedy School, 2025.]

埃莉卡·切诺维斯
马修·塞布尔 文

言晓义 译

本文作者埃莉卡·切诺维斯 (Erica Chenoweth) 是哈佛大学肯尼迪政府学院“法兰克·斯塔顿”(Frank Stanton) 第一修正案讲席教授, 同时也是哈佛大学拉德克利夫研究所 (Radcliffe Institute) “苏珊与肯尼斯·瓦拉赫”(Susan S. and Kenneth L. Wallach) 讲席教授。其最新著作是《论革命: 当今世界中脱缰的政治》(On Revolutions: Unruly Politics in the Contemporary World, 2022)

马修·塞布尔 (Matthew Cebul) 是哈佛肯尼迪学院阿什民主治理与创新中心“非暴力行动实验室”领衔研究员, 主要从事当代非暴力抗争运动动态的应用研究

译者言晓义毕业于北京大学, 曾获台湾梁实秋文学翻译奖

中国民主季刊

第4卷 第4期

2026年10月

CC-BY-NC-ND

自由、免费转载分发, 但须注明
作者与出处, 并不得修改及商用

Z世代为何崛起

编按：本文分析了全球范围内 Z 世代主导的社会动员浪潮，认为青年抗议运动的主要驱动力包括经济不安全、被排斥在权力体系之外，以及突出的腐败问题。青年广泛参与的运动可以形成强大的动员力量，但同时也面临严厉的镇压，而且往往难以在短期内解决深层次的结构性质诉求。这些运动通常采取分散、没有统一领导的组织方式，这有助于运动迅速扩散，但也使谈判以及政治转型更加复杂。本文原载于美国《民主季刊》2026 年第 1 期（*Journal of Democracy*, Volume 37, Number 1, January 2026, pp. 5-14）。

调查数据中有一个常见的发现，就是 Z 世代（Gen Z，通常指年龄在 13 岁至 28 岁之间的年轻人）正日益与政治疏离，并对民主制度感到幻灭。¹ 然而，青年动员一直是近年来世界不同地方推动政府变革、对政府治理失效问责的重要力量。仅在今年，从印度尼西亚到肯尼亚，从马达加斯加到秘鲁，一股“Z 世代抗争”的浪潮席卷全球。这类由青年主导的抗争，其结果有成有败：有的推翻了政府，有的迫使政府进行改革，也有的引发了暴力冲突和镇压。在塞尔维亚等地，青年主导的抗争至今仍在持续。

当今 Z 世代的抗争运动，其实延续了一种更为久远的历史趋势，就是社会抗争运动往往由年轻人推动。一项研究指出，在 1990 年至 2020 年间，以推翻在任国家领导人或实现民族自决为目标的非暴力运动中，高达 80% 均有大规模的青年参与。据估算，30 岁以下的年轻人至少占一线参与者的四分之一。² 2001 年推翻约瑟夫·埃斯特拉达（Joseph Estrada）的菲律宾“第二次人民力量运动”、2005 年将叙利亚军队驱逐出境的黎巴嫩“雪松革命”（the Cedar Revolution），

以及 2019 年促成奥马尔·巴希尔（Omar al-Bashir）垮台的苏丹起义，无一不是青年人积极参与一线行动、并在很大程度上由青年力量推动的典型例证。青年对既有治理模式的幻灭感看来是真切的，青年动员在挑战根深蒂固的权力结构上所展现的潜力也同样明显。

表——青年参与与非暴力运动成败，1990-2020

青年前线参与程度	无	有限	中等	广泛	合计
失败	3 (50%)	18 (60%)	43 (68%)	44 (44%)	108 (54%)
成功	3 (50%)	12 (40%)	20 (32%)	56 (56%)	91 (46%)
合计	6	30	63	100	199

资料来源：埃莉卡·切诺维斯等，《青年与 LGBTQ+ 群体在非暴力行动中的参与》（美国国际开发署，2023）。

注：观察单位为目标最大化类型非暴力运动。卡方检验 $\chi^2(3) = 9.655$, $Pr = 0.022$ 。

当今，Z 世代正在崛起。究竟是什么在驱动 Z 世代的抗争？这些运动与过往的抗争浪潮相比又有何不同？年轻的行动者们能否抵得住威权主义的逆流、催生新一轮的民主变革，抑或终将受到阻挠而难以有所作为？这些运动到底算不算民主运动？日益高涨的青年动员，又为未来的岁月预示了怎样的图景？本文将当前的 Z 世代抗争置于比较的视角之下，就青年主导的大规模动员所蕴含的希望与风险进行简要的观察。若历史可作指引，那么扭转全球民主倒退的努力，极有可能与世界各地青年运动的命运息息相关。

贪腐激起全球民怨

驱使年轻人走上街头的诱因不一而足。青年群体在经济动荡时尤为易受损害；经济停滞不前使年轻人长期陷入失业的泥淖。即使是受过良好教育的大学毕业生，也难以实现向上流动的梦想。2010 年的突尼斯就是如此：经济低迷、血腥镇压，以及政府滥权枉法等多重打击交织，令挫败不已的年轻人因小摊贩穆罕默德·布瓦吉吉（Mohamed Bouazizi）的自焚而群情激愤，最终引发了“阿拉伯之春”。³ 此外，众所周知，青年在正式的建制政治中缺乏代表性，老人政治体系（gerontocratic political systems）将年轻一代排斥在掌控其命运的权力殿堂之外。⁴

Z 世代的抗争运动正是这些趋势的集中体现。许多爆发 Z 世代抗争的国家都存在经济持续疲软、青年失业率高、年轻人口不断膨胀的问题，而当权者却是一群年龄远高于国民平均线的政坛老朽。例如，尼泊尔的青年失业率高达约 20%，摩洛哥更是达到了惊人的 35%，两国经济都曾遭受新冠疫情冲击而陷入困境。⁵

然而，Z 世代抗争真正的集结号似乎是“反腐败”：年轻人对政治精英感到怒不可遏——这些精英在国家深陷苦难之际，竟毫无廉耻地收割着奢华特权带来的红利。在尼泊尔，网络上疯传的“官商二代”（nepo kids）炫富照片激起了民愤。在印度尼西亚，议员每月领取的住房津贴竟是最低工资的数倍之多，这一消息曝光后激起了抗议。2024 年在塞尔维亚，造成十多人死亡的诺维萨德火车站坍塌事件发生

后，年轻人纷纷走上街头，抗议者将这场悲剧归咎于政府腐败和玩忽职守。在秘鲁，针对长期存在的敲诈勒索和暴力犯罪现象，抗议者要求政府执法申冤，而这类犯罪之所以猖獗，正是因为政府与有组织犯罪明目张胆地勾结。在菲律宾，政府 7 月发布的一份报告披露，原本用于防洪救灾项目的资金中，有超过 20 亿美元因腐败而流失，随后当地爆发了抗议。包括上述这些案例在内的许多迹象都表明，Z 世代正在反抗那些贪婪攫取、将自身特权置于公共利益之上的精英。

2025 年的反腐浪潮首先在亚洲蔓延开来：印度尼西亚、菲律宾和尼泊尔的抗议者，他们都从此前斯里兰卡（2022 年）和孟加拉国（2024 年）的起义中获得了启发。⁶ 反腐败运动在全球范围内引燃。从马达加斯加到秘鲁，全世界的 Z 世代抗争者纷纷以日本热门漫画《航海王》（*One Piece*）中的草帽海贼团旗帜为创意元素，加以变体，将其作为普通人摆脱腐败统治者、实现解放的象征。反腐败是一个能引发广泛共鸣的普世主题。而通过互联网与社交媒体日益加深的全球互联互通，很可能进一步促进了抗争跨越地域界限迅速扩散。

