arXiv:2302.00672v1 [cs.HC] 1 Feb 2023

通过集体响应系统实现“生成式集体智能”

Aviv Ovadya aviv@aviv.me 美国哈佛肯尼迪学院贝尔弗科学与国际事务中心技术与公共目的研究员；英国剑桥大学Leverhulme未来智能研究中心访问学者

摘要

本文阐述一种特定类型的生成式集体智能（CI）促进系统：集体响应系统——以及它所促成的集体对话。本文界定了这类系统的结构、流程、关键属性和关键原则，旨在建立一套实用的共同语言，并借此推动人们认识此类系统的潜在益处，同时提供一份简明的参考文本。

集体响应系统支持一种“生成式投票”形式——“票”以及投票议题的选择均由集体提供。这使不同群体能够表达自己的观点，并聆听他人的观点，以此克服民意调查、市民大会、投票、公民投票等传统方式的局限。这有助于以非对抗的方式探讨分歧议题，帮助找出共同点，并从最接近议题的人那里发掘见解；从而化解围绕冲突和治理挑战的僵局，增进信任，并形成授权。[1]

现有集体响应系统的典型例子包括 Polis 和 Remesh。台湾及世界各地的政府曾在不同层级的政策制定中使用 Polis [1]。联合国部署了 Remesh——一种由人工智能支持的集体响应系统，用于了解饱受战乱的国家中普通民众面临的挑战和需求；数千名直接参与者参与其中，社交媒体上的参与者也占了相当大的比例 [2–4]。

现有的集体响应系统建立在推荐系统、语言模型和人机交互等多个学科的技术工作之上。[2] 持续推进此类系统的开发与应用，有助于重振民主制度、重新构想公司治理、化解冲突，并治理强大的人工智能系统——既可作为深化审议式民主进程的补充，也可在无法开展更深入进程时作为一种选择。

1 动机

当前大多数用于群体理解和决策的方法难以大规模奏效，尤其是在现代注意力经济扭曲的激励机制下：

- **深思熟虑的审议无法扩展**：如果一千人的“市民大会”都试图同时发言和倾听，结果只会是一片嘈杂。反过来，如果人们一个接一个地发言，那可能要花上好几天。若有一百万人，针对单个议题就要花上几十年，因此，大多数受政策或冲突影响的人都无法真正发言或倾听。
- **人们感到自己没有发言权，也不受尊重：**人们往往觉得自己没有能力表达观点，也看不到自己的看法和经历得到反映。这会导致不信任，使人们不再参与治理和冲突解决流程。
- **有价值的洞见被错过**：“知识钥匙”——信息 n, — 思想和洞见，有助于克服根深蒂固的冲突往往会被忽视。这些观点通常来自最接近问题的人，而他们中的许多人通过标准渠道所能触及的受众极为有限。
- **制造分歧胜过寻找共同点：**煽动冲突所获得的回报，远远高于发现并凸显共同点。

然而，连接性、机器学习和民主实践的进步[3] 或许能让我们通过“规模化同步沟通”克服其中一些挑战。

集体响应系统旨在使任意规模的群体都能做出生成式决策（例如，由参与者共同制定选项空间的决策）。它们的设计目标是尽可能接近某种“民主理想”——即在集体决策中：

1. 每个人都能表达自己的观点。2. 每个人的回应都能得到倾听并被纳入考量。3. 选择最能代表群体的回应。[4]

集体响应系统以决策和审议为核心，旨在避免娱乐导向或政治导向沟通环境（例如社交媒体）中的不良激励；在这些环境中，博取关注的行为会受到奖励。它们支持一种聚焦的规模化同步沟通方式。它们还使集体能够被视为单一代理，进而与其他此类集体代理互动。

贡献

这篇框架性论文定义了集体响应系统的结构、流程、关键属性和关键原则，旨在为集体智能促进和大规模民主决策的研究与实践提供共同语言，

[1]换言之，在选项众多、价值观相互冲突且利益相关者多元的环境中，它们有助于解决“对齐”和“优化”问题。[2]与任何社会技术一样，非技术层面的因素在很多方面都更为重要、影响也更深远；但技术组件决定并约束着这一媒介的关键组成部分。本文尤其旨在为技术领域的研究人员和实践者提供简明概述，因此我们聚焦于与这一目标最相关的方面；不过，开展更全面的跨学科探索将极具价值，而且这类探索还可以借鉴许多其他领域的成果。如果您有兴趣参与此类项目，请联系作者。

