arXiv:2304.13556v1 [cs.HC] 22 Apr 2023

系统综述革命：一份旨在提升研究严谨性与包容性的宣言

Katja Rogers
k.s.rogers@uva.nl 阿姆斯特丹大学 阿姆斯特丹，荷兰

Katie Seaborn
seaborn.k.aa@m.titech.ac.jp
东京工业大学 东京，日本



图 1：Dall-E 根据提示词“一个机器人在由论文、数字和字母组成的海浪中溺水，由莫奈绘制”生成的图像（https://openai.com/dall-e-2/），呈现了研究人员在 2022 年有时所感受到的状态。

摘要

人机交互（HCI）领域正在逐渐成熟。系统综述是许多学科的基础组成部分，在各领域对人类知识作出贡献的过程中发挥着重要且往往不可或缺的作用。针对这一前景，我们认为，人机交互领域内部的研究元层面方法需要一场革命。首先，为支持二手研究中的知识综合，我们呼应此前关于提高一手研究报告严谨性的倡议。其次，鉴于人机交互领域中知识产生方式的多样性，我们必须以学术共同体的身份决定如何开展系统综述工作（即二手研究方法的严谨性与认识论包容性）。简而言之，我们的宣言是：通过仔细考量一手研究和二手研究的方法、预期与基础设施，我们需要制定并为一套兼具包容性与严谨性的标准留出空间，以支持人机交互领域的系统综述工作。我们呼吁所有系统综述革命者加入我们。

引文

Katja Rogers 和 Katie Seaborn。2023 年。《系统综述宣言：推动研究综合的严谨性与包容性》。载于 *2023年CHI人机交互系统中的人为因素会议扩展摘要（CHI EA '23）*。计算机协会，美国纽约州纽约市，第 424 号文章，第 1–11 页。https://doi.org/10.1145/3544549.3582733

> “我们有理由担心，数量以惊人速度日益增长的书籍，将使后世陷入一种如同罗马帝国覆灭后那些世纪般的野蛮状态。”
>
> Adrien Baillet，1685年（Ann Blair 引述） [9]

CCS概念

• **以人为中心的计算 → 人机交互（HCI）**；• **通用与参考 → 综述与概览**。

关键词

研究综合、系统综述、严谨性、文献、认识论

出版物

最终出版物可通过 ACM 获取，网址为 https://dl.acm.org/doi/10.1145/3544549.3582733。

为什么研究综合很重要



我们在人机交互（HCI）领域发表了大量论文——甚至可以说太多了？ACM数字图书馆（DL）中的论文数量就是一个例证（图2）。单个人，甚至一个团队，都无法跟上我们所产出的庞大研究量。然而，我们也需要综合这些文献。在一些领域，研究产出的热情步伐促使部分研究人员探索自动化。机器学习等现代计算技术 [6, 53, 82] 或许能够帮助我们跟上并综合各类出版物。但在人机交互领域，这是否必要或符合期望，或者我们是否需要重新思考知识生产与报告流程?

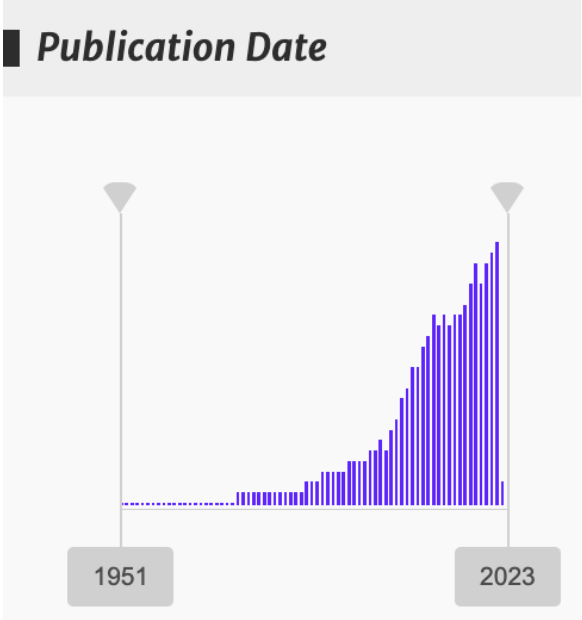


图 2：ACM Digital Library 中关键词“人机交互”按年份统计的结果。截至 2023年2月7日 14:24，共找到 658,883 条结果。仅 2000 年就有 8,519 条结果，2010 年为 27,009 条；2020 年为 35,183 条；2022 年为 37,215 条。

在其他较大（或更大）的领域中，Chu 和 Evans [18] 警告称，出版产出的增加可能导致“僵化”，因为新颖想法无法与最常被引用的论文所构成的根深蒂固的经典文献相抗衡。这可能给整个领域带来严重后果：“每年发表过多论文，可能导致停滞，而非推动 *[*知识创造*]*” [18]。在人机交互领域，已经出现了这一趋势的迹象：被引用次数最多的论文，其引用次数远高于平均水平，而论文每年获得的引用数量总体上正在下降 [66]。根据其他领域的情况，这表明在人机交互领域，新想法越来越难以突破并撼动既有观点。然而，在人机交互领域，我们实际上可能存在相反的问题，即由于我们以如此快的速度发表的论文的类型 而产生的问题。近年来，一些研究人员开始担心，我们过于关注新颖性 [5, 62]。以最近的（2022年）计算系统中的人为因素会议（CHI）论文集为例，搜索“novel”会得到602条结果，而论文总数为697篇。当86%的论文都被描述为“新颖”时，我们需要思考这对知识增益和共识构建意味着什么。作为唱反调者，我们可以说，本领域激励我们，甚至要求我们，源源不断地发表华而不实的短篇成果。我们没有投入精力开展严谨的渐进式研究，以通过多项研究确认现有证据，反而发现自己在追逐新颖性这根胡萝卜[1]。

这种状况所带来的影响值得我们谨慎而批判性地加以关注。过于强调新颖性，或许正是导致我们难以明确回答“我们迄今在人机交互领域究竟知道什么”的部分原因所在。 [41]这也使人机交互难以将自身置于其他研究领域之中并与之共同开展研究，同时限制了我们所进行的研究类型。对于可持续人机交互，我们认同DiSalvo等人 [24]的表述，并认为其同样适用于整个人机交互领域：“*[*为*]*避免重复造轮子，该领域有必要盘点已有知识，并确定由既有成果引发的重大未知问题或议题，以此作为未来工作的基础。”Whittaker等人 [84]也提出了类似的批评，指出人机交互社群“*[*过度强调*]* ‘彻底创新’，代价是无法形成共同的研究重点。”他们进一步指出，如果缺乏“这样的重点，就很难在既有工作基础上继续推进，客观比较不同的交互技术，也难以在理论发展方面取得进展。”这不仅是研究实践层面的问题；当我们作为一个研究领域无法提供清晰的指南或启示时，它也会产生实际影响。我们发现自己处于一个过渡性空间：我们都在开展研究并产出各种成果，但从外部观察的人可能很难看清整体图景。这样的外部观察者随后可能转向一个定义更加清晰的领域，使我们的工作无人认可、遭到忽视。从内部来看，我们可能会只见树木、不见森林，因而找不到明确的前进路径。研究综合能够厘清一个研究领域所开展的工作，不仅有助于他人，也同样有助于我们自己。

然而，重复研究、后续工作、更正与扩展，以及其他明确属于非新颖 形式的探究，仍然处于边缘地位，尽管早在十多年前就已有行动呼吁 [25, 42, 55, 85, 86]。鉴于研究实践中的更大趋势与问题，包括 p-hacking [38], 、复现 [2, 72] 和发表偏倚 [60]危机，以及预印本服务器激增所带来的不良影响 [1]，这具有严重的影响。这不仅是实验性或定量研究的问题，而这类研究通常被置于实证主义框架之中。我们认识到，并非所有人机交互研究项目都旨在实现概化或达成共识。尽管如此，其中许多项目，甚至可以说几乎所有项目，本身都蕴含着有价值的洞见，而这些洞见可以得到富有成效的综合。若要呈现人机交互研究的完整图景，知识综合工作就应当接纳所有认识论与方法论视角。

全球化的基于计算机的信息技术——我们这一学科的核心——为学术成果创造了新的驱动力，也带来了新的张力。应对这一现象的解决方案，或许可以在一定程度上通过拥抱慢科学 [77], 找到。然而，即使我们明天彻底停止发表成果——并拔掉互联网的插头——我们仍然需要筛选并综合迄今为止已发表的现有研究。我们中的许多人在人机交互领域正在承担这项任务，但几乎没有可遵循的指南或标准化做法。这并不是因为不存在指南或标准——它们确实存在 [13, 37, 45, 63, 67, 73, 75, 79, 80]。然而，这些指南和标准是以其他领域所发展（并重视）的工作为前提的，例如医学领域中的随机对照试验 [76]。人机交互固有的部分挑战在于，可获得的研究工作种类繁多（这甚至可能导致 Reeves [69] 以及 Fallman 和 Stolterman [27] 所称的“学科焦虑”）。与此密切相关的另一个关键挑战，是缺乏关于一般应如何开展研究综合，尤其是如何进行系统综述的共识：我们的领域尚未就我们对系统综述的期望，以及如何处理本领域所产出的不同类型知识展开明确的讨论。