青年运动常胜，但有代价

过往的研究发现，青年对政治抗争的参与，与目标最大化类型（maximalist）非暴力运动的成功之间存在着强烈的正相关——所谓目标最大化类型非暴力运动，是指那些“旨在推翻现任国家领导层或实现领土独立”的运动。⁷ 事实上，大多数拥有大规模青年参与的抗争最终都取得了成功，而没有大规模青年参与的案例则大多以失败告终（见表）。一项最新研究的结果显示，若一场非暴力运动拥有大规模

的青年参与（即所观察到的一线参与者中有超过半数年轻人），那么其成功的几率会是青年参与度低的运动（青年占比低于 10%）的两倍以上。⁸

我们还观察到，青年参与程度与运动结束后的平等主义、自由主义民主程度之间存在正向关联。西里安妮·达鲁姆（Sirianne Dahlum）与亚当·波尼卡（Adam Bonica）特别强调指出，过去 70 年来，从韩国、台湾到东德，再到巴西，学生抗争在各国民主运动中一直发挥着至关重要的作用。⁹

现代社会运动往往将去中心化奉为核心优势，因为年轻行动者倾向于摆脱僵化的层级领导体制，转而通过数字化组织建立广泛而松散的横向网络联接

为什么青年参与度高的运动更容易获胜？学者们认为，年轻人是投入度极高的行动者，能够运用新颖、富有创意的抗争战术，而且更有能力跨越固化的社会文化藩篱、建立起广泛的联盟。¹⁰ 这种态势的例证不胜枚举：2013—2014

年“广场革命”（Euromaidan）期间，乌克兰学生在严寒彻骨的冬日里占领了基辅独立广场；2019 年“十月革命”期间，黎巴嫩青年跨越根深蒂固的宗派界限发起动员；2019 年香港抗争期间，青年以独具匠心的智慧不断创新抗议方式，以规避中国当局的监控网络。

然而，年轻人同样面临着严峻的挑战。最关键的一点在于，青年主导的运动往往与更为严厉的镇压呈正相关，尽管这些运动本身并不比其他运动更倾向于使用暴力。¹¹ 这一发现与拉格希尔德·诺达斯（Ragnhild Nordås）、克里斯蒂安·达文波特（Christian Davenport）

以及阿亚尔·费恩伯格（Ayal Feinberg）、伊德安·萨勒扬（Idean Salehyan）的研究遥相呼应——他们指出，威权政权及其安全部队将青年视为尤其具有威胁性的力量，因而倾向于对他们实施极其残暴的镇压。¹²

遗憾的是，当下 Z 世代的抗争完完全全印证了这一模式，其中许多运动均遭遇了致命的流血镇压。2024 年 7 月在孟加拉国，政府发动的血腥屠杀导致数百人遇害——然而，这场镇压很快引发了反弹（backfired），沸腾的民怨最终迫使总理谢赫·哈西娜（Sheikh Hasina）于 8 月仓皇流亡海外。在肯尼亚，抗议 2024 年《财政法案》的示威导致数十人丧生，次年又有数十人遇害，许多年轻行动者遭到安全部队绑架。当局企图借此压制异议，却最终未能奏效。摩洛哥安全部队残酷镇压了“Z 世代 212 运动”（Gen-Z 212 movement），逮捕了数千人并判处重刑。就在 2025 年底，坦桑尼亚大选后，当局对青年抗争进行了严酷镇压，酿成了逾千人死亡的惨剧。¹³

此外，即使青年主导的抗争成功推翻了独裁政权，年轻人的物质生活状况也未必因此得到改善。例如，一项研究揭示，青年参与抗争与随后几年青年失业率的改善之间，并不存在关联。¹⁴ 尽管青年抗争与政治自由化呈正相关，但更换国家元首往往远比疏解最初引发抗争的深层次社会经济积怨要容易得多。当下的 Z 世代抗争，或许正走向与这一现实猛烈相撞的轨道——如果新政府最终不能兑现承诺、改善民生，那么在未来的岁月中，他们恐怕仍将面临一波又一波的社会抗争动员。

Z 世代是为了民主吗？

近期的多项研究表明，一个积极动员起来的公民社会，是抵御全球民主持续倒退趋势的重要力量。¹⁵ 据此，熟悉相关研究的读者或许会推断，Z 世代的抗争是民众抵抗威权主义的一个积极信号，是正在崛起的、抵御全球非自由主义逆流的坚固堡垒。

不过，在预测 Z 世代抗争的民主化效应时，我们既有理由保持乐观，也有理由保持谨慎。民主的深化与巩固，固然是 Z 世代大规模动员可能带来的成果之一，但这场 Z 世代抗争中至少有四个层面的特质值得注意，它们应当促使我们冷静下来，避免对民主化前景抱有过度理想化的期待。

非暴力纪律

首先，近期的 Z 世代抗争能否完全秉持或恪守非暴力纪律，目前仍未可知。大量学术研究都指出，相较于诉诸暴力手段，非暴力抗争更能有力地促进民主。非暴力对于公众来说更具道德感召力与正当性，从而更能显著提升民众的参与度。以非暴力方式建立联盟，本身也体现了多元主义与宽容等民主核心价值，使亲民主网络能够跨越不同的重要社会界别进一步扩大，并且与持久的民主化进程相关。¹⁶ 相比之下，暴力抗争则往往会削弱大众对运动的认同感，排斥潜在的参与者，并进一步强化社会对当局压制运动、打击异见行为的支持。¹⁷

需要说明的是，许多 Z 世代行动者在开展动员之初，都明确以和平抗

争为号召。例如，最初参与组织抗议、与“Z世代212”并肩行动的团体“摩洛哥青年之声”（Morocco Youth Voice），就曾发布一份宣示支持非暴力的行为准则，明确反对暴力与破坏公物行为。¹⁸在马达加斯加，抗争最初也是和平进行的，但随着警方开始使用催泪瓦斯与橡胶子弹等更为强硬的清场手段，游行示威陷入了混乱。

无论运动组织者的初衷多么美好，以非暴力开局的抗争未必总能贯彻始终。在尼泊尔，一场声势浩大的非暴力游行迅速被暴力骚乱者裹挟。第二天，一群人纵火焚烧议会大厦以及一些知名政治人物的住宅，造成数十人死亡。类似的场景也在摩洛哥、肯尼亚和印度尼西亚出现——抗争者与安全部队爆发激烈冲突，纵火者则将政府大楼与警局付之一炬。许多抗争者实施的暴力显然是由国家的暴力镇压所激起的。然而，无论最初的导火索是什么，一旦与运动站在同一阵营的行动者开始对手使用暴力，局势就可能迅速失去组织者的控制，使和平行动者被边缘化；与此同时，运动要么被更为激进的力量“劫持”，要么被反革命势力或政治温和派所孤立——后者或许支持变革，却绝不愿以暴乱为代价。¹⁹

在短期内，暴力抗争或许仍能迫使当局让步，正如在尼泊尔与马达加斯加所见的那样；但从长远来看，此类手段通常不利于建立持久而稳定的民主制度。²⁰

去中心化的运动

Z世代运动的另一个显著特征，是其“无人领导”（leaderless）的

组织结构。现代社会运动往往将去中心化奉为核心优势，因为年轻行动者倾向于摆脱僵化的层级领导体制，转通过数字化组织建立广泛而松散的横向网络联接。²¹这使得现代抗争运动能够更迅速地动员起来，更及时地回应自发形成的基层民意，也更难以被彻底镇压。近年来的一些运动明确采取了这种组织方式，例如 2019 年香港反修例运动中的行动者们即是如此。