[3]具体而言，是协同过滤和语言建模方面的改进，以及有效的产品设计和促进协作的洞见。[4]例如，可以选取得票总数最多的回应，或跨越不同阵营获得最多选票的回应（例如 Polis [1] 及采用基于桥接的排名的相关系统 [5]），也可以采用截然不同的方式。还可以针对参与群体与其所代表的更大群体在特征上的差异进行调整（类似于民意调查中的做法）。

并借鉴人机交互与人工智能领域的进展。

2 结构与流程

集体响应系统是一种集体智能促进系统，符合下文第 2 节和第 3 节所述的结构、流程、属性和原则。

2.1 输入：一个群体和一个提示

系统有两个输入：提示[5]，提示可以是一个问题、需要完成的内容等；另一个输入是群体，即参与者的集合。

示例：

- 对于一座城市的公民： 我们这座城市在教育方面面临的最重要挑战是.....
- 对于像 Facebook 这样的平台： 在这个平台上，哪些行为不应被允许?
- 对于即将毕业的学生： 谁应该担任我们的毕业典礼演讲嘉宾?
- 对于一个组织： 我们组织的核心价值观应该是:
- 对于全球人口： 我们应采取哪些最重要的措施来应对气候变化.....

2.2 输出：对评估结果的代表性提炼，以及可选的原始和/或派生数据。

这类集体响应系统的核心输出，是对响应的近似或汇总，以有意义且具有代表性地传达“最佳”响应。这通常可统称为**代表性提炼**。

“最佳”的定义很大程度上取决于如何解释，但受到下文所述关键原则的约束。例如，获得最多认可的响应可能被视为最佳响应，不过也可以采用其他方法（例如 Polis 团队的知情共识所采用的基于桥接的排名 [5] [1]，或 Twitter 的社区注释 [7]）。诸如“最佳响应”之类的输出也可用于下游系统，包括下文所述的未来集体响应流程。

此外，还可以输出有助于向群体反馈其响应和评估的派生数据。例如，Polis 这类系统可以向参与者展示议题和人们在观点地图中的位置。最后，还可以输出响应和评估等原始结果，以供进一步处理和分析。

[5]虽然集体响应系统中的提示主要面向人类，但这种框架在某些方面与 GPT-3 [6] 等生成式语言模型类似。不过，生成式机器学习模型还可以通过多种方式参与此类系统。

2.3 流程：响应、评估、提炼

由集体响应系统促成的过程称为**集体响应过程**。这类过程由三种子流程组成：[6]

- **响应子流程**：群组中的每个人都可以（选择性地）回应提示。
- **评估子流程**：群组中的每个人都会评估部分回应（可能由系统选定）。
- **提炼子流程**：系统会近似计算和/或汇总对每个回应的评估，以生成有用的输出。

将这些结合起来，设想一下这样一种情况：城市中的五千人参与了一场关于教育政策的集体对话。

在这种情况下，**集体响应流程** 可能如下所示：

1. 回应： 参与者会被问到：“我们城市在教育方面面临的最重要挑战是什么？”他们可以用简短的回答作出回应。在 Polis [1] 和 Remesh [8] 这类系统中，回答通常很短，例如一到三句话。2. 评估：参与者会收到“投票任务”，例如评估自己是否赞同某条回答。3. 提炼： 系统随后可以向 参与者展示 哪些回答获得的支持 最多，以及哪些回答在不同（例如政治）阵营之间拥有最多共同点 。

这一过程最快只需 5 分钟即可完成（例如通过 Remesh 进行同步交流），也可以历时数天或数周（例如通过 Polis [1]进行异步、迭代式交流）。

3 属性与原则

关键属性。 系统必须满足以下属性，才能成为集体回应系统：

(1) 参与者能动性：参与者可以自行提出回应，而不必受限于固定选项集——这可称为“参与式视角”。[7][8]**(2) 并行沟通**：为了推动流程继续进行，并非每位参与者都需要与其他所有人交谈或倾听他们的意见。这可能会采用诸如分组法（即大致类似于分组讨论 [10]）和引出推断（即根据参与者对部分回应的评估，近似推断他们会如何评估所有回应的方法） [11]等方法。**(3) 代表性提炼**：通过某种机制，近似处理和/或汇总所有参与者的输入，以获得能够代表整体的评估。