[1]请牢记，我们的领域未必对“新颖性”有统一的理解。

这就是我们的宣言。我们提议发起一场由共同体推动的对话，以确定如何偏离我们典型的研究实践，从而在元层面支持研究综合。我们主张，对人机交互领域的研究综合而言，严谨且包容的系统综述方法是未来的发展方向。在这一节点，我们提出两个关键问题：*1)* 我们应如何对研究工作进行组织，使得研究综合能够建立在我们在 *CHI* 及相关领域所产出成果的丰富多样性之上并得以实践？其次，鉴于本领域产生的多种知识形式，我们相信并希望说服读者：我们必须作为一个共同体共同努力，针对在人机交互领域内规划、开展和报告研究综合，制定一套共享的研究实践：*2)* 当研究综合以多元性为基础时，它应呈现何种面貌：定量研究、定性研究、设计研究、民族志，以及交互式制品与系统的开发？我们在学术传统中提出这些担忧、挑战，以及以不同方式开展工作的愿望，以回应对文献“令人困惑且有害的过度丰富”的谴责；这种自我反思式的话语形式可以追溯到几个世纪以前 [9]。我们的目标是提醒自己关注更大的图景，并作为实践共同体的成员彼此重新定位。

基于以上内容，我们阐述了开展研究综合的理由、它为何对人机交互领域至关重要，以及我们对这一问题的回答为何可能不同于其他领域。接下来，我们将介绍研究综合的成熟方法论，重点讨论目前在大多数研究领域通行的全球标准：系统综述。随后，我们将提出依赖系统综述方法开展人机交互研究所涉及的关键问题，以及围绕其在主题、方法和认识论观点方面的丰富多样性所提出的倡议。最后，我们将呼吁以一种新颖的框架来理解研究综合：邀请亲爱的读者以及所有参与人机交互研究的人士共同加入。我们的目标是在今年的会议上开启一场对话，并期望由此推动未来研讨会、特别兴趣小组、委员会和/或合作项目的发展，最终建立一个致力于人机交互研究综合的实践共同体。通过汇聚我们共同体中众多学科的声音和认识论视角，我们希望在知识创造和方法论方面对本学科产生影响，并将这种影响回馈给更广泛的研究共同体。

启程台：什么是系统综述？或者：“但我有一张PRISMA流程图，这当然就说明它是系统的吧”

系统综述ystematic 是许多研究领域中研究综合的常见组成部分。在医学中，它们被认为是“不可或缺的”[44] “黄金标准” [59] 手段，用于就某一特定主题（通常是某种干预）在各项独立研究之间达成共识，从而支持以循证工作为基础的决策 [44]。然而，对于系统综述作为成果应是什么，以及作为过程应包含哪些内容，目前并没有明确且获得共识的定义 [56]。正如 Martinic 等人 [56]所解释的那样，“系统综述的 *[定义]* 模糊且含义不清，常使用诸如清晰、明确和系统等术语，却不作进一步阐释。”这给人机交互领域的我们带来了挑战，也带来了机遇：我们可能难以理解和应用系统综述将方法论应用于我们的工作，尽管在系统综述的历史与实践方面，我们还有很多东西需要向其他学科学习。然而，也许我们不需要原样采用所有这些方法，在某些情况下，尝试这样做甚至可能并不合适；相反，我们可以开辟一条新的前进道路。

理解与进行富有成效的偏离，第一步是给出定义。让我们先从系统综述不是什么说起：它不只是对某一主题、某一研究领域，或围绕某个研究问题或假设开展的既有工作的描述。它不是一份带注释的参考文献目录，在其中我们评论读过的论文。它也不是在没有先验计划的情况下临时开展的，尤其不是在过程中途改变程序，最终不得不调整研究问题以适应方法。它也不只是摘要；如果没有说明结果是如何构建的，仿佛施展魔法般从检索过程直接跳到结果，中间没有任何解释，它就不成立。它不是对经过筛选的作品进行叙述。

那么，系统综述是什么？换句话说，作为学术成果的产出与方法，它的本质是什么？我们从一对概念开始：一手研究与二手研究——它们之间的关系，以及二者之间的基本区别。一手研究是系统综述的素材。我们将一手研究定义为任何直接报告所收集和分析数据的论文[2]，例如，报告用户研究的论文。二手研究则更进一步：它是报告所收集和分析的一手研究论文样本的论文；系统综述就是二手研究的一个例子。我们认为，在人机交互领域进行研究综合时，沿用这些概念并无不妥。

掌握了这些基础概念后，我们应如何将作为一手研究的材料处理为作为二手研究的成果？遗憾的是，典型的系统综述流程结构存在的争议超出了理想范围 [56]。Martinic 等人 [56] 提出了以下组成部分，并按程序顺序列出：“*i)* 研究问题；*ii)* 进行检索的来源，并采用可复现的检索策略（列明数据库、检索平台/引擎、检索日期和完整的检索策略）；*iii)* 纳入和排除标准；*iv)* 选择（筛选）方法；*v) [对]* 纳入研究的质量/偏倚风险进行批判性评价；*vi)* 关于数据分析和综合的信息，以确保结果可复现。” Haddaway 和 Bilotta [36] 则比较了倡导通过系统综述开展循证研究的机构（例如 Cochrane 协作组织）所提出的要求。他们提出了三项基本标准：“*(i) [...]*方法应予以充分详细的描述，以实现完全的可重复性和可追溯性；*(ii) [...]* 采用系统方法识别和筛选相关的学术文献与灰色文献；*(iii) [...]* 对纳入研究的有效性（质量和可推广性）进行批判性评价，从而赋予更可靠的研究更大的权重。”考虑到我们领域的多元性，我们归纳出以下一般特征：(i) 预先制定并预注册的方案，即对计划中的综述程序进行完整记录，同时清晰、全面地阐述研究问题、检索流程、筛选流程、数据提取流程，以及.....的方式

[2]虽然我们在此使用“论文”一词，但并不是说论文本身就是“研究”。为了行文简洁，并考虑到大多数学术成果都以论文形式呈现，尤其是在系统综述和研究综合领域，我们使用“论文”一词。

质量评价（或说明省略质量评价的理由）；(ii) 数据分析和综合方法；以及 (iii) 透明地讨论检索和综合两方面的局限性，以及方法选择和结果方面局限性的讨论。

在 人机交互领域，综述要具备哪些要素才算“系统性”，仍是一个开放性问题。是否有必要进行预注册，尤其是在已经存在类似注册的情况下？每篇综述都需要评估一手研究的质量吗？是否必须使用某些工具或平台，例如 ACM数字图书馆或 IEEE Xplore？它们是发表一手研究和二手研究的基础平台，但也并非没有各自的怪癖和明显故障。此外，实践中应当如何 开展这些步骤，其具体做法同样尚不明确，且杂乱无章。以人机交互领域为例，对于选择单人筛选还是双重筛选所涉及的权衡，尚无共识，甚至没有人对此进行权衡——由一个人做决定是否足够，还是至少需要另一个人参与？其他人是否只需审查被拒绝的条目，以避免假阴性？能否将工作分配给不同的人？是否应该经历一个“震荡与规范”过程，让大家达成一致，或者采用某种评定者间信度指标？我们是否应该放弃可推广性，接受认识论上的多样性？是否应该采纳Braun和Clarke [10] 在其反思性 主题分析方法中倡导的理念，力求实用、灵活且透明？由此而言，我们能否有意义且恰当地借鉴其他方法论，为研究综合提供参考？这些只是近年来其他领域的研究人员一直在探索的众多方法论问题中的一小部分 [32, 52],，而其中任何一个问题单独来看，都已足以引发一系列关于人机交互研究情境的疑问与倡议。

系统综述。 /sɪstəˈmatɪk rɪ ˈvjuː/。“定义错误：术语 ’系统综述’ 未定义”。

人机交互，或许如此

现代形式的系统综述主要可以追溯至医学领域，其目标是综合多项随机对照试验的结果，从而更准确地估计特定干预的效果 [76]。当通过统计方法综合非常相似的研究中的效应量时，这被视为荟萃分析系统综述 [22]。“荟萃分析”一词在 人机交互中有时被用来指代不对效应量进行统计汇总的综述工作，这可能是对“荟萃”一词更字面化的理解，用于描述一篇报告一项或多项分析的论文，例如 [20, 83]。其他领域则调整了综合方法论，或创建了适合自身需求的方法，例如那些不开展（许多）随机对照试验的领域和子领域 [79, 80]。多个领域中这种平行的方法论演化，催生了大量密切相关但各不相同的综合方法和综述类型：范围综述、快速综述、映射综述、综述的综述、（最佳匹配）框架综合、混合方法综合，以及许多其他类型 [78]。这些方法的采用情况及其使用指南各不相同，对于其中哪些属于或不属于“系统性”方法，人们的看法也大相径庭。随着这些领域逐渐成熟，它们开始面对另一波论文洪流，这一次涉及的不是一手研究，而是二手研究 [44]。瀑布并不会在池塘边停下，而是继续向更远处层层倾泻：长期以来，学术研究文献中一直包含三级研究 [4, 23, 74] ，甚至偶尔还会出现“四级”研究的例子 [58]。我们没有理由认为人机交互不会如此；现在正是采取行动、寻找新前进道路的时候。

输入坐标：梳理开放性问题并回应质疑

The alt.chi网站希望投稿具有“争议性、敢于冒险并突破边界” [15]——那么，既然系统综述是一种成熟的方法论，甚至是黄金标准，既能凸显现有知识（“回顾性” [61]），也能从已有成果中创造新的知识形式（“前瞻性” [61]），我们为什么还要撰写关于系统综述的文章？这显然不是一个有争议的话题吧？然而，事实却并非如此：根据我们在提交和评审人机交互论文时的亲身经历，我们遇到过各种各样的期望和观点，包括：