然而，即便是分散、没有统一领导（leader-full）的运动，也并非完全能够免受镇压，香港起义中许多核心组织者的亲身经历便说明了这一点。而且，这类运动虽然在适应性、灵活性方面可以有所得，却会牺牲联合的凝聚力与谈判的力量。缺乏一个集中的权威机构，运动可能难以确立并执行明确的行为规范，难以及时遏制暴力冲突，以免局势失控，也难以在谈判桌上作出具有可信度的承诺。鉴于通过谈判实现的民主转型往往最为持久，在这类谈判中拥有较为明确组织代表的公民社会团体，更有可能参与塑造国家未来的走向。²²

算是“民主”运动吗？

第三，Z 世代的抗争运动的发起未必是为了民主本身（qua democracy）。诚然，某些 Z 世代运动明确以民主变革为号召。孟加拉国在 2024 年起义之后，正在进行大规模宪法修订。在秘鲁，一些知名行动者严厉抨击政府，称其为“披着民主外衣的独裁体制”，将去除腐败与建立民主制度视为休戚相关、不可分割的整体。²³在尼泊尔，Z 世代行动者以抗议为契机推动政党改革，希望让年长的政治精英退出政治舞台，提高青年人在议会中的代表性。

尽管如此，Z 世代抗争的核心驱动力，本质上仍源于社会经济不平等、贪腐以及用人唯亲等问题——这些诉求固然与“良好治理”（good governance）息息相关，却并非以民主本身为中心。例如，在摩洛哥，Z 世代抗争者要求的是政府加大投资、改善已经破败不堪的医疗和教育服务，而不是要求废除君主制。在马达加斯加，Z 世代抗议最初是针对总统安德里·拉乔利纳（Andry Rajoelina）而展开的，最终却以军事政变告终。如果新的军方领导政府能够解决供水和停电问题，抗议者是否还会抵制这一军事政权，这一点尚难以确定。²⁴

简言之，尽管一些 Z 世代的抗议运动可能会以民主为主题，但民主本身未必是这些运动的首要诉求或根本动因（raison d'être）。在大众叙事中，“民主运动”往往是一个笼统的说法，泛指反抗腐败政府的抗争，但观察者们应当保持谨慎，切勿在抗争者明确表达的不满之外，将特定的意识形态立场强加于 Z 世代运动上。²⁵

转型期的引航

最后，政权转型期充满了不确定性。转型过程涉及权力斗争和复杂的政治互动，即使在形势喜人的情况下，前景也难以预测。反动势力更是常常会打断民主改革进程，使已经取得的进展出现倒退。

尽管一些 Z 世代抗议已经迫使一些国家的领导人下台，但这些国家未来的走向仍未确定。孟加拉国在总理哈西娜逃往印度后陷入宪政危机，不过，临时政府目前似乎正朝着明年举行大选以及就宪法修订举行全国公投的方向推进（2026 年 2 月孟加拉国举行了议会选举和

宪法改革公投——译者）。在尼泊尔，由 Z 世代行动者通过 Discord 平台协调选出的临时政府，正面临着巨大的压力，必须及时恢复秩序，以确保能在 2026 年 3 月举行安全、自由和公正的大选（尼泊尔于 2026 年 3 月 5 日举行了联邦议会众议院大选——译者）。²⁶ 在马达加斯加，新上台的军政府可能会巩固权力，但也可能再度面临新一轮的群众抗争。而在秘鲁，被废黜的总统迪娜·博鲁阿尔特（Dina Boluarte）的继任者何塞·赫里（José Jerí）虽承诺将严厉打击犯罪，但 Z 世代抗争者认为，整个秘鲁政治阶层都腐败、不值得信赖，要求赫里辞职。

公民抗争领域的研究表明，在政治转型期内持续进行非暴力动员，是成功实现民主化的关键所在。²⁷ 若民主运动过早停止动员，便会为根深蒂固的精英阶层打开方便之门，使其得以破坏转型议程，并将青年行动者边缘化。沿着这一思路，Z 世代行动者可以从塞尔维亚的同龄人那里获得启发——后者已经持续动员抗议一年多，以持续不断的抗争发出一个明确而坚定的信号：在问责诉求得到充分回应之前，他们不会保持沉默。

与此相关的是，民主转型能否成功，也取决于民主运动的内部构成。例如，西里安妮·达鲁姆的研究指出，将不同社会阶层的人们凝聚在一起、具备高度社会多样性的非暴力运动，不仅与民主突破呈正相关，亦与可持续的民主化息息相关。²⁸ 因此，Z 世代能否成功应对政权转型期的挑战，很大程度上将取决于其能否在包括青年、女性及少数民族在内的不同社会群体之间，建立起实质且持久的紧密联系。²⁹

Z 世代崛起

现有研究所揭示的核心脉络,正如一句广为流传的话所说:非暴力行动“是持续的运动,而非转瞬的时刻”(movements, not moments)。运动的成功,离不开战略布局、严明的纪律,以及长期坚持的坚韧不拔。在全球政治体制屡屡辜负年轻人期望之际,人数庞大的 Z 世代已经厌倦了置身事外、袖手旁观,这一点已经十分明显。尤为引人注目的是,在全球民主倒退、政府面对大众运动通常占据优势的暗夜里,诸多 Z 世代运动竟成功撼动了既有体制,造成了巨大的政治震荡。³⁰然而,这些年轻的抗争者能否真正展现并长期保持成功的非暴力运动所具备的核心特质,仍有待观察。

展望未来,一个关键问题在于:Z 世代能否将街头抗争所积聚的势头,导入到正式的制度性政治参与之中。至少在民主制度下,选举政治是连接民主运动与持久政治变革的关键环节:民主运动正是通过选举政治,将以扰乱性行动所获得的短期政治杠杆,转化为能够长期延续的政治变革。然而,许多行动者对政党乃至政治本身都抱有深刻的怀疑。这完全可以理解——当今世界的年轻人早已对那些看似腐败到无可救药、又蓄意将他们排除在权力之外的既有政党与政治机构感到彻底幻灭。

就这一点而言,政治人物和政党在如何将 Z 世代的能量转化为积极的政治力量方面,扮演着重要角色。那些真诚地与青年展开互动、赢得 Z 世代支持的政府或政治领导人,有可能获得巨大的政治收益。民主社会主义者佐赫兰·曼达尼 (Zohran Mamdani) 2025 年竞选纽约市

市长的历程，便生动诠释了这种潜在的青年政治力量。起初，曼达尼只是一名知名度不高的候选人；但他直接面向年轻人发声，为他们描绘了一幅“我们负担得起的城市”（a city we can afford）的动人愿景，严厉指责其主要竞争对手腐败透顶，并通过一场运用数字媒体的竞选活动与选民建立联系，从而与那些脱离现实的年长候选人形成鲜明区别。到了投票日，曼达尼面向青年开展的动员取得了显著成效，充满活力的年轻选民帮助他赢得了胜利。

归根结底，Z 世代或许正在经历一场全球性的政治觉醒。如今，年轻人构成了一个规模庞大且不断增长的人口群体——据估计，全球 10 岁至 29 岁的人口约有 28 亿，占世界总人口的三分之一左右。如果年轻人能够继续借势而上，将这种势头转化为持续的政治参与，而僵化迟滞的政府也能够敞开胸怀，将年轻人纳入政治体系之中，那么 2025 年或许会成为世界事务发生深刻代际转型的起点。为了全球民主的未来，更广泛、更深入的青年参与已经刻不容缓。

注释.....