[6]这些子流程不一定严格按照“瀑布式”方式进行；具体取决于系统，它们也可能以异步方式部分交错进行。[7]这就是 WikiSurvey 论文所称的协作性。

[8]与者以外的其他方式提出（例如由引导者、生成式模型等提供），但这些回答会与参与者的回答同等对待，并且可能会标注其来源信息。[9] 回答也可以通过参

提炼后的结果（可能包含输出的“映射” [1]）会分享给参与者。

关键原则。 此外，集体响应系统的设计必须满足以下关键原则，才能符合前文所述的民主理想：

(1) **集体导向**：系统设计优先支持和理解群体及其决策，而非支持和理解个人及其决策。[9][10]

(2) **提示导向**：系统设计优先创建能够最佳回应输入提示的输出，而非追求其他目标（例如娱乐）。[11]

(3) **审议导向**：系统设计优先提供支持参与者理解潜在选择以及彼此对这些选择的观点的功能与激励，而非追求其他目标（例如参与度）。[13] （例如，这可以包括支持识别共同点、呈现洞见、对群体经验进行自我反思以及绘制观点图谱等行动。）

(4) **公平聆听导向**：系统设计必须确保所有参与者都能得到公平聆听——让所有回应都有机会被群体听见，从而有可能最终被视为最佳回应。[14]

属性是非此即彼、界限分明的，而原则则留有更多解释空间；或许还需要开展更多哲学、实证和规范性研究，以进一步完善这些原则。

3.1 集体对话

集体对话系统（CDS）是一种集体响应系统，包含某种形式的迭代或反馈循环，使参与者能够反复与系统互动，并根据先前的输出生成新的回应（针对相同或新的提示）。**集体对话**可作为集体对话过程（由集体对话系统促成的过程）的简称。

这样的对话流程能够进一步深入探究最有希望的解决方案的细节、动机和障碍。它们也可用于任何其他意义建构或迭代式决策任务。

回想一下前文的例子：“我们城市面临的最重要的教育挑战是什么？”如果这是集体对话的一部分，那么在获得第一个答案后，可以根据被认为最具代表性的回答提出后续问题，并让相同的参与者参与新的集体响应流程。

这可能看起来像是在询问：“你认为应对 *[这一挑战]*的最佳解决方案是什么？”

4 现有工作与相关概念

目前已有许多集体响应系统的部分实现，也有少数完整实现。其中一些最值得注意的包括*WikiSurvey* 的，例如*Polis*和*All Our Ideas* [1, 9]。[15] *Remesh*及类似工具专注于近似组件，从而实现更大的规模 [4]。另一种方法见于*PSiApp* 等分组法系统 [10]。为公民大会 [12] 和审议式民意调查 [13] 创建的某些流程，也可能符合集体响应系统的标准。类似地，在对整个群体使用集体响应系统并不实际的情况下，可以利用基于抽选的参与者选拔，从规模大得多的群体中选出人员。

集体响应系统可能会模糊治理系统与推荐系统之间的界限。例如，集体响应系统可能会使用矩阵分解等方法来支持近似 [14]。[16] 近期的方法还涉及使用大语言模型 [11, 15]。

其他类型的在线机器学习系统也可能与集体响应系统有关（尤其取决于训练数据的来源和获取方式）。将集体响应系统串联起来，构成复杂的决策网络，在某种程度上类似于提示链 [16] 和多主体抽签遴选 [17]。[17] 生成式人工智能的提示“工程”与生成式集体智能的提示设计之间，也可能存在相关的相似之处。

更一般地说，集体响应系统通常旨在解决多委托人—代理人问题 [11],，因此，与这类问题相关的既有研究很可能也具有参考价值。最后，包括社会选择理论在内的学科，对于如何定义“最佳”有许多见解——当然，也可以融入其他学科的洞见。这显然并非一份详尽的清单，只是为了说明相关研究的广泛范围。

[9]例如，普通推荐系统无法通过这项检验——它旨在向每位用户提供他们想要的内容。[10]当前的审议式民意调查也可能无法通过这项检验，因为它们关注的是个人选择的变化，而非集体结论。[11]按照这一原则，某些系统或许也会优先考虑建立联系、增进信任、练习审议式沟通等，但这些都只是为了更好地回应提示。这一点或许值得进一步探讨，或为其设定例外。[12]其中可能包括不同观点所占比重的相对大小。[13]换言之，其目的是针对提示输出集体的回应，以促进相互理解、信任和明智的决策；或者说，“兼顾多元性与智慧”。[14]集体响应系统的理想状态或许是实现完全平等的聆听，但具体取决于系统：更受关注的回应可能会得到更多展示，有些回应也可能因其他原因而展示，例如更好地推断哪些内容具有最大的共同点。（Polis 就是如此。）在某些情境下，让声誉评分对这一过程产生一定影响，也可能有所帮助。