- 系统综述作为一种高度依赖一手研究的二手研究形式，是否在HCI中占有一席之地，因为这类工作并不总能以传统意义上的新颖成果告终；
- 系统综述的用途 （提供某一子领域的客观且全面概述，还是提供带有主观观点的叙述，抑或提供能够回答非常具体问题的估计；建立共识，还是提供主观但有依据的观点），
- 应如何开展系统综述 （基于一系列具体指南；忽略或纳入定性研究；进行或不进行荟萃分析；进行或不进行批判性评价、双重筛选、使用数据提取表，或……），
- 能够且应当产生何种形式的知识 （“地图”与综合效应量估计；分类体系、理论或框架；新的研究问题与方向；新的一手研究、工具或原型），以及
- 基本术语和定义 （何时应将一项综述视为系统综述；什么是荟萃分析；等等）

让我们设想一位虚构的人机交互研究者，他看不出系统综述有任何益处，并认为系统综述是普罗克鲁斯忒斯式的：

普罗克鲁斯忒斯式的。 /prəʊˈkrʌstɪən/。属于、涉及或类似于普罗克鲁斯忒斯的做法（参见 Procrustes n.）；（因此）不顾自然差异或个体特性，强制实行统一或一致。

《牛津英语词典》

正如我们上文所述，在人机交互领域持有这样的立场是有理由的，因此我们为这一观点提供一个平台，并梳理可能引发的担忧。研究人员或许会合理地提出这样的疑问：系统综述会不会导致再也没人阅读原始论文？我们再次强调，通常不可能及时掌握人机交互领域的最新进展并阅读所有论文。很遗憾，时机已过。不过，尽管看起来我们并没有像应有的那样，以批判且全面的方式审视被引用的研究

深入研读 [54]，但回应说“反正没人读任何东西”可能过于简单化，也过于悲观。系统综述可能 确实会使引用从原始研究论文转移或分流。综述便于在概述性讨论中引用；若没有系统综述，同一批作者可能会转而引用几篇精心挑选的原始研究论文。然而，研究某个特定主题或在同一领域开展工作的研究人员，仍会直接引用最相关的论文——或者说，他们理应如此。尽管如此，我们承认，系统综述数量的增加可能会影响引文实践，尤其是当我们考虑到谁 在撰写它们（以及谁没有撰写），以及如何撰写时：“引用有其政治性”[19]。正如 Kumar 和 Karusala 所指出的：“如何书写对某项工作的描述同样重要，因为随着时间推移，这可能会扭曲甚至抹去其贡献”[46]。然而，一篇撰写得当的系统综述，应基于全面的检索策略和反思性的质量评估，搜集并呈现广泛且无偏的论文遴选结果。因此，它有助于减少我们引用和关注既有研究时产生的偏见，也就是说，它能够像科学方法一样发挥自我纠正作用。系统综述程序并非鼓励我们引用自己熟悉的研究（或研究者）——这些对象未必能体现该领域的多样性，反而可能只是反映我们的社交网络 [31], ；它能够拓宽视野，让引文实践更具包容性。此外，一篇撰写得当的系统综述本身就是对其收录的一手研究进行深入批判性反思和参与的一种形式。尽管它可能会“抢走”一些引用，但它本身应当引用原始研究，也很可能为其带来未来的引用[3]。

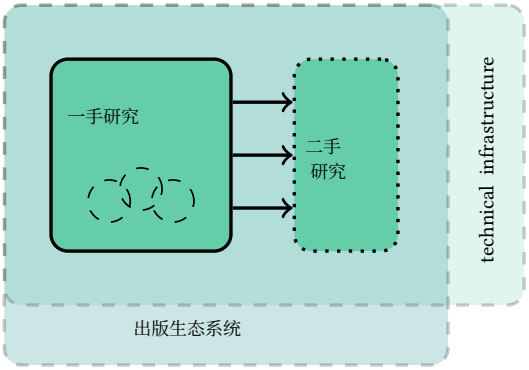


图 3：学术生态系统由基于广泛多样的方法和研究范式的一手研究构成，而这些研究随后可以在二手研究中得到综合。所有研究都由学术出版生态系统——例如出版场所、审稿人和会议（分）委员会——以及技术基础设施——例如数据库及其搜索功能——共同创造并塑造。

我们的想象研究者接下来可能会问：系统综述是否会使我们在综合结果中强行采用一种普罗克鲁斯忒斯规范，完全忽视每篇论文中那些美妙的差异？这或许确实如此。但也许这种差异并不总能帮助我们实现当下的目标。当我们试图基于该领域目前已有的研究，为一个具体问题寻求一个足够好的答案时，这些差异在元层面上或许并不总是有用或相关。事实上，将普罗克鲁斯忒斯的隐喻延伸到用户研究，可以说明为什么这些质疑不应成为问题。作为一个研究领域，我们通常不会在借助某项用户研究的结果来推断其对整个用户群体可能如何运作时，回避个体用户。我们基于单项用户研究，就具体研究问题提出启示和结论，并不是要界定或强行实施一种忽视个体参与者美妙差异的规范。相反，我们利用手头的资源，呈现可获得经验中的一个切片，然后转而采用其他方法，探索并展示我们在一次研究中无法触及的差异。同样，我们可以将系统综述与具体的分析方法结合起来，以得出结论，并获得细致入微且丰富、立足于具体情境的理解。

最后，我们转向 Blackwell 对人机交互这一研究领域的看法：或许，人机交互的目标不应是“构建并维护一套稳定的知识体系，而应成为创新的催化剂或源泉”。这意味着，我们作为人机交互研究人员，应从事“质疑、挑衅、颠覆且令人不适”的学术成果 [8]。这或许可以成为反对系统综述的一个论据，因为综合的目标往往是在我们领域的流变中求稳并找到坚实的立足点。尽管如此，Blackwell [8]也强调了“反思性实践”的重要性，而系统综述中的知识综合本身就能实现并构建这种实践。我们认为，系统综述虽在方法上各有不同，却有潜力既为人机交互领域知识生产奠定稳固基础，也推动其中的颠覆性实践。

描绘新轨迹：关键问题与倡议

Based 基于我们规划、开展和发表多项系统综述的经验（按照我们目前对该术语的理解，其中也包括一些不够系统的综述），我们在此提出在人机交互这一领域为回答本宣言提出的问题与呼吁而继续前行的过程中，需要认真应对的关键问题与倡议。这些问题涉及围绕人机交互中一手研究与二手研究的知识构建生态，而这一生态受到人机交互研究社群内部认识论多样性的塑造。然而，我们也预见，在人机交互出版生态系统中支持二手研究将面临摩擦（会议论文集与期刊及其审稿人、编辑/副主席和分委员会的期望与要求），技术基础设施方面也存在类似问题（例如现有且相关的数据库及其搜索功能），详见图 3。接下来我们将对这些问题进行梳理。

一手研究报告。一般而言，研究综合，尤其是系统综述，建立在一手研究报告的基础之上。然而，从整个人机交互领域来看，实证工作各不相同；我们的报告方式多种多样，有时也不够完善。这当然并非新颖的批评；我们呼应了其他研究人员的呼吁，他们正在批判性地反思

[3]我们还可以追问，鉴于引用既可能表明学术参与，也同样可能反映社会权力差异，引用何时才真正具有意义或价值。系统综述或许能帮助我们避开作为社会性动物对流行度指标的自然倾向，而这种倾向正是通过引用次数操作化的。

我们研究报告实践，例如，种族与族裔数据的报告 [14],脑信号实验数据 [68], 参与者补偿数据 [65], 定性研究中的评定者间可靠性 [57],具体测量工具 [71]和问卷[43, 47], 对自我决定理论的运用 [81],制品描述 [33], 以及推断统计 [12], , 这里只列举少数几个例子。这些问题的出现，部分可能源于页数限制或确保论文篇幅与其预期贡献相匹配的努力，但也可能是由于缺乏由共同体推动的标准化和教育。

这使二手研究中的研究综合变得更加复杂，因为研究结果难以比较和权衡。我们再次承认，这可能并不总是目标，但在人机交互领域往往确实如此。然而，如果无法以高度的严谨性或系统性综合研究结果，我们又如何指出哪些方法有效、哪些方法无效呢？好消息是，更多地采用现有的报告指南，也很可能通过使报告的一手研究更加全面且更具可比性，来助益二手研究。尽管如此，或许有必要考察现有的一手研究报告指南在多大程度上能够支持后续的二手研究。

此外，鉴于多年来人们一直在批评人机交互领域的报告方式，或许现在是时候考虑更明确地向人机交互领域的作者和审稿人介绍此类指南了。如今已经有迹象表明，一些会议正在为投稿流程增加更多结构。例如，自2021年起，至少有一个 ACM 会议要求作者注明“其论文的一手和二手贡献类型”（实证性—定性；实证性—定量；实证性—混合方法；制品—技术；制品—设计；理论性；或元研究[4]），以帮助将论文分配给匹配度合适的审稿人 [17]。我们还可以要求论文将某种结构化摘要作为补充材料的一部分，且根据贡献类型进行定制。这不仅能够支持未来的二手研究，也能通过提供对所开展工作的简明概述，帮助审稿人快速、轻松地理解研究内容，从而促进审稿过程本身。