- 1 罗伯特·斯蒂芬·福阿等：《青年与对民主的满意度：扭转民主脱节？》（Roberto Stefan Foa et al., “Youth and Satisfaction with Democracy: Reversing the Democratic Disconnect?,” Cambridge: Centre for the Future of Democracy, October 2020; Gerardo Berthin, “Why Are Youth Dissatisfied with Democracy?” Freedom House, 14 September 2023, <https://freedomhouse.org/article/why-are-youth-dissatisfied-democracy>.)
- 2 这些主张基于一项针对 1990 年至 2020 年间 209 起目标最大化类型非暴力运动中青年前线参与情况的最新跨国研究。数据在此处提供。详尽分析见切诺维斯等：《非暴力行动中的青年与 LGBTQ+ 的参与》[Chenoweth et al., *Youth and LGBTQ+ Participation in Nonviolent Action* (Washington, D.C.: U.S. Agency for International Development, 2023), <https://www.hks.harvard.edu/publications/youth-and-lgbtq-participation->

nonviolent-action.]

- 3 马克·贝辛格、阿玛尼·贾马尔与凯文·马祖尔：《分道扬镳的革命联盟之解析》[Mark R. Beissinger, Amaney A. Jamal, and Kevin Mazur, “Explaining Divergent Revolutionary Coalitions,” *Comparative Politics* (2015).]
- 4 布里特·安拉尔等：《2025 年全球青年参与指数报告》（Brit Anlar et al., “The Global Youth Participation Index Report 2025,” Brussels: European Partnership for Democracy, 2025.）
- 5 鲁达贝·沙希德、尼什查尔·邓格尔与沙克蒂·德西尔瓦：《尼泊尔起义的经济根源——及其对该地区的意义》（Rudabeh Shahid, Nischal Dhungel, and Shakthi De Silva, “The Economic Roots of Nepal’s Uprising—And What It Means for the Region,” *Atlantic Council*, 19 September 2025, <https://www.atlanticcouncil.org/blogs/new-atlanticist/the-economic-roots-of-nepals-uprising-and-what-it-means-for-the-region/>）
- 6 保罗·斯坦尼兰：《南亚政权何以接连垮台》（Paul Staniland, “Why South Asia’s Regimes Keep Falling,” *Journal of Democracy*, September 2025, <https://www.journalofdemocracy.org/online-exclusive/why-south-asias-regimes-keep-falling/>）
- 7 埃莉卡·切诺维斯：《非暴力抗争的未来》[Erica Chenoweth, “The Future of Nonviolent Resistance,” *Journal of Democracy* 31 (July 2020): 70.]
- 8 埃莉卡·切诺维斯等：《非暴力行动中的青年与 LGBTQ+ 的参与》（Erica Chenoweth et al., *Youth and LGBTQ+ Participation in Nonviolent Action*）
- 9 西里安妮·达鲁姆：《街头的学生：教育与非暴力抗争》；亚当·波尼卡：《学生抗争在争取民主斗争中的力量：为何川普与全球威权主义者害怕校园行动主义》[Sirianne Dahlum, “Students in the Streets: Education and Nonviolent Protest,” *Comparative Political Studies* 52, no. 2 (2019): 277–309; Adam Bonica, “The Power of Student Protests in the Fight for Democracy: Why Trump and Authoritarians Worldwide Fear Campus Activism,” *On Data and Democracy*, Substack newsletter, 7 March 2025.]
- 10 乔纳森·平克尼与米兰达·里弗斯：《脆弱性与力量：对非暴力行动中女性与青年的思考》；西汗·埃尔达尔与杰奎琳·肯内利编：《德古意特青年行动主义手册》[Jonathan Pinckney and Miranda Rivers, *Precarity and Power: Reflections on Women and Youth in Nonviolent Action* (Washington, D.C.: U.S. Institute of Peace, 2021), <https://www.usip.org/publications/2021/09/precarity-and-power-reflections-women-and-youth-nonviolent-action>; Cihan Erdal and Jacqueline Kennelly, eds., *De Gruyter*

Handbook of Youth Activism (Berlin: Walter de Gruyter, 2026).]

- 11 切诺维斯等：《非暴力行动中的青年与 LGBTQ+ 的参与》；达鲁姆：《街头的学生》（Chenoweth et al., *Youth and LGBTQ+ Participation in Nonviolent Action*; Dahlum, “Students in the Streets.”）
- 12 拉格希尔德·诺达斯与克里斯蒂安·达文波特：《打击青年：青年人口膨胀与国家镇压》；阿亚尔·费恩伯格与伊德安·萨勒扬：《学生抗争、暴力互动与国家镇压》[Ragnhild Nordås and Christian Davenport, “Fight the Youth: Youth Bulges and State Repression,” *American Journal of Political Science* 57, no. 4 (2013): 926–40; Ayal Feinberg and Idean Salehyan, “Student Protest, Violent Interactions, and State Repression,” *Journal of Peace Research* 62, no. 1 (2025): 85–101.]
- 13 尼古拉斯·巴里约：《Z 世代抗争者的力量在坦桑尼亚遭遇致命武力》（Nicholas Bariyo, “The Power of Gen Z Protesters Meets Lethal Force in Tanzania,” *Wall Street Journal*, 9 November 2025, <https://www.wsj.com/world/africa/the-power-of-gen-z-protesters-meets-lethal-force-in-tanzania-f298adfb>.)
- 14 切诺维斯等：《非暴力行动中的青年与 LGBTQ+ 的参与》（Chenoweth et al., *Youth and LGBTQ+ Participation in Nonviolent Action*.)
- 15 斯蒂芬·哈格德：《倒退：当代世界的民主衰退》；马修·塞布尔：《反抗民主侵蚀的非暴力行动：比较视野下的美国》[Stephan Haggard, *Backsliding: Democratic Regress in the Contemporary World* (Cambridge: Cambridge University Press, 2021); Matthew Cebul, *Nonviolent Action Against Democratic Erosion: The United States in Comparative Perspective* (Cambridge: Harvard University, Ash Center for Democratic Governance and Innovation, 2025), <https://ash.harvard.edu/resources/nonviolent-action-against-democratic-erosion-the-united-states-in-comparative-perspective/>.]
- 16 马库斯·拜尔、菲利克斯·贝特克与丹尼尔·朗巴赫：《非暴力抗争的民主红利》；丹尼尔·朗巴赫等：《非暴力抗争与民主巩固》；卢克·阿布斯：《非暴力抗争对内战和平转型的影响》[Markus Bayer, Felix S. Bethke, and Daniel Lambach, “The Democratic Dividend of Nonviolent Resistance,” *Journal of Peace Research* 53, no. 6 (2016): 758–71; Daniel Lambach et al., *Nonviolent Resistance and Democratic Consolidation* (New York: Palgrave Macmillan, 2020); Luke Abbs, *The Impact of Nonviolent Resistance on the Peaceful Transformation of Civil War* (Washington, D.C.: International Center on Nonviolent Conflict, 2021).]
- 17 埃莉卡·切诺维斯与玛丽亚·斯蒂芬：《公民抗争何以有效：非暴力冲突的战略逻辑》；约纳坦·卢普与杰弗里·华莱士：《暴力、非暴力与国际人权法的效力》；皮尔斯·爱德华兹与丹尼尔·阿农：《多