[15]与 WikiSurvey 不同，集体响应系统不必是“贪婪”的，而且有些集体响应系统会有意限制可评估的内容。[16]作为比较，Reddit 的评论及其排序满足集体响应系统的三项特性，但不符合其中大多数原则。[17]这种多重集体对话 可以将多个输出串联起来。例如，第 1 组确定最需要解决的问题，第 2 至第 4 组针对该问题提出解决方案，第 5 组找出所提方案的问题和关键权衡，第 6 组将前面各组最重要的成果整合为一个解决方案，第 7 组则对所提方案进行最后的“质量控制”检查。根据具体情境，这些集体对话也可以由其他类型的流程取代，从公民大会到基础模型推理皆可。

5 结论

集体响应系统背后的理念与理想并非新事物——本文主要旨在简明扼要地阐述这些内容，以期建立一种实用的共享语言。

集体响应系统通过结合参与者主导的回应与评估，并逆转注意力经济的激励机制，来支持理解群体的价值观、需求和愿望——进而帮助形成具有决断力的授权。这类系统可用于识别并奖励能够弥合分歧的回应——这与大多数现代政治、媒体和社交媒体的做法恰恰相反。根据轶闻，集体响应系统在发掘最接近相关议题的人所提出的新见解方面也能极其高效；当人们看到自己的回应被纳入流程时，会感到自己的声音得到倾听和尊重。

5.1 局限性

集体响应系统并非万能药。

孤立的集体响应流程只能引 elic 出直觉性的意见，而不是经过深思熟虑的判断。通过提供教育材料、专家、利益相关者、带薪参与时间以及谨慎的中立主持，更为复杂（也更昂贵）的流程，例如公民大会和审议式民意调查，在一定程度上解决了这一问题 [12, 13, 18]。对于具有重大意义的决策而言，这样的资源投入和审议深度至关重要。

集体响应系统的部分（但并非全部）缺点可以通过迭代，即集体对话，得到克服。这种迭代能够促成更深入、更具审议性的探索。经过迭代或串联的集体对话甚至可以用来确定，例如哪些问题对于向专家、利益相关者或整个群体提问最为重要。[18]

然而，即使使用设计完善的集体对话系统，谨慎的流程设计仍至关重要。具体的集体响应系统实现也可能存在显著的权衡，例如将其串联成集体对话、在流程中途纳入参与者提供的新信息、采用同步还是异步方式、使用文本还是音频或视频等方面的能力。所有这些因素都必须纳入考量。

许多由数字媒介支持的集体响应系统默认并不能促成正常的人际关系建立（类似于普通的、非生成式的投票系统）。然而，尽管如此，它们确实（出人意料地）似乎能够促成一种有价值的联结与信任，这很可能是因为它们通过代表性提炼促进了某种形式的< शैली id='1'>集体反思，并帮助识别共同点领域。

在可能的情况下，理想而言，集体响应流程应作为更深入审议流程的补充，而非替代。例如，公民大会可以利用这些流程，从公众那里收集开放式意见，以支持其决策。由于开销较低，集体响应流程也可以定期运行，频率远高于公民大会，用于不太重要的决策，或用于治理流程中无需深入个人审议的特定环节。

[18]它们也有望用于识别权衡和“关键难题”，进而加以解决。

5.2 机遇

虽然并非适用于所有集体决策，但集体响应系统或许有助于在社会的各个层面融入新形式的意义建构、治理、民主和能动性：

- **治理与冲突**：代表、公务员、公民议会和冲突调解员可以运用它们，让更广泛的公众（即使他们忙碌）切实参与进来。
- **“企业民主”**：它们支持一种新型企业治理，其涵盖范围不再局限于股东和精英利益相关者，还包括用户、员工以及受到影响的公众。
- **媒体与理解**：它们能够支持一种新型集体反思，帮助公众“认识自己”、找出共同点，从而更好地应对内部和外部挑战。

研究有一个重要机遇：为此类系统制定评估规程和指标，了解不同设计决策的影响，从而针对不同目的和情境，开发出日益完善的系统（以及系统链）。

正如机器学习的进步通过实现“生成式人工智能”，为人们提供了与计算机互动的新方式，集体响应系统的进步也能通过支持“生成式集体智能”（集体智能），为人们提供与群体互动的新方式。