*CHI*的认识论多样性。接下来，我们将探讨人机交互中不同的认知方式在综合一手研究时所发挥的作用。这既需要从人机交互一手研究中研究多样性的角度加以考量，也需要考量我们在二手研究中用于综合这些研究的多种方法。

对于人机交互这样领域广泛的学科，我们该如何着手对系统综述进行形式化？近年来，其他领域对混合方法综述的方法论进行了更为详细的探讨 [40, 70]。若要开展一项能够兼顾定量与定性方法研究结果的综述，我们可以将 Joanna Briggs研究所（JBI）关于混合方法系统综述的指南 [50] 作为涵盖部分方法的起点。目前，人机交互领域仅有少数综述采用了这些或相关指南。尽管如此，我们认为，混合方法取向的研究领域值得对此给予更多关注，这也能极大地助益于我们开展综合的方式。

但人机交互领域也包含更多立足于设计研究方法的路径，例如参与式设计和通过设计进行设计。Wolf 等人 [88] 将该领域描述为具有“内在张力 *[，即]* 工程与创意设计各自鲜明的实践和学科取向中所体现的张力。”我们认同这一观点，并且也强调，这“并非无法克服的冲突 *[.....]* 两种视角都合理” [88]。在进行综合时排除通过设计研究方法创造的知识，会使我们领域中相当重要的一部分边缘化，并妨碍我们全面了解人机交互实践。然而，据我们所知，目前还没有专门用于处理和综合通过设计研究方法创造的证据与知识的方法或指南。我们可能需要开发新方法，将这类工作纳入系统综述。我们呼吁熟悉各类方法的研究人员谨记：“任何关于严谨性的概念，都必须建立在‘充分理解每种方法的具体目的’的基础之上” [30] （Frauenberger 等人 [30] 引述 Fallman 和 Stolterman [27]）。

这些领域并不缺少探讨严谨性在不同人机交互研究路径中意味着什么，以及如何为综合目的评估此类工作的研究。例如，Wolf 等人[88] 概述了设计实践中旨在实现“设计严谨性”的特质，其中包括设计批评：“设计师对其设计判断及所参与活动所作的反思性、评估性和沟通性阐释”。同样，Zimmerman 等人 [89]提出的评估交互设计研究的标准或视角（过程、创新、相关性和可扩展性）也可能是开展综合的一种有用工具。对于参与式设计等研究路径，Frauenberger 等人 [30] 谈到，传统的实证主义严谨性理解需要重新诠释：“在后现代科学语境中，问责性与严谨性是通过辩论、批评和反思来实现的”，并主张“承认不同的认知方式” [30]。我们进一步拓展这一论点：我们不仅需要承认这些不同的认知方式，还需要开发方法，对不同的认知方式加以综合和整合。

二手研究报告。 或许，现有方法论指南中，与人机交互领域综合指南最为契合、可作为其参考方向的，是软件工程领域的相关工作（例如 Kitchenham 和 Charters [45] 的研究）、定性健康研究领域的工作（例如 Tong 等人 [79] 的 ENTREQ，或 Cooke 等人 [21] 的 SPIDER框架），以及最近 Topor 等人 [80] 针对非干预研究提出的 NIRO-SR（目前仍为预印本）。不过，我们坚信，人机交互领域还需要借鉴能够更明确地结合 定量与定性研究的综合方法：借用 Reeves[69], 的话，我们需要“更多地评述和反思人机交互中不同推理形式的全貌，并通过这种方式，更好地管理潜在相互竞争的学科视角如何彼此汇聚。”关于如何整合来自不同学科和方法论的证据，现有确实 相关指南（例如 [50]），尽管这类指南并不常见。然而，这些方法在人机交互领域的实际成效仍是一个悬而未决的问题；目前，它们在我们这一领域基本上还没有得到采用。

当CHI现有的系统综述在说明其方法时引用指南，它们主要引用 PRISMA [63]（例如， [7, 39, 51]）。具体而言，PRISMA流程图很受欢迎，因为它可以很好地呈现检索流程和筛选过程。然而，PRISMA指南整体上是为报告医学领域中的干预研究的系统综述和荟萃分析而制定的 [63]。大多数人机交互综述，即使声称遵循了PRISMA

[4]改编自 Wobbrock 和 Kientz [87]对人机交互领域研究贡献的分类。

指南！——实际上也没有回答PRISMA清单中的所有条目（甚至大多数条目） [63, 64]。例如，你有多久没见过CHI上的系统综述报告每项研究的质量评估和/或偏倚风险[5]，或报告整体证据确定性[6]？此外，由于医学界关注荟萃分析，PRISMA的若干条目是为统计综合方法设计的，而人机交互综述极少采用这些方法（例如探讨统计异质性的成因[7]），因此这些条目并不适用于我们能够开展的综述类型。PRISMA流程图可能有用，但就我们这一领域而言，这些指南在大多数情况下并不真正合适——至少当涉及综述核心的综合方法时，检索流程之后的部分并不合适。

在 ACM数字图书馆中快速搜索“系统综述”，会发现发表的系统综述数量急剧增加。这意味着现在正是一个重要时刻，我们需要**停下来**反思本领域系统综述所采用的方法。我们需要弄清楚，在综述工作中使用“系统性”一词时究竟意味着什么，以及我们对最佳实践有何期待。我们需要清晰、全面地报告方法，包括如何调整指南以供自己使用。我们需要更深入地探讨综合方法，并谨慎选择、命名和说明选择这些方法的理由。我们也可以考虑将结构化摘要作为二手研究的补充材料（例如，Haddaway 等人 [37] 的 ROSES 可直接借用，也可根据人机交互研究的需要加以调整）。

Oulasvirta 和 Hornbæk [62] 提出，人机交互领域需要更多“将实证发现与技术设计联系起来的概念性贡献”，以使我们的研究发现能够付诸实践，并创造“整合型知识。”我们认为，如果投入精力制定并维护专为人机交互领域设计的综述综合及其报告指南和标准，就能提升人机交互作为一个领域所能作出的概念性贡献。如果我们将人机交互视为一个由其问题解决能力 [62], 所界定的领域，那么严格开展的系统综述就能直接帮助我们改善他们提出的若干重要问题解决标准：它能够帮助我们更好地理解解决方案的可迁移性，并增强我们对其的信心。

会议与分委员会。 Marshall 等人 [55] 感叹，人机交互领域鲜有明确的出版形式能够鼓励批判性讨论：“主要的 *[出版场所]* 都没有针对已发表文章进行批判性回应的形式 *[.....]* 人机交互领域的研究一经发表，几乎就不会引发任何批判性讨论。”相较之下，批判性讨论更可能发生在社交媒体、Slack 工作区、Discord 频道以及其他非正式场所。系统综述可以严谨而规范地发挥批判性讨论的作用，并让整个实践共同体都能参与其中。然而，目前也没有明确的发表渠道。或许在人机交互领域，唯一明确欢迎综述（“综述论文”）的出版场所是 ACM Computing Surveys (CSUR) 期刊，但其作者指南 [3]并未提及系统性要求。

作为“该学科的旗舰会议” [49]，CHI 仅有一个分委员会——健康——在其介绍中提及（系统）综述是一种方法 [16]。即便是自称“认识论多元主义 *[且]* 欢迎多种观点、方法和贡献” [16] 的分委员会，也可能招募不把系统综述视为一种方法论本身的副主席和审稿人，并可能仅因这一点就倾向于拒稿。就综述而言，整个人机交互领域的审稿人对其抱有截然不同的期待，方法学专长也大相径庭；若能达成更多共识，将大有裨益。

认为系统综述无法产出“新颖”成果的看法，或许是造成这一问题的部分原因。例如，TOCHI期刊指出，他们“很少发表*[...]* 综述论文，除非其具有重大的原创性贡献。” 我们指出，综述完全可以通过综合形成原创性贡献，例如分类体系等中层次知识 [11]。何种贡献应被视为“重大”，以及对现有研究的系统综述是否、或何时应被视为“原创性贡献”，都是TOCHI可以为潜在作者进一步阐明的问题，也确实是我们作为一个领域应当讨论的问题。

基础设施：数字图书馆与机器学习方法。 我们的数字图书馆文档记录不完善，评估也寥寥无几。结果可能随时间发生很大变化。有时这是意料之中的（例如，随着发表的研究增多，数量也会增加）；但有时数量也会因元数据调整而减少[8]。出版物数据库中的元数据常有错误，因此未必可靠 [28, 29]。此外，数据库的覆盖范围并不总是十分明确，而且可能因机构访问权限而异——例如，“*Web of Science*核心合集”会根据大学订阅包含不同的子数据集 [48]。这使系统综述的一项基本目标——即结果应当能够复现——变得相当难以实现。在其他领域，在多个数据库中进行检索被视为最佳实践。或许我们需要考虑在数天内进行多次检索，以尽量减轻数据库波动的影响。不过，我们或许也需要重新审视，或更明确地说明系统综述需要满足什么条件才算“可复现”：我们所说的复现结果究竟是什么意思？例如，只要报告了搜索查询本身，以及每个步骤中找到的论文记录，我们或许就不应要求检索结果数量完全相同，因为数据库的一致性并不可靠。

不过，要设计能够在不同数据库之间具有可比性的多组检索策略，还存在其他问题。数据库采用各种不同的关键词和筛选选项，而且相关文档往往不够完善。若能提供指导，说明如何在例如 ACM 和 Scopus 之间创建可比的检索策略，将极大有益于我们领域的综合研究。现有的ACM数字图书馆教程不足以满足这一目的，而联系 ACM数字图书馆团队咨询数据库和检索方面的具体问题也没有成效。一种选择是与出版商合作开展参与式设计项目，由我们作为目标“最终用户”，引导这些