- 方暴力：框架效应对抗争及支持镇压的影响》；西里安妮·达鲁姆、乔纳森·平克尼与托尔·威格：《支持非暴力抗争的道德逻辑：来自跨国问卷实验的证据》[Erica Chenoweth and Maria J. Stephan, *Why Civil Resistance Works: The Strategic Logic of Nonviolent Conflict* (New York: Columbia University Press, 2011); Yonatan Lupu and Geoffrey P. R. Wallace, “Violence, Nonviolence, and the Effects of International Human Rights Law,” *American Journal of Political Science* 63, no. 2 (2019): 411–26; Pearce Edwards and Daniel Arnon, “Violence on Many Sides: Framing Effects on Protest and Support for Repression,” *British Journal of Political Science* 51, no. 2 (2021): 488–506; Sirianne Dahlum, Jonathan Pinckney, and Tore Wig, “Moral Logics of Support for Nonviolent Resistance: Evidence from a Cross-National Survey Experiment,” *Comparative Political Studies* 56, no. 3 (2023): 326–62.]
- 18 费尔道斯·纳伊姆：《摩洛哥青年对破败体制说“够了”，计划于9月27日举行全国抗争》（Firdaus Naim, “Morocco’s Youth Say Enough to Broken System, Plan Nationwide Protest September 27,” *Morocco World News*, <https://www.moroccoworldnews.com/2025/09/259678/moroccos-youth-say-enough-to-broken-system-plan-nationwide-protest-september-27/>.)
- 19 汉娜·比奇：《一场Z世代革命如何失控恶化》（Hannah Beech, “How a Gen Z Revolution Spiraled Out of Control,” 8 October 2025, *New York Times*, <https://www.nytimes.com/2025/10/08/world/asia/nepal-gen-z-revolution.html>.)
- 20 切诺维斯等：《非暴力行动中的青年与LGBTQ+的参与》（Chenoweth et al., *Youth and LGBTQ+ Participation in Nonviolent Action*.)
- 21 塞伊内普·图费克奇：《推特与催泪瓦斯：网络化抗争的力量与脆弱性》；郑炜与袁玮熙：《多领袖动员的形成》；康秀妍：《无领袖运动的政治学》[Zeynep Tufekci, *Twitter and Tear Gas: The Power and Fragility of Networked Protest* (New Haven and London: Yale University Press, 2017); Edmund W. Cheng and Samson Yuen, *The Making of Leaderful Mobilization* (Cambridge University Press, 2025); Sooyeon Kang, *The Politics of Leaderless Movements*, unpubl. ms., Loyola University Chicago, 2025).]
- 22 康秀妍：《无领袖运动的政治学》（Kang, *The Politics of Leaderless Movements*.)
- 23 丹·柯林斯：《“如有必要，我们将游行无数次”：秘鲁青年抗争者蔑视紧急状态》（Dan Collins, “‘We’ll March as Many Times as Necessary’: Peru’s Youth Protesters Defy State of Emergency,” *Guardian*, 25 October 2025, <https://www.theguardian.com/world/2025/oct/25/peru-youth-protesters-state-of-emergency-gen-z>.)

- 24 关于 2024 年马达加斯加对军人统治支持率的民意调查数据，见 *Afrobarometer* (2025)。
- 25 马克·贝辛格：《民主革命的外表：乌克兰橙色革命中的联盟》[Mark R. Beissinger, “The Semblance of Democratic Revolution: Coalitions in Ukraine’s Orange Revolution,” *American Political Science Review* 107, no. 3 (2013): 574–92.]
- 26 提尔西·劳尼亚尔：《Z 世代如何在 Discord 上推翻尼泊尔领导人并选出新人的内幕故事》(Tulsi Rauniyar, “The Inside Story of How Gen Z Toppled Nepal’s Leader and Chose a New One on Discord,” *Wired*, 4 November 2025, <https://www.wired.com/story/nepal-discord-gen-z-protests-vote-prime-minister-election/>.)
- 27 乔纳森·平克尼：《从异见到民主：公民抗争转型的希望与危险》[Jonathan Pinckney, *From Dissent to Democracy: The Promise and Peril of Civil Resistance Transitions* (Oxford University Press, 2020).]
- 28 西里安妮·达鲁姆：《合力：社会联盟与民主革命》[Sirianne Dahlum, “Joining Forces: Social Coalitions and Democratic Revolutions,” *Journal of Peace Research*, 60 (January 2023): 42–57.]
- 29 平克尼：《从异见到民主》；达鲁姆：《合力》(Pinckney, *From Dissent to Democracy*; Dahlum, “Joining Forces.”)
- 30 切诺维斯：《非暴力抗争的未来》(Chenoweth, “The Future of Nonviolent Resistance.”)

简记

林 秦 | 作者林秦为加拿大多伦多大学留学生

中国民主季刊

第 4 卷 第 4 期

2026 年 10 月

CC-BY-NC-ND

自由、免费转载分发，但须注明
作者与出处，并不得修改及商用

全球民主 人权事件

1. 多个国家相继退出国际刑事法院

从 2026 年 7 月至 9 月，乍得、委内瑞拉和瑙鲁相继启动退出国际刑事法院（ICC）的程序。此前，在 2025 年 9 月，还有尼日尔、布基纳法索、马里同时启动这一程序。

2026 年 7 月 27 日，美国国务院非洲事务局在 X 平台上发帖“欢迎乍得决定退出国际刑事法院，加入越来越多从这个有缺陷的机构手中夺回主权的国家行列”，呼应国务卿卢比奥相关表态，美国正推动削弱国际刑事法院对各国“主权”的影响。

四个非洲国家及委内瑞拉在退出通知中指责国际刑事法院偏离创立初衷、近年被频繁政治化；乍得还重点批评国际刑事法院调查活动过度针对非洲国家。大赦国际西非和中非区域主任马尔索·西维厄德（Marceau Sivieude）表示：“布基纳法索、马里和尼日尔三国政府的这项决定有可能剥夺成千上万受害者获得真相、正义和赔偿的机会。”

国际刑事法院基于 1998 年《罗马规约》于 2002 年设立，管辖灭绝种族罪、危害人类罪、战争罪和侵略罪。不同于联合国下设的国际法院（ICJ），国际刑事法院审判针对个人而非国家。退出程序由《罗马规约》第 127 条规定，缔约国向联合国秘书长提交书面通知后需经过一年等待期方生效。

撒哈拉以南西非的尼日尔、布基纳法索和马里在 1990 年代第三波民主化浪潮中曾尝试转型。三国于 2023 年形成萨赫勒国家联盟，2025

年已集体退出西非国家经济共同体（ECOWAS），并逐步疏远西方、加强与俄罗斯联系。

2. 孟加拉国围绕《七月宪章》如何实施发生争议

2024 年 8 月哈西娜政府倒台后，孟加拉国国家共识委员会于 2025 年 10 月协商通过《七月宪章》，临时政府 11 月签发《实施令》。2026 年 2 月 12 日全民公投通过《七月宪章》后，按照《实施令》原定安排，新当选国民议会议员应同时宣誓成为宪法改革理事会成员，获得开展宪法改革、落实宪章的授权。

在孟加拉国民族主义党（BNP）坚持下，《七月宪章》签署前补入一项关键条款：“若任何政党或联盟在其竞选宣言中明确阐明并获得公众授权，可以按其宣言采取必要措施。”2 月大选中以压倒优势获胜的新执政党孟加拉国民族主义党拒绝参与改革理事会宣誓，理由是现行宪法中不存在该机构。2 月 17 日新政府成立后，强调不会履行其最初留下异议的相关条款。原本共同推翻哈西娜政府的反对派力量，现已分化为孟加拉国民族主义党与以孟加拉国伊斯兰大会党（Jamaat）和国家公民党（NCP）为核心的 11 党联盟对峙局面。后者要求尽快按原定方式落实《七月宪章》，孟加拉国伊斯兰大会党甚至提出再次起义的威胁。

7 月 13 日，由孟加拉国民族主义党控制的第 13 届国民议会决定绕开该党不赞成的理事会机制，设立“宪法修正案特别委员会”推进宪法改革，其中 12 席由孟加拉国民族主义党指定，预留 5 席给反对党。