致谢

感谢 Anna Waldman-Brown、Andrew Konya、Colin Megill、Divya Siddarth、Ted Suzman、Bruce Schneier、Luke Thorburn、An- toine Vergne、Glen Weyl、Kimbra White、CathyWu、Jessica Yu 等人参与讨论并提供反馈，促成了本文的完成。

参考文献

[1] Christopher Small、Michael Bjorkegren、Timo Erkkilä、Lynette Shaw 和 Colin Megill。Polis：通过绘制高维意见空间来扩大审议规模。*Recerca. Revista de Pensament i Anàlisi*，26(2):1–26。[2] Daanish Masood Alavi、Martin Wählisch、Colin Irwin 和 Andrew Konya。将人工智能用于和平建设。*Journal of Peacebuilding & Development*，17(2):239–243，2022年8月。[3] 让人民的声音得到重视 | 联合国网络电视，6月22日。[4] ASRSG Williams 与 1000 名利比亚人进行数字对话。https://unsmil.unmissions.org/asrsg-williams-conducts-digital-dialogue-1000-libyans，2021年1月。[5] Aviv Ovadya。基于桥接的排名。技术报告，哈佛肯尼迪学院贝尔弗科学与国际事务中心，2022年5月。[6] Tom Brown、Benjamin Mann、Nick Ryder、Melanie Subbiah、Jared D Kaplan、Prafulla Dhariwal、Arvind Neelakantan、Pranav S hyam、Girish Sastry、Amanda Askell、Sandhini Agarwal、Ariel Herbert-Voss、Gretchen Krueger、Tom Henighan、Rewon Child、Aditya Ramesh、Daniel Ziegler、Jeffrey Wu、Clemens Winter、Chris Hesse、Mark Chen、Eric Sigler、Mateusz Litwin、Scott Gray、Benjamin Chess、Jack Clark、Christopher Berner、Sam McCandlish、Alec Radford、Ilya Sutskever 和 Dario Amodei。语言模型是少样本学习者。载于神经信息处理系统进展，第33卷，第1877–1901页。Curran Associates, Inc.，2020年。[7] Stefan Wojcik、Sophie Hilgard、Nick Judd、Delia Mocanu、Stephen Ragain、Keith Coleman、M B Fallin Hunzaker 和 Jay Baxter。Birdwatch：群体智慧和桥接算法有助于理解并减少错误信息的传播。第24页。[8] Remesh——以量化规模获取定性见解。https://www.remesh.ai/。[9] Matthew J. Salganik 和 Karen E. C. Levy。维基调查：开放且可量化的社会数据收集。《公共科学图书馆·综合》，10(5):e0123483，2015年5月。[10] PSi。https://www.thepsiapp.com/。[11] Andrew Konya、Yeping Lina Qiu、Michael P Varga 和 Aviv Ovadya。面向多委托方—代理人对齐的引导、推断与优化。载于*NeurIPS 2022* 决策制定基础模型研讨会，2022年。

[12] Kimbra White、Nicole Hunter 和 Keith Greaves。促进审议：实用指南。2022年。[13] James Fishkin。当人民思考时的民主：审议与民主革新。美国哲学学会会刊，163(2):108–121，2019年。[14] Jordan Bilich、Michael Varga、Daanish Masood 和 Andrew Konya。通过包容性实现更快的和平：理解冲突地区人口的高效范式。第6页。[15] Michiel A. Bakker、Martin J. Chadwick、Hannah R. Sheahan、Michael Henry Tessler、Lucy Campbell-Gillingham、Jan Balaguer、Nat McAleese、Amelia Glaese，John Aslanides、Matthew M. Botvinick 和 Christopher S ummerfield。《微调语言模型，以寻求具有不同偏好的人类之间的共识》，2022年11月。[16] Tongshuang Wu、Ellen Jiang、Aaron Donsbach、Jeff Gray、Alejandra Molina、Michael Terry 和 Carrie J. Cai。《PromptChainer：通过可视化编程串联大型语言模型提示》，2022年3月。[17] Terrill Bouricius。《为什么混合两院制不适用于抽选》。《政治与社会》，46(3):435–451，2018年9月。[18] 经济合作与发展组织。《创新公民参与和新的民主制度：迎接审议浪潮》。经济合作与发展组织，2020年6月。