[5]“条目 *18* [64] 。报告对每项纳入研究的偏倚风险评估” “条目*15*。描述用于评估某个结局的证据总体确定性（或可信度）的任何方法” [64], 例如，通过用于评估综述中证据质量的GRADE框架 [35]。 [7]“条目*13e*。描述用于探究研究结果之间异质性的可能原因（如亚组分析、元回归）的任何方法” [64]

[8]有些日子，数据库就是会出现故障：有一次令人难忘的经历是，在同一天内，我们发现ACM数字图书馆报告0，然后是 200+，然后是 0，最后是 500+ 条结果，而这些结果都来自同一次搜索。

系统朝着更有助于系统综述工作的方向发展。另一个值得考虑的选择是自动化。随着人工智能和机器学习的发展，围绕数据库和出版物检索的数字基础设施如今也包含用于（半）自动化检索（例如 Research Rabbit[9]）或筛选（例如 ASReview[10]）的工具。这些工具或许有助于梳理该领域，但在正式的系统综述中，能够并且应该在多大程度上使用它们，仍是一个悬而未决的问题——尤其是数据源的具体内容及其更新频率往往没有明确说明。或许，参与式设计方法仍可作为一个有益的起点。

最后，我们注意到，目前关于与人机交互相关的哪些类型出版物分布在哪些数据库中的信息十分有限。Gusenbauer [34]创建了涵盖广泛学术数据库的基于学科的覆盖图，为我们提供了初步线索。然而，人机交互并未被纳入该学科覆盖图；或许值得专门为人机交互创建数据库的学科覆盖图。这将使我们更清楚地了解哪些人机交互研究可以在哪些数据库中找到，并为针对具体研究问题选择数据库提供指导。

ENGAGE：呼吁在人机交互领域开展研究综合

As 作为一个研究群体，我们需要共同努力，决定应采取哪些行动，以建立一套既严谨、又包容我们在人机交互领域所开展工作多样性的标准。在这份宣言中，我们无意规定具体做法，但基于目前的讨论，我们确实提出了一些值得努力的方向：

- 对系统综述应包含的内容、系统综述所期望和可能产生的结果，以及在探索不同研究范式（定量、定性、混合方法以及设计研究方法）产生的多样证据时，系统综述可以采用的形式形成共识
- 对我们希望在二手研究方法中推广的最佳实践形成共识：双重筛选、数据提取、批判性评价、方案制定和预注册等，具体而言，是通过就标准达成一致（例如，针对一手研究的批判性评价：采用何种评价以及何时进行评价）
- 发掘与人机交互工作相关的数字图书馆（例如查询过滤器）及其涵盖的内容
- 为我们的出版生态系统提供更完善的基础设施：现在是否应设立一个负责研究综合和元科学的分委员会或专题方向？我们是否应要求一手研究和/或二手研究采用结构化摘要或核对清单？
- 对一手研究论文中报告的交互式工件进行充分描述和/或提供访问权限，以支持对这些工件开展研究综合
- 探索动态综述的设计与应用 [26]——作为交互系统，人机交互领域的专业知识在此尤其有益

[9]https://www.researchrabbit.ai/，最后访问日期：2022年11月23日，https://asreview.nl/，最后访问日期：2022年12月15日

我们的目标是开启一场讨论，汇集研究人员对于系统综述应发挥何种作用、应呈现何种形式，以及系统综述目前在 CHI 社群乃至整个更广泛的人机交互领域中如何受到评价和接纳的不同经验与观点。

致谢

非常感谢 Maximilian Altmeyer 对这篇 alt.chi 论文早期草稿提出的意见，也感谢我们所有同事多年来围绕这些主题展开的诸多热烈讨论。

参考文献

[1] 2020。预印本的兴起。*Nature Cancer* 1, 11（2020年11月），1025–1026。https://doi.org/10.1038/s43018-020-00151-y 编号：11 出版商：自然出版集团。[2] 2022。重复研究是实现概化的关键。*Nature Communications* 13, 1（2022年11月）。https://doi.org/10.1038/s41467-022-34748-x [3] ACM Computing Surveys (CSUR)。2022年。CSUR 作者指南。2022年11月23日检索自 https://dl.acm.org/journal/csur/author-guidelines [4] Edoardo Aromataris、Ritin Fernandez、Christina M. Godfrey、Cheryl Holly、Hanan Khalil 和 Patraporn Tungpunkom。2015年。总结系统综述：伞式综述方法的方法学发展、实施与报告。*International Journal of Evidence-Based Healthcare* 13, 3（2015年9月），132 140 10 1097 0000000000000055 – 。https://doi.org/ 。/xeb。[5] Mara Balestrini、Yvonne Rogers 和 Paul Marshall。2015年。公民参与式人机交互：新颖性与社会影响之间的张力。载于*2015*年英国人机交互会议论文集（英国林肯郡林肯）（*British HCI '15*）。计算机协会，纽约州纽约，美国，35–36。https://doi.org/10.1145/2783446.2783590[6] Elaine Beller、Justin Clark、Guy Tsafnat、Clive Adams、Heinz Diehl、Hans Lund、Mourad Ouzzani、Kristina Thayer、James Thomas、Tari Turner、Jun Xia、Karen Robinson 和 Paul Glasziou。2018年。通过系统综述自动化取得进展：系统综述自动化国际协作组织（ICASR）的原则。*Systematic Reviews* 7, 1（2018年5月）。https://doi.org/ 10 1186 13643 018 0740 7 。/s - - - [7] Joanna Bergström、Tor-Salve Dalsgaard、Jason Alexander 和 Kasper Hornbæk。2021年。如何评估虚拟现实中的对象选择与操控？来自20年研究的指南。载于*2021*年*CHI*计算系统人为因素会议论文集（日本横滨）（*CHI '21*）。计算机协会，纽约州纽约，美国，文章533，20页。https://doi.org/10.1145/ 3411764.3445193[8] Alan F. Blackwell。2015年。人机交互作为跨学科领域。载于第*33*届年度计算机系统中的人为因素*ACM*会议扩展摘要（韩国首尔）（*CHI EA '15*）。计算机协会，纽约州纽约，美国，503 516 10 1145 2702613 2732505 – 。https://doi.org/ 。/ [9] 。Ann Blair。2003。应对信息过载的阅读策略，约 1550—1700 年。*Journal of the History of Ideas* 64, 1（2003年1月），11。https://doi.org/10. 2307/3654293[10] Virginia Braun 和 Victoria Clarke。2019年。反思反思性主题分析。*Qualitative Research in Sport, Exercise and Health* 11, 4（2019年8月），589–597。https://doi.org/10.1080/2159676X.2019.1628806 出版商：Routledge _电子印本：https://doi.org/10.1080/2159676X.2019.1628806。[11] Frederik Brudy、Christian Holz、Roman Rädle、Chi-Jui Wu、Steven Houben、Clemens Nylandsted Klokmose 和 Nicolai Marquardt。2019年。跨设备分类：跨多个设备的交互综述、机遇与挑战。载于*2019*年*CHI*计算系统中的人为因素会议论文集（苏格兰格拉斯哥）（*CHI '19*）。计算机协会，纽约州纽约，美国，1 28 10 1145 3290605 3300792 – 。https://doi.org/ 。/ [12] 。2007。Paul Cairns。。人机交互……并非应有的样子：人机交互研究中的推断统计。载于*HCI 2007*：第*21*届英国人机交互小组年会，英国兰卡斯特大学。1–7。[13] Mhairi Campbell、Joanne E McKenzie、Amanda Sowden、Srinivasa Vittal Katikireddi、Sue E Brennan、Simon Ellis、Jamie Hartmann-Boyce、Rebecca Ryan、Sasha Shepperd、James Thomas、Vivian Welch 和 Hilary Thomson。2020年。系统综述中的无荟萃分析的综合（SWiM）：报告指南。*BMJ*（2020 6890 10 1136 6890 年1月），l 。https://doi.org/ 。/ [14] bmj.l Yiqun T. Chen、Angela D. R. Smith、Katharina Reinecke 和 Alexandra To。2022年。在人机交互领域收集和报告种族与族裔数据。载于*2022*年计算机系统中的人为因素会议扩展摘要（美国路易斯安那州新奥尔良）（*CHI EA '22*）。计算机协会，纽约州纽约，美国，文章327，8页。https://doi.org/10.1145/3491101.3519685