反对党当天以退场和拒绝提名表示反对，但委员会仍于 14 日举行首次会议。

3. 香港国安搜查两家书店，持续收紧言论自由

7 月 15 日，香港国安处突袭搜查留下书舍和田园书屋，加上此前 3 月被搜查的一拳书馆和 6 月负责人被捕的猎人书店，今年已有 4 家独立书店遭国安处查办。此外，榆林书店和乐文书店虽未公开遭搜查，但于 6 月被取消参加香港书展资格。

截至目前，留下书舍已于原先宣布的 8 月 30 日结业；榆林书店宣布将在明年 4 月租约期满后“光荣结束门市”；乐文书店已从 9 月 19 日开始清仓甩卖；猎人书店仍顶着各方压力继续营业；田园书屋营业时间不稳。

港府在 15 日当晚发布声明，称搜查起因是接获海关转介，在一批由境外寄到香港的货物中发现具煽动意图的书籍。当天依据《维护国家安全条例》第 24 条“出于煽动意图作出具煽动意图的作为罪”拘捕留下书舍 3 人及田园书屋 2 人，并从两家书店各带走数箱证物。

17 日，被扣查的 5 人均获准取保候审。留下书舍在 8 月 30 日最后一天营业时仍有大量购书人潮。猎人书店曾表示，房东收到大量恐吓信，因此决定租约到期后不再续约。

4. 法国通过对 15 岁以下青少年的社交媒体禁令

7月21日，法国参议院以24票对2票、国民议会以279票对81票通过禁止15岁以下儿童使用社交媒体的法案。该法案最初由复兴党议员劳拉·米勒（Laure Miller）提出，旨在减少社交媒体对儿童的负面影响。

法案原定于9月1日生效。此前仅澳大利亚有类似禁令（针对16岁以下）。但左翼政党不屈法国（LFI）和社会党（PS）在法案通过后，很快联合约60名国民议会议员，以侵犯隐私和言论自由为由提交联名申请，触发法国宪法委员会事前合宪性审查。8月14日，宪法委员会作出违宪裁决，表示该立法第1条“构成对15岁以下人士言论和交流自由的不适当、不必要、也不相称的侵犯”，并对要求提供年龄证明才能访问在线服务的规定表示担忧，称其未能建立“必要的法律保障”。总统办公室随后表示，政府将“尽快”提出新草案。

5. 刚果（金）反对派持续示威，反对修宪延长总统任期

从2026年6月起，刚果民主共和国现任总统费利克斯·齐塞克迪（Félix Tshisekedi）寻求修改与任期限制相关的宪法，引发民众强烈抗议。抗议主要组织者为“维护宪法秩序第64条联盟”（C64）。宪法第64条规定：“任何刚果公民均有责任反对任何个人或群体企图通过暴力、违宪手段夺取或行使权力的行为”，该联盟包含多个反对修宪党派。

齐塞克迪5月提议修改宪法，并表示“我没有要求连任第三届，但如果人民想要连任第三届，我会接受。”6月，国民议会通过法案，允

许将宪法修正案提交全民公投。随后，组建约一个月的维护宪法秩序第 64 条联盟于 6 月组织议会大楼外静坐示威。示威当日，安全部队和亲政府的“进步力量”（Force du Progrès）准军事组织拦截并袭击正在和平游行的人群及反对派部分办公室。据媒体报道，数十名抗议者被拘留或失踪。

7 月 20 日，维护宪法秩序第 64 条联盟发表倡议举行包容性全国对话的声明，并要求对话最迟于今年 8 月 15 日举行。7 月 28 日，抗议活动呈现全国蔓延趋势。

8 月 2 日，政府“青年与爱国觉醒部”部长格蕾丝·埃米·库蒂诺发起动员修宪支持者的“是的，我愿意”（« Oui, je le veux »）宣传运动。10 日，包括维护宪法秩序第 64 条联盟在内的 65 个民间社会组织共同签署并向齐塞克迪提交备忘录，要求总统：1）放弃公投计划；2）不要颁布公投组织法。支持方则发起针锋相对的游行，并得到选择性宽容对待和总统迅速接见。

8 月 15 日，政府未正式启动对话。27 日，齐塞克迪宣布全国对话计划，表示具体时间和条件将后续公布。但维护宪法秩序第 64 条联盟不接受此结果，继续举行示威。

6. 德国极右政党在州级选举中获得重大突破

2026 年 9 月，德国举行三场州级议会选举，被广泛视为极右翼的另类选择党（AfD）在多场选举中取得重大突破。其中两场发生在德国

东北部较贫困的萨克森 - 安哈尔特州和梅克伦堡 - 前波莫瑞州，另类选择党分别获得约 44% 和 38% 的选票，成为最大党。最后一场为柏林州议会与区议会合并选举，另类选择党获得 16.3% 的选票，接近主要执政党德国基督教民主联盟（CDU）的 18.8%（柏林选举由左翼党 Die Linke 以约 25.7% 领先）。

德国民调机构 Infratest dimap 数据显示，萨克森 - 安哈尔特州选民支持极右的前三位原因包括：不安全感、认为另类选择党遭到其他政党不公平对待，以及赞成该党更严格限制外国人和难民涌入的主张。在德国，通常需席位超过 50% 或组成超过 50% 的政党联盟才能执政，少数派政府极易受议会掣肘。但由于二战后反纳粹传统，主流政党长期默认不与极右党派结盟，而另类选择党早前已被德国国内情报机构认定为右翼极端主义政党。德国总理默茨承认，这是其所在政党几十年来最糟糕的选举结果之一，但表示“另类选择党的胜利并非改变路线的理由”。

7. 美国最高法院对邮寄选票、选民身份验证做出裁决

2026 年 9 月 14 日与 9 月 25 日，美国最高法院分别就川普政府声称加强选举“诚实性”的两项行政令相关争议作出紧急程序裁决。一项裁决限制美国邮政局在 11 月中期选举前实施新的邮寄选票规则，另一项则允许升级版选民资格核验系统继续运行。

根据宪法第一条第四款，各州有权规定众议员和参议员选举的时间、地点与方式，但国会可制定或修改相关法规。宪法并未赋予总统规范

选举事务的任何权力。

川普于 2025 年 3 月 25 日签发 14248 号行政令《维护和保护美国选举完整性》（EO），要求国土安全部（DHS）与社会保障局（SSA）打通数据，扩展原本主要用于公共福利审核的 SAVE 系统，使州与地方选举官员能够用于核验注册选民及申请注册者的公民身份，防范非公民非法投票。

2026 年 3 月 31 日，川普再签发 14399 号行政令《确保联邦选举中的公民身份核查与选举诚信》，要求美国邮政局（USPS）重新制定接收有效邮寄选票的规定，包括要求州提交参与邮寄投票的选民名单、使用特定信封格式与条形码等标准，否则邮政局可拒收或拒寄不符合要求的选票，从而收紧邮寄选票合规要求。USPS 随后于 8 月正式发布相关新规。

美国各州对邮寄选票标准原本存在差异。加州等 23 州、哥伦比亚特区及妇女选民联盟等团体挑战 14399 号行政令相关措施，认为临近选举强加复杂信封格式和数据上传要求会造成行政混乱、可能误伤合法选民，并干涉州选举权。最高法院 9 月 14 日作出有利公民团体一方的裁决，禁止 USPS 在 11 月中期选举前实施邮寄选票新规。

同时，川普团队主张邮寄投票缺乏现场身份核验，必须通过严格数据库比对来防范非公民非法投票；反对方则认为政府数据库常有滞后或错误，严格比对会大量误伤合法公民。妇女选民联盟等团体起诉 DHS 与 SSA 后，联邦法官在 6 月裁定撤销并废止该改版数据库及其相关通