[15] CHI 2023 alt.chi 委员会。2022年。CHI 2023网站上的alt.chi页面。2022年9月29日检索自 https://chi2023.acm.org/for-authors/alt-chi/ [16] CHI 2023 alt.chi 委员会。2022年。选择分委员会；CHI 2023网站。2022年11月23日检索自 https://chi2023.acm.org/subcommittees/ selecting-a-subcommittee/ [17] CHI PLAY 2021 委员会。2021年。CHI PLAY 2021网站上的完整论文页面。2022年11月23日检索自 https://chiplay.acm.org/2021/full-papers/ [18] Johan S. G. Chu 和 James A. Evans。2021年。大型科学领域中规范性进展的放缓。《美国国家科学院院刊》118, 41（ 2021 10 1073 2021636118 年10月）。https://doi.org/ 。/pnas。引文[19]
正义集体、Gabriela Molina León、Lynn Kirabo、Marisol Wong-Villacres、Naveena Karusala、Neha Kumar、Nicola Bidwell、Pedro Reynolds-Cuéllar、Pranjal Protim Borah、Radhika Garg、Sushil K. Oswal、Tee Chuanromanee 和 Vishal Sharma。2021年。沿着引文正义的轨迹：批判性审视人机交互领域的知识生产。载于*2021*年计算机支持的协同工作与社会计算会议伴随出版物（美国，线上活动）（*CSCW '21*）。计算机协会，纽约州纽约，美国， 360 363 10 1145 3462204 3481732 – 。https://doi.org/ 。/ 。[20] Mark Colley、Taras Kränzle 和 Enrico Rukzio。2022。基于荟萃分析的人机交互领域无障碍相关出版物分布。载于*2022*年计算机系统中的人为因素会议扩展摘要（美国路易斯安那州新奥尔良）（*CHI EA '22*）。计算机协会，纽约州纽约，美国，文章 299, 28 10 1145 3491101 3519701 页。https://doi.org/ 。/ 。[21]
Alison Cooke、Debbie Smith 和 Andrew Booth。2012。超越PICO：用于定性证据综合的SPIDER工具。《定性健康研究》22, 10（ 2012 1435 1443 10 1177 1049732312452938 年7月），– 。https://doi.org/ 。/[22]
Harris Cooper。2015。研究综合与荟萃分析：分步方法。第 2 卷。Sage 出版社。[23] Fabio Q.B. da Silva、André L.M. Santos、Sérgio Soares、A. César C. França、Cleviton V.F. Monteiro 和 Felipe Farias Maciel。2011年。软件工程领域六年的系统性文献综述：更新后的三级研究。《信息与软件技术》53, 9（2011年9月），899–913。https://doi.org/10.1016/j.infsof. 2011.04.004[24] Carl DiSalvo、Phoebe Sengers 和 Hrönn Brynjarsdóttir。2010年。绘制可持续人机交互的图景。载于*SIGCHI*计算系统人为因素会议论文集（美国佐治亚州亚特兰大）（*CHI '10*）。计算机协会，纽约州纽约，美国，1975–1984。https://doi.org/10.1145/ 1753326.1753625[25] Florian Echtler 和 Maximilian Häußler。2018年。开源、开放科学与人机交互领域的重复危机。载于*2018*年计算机系统中的人为因素会议扩展摘要（加拿大魁北克省蒙特利尔）（*CHI EA '18*）。计算机协会，纽约州纽约，美国，1–8。https://doi. org/ 10 1145 3170427 3188395 。/ 。[26]
Julian H. Elliott、Anneliese Synnot、Tari Turner、Mark Simmonds、Elie A. Akl、Steve McDonald、Georgia Salanti、Joerg Meerpohl、Harriet MacLehose、John Hilton、David Tovey、Ian Shemilt、James Thomas、Thomas Agoritsas、John Hilton、Caroline Perron、Elie Akl、Rebecca Hodder、Charlotte Pestridge、Lauren Albrecht、Tanya Horsley、Joanne Platt、Rebecca Armstrong、Phi Hung Nguyen、Robert Plovnick、Anneliese Arno、Noah Ivers、Gail Quinn、Agnes Au、Renea Johnston、Gabriel Rada、Matthew Bagg、Arwel Jones、Philippe Ravaud、Catherine Boden、Lara Kahale、Bernt Richter、Isabelle Boisvert、Homa Keshavarz、Rebecca Ryan、Linn Brandt、Stephanie A. Kolakowsky-Hayner、Dina Salama、Alexandra Brazinova、Sumanth Kumbargere Nagraj、Georgia Salanti、Rachelle Buchbinder、Toby Lasserson、Lina Santaguida、Chris Champion、Rebecca Lawrence、Nancy Santesso、Jackie Chandler、Zbigniew Les、Holger J. Schünemann、Andreas Charidimou、Stefan Leucht、Ian Shemilt、Roger Chou、Nicola Low、Diana Sherifali、Rachel Churchill、Andrew Maas、Reed Siemieniuk、Maryse C. Cnossen、Harriet MacLehose、Mark Simmonds、Marie-Joelle Cossi、Malcolm Macleod、Nicole Skoetz、Michel Counotte、Iain Marshall、Karla Soares-Weiser、Samantha Craigie、Rachel Marshall、Velandai Srikanth、Philipp Dahm、Nicole Martin、Katrina Sullivan、Alanna Danilkewich、Laura Martínez García、Anneliese Synnot、Kristen Danko、Chris Mavergames、Mark Taylor、Emma Donoghue、Lara J. Maxwell、Kris Thayer、Corinna Dressler、James McAuley、James Thomas、Cathy Egan、Steve McDonald、Roger Tritton、Julian Elliott、Joanne McKenzie、Guy Tsafnat、Sarah A. Elliott、Joerg Meerpohl、Peter Tugwell、Itziar Etxeandia、Bronwen Merner、Alexis Turgeon、Robin Featherstone、Stefania Mondello、Tari Turner、Ruth Foxlee、Richard Morley、Gert van Valkenhoef、Paul Garner、Marcus Munafo、Per Vandvik、Martha Gerrity、Zachary Munn、Byron Wallace、Paul Glasziou、Melissa Murano、Sheila A. Wallace、Sally Green、Kristine Newman、Chris Watts、Jeremy Grimshaw、Robby Nieuwlaat、Laura Weeks、Kurinchi Gurusamy、Adriani Nikolakopoulou、Aaron Weigl、Neal Haddaway、Anna Noel-Storr、George Wells、Lisa Hartling、Annette O'Connor、Wojtek Wiercioch、Jill Hayden、Matthew Page、Luke Wolfenden、Mark Helfand、Manisha Pahwa、Juan José Yepes Nuñez、Julian Higgins、Jordi Pardo Pardo、Jennifer Yost、Sophie Hill 和 Leslea Pearson。2017年。动态系统综述：1. 引言——为何、何事、何时以及如何。《临床流行病学杂志》 91 2017 23 30 10 1016 2017 08 010 （年11月），– 。[27] https://doi.org/ 。/j.clinepi。。Daniel Fallman 和 Erik Stolterman。2010。在交互设计研究中确立严谨性与相关性的标准。《数字创意》21, 4（2010年12月）， 265 272 10 1080 14626268 2010 548869 – 。https://doi.org/ 。/。。[28]
Fiorenzo Franceschini、Domenico Maisano 和 Luca Mastrogiacomo。2016 年。Scopus 和 Web of Science 中数据库错误的实证分析与分类。《信息计量学杂志》10，4（2016 年 11 月），933–953。https://doi.org/10.1016/j.joi. 2016.07.003[29] Fiorenzo Franceschini、Domenico Maisano 和 Luca Mastrogiacomo。2016 年。Scopus 中的错误/恐怖博物馆。《信息计量学杂志》10，1（2016 年 2 月），174–182。https://doi.org/10.1016/j.joi.2015.11.006[30] Christopher Frauenberger、Judith Good、Geraldine Fitzpatrick 和 Ole Sejer Iversen。2015 年。追求参与式设计中的严谨性与问责性。《国际人机研究杂志》74（2015 年 2 月），93–106。https: //doi.org/ 10 1016 2014 09 004 。/j.ijhcs.。。Riccardo Gallotti 和 Manlio De [31]
Domenico。2019 年。同质性与学术声誉对诺贝尔奖得主提名和遴选的影响。《科学报告》9，1（2019 年 11 月），17304。https://doi.org/10.1038/s41598-019-53657-6 编号：1 出版者：自然出版集团。[32] Gerald Gartlehner、Lisa Affengruber、Viktoria Titscher、Anna Noel-Storr、Gordon Dooley、Nicolas Ballarini 和 Franz König。2020 年。单名评审者进行摘要筛选会遗漏 13% 的相关研究：一项基于众包的随机对照试验。《临床流行病学杂志》121（2020 年 5 月），20–28。https://doi.org/ 10 1016 2020 01 005 。/j.jclinepi.。。Kathrin Gerling 和 Max V. [33]
Birk。2022。严谨性与可复现性反思：迈向实验游戏研究中制品描述的社区标准。载于《2022 年计算机人机交互与游戏年度研讨会扩展摘要》（德国不来梅）（CHI PLAY '22）。计算机协会，美国纽约州纽约，266–267。https://doi.org/ 10 1145 3505270 3558360 。/。[34]
Michael Gusenbauer。2022。搜索最有可能找到目标的地方：比较 56 个书目数据库的学科覆盖范围。《科学计量学》127，5（5 月 2022 2683 2745 10 1007 11192 022 04289 7 ），–。https://doi.org/ 。/[35]
s---Gordon H Guyatt、Andrew D Oxman、Gunn E Vist、Regina Kunz、Yngve Falck-Ytter、Pablo Alonso-Coello 和 Holger J Schünemann。2008 年。GRADE：关于证据质量和推荐强度分级的新兴共识。《BMJ》336，7650（2008 年 4 月），924–926。https://doi.org/10.1136/bmj.39489.470347.ad [36] Neal R. Haddaway 和 Gary S. Bilotta。2016 年。系统综述：区分事实与虚构。《国际环境》92-93（2016 年 7 月），578–584。https: //doi.org/ 10 1016 2015 07 011 。/j.envint.。。Neal R. [37]
Haddaway、Biljana Macura、Paul Whaley 和 Andrew S. Pullin。2018 年。ROSES 系统证据综合报告标准：环境系统综述和系统地图的计划与实施标准表格、流程图及描述性摘要。《环境证据》7，1（2018 年 3 月 2018 10 1186 13750 018 0121 7 ）。https://doi.org/ 。/s---Megan L. Head、Luke Holman、Rob Lanfear、Andrew T. Kahn 和 Michael D. Jennions。2015 年。科学研究中 p 值操纵的程度及其后果。《公共科学图书馆生物学》13，3（2015 年 3 月），e1002106。https://doi.org/10.1371/journal.pbio.1002106 出版者：公共科学图书馆。[39] Teresa Hirzle、Maurice Cordts、Enrico Rukzio、Jan Gugenheimer 和 Andreas Bulling。2021 年。对使用 SSQ 衡量虚拟现实头戴式显示器中一般不适的批判性评估。载于《2021 年 CHI计算系统人为因素会议论文集》（日本横滨）（CHI '21）。计算机协会，美国纽约州纽约，文章 530，14 页。https://doi.org/ 10 1145 3411764 3445361 。/。[40]
Quan Nha Hong、Pierre Pluye、Mathieu Bujold 和 Maggy Wassef。2017 年。汇聚式与序贯式综合设计：对定性和定量证据系统综述实施与报告的影响。《系统综述》 6, 1 2017 10 1186 13643 017 0454 2 （3 月 ）。https://doi.org/ 。/[41]
s---Kasper Hornbæk。2015。我们必须在人机交互研究中更多地犯错。《交互》 22, 6 2015 20 21 10 1145 2833093 （10 月 ），–。https://doi.org/ 。/[42]
Kasper Hornbæk、Søren S. Sander、Javier Andrés Bargas-Avila 和 Jakob Grue Simonsen。2014 年。一次就够了吗？论人机交互中重复研究的范围与内容。载于《SIGCHI计算系统人为因素会议论文集》（加拿大安大略省多伦多）（CHI '14）。计算机协会，美国纽约州纽约，3523–3532。https: //doi.org/ 10 1145 2556288 2557004 。/。[43]
Nathan G.J. Hughes、Josephine R. Flockton 和 Paul Cairns。2023 年。共同成长：对玩家动机问卷 15 年测量透明度的分析。《国际人机研究杂志》169（2023 年 1 月 2023 102940 10 1016 2022 102940 ），。https://doi.org/ 。/[44]
j.ijhcs.。。John P.a. Ioannidis。2016。冗余、误导性和相互冲突的系统综述与荟萃分析的大规模生产。《米尔班克季刊》94，3（2016 年），485–514。https://doi.org/10.1111/1468-0009.12210 _电子版：https://onlinelibrary.wiley.com/doi/pdf/10.1111/1468-0009.12210。[45] Barbara Kitchenham 和 Stuart Charters。2007 年。软件工程中开展系统文献综述的指南。技术报告。EBSE技术报告，第 2.3 版。。[46] Neha Kumar 和 Naveena Karusala。2021 年。在人机交互中勇于践行引文正义。载于《2021 年 CHI计算系统人为因素会议扩展摘要》（日本横滨）（CHI EA '21）。计算机协会，美国纽约州纽约，文章 11，9 页。https://doi.org/

10.1145/3411763.3450389[47] Effie L.-C. Law、Florian Brühlmann 和 Elisa D. Mekler。2018 年。游戏体验问卷（GEQ）的系统综述与验证：对引文与报告实践的启示。载于*2018* 年计算机人机交互与游戏年度研讨会论文集（澳大利亚维多利亚州墨尔本）（*CHI PLAY '18*）。计算机协会，美国纽约州纽约，257–270。https: //doi.org/ 10 1145 3242671 3242683 。/。Weishu Liu。。本研究的数据源是 Web of Science[48] 2019核心合集吗？还不够。《科学计量学》121，3（2019 年 9 月），1815–1824。https://doi.org/10. 1007/s11192-019-03238-1[49] Yong Liu、Jorge Goncalves、Denzil Ferreira、Bei Xiao、Simo Hosio 和 Vassilis Kostakos。2014 年。CHI 1994–2013：通过共词分析绘制二十年思想进步图谱。载于*SIGCHI*计算系统人为因素会议论文集（加拿大安大略省多伦多）（*CHI '14*）。计算机协会，美国纽约州纽约，3553–3562。https://doi.org/10. 1145/2556288.2556969[50] Lucylynn Lizarondo、Cindy Stern、Judith Carrier、Christina Godfrey、Kendra Rieger、Susan Salmond、Joao Apostolo、Pamela Kirkpatrick 和 Heather Loveday。2020 年。第 8 章：混合方法系统综述。载于 *JBI*证据综合手册。JBI。https://doi.org/10.46658/jbimes-20-09 可从 https: // synthesismanual.jbi.global 获取。[51] Cayley MacArthur、Arielle Grinberg、Daniel Harley 和 Mark Hancock。2021 年。你让我感到不适：虚拟现实研究如何考量性别与晕动症的系统综述。载于*2021* 年 *CHI*计算系统人为因素会议论文集（日本横滨）（*CHI '21*）。计算机协会，美国纽约州纽约，文章 401，15 页。https: //doi.org/ 10 1145 3411764 3445701 。/。Kamal R Mahtani、Carl Heneghan 和 Jeffrey Aronson[52]。2019 年。系统综述中的研究筛选应采用单次筛选还是双重筛选？《BMJ循证医学》 25, 4 2019 149 150 10 1136 2019 111269（2019 年 11月），–。https://doi.org/ 。/bmjebm–Iain J. Marshall 和 Byron C. Wallace[53]。2019。迈向系统综述自动化：在研究综合中使用机器学习工具的实用指南。《系统综述》 8, 1 2019 10 1186 13643 019 1074 9（7 月 ）。https://doi.org/ 。/s—Joe Marshall、Conor Linehan、Jocelyn Spence 和 Stefan Rennick Egglestone[54]。2017 年。对先前工作的随意引文会带来糟糕人机交互研究的风险。载于*2017* 年 *CHI*计算系统人为因素会议扩展摘要论文集（美国科罗拉多州丹佛）（*CHI EA '17*）。计算机协会，美国纽约州纽约，827–836。https://doi.org/10.1145/ 3027063.3052751[55] Joe Marshall、Conor Linehan、Jocelyn C. Spence 和 Stefan Rennick Egglestone。2017 年。一点尊重：人机交互忽视其他学科的四个案例研究。载于 *2017* 年 *CHI*计算系统人为因素会议扩展摘要论文集（美国科罗拉多州丹佛）（*CHI EA '17*）。计算机协会，美国纽约州纽约，848–857。https://doi.org/10.1145/ 3027063.3052752[56] Marina Krnic Martinic、Dawid Pieper、Angelina Glatt 和 Livia Puljak。2019 年。在系统综述概览、荟萃流行病学研究和教科书中使用的系统综述定义。《BMC医学研究方法学》 19，1（2019 年 11 月 2019 10 1186 12874 019 0855 0）。https://doi.org。/s—Nora McDonald、Sarita Schoenebeck 和 Andrea Forte[57]。2019 年。定性研究中的可靠性与评定者间可靠性：计算机支持的协同工作与人机交互实践的规范和指南。《ACM人机交互汇刊》3，计算机支持的协同工作，文章 72（11 月 2019 23 10 1145 3359174），页码。https://doi.org。/ Alexios-Fotios A. Mentis、Efthimios Dardiotis、Vasiliki Efthymiou 和 George P. Chrousos[58]。2021 年。神经系统疾病的非遗传风险因素、保护因素和生物标志物：伞式综述的荟萃伞式系统综述。《BMC医学》 19, 1 2021 10 1186 12916 020 01873 7（2021 年 1 月）。https://doi.org。/s—Robert Andrew Moore、Emma Fisher 和 Christopher Eccleston[59]。2022。。系统综述尚未代表证据的“黄金标准”：一篇立场论文。《欧洲疼痛杂志》 26, 3 2022 557 566 10 1002 1905（ ），–。https://doi.org。/ejp。电子版：https://onlinelibrary.wiley.com/doi/pdf/10.1002/ejp.1905。。[60]《自然·人类行为》。2019 年。无证据的重要性。《自然·人类行为》3，3（2019 年 3 月），197–197。https://doi.org/10.1038/s41562-019-0569-7 编号：3 出版者：自然出版集团。[61] Ilkka Niiniluoto。2019 年。科学进步。载于斯坦福哲学百科全书（2019 年冬季版），Edward N. Zalta（编）。形而上学研究实验室，斯坦福大学。[62] Antti Oulasvirta 和 Kasper Hornbæk。2016 年。作为问题解决的人机交互研究。载于*2016* 年 *CHI*计算系统人为因素会议论文集（美国加利福尼亚州圣何塞）（*CHI '16*）。计算机协会，美国纽约州纽约，4956 4967 10 1145 2858036 2858283 –。https://doi.org。/。Matthew J. Page、Joanne E. McKenzie、Patrick M. Bossuyt、Isabelle Boutron、Tammy C. Hoffmann、Cynthia D. Mulrow、Larissa Shamseer、Jennifer M. Tetzlaff、Elie A. Akl、Sue E. Brennan、Roger Chou、Julie Glanville、Jeremy M. Grimshaw、Asbjørn Hróbjartsson、Manoj M. Lalu、Tianjing Li、Elizabeth W. Loder、Evan Mayo-Wilson、Steve McDonald、Luke A. McGuinness、Lesley A. Stewart、James Thomas、Andrea C. Tricco、Vivian A. Welch、Penny Whiting 和 David Moher[63]。2021年。PRISMA 2020声明：报告系统综述的更新指南。系统综述10, 1（2021年3月）。https://doi.org/10.1186/s13643- 021-01626-4[64] Matthew J Page、David Moher、Patrick M Bossuyt、Isabelle Boutron、Tammy C Hoffmann、Cynthia D Mulrow、Larissa Shamseer、Jennifer M Tetzlaff、Elie A Akl、Sue E Brennan、Roger Chou、Julie Glanville、Jeremy M Grimshaw、Asbjørn Hróbjartsson、Manoj M Lalu、Tianjing Li、Elizabeth W Loder、Evan Mayo-Wilson、Steve McDonald、Luke A McGuinness、Lesley A Stewart、James Thomas、Andrea C Tricco、Vivian A Welch、Penny Whiting 和 Joanne E McKenzie[65]。2021年。PRISMA 2020解释与详述：报告系统综述的更新指导与范例。*BMJ*（2021年3月），n160。https://doi.org/10.1136/bmj.n160[65] Jessica Pater、Amanda Coupe、Rachel Pfafman、Chanda Phelan、Tammy Toscos 和 Maia Jacobs。2021年。在人机交互中规范化参与者报酬报告：系统文献综述与领域建议。载于*2021*年*CHI*计算系统人为因素会议论文集（日本横滨）（*CHI '21*）。计算机协会，美国纽约州纽约，文章 141, 16 10 1145 3411764 3445734 页码。https://doi.org/ . / . Henning Pohl 和 Aske Mottelson[66]。2019。如何在CHI中进行指导、写作与引用。载于*2019*年*CHI*计算系统中的人为因素会议扩展摘要（苏格兰格拉斯哥，英国）（*CHI EA '19*）。计算机协会，美国纽约州纽约， 1 11 10 1145 3290607 3310429 –。https://doi.org/ . / . Andrew S. Pullin 和 Gavin B. Stewart[67]。2006。系统综述在保护与环境管理中的指南。《保护生物学》20, 6（2006年12月），1647–1656。https://doi.org/10.1111/j.1523-1739.2006.00485.x [68] Felix Putze、Susanne Putze、Merle Sagehorn、Christopher Micek 和 Erin T. Solovey。2022年。理解脑信号实验报告的人机交互实践与挑战：迈向可复现性与再利用。*ACM*计算机与人机交互汇刊29, 4，文章 31（2022年3月），43页。https://doi.org/10. 1145/3490554[69] Stuart Reeves。2015年。作为科学的人机交互。奥胡斯人本计算系列1, 1（2015年10月），12。https://doi.org/10.7146/aahcc. v1i1.21296[70] Margarete Sandelowski、Corrine I Voils 和 Julie Barroso。2006年。定义和设计混合研究综合研究。《学校研究：由中南部教育研究协会和阿拉巴马大学主办的全国性同行评审期刊13, 1（2006年），29。https://www.ncbi.nlm.nih.gov/pmc/ articles/PMC2809982/[71] Katie Seaborn 和 Jacqueline Urakami。2021年。定量测量语音用户体验：快速综述。载于*2021*年*CHI*计算系统中的人为因素会议扩展摘要（*CHI EA '21*）。计算机协会，美国纽约州纽约， 1 8 10 1145 3411763 3451712 –。https://doi.org/ . / . Patrick E. Shrout 和 Joseph L. Rodgers[72]。2018。心理学、科学与知识建构：从重复危机拓展视角。《心理学年度评论》69, 1（2018年），487–510。https://doi.org/10.1146/annurev-psych-122216-011845 _电子预印本：https://doi.org/10.1146/annurev-psych-122216-011845.[73] Andy P Siddaway、Alex M Wood 和 Larry V Hedges。2019年。如何开展系统综述：进行和报告叙述性综述、荟萃分析与荟萃综合的最佳实践指南。《心理学年度评论》70（ 2019 747 770 10 1146 010418 102803），–。https://doi.org/ . /annurev-psych- - K. Slim 和 T. Marquillier[74]。2022。伞式综述：综合外科手术科学证据的新工具。《内脏外科杂志》159, 2（2022年4月）， 144 149 10 1016 2021 10 001 –。https://doi.org/ . /j.jviscsurg. . . Hannah Snyder[75]。2019。文献综述作为研究方法论：概述与指南。《商业研究杂志》104（2019年11月），333–339。https: //doi.org/ 10 1016 2019 07 039 . /j.jbusres. . . Mark Starr、Iain Chalmers、Mike Clarke 和 Andrew D. Oxman[76]。2009年。Cochrane系统综述数据库的起源、演变与未来。《国际卫生技术评估杂志》25, S1（2009年），182–195。https://doi.org/10.1017/S026646230909062X [77] Isabelle Stengers。2018年。另一种科学是可能的：慢科学宣言。John Wiley & Sons。Google-Books-ID：oxJSD wAAQBAJ。[78] Anthea Sutton、Mark Clowes、Louise Preston 和 Andrew Booth。2019年。认识综述家族：探索综述类型及其相关信息检索要求。《健康信息与图书馆学杂志》36, 3（2019年9月），202–222。https://doi.org/10.1111/hir.12276[79] Allison Tong、Kate Flemming、Elizabeth McInnes、Sandy Oliver 和 Jonathan Craig。2012年。提升定性研究综合报告的透明度：ENTREQ。《*BMC*医学研究方法论》12, 1（2012年11月）。https: //doi.org/10.1186/1471-2288-12-181[80]Marta Topor、Jade Pickering、Ana Barbosa Mendes、Dorothy Bishop、Fionn Cléirigh Büttner、Mahmoud Elsherif、Thomas Rhys Evans、Emma L Henderson、Tamara Kalandadze、Faye Nitschke 等。2020年。规划和开展非干预、可复现、开源系统综述（NIRO-SR）的整合框架。（2020年）。预印本/尚未发表。https://doi.org/10.31222/osf.io/8gu5z。[81] April Tyack 和 Elisa D. Mekler。2020年。人机交互游戏研究中的自我决定理论：当前应用与开放问题。载于*2020*年*CHI*论文集