知。9月25日，最高法院裁决暂时允许川普政府及各州在11月中期选举前使用改版后的SAVE选民数据库核查选民资格。

两项裁决均为最高法院紧急程序（shadow docket）中的临时中止决定，并非对实体问题的终局裁判。相关诉讼仍在下级法院继续。

8.130 多名联合国人权专家联合呼吁应对跨国镇压

2026年9月23日，130多名联合国人权专家（含特别程序任务负责人与条约机构成员）签署联合声明，呼吁就跨国镇压加强国际合作。

声明将跨国镇压界定为一个国家在其境内或境外采取的措施，旨在基于个人、群体或社群真实或被认定的意见、信念、身份或行为（包括异议表达，或行使、捍卫或倡导人权、民主与法治），对其在境外的活动进行压制、威慑、惩罚、胁迫、控制或以其他方式造成伤害。相关行为可由国家当局直接实施，也可由受其指挥或控制的非国家行为体实施，或由一国协助另一国实施。

常见手段包括法外处决、酷刑或虐待、人身攻击、强迫失踪、威胁与骚扰、刑事化、绑架、任意拘留、任意驱逐、旅行禁令、剥夺国籍，以及滥用国际刑警组织（INTERPOL）红色通缉令与扩散通告等；数字监控与人工智能正在放大这一问题。移民、寻求庇护者、难民、人权捍卫者、记者、律师、异议人士及与联合国合作的人员等被列为特别易受伤害群体，镇压还常延伸至其家人与关联人员。

声明敦促各国制定法律与政策预防、调查和惩罚此类行为，加强对高风险人群的保护（含加快庇护与重新安置）；要求国际刑警组织打击滥用通告行为；呼吁科技公司开展人权尽职调查并加强隐私与数据保护。专家们认为跨国镇压正在加剧，对人权、民主与法治构成严重挑战。

9. 委内瑞拉反对派要求允许其领导人安全回国

自委内瑞拉北部于 2026 年 6 月 24 日发生约 7.5 级大地震后，反对派领袖玛丽亚·科里纳·马查多（María Corina Machado）再次反复要求当前政府允许她回国，并谴责罗德里格斯政府关闭领空以阻止她返回。马查多是委内瑞拉反对派主要领导人之一，2024 年曾作为反对派候选人参与总统竞选，后因主张民主选举与人权而长期面临打压；2025 年底秘密离境赴挪威领取诺贝尔和平奖，此后一直流亡海外（近期多在巴拿马）。

6 月 29 日，马查多表示“愿意不惜一切代价”进入委内瑞拉，帮助该国在连续地震后进行灾后重建。她已为此联系多位美国政府官员寻求支持，但华盛顿方面要求她推迟回国，表示担心其过早返回可能干扰灾后救援与政治稳定。9 月 22 日，第 81 届联合国大会在纽约召开期间，川普利用会程间隙与临时总统罗德里格斯进行闭门会晤。

截至目前，委内瑞拉国内物资匮乏和经济窘况仍未改善。9 月 24 日和 27 日，委内瑞拉爆发游行，要求马查多安全返回、举行总统选举，以实现民主过渡。马查多的政党表示，她近期又进行了多次回国尝试（包括海路与空路），但均因各种原因未能成功。

ABSTRACTS IN ENGLISH

ROUNDTABLE

Democratic Opportunities in the Age of Artificial Intelligence

Genia Kostka, Qiang Xiao, Alex Gladstein, Zhiqiang Ji, and Shuang Liu

The impact of artificial intelligence (AI) on authoritarian regimes has drawn intense attention from scholars and observers. Authoritarian regimes are actively using AI to consolidate their rule—deploying facial recognition and smart surveillance to tighten social control, for example, and large language models to conduct propaganda, steer public opinion, and carry out ideological indoctrination. Scholars and observers have accordingly put forward concepts such as “digital authoritarianism” and “data-centric authoritarianism” to capture the new features of authoritarian rule in the age of AI. Recent research, however, has begun to reveal the other side of the story: authoritarian regimes face inherent limits in harnessing AI, and the same technology can also be put to use by democratic forces.

Even as AI provides authoritarian regimes with new instruments of rule, does it also confront them with significant shocks and challenges, opening up new opportunities for the opposition, civil society, and pro-democracy forces? Judging from current trends, where do these shocks and challenges

mainly lie?

China Journal of Democracy invited five scholars and practitioners—Genia Kostka, Qiang Xiao, Alex Gladstein, Zhiqiang Ji, and Shuang Liu—to discuss these questions.

INTERVIEWS

“Gagging” and “Brainwashing”: Can China’s AI Censorship Be Taken Apart?

Hannah Cyberey / Weifeng Zhong

“It skips the thinking process entirely and goes straight to the answer—either it refuses to answer, or it talks around the question, or it gives you the Party line.” This is how Hannah Cyberey describes the censorship mechanism built into DeepSeek’s models. Cyberey is a postdoctoral researcher at the School of Data Science at the University of Virginia. In *Steering the CensorShip*, a paper co-authored with her supervisor, Professor David Evans, they located “steering vectors” inside large language models that control the degree of censorship and can be turned up or down like a dial. The paper also found that when the DeepSeek-R1 series of reasoning models encounters sensitive topics such as Taiwanese independence or Tiananmen Square massacre, it skips the reasoning process altogether—a phenomenon the authors call “thought suppression.”

Born and raised in Taiwan, Cyberey received her PhD in computer science from the University of Virginia in 2025 and studies the evaluation and

auditing of AI models. On September 15, 2026, Dr. Weifeng Zhong, a scholar at the Mercatus Center at George Mason University and an advisory editor of *China Journal of Democracy*, interviewed Dr. Cyberey. The conversation explores the censorship mechanisms built into Chinese large language models, white-box and black-box auditing methods, so-called “uncensored” models, how the safety restrictions in Western models resemble and differ from censorship, the debate over slowing down AI development, and how authoritarian governments control information in the age of AI.

Censorship on the Assembly Line: Political Control Inside China’s AI Models

Owen Au / Eddy Ji

Generative AI is “a very long content-generation pipeline. The longer the pipeline, the more points at which external actors such as the state or the authorities can intervene in what it generates,” says Owen Au. It is precisely along this pipeline that Beijing has systematically translated its political demands on content and information into technical procedures. Born in Hong Kong, Au is now an independent China researcher and analyst based in the United Kingdom, working for various think tanks, consultancies, and NGOs. In July 2026, he and Aiden Gu published the report *Authoritarian Innovation: How China Is Building Control into Generative AI Systems*. Commissioned by *China Journal of Democracy*, Eddy Ji, a researcher in computational politics, interviewed Au on October 3, 2026, about the report: why Beijing both relies on and guards

against generative AI; how political requirements are written, by way of technical standards, into every stage of data, training, and testing; how this system of control spreads abroad along with Chinese models; and how democracies should respond.

A Decade of China's Feminist Movement in the Context of Authoritarian Rule

Pin Lü / Xin Ai

“Thirty years ago, when I threw myself into the feminist movement, the term ‘feminism’ (女权主义) was shocking and frightening in the Chinese context,” said Pin Lü, a prominent Chinese feminist advocate currently at Rutgers University in the United States. Pin Lü was formerly the lead writer for the Women’s Voice e-newsletter and the founder of the Feminist Voice self-media outlet—of course, what she has done goes far beyond these. From the mid-1990s to the present, she has been involved in and witnessed the contemporary Chinese feminist movement. Commissioned by China Journal of Democracy, Xin Ai, a feminist activist currently in mainland China, interviewed Pin Lü. The theme of this interview is as its title indicates, but the richness of its content and the intellectual tension within it far exceed what the title alone can convey.

POLITICS AND SOCIETY

A Historical Examination of the Chinese Communist Party's “External Propaganda” (Part II)

Yiran Pei

“Propaganda” has been the foundation of the Chinese Communist Party and its vanguard from the very start. The CCP’s “big external propaganda” (*dawaixuan*), short for the “grand layout of external propaganda,” has attracted wide attention only relatively recently, yet its history is in fact quite long. Since 1949, external propaganda has remained an important wing of the CCP regime, seeping steadily overseas: before the Cultural Revolution it effectively shaped Western public opinion, and after the Cultural Revolution it quietly shifted the positions of some overseas Chinese-language media. Over the past two decades, the CCP has kept pouring funds into “big external propaganda,” broadcasting red ideology abroad with full force in an attempt to contest the ideological high ground with Europe and the United States.

The Downstream Yarlung Tsangpo Hydropower Project: Goals, Political Intent, and Environmental Issues

Weiluo Wang

On July 19, 2025, the groundbreaking ceremony for the hydropower project on the lower reaches of the Yarlung Tsangpo River (hereafter “the Yaxia Project”) was held at the Mainling hydropower station dam site in Nyingchi. The Chinese government describes the Yaxia Project as a green project that harnesses hydropower, with a projected annual output of 300 billion kilowatt-hours—three times that of the Three Gorges Project. The author argues that this generation target cannot be achieved. The project’s

real purpose lies not primarily in electricity but in the “South Tibet” question and the political dispute between China and India. The Yaxia Project will inflict severe damage on the ecological environment of the Tibetan Plateau, and in particular on its biodiversity.

FEATURE

Research That Does Not Falter: A Review of the North American Symposium on the 60th Anniversary of the Cultural Revolution

Xiao Xiao

On August 18, 2026, the editorial board of the *Chinese Cultural Revolution Database* in the United States convened an online symposium titled “Sixty Years of the Cultural Revolution: Looking Back and Looking Around at China and the World.” The symposium posed new challenges to established conclusions and entrenched views in Cultural Revolution studies. Yet it was not confined to the historical facts and research questions of the Cultural Revolution itself; participants also discussed Cultural Revolution-like phenomena in China and abroad today. In this account, Xiao Xiao summarizes the main content of the symposium and offers personal reflections on it.

CIVIL RESISTANCE

The Limits of Authoritarian AI

L. Jason Anastasopoulos and Jie Lian

Translated by Fan Hong

The advent of artificial intelligence is widely seen as handing authoritarian rulers an exceptionally powerful tool of control. In this article, L. Jason Anastasopoulos and Jie Lian argue that authoritarian regimes face inherent limits in using AI. AI systems force rulers to confront an inescapable “calibration dilemma”: lowering the threshold of judgment reduces false negatives and catches more potential challengers, but it also increases false positives, sweeping in more innocent people and provoking backlash. This structural instability produces what the authors call “threshold whiplash,” a cycle that swings between continual tightening and sudden loosening—China being a case in point. Democratic forces can exploit these vulnerabilities by demystifying algorithmic power, building protective norms, and using strategic resistance to challenge the illusion of panoptic surveillance. The article was originally published in the *American Journal of Democracy*, Volume 37, Number 2, April 2026, pp. 5–17.

WHY GEN-Z IS RISING

Erica Chenoweth and Matthew Cebul

Translated by Xiaoyi Yan

This essay analyzes the global wave of social mobilization led by Generation Z, arguing that the main drivers of youth protest movements include economic insecurity, exclusion from the structures of power, and glaring corruption. Movements with broad youth participation can

generate formidable mobilizing power, but they also face harsh repression and often struggle to resolve deep-seated structural grievances in the short term. These movements typically adopt a decentralized, leaderless mode of organization, which helps them spread rapidly but also complicates negotiation and political transition. The article was originally published in the *American Journal of Democracy*, Volume 37, Number 1, January 2026, pp. 5–14.

NOTES

Global Events in Democracy and Human Rights

Qin Lin

《中国民主季刊》征稿启事

由中国民主转型研究所主办的《中国民主季刊》（China Journal of Democracy）旨在为中文读者提供关于中国与世界的最新观点、新知识，特别是关于中国政治、社会、民主转型的深入探讨。它注重现实性、学术性与可读性。敬请中国大陆和台湾、香港、美欧、澳洲等广泛地域范围以及比较政治、国际政治、经济、社会学、历史等专业背景不同的专家学者不吝赐稿。

欢迎就下列几个方面的问题赐稿——

1. 关于中国重要的政治、社会、经济、文化、外交问题；
2. 关于中国民主转型可能前景、路径、风险、战略等问题；
3. 关于其他国家和地区的民主转型过程、经验、教训；
4. 关于民主理论、宪法设计与民主建构；
5. 关于民族与国家性问题；
6. 新书与重要著作的推荐、评论。

投稿要求如下：

1. 稿件应为汉语。也欢迎用藏语、维吾尔语投稿，我们将协助译成汉语发表。来稿请标明篇名、关键词、摘要（不超过 200 字）、作者姓名及简介、联系方式。
2. 国内作者：由于是学术刊物，鼓励国内作者用实名，但出于安全考虑，国内作者基于自己的判断选择用笔名。
3. 稿件字数：论文，以 5000-8000 字为最适宜，建议不超过 10000 字；更长的文章可以考虑分期发表。书评，建议在 2000-3000 字之间。译作，建议在 4000—10000 字之间。
4. 来稿须符合写作规范要求，详细注释体例在中国民主转型研究所网站上：<https://chinademocrats.org/?p=24>。
5. 投稿邮箱：cjd@chinademocrats.org。您也可以通过此邮箱向我们提出选题策划、建议意见、咨询或者合作方案。
6. 《中国民主季刊》拟每年 1 月、4 月、7 月、10 月中出版，截稿日期分别为每年 3 月 1 日，6 月 1 日，9 月 1 日与 12 月 1 日。
7. 稿件刊发后将按字数支付稿酬。

《中国民主季刊》订阅方式

How to Subscribe to China Journal of Democracy

《中国民主季刊》是电子出版物，可以登录我们的网站免费下载，也可以提供您的电子邮箱地址，我们免费通过电子邮件定期发送给您。

China Journal of Democracy is an electronic publication that can be downloaded for free from our website, or you can provide your email address and we will send it to you periodically by email free of charge.

我们也印刷少量纸质版，以满足个人、机构、图书馆收藏需要。纸质版每期价格为 38 美元。

In addition, we print a small number of paper copies to meet the needs of individual, institutional, and library collections. Paper copies are available for \$38 per issue.

下载电子版：请登录中国民主转型研究所网站 <https://chinademocrats.org/>。

To download an electronic copy: Please visit the China Institute for Democratic Transition website at <https://chinademocrats.org/> .

订阅电子版：请登录我们的网站 <https://chinademocrats.org/?p=821> 或将您的电子邮箱发送至 cjd@chinademocrats.org .

To subscribe to the electronic edition: Please visit our website <https://chinademocrats.org/?p=821> or send your email address to cjd@chinademocrats.org .

订购纸质版：请电邮 cclassics1994@gmail.com 联系代理商 China Classics Inc.。

To order paper copies: Please contact our agent China Classics Inc. by email at cclassics1994@gmail.com , or by phone at 858 229 9677.

中国民主季刊

CHINA JOURNAL OF DEMOCRACY

回顾历史，从晚清到当代，中国发生的、争取自由民主的每一次重要努力，知识分子都有着特殊重要的作用。当今中国最重要的事务莫过于实现民主宪政，为此而探索、写作，以及传播相关的知识和思想，是知识分子的第一责任。



中国民主转型研究所
INSTITUTE FOR CHINA'S DEMOCRATIC TRANSITION

ISBN 979-8-9875925-0-2



9 798987 592502