计算系统中的人为因素会议（美国夏威夷州檀香山）（*CHI '20*）。计算机协会，美国纽约州纽约，1–22。https: //doi.org/10.1145/3313831.3376723[82] Rensvande Schoot、JonathandeBruin、RaoulSchram、Parisa Zahedi、JandeBoer、Felix Weijdema、Bianca Kramer、Martijn Huijts、Maarten Hoogerwerf、Gerbrich Ferdinands、Albert Harkema、Joukje Willemsen、Yongchao Ma、Qixiang Fang、Sybren Hindriks、Lars Tummers 和 Daniel L. Oberski。2021年。用于高效、透明系统综述的开源机器学习框架。《自然·机器智能》3, 2（2021年2月），125–133。https://doi.org/10.1038/ s42256-020-00287-7[83] Sarah Theres Völkel、Christina Schneegass、Malin Eiband 和 Daniel Buschek。2020年。“智能用户界面”中的“智能”是什么？对25年IUI研究的荟萃分析。载于第*25*届智能用户界面国际会议论文集（意大利卡利亚里）（*IUI '20*）。计算机协会，美国纽约州纽约，477 487 10 1145 3377325 3377500 –。https://doi.org/ . / .[84] 2000 Steve Whittaker、Loren Terveen 和 Bonnie A. Nardi。让我们停止突破界限，开始应对界限：人机交互的参考任务议程。人机交互15, 2-3（2000年9月），75–106。https://doi.org/10.1207/ s15327051hci1523 2[85] – Max Wilson、Wendy Mackay、Ed Chi、Michael Bernstein 和 Jeffrey Nichols。2012年。RepliCHI特别兴趣小组：从专题讨论会到新的复现投稿场所。

载于*CHI '12*计算机系统中的人为因素扩展摘要（美国得克萨斯州奥斯汀）（*CHI EA '12*）。计算机协会，纽约，纽约州，美国，1185 1188 10 1145 2212776 2212419 – 。https://doi.org/ 。 / 。Max[86] L. Wilson、Wendy Mackay、Ed Chi、Michael Bernstein、Dan Russell 和 Harold Thimbleby。2011。RepliCHI——CHI 应该更多地复现和验证结果：讨论。载于*CHI '11*计算机系统中的人为因素扩展摘要（加拿大不列颠哥伦比亚省温哥华）（*CHI EA '11*）。计算机协会，纽约，纽约州，美国，463 466 10 1145 1979742 1979491 – 。[87] 2016 https://doi.org/ 。 / 。Jacob O. Wobbrock 和 Julie A. Kientz。 。人机交互领域的研究贡献。交互23, 3（2016年4月），38–44。https://doi.org/10. 1145/2907069[88] Tracee Vetting Wolf、Jennifer A. Rode、Jeremy Sussman 和 Wendy A. Kellogg。2006。打破将“设计”视为 CHI 黑箱技艺的迷思。载于*SIGCHI*计算系统人为因素会议论文集（加拿大魁北克省蒙特利尔）（*CHI '06*）。计算机协会，纽约，纽约州，美国，521–530。https://doi.org/10 1145 1124772 1124853 . / 。John Zimmerman、Jodi Forlizzi 和[89] Shelley Evenson。2007。将通过设计进行研究作为人机交互领域交互设计研究的方法。载于*SIGCHI*计算系统人为因素会议论文集（美国加利福尼亚州圣何塞）（*CHI '07*）。计算机协会，纽约，纽约州，美国，493–502。https://doi.org/10.1145/1240624.1240704