对事实核查工作的若干观察及其对计算机辅助的启示

Rob Procter[1, 5], Miguel Arana-Catania[2], Yulan He[3, 5], Maria Liakata[4, 5], Arkaitz Zubiaga[4], Elen **a Kochkina[4],Runcong Zhao[3]**[1]华威大学[2]克兰菲尔德大学 [3]伦敦国王学院[4]伦敦玛丽女王大学[5]艾伦·图灵数据科学与人工智能研究所 rob.procter@warwick.ac.uk, miguel.aranacatania@cranfield.ac.uk, yulan.he@kcl.ac.uk, m.liakata@qmul.ac.uk, a.zubiaga@qmul.ac.uk,e.kochkina@qmul.ac.uk, runcong.zhao@kcl.ac.uk

arXiv:2305.02224v4 [cs.HC] 6 Jul 2023

摘要

社交媒体和用户生成内容（用户生成内容）已通过多种不同方式日益成为新闻工作的重要组成部分。然而，错误信息的增长意味着，新闻机构不得不投入越来越多的资源来确定其真实性，并在发现其具有误导性时发布更正。在本文中，我们介绍了对来自两个组织的八名事实核查团队成员进行访谈的结果。团队成员描述了他们的事实核查流程，以及当前在以稳健且及时的方式完成事实核查时所面临的挑战。前者揭示了事实核查实践之间存在的显著差异，以及团队成员之间协作所发挥的作用。最后，我们讨论了对计算工具开发与应用的启示，包括目前缺乏计算工具支持的领域，以及能够适应不同事实核查实践的重要性。

引言

本文所报告的研究是在 PANACEA 项目“大流行期间人工智能主张真实性评估：大流行病期间主张真实性评估的人工智能赋能循证框架”背景下开展的[1]。该项目的主要目标是从信息与传播视角分析大流行病，并开发分析与核验工具，以应对错误信息危机。项目成果包括新的假新闻数据集（Arana-Catania 等，2022；Dougrez-Lewis 等，2022）、对现有事实核查技术的评估以及新型事实核查技术的开发（Kochkina 等，2023；Arana-Catania 等，2022；Dougrez-Lewis 等，2021；Zhu 等，2022；Abumansour 和 Zubiaga，2021；Zhang、Gui 和 He，2021；Si 等，2021；Zhang、He 和 Procter，2023），以及事实核查工具原型（Zhao 等，2023）。

本文介绍了我们通过访谈事实核查员所得到的一些调查结果，内容涉及他们目前如何应对错误信息问题、采用的方法以及使用的工具。

版权所有 © 2022，人工智能促进协会（www.aaai.org）。保留所有权利。[1]https://panacea2020.github.io/

目前使用的方式，以及他们认为基于自然语言处理的技术和工具如何在他们面临的压力和可用人力资源条件下，帮助他们更有效地完成这项工作。

相关工作

社交媒体和用户生成内容（UGC）通过多种不同方式，日益成为新闻工作中重要的组成部分（Tolmie 等，2017）。例如，它们被用作寻找新闻线索和发现报道的一手来源（Brandtzaeg 等，2016）。然而，错误信息的增长意味着，新闻机构不得不投入越来越多的资源，以确定此类内容的真实性，并在发现其具有误导性时发布回应（Bontcheva 等，2015）。

面对不断升级、数量日益庞大的错误信息，如何及时应对这一挑战促使人们开发出各种不同的基于自然语言处理的技术和工具，以支持事实核查工作（Lazarski、Al-Khassaweneh 和 Howard，2021；Zeng、Abumansour 和 Zubiaga，2021；Guo、Schlichtkrull 和 Vlachos，2022；Krishna、Riedel 和 Vlachos，2022）。然而，尽管已有多篇关于事实核查研究文献的综述发表（例如 Demartini、Mizzaro 和 Spina（2020），Das 等（2023）），除少数值得注意的例外（例如 Micallef 等（2022），Juneja 和 Mitra（2022））外，旨在详细了解事实核查员在应对新冠错误信息或“信息疫情”（Diseases，2020）等情况时所面临日常挑战的一手研究仍相对较少。B´elair-Gagnon 等（2023）采访了事实核查员和主要社交媒体平台工作人员，探究双方如何在事实核查工作中合作。Micallef 等（2022）共采访了 21 位事实核查员，了解他们的方法、工具使用情况和面临的挑战。他们发现，目前的事实核查实践在很大程度上仍依赖人工，缺乏系统性，且规模、范围和覆盖面有限。

在此前的研究中，我们对新闻编辑部工作的记者进行了深入的民族志研究，特别关注他们在寻找可供发表的候选报道线索时如何使用用户生成内容（Tolmie 等，2017）。其中一项

这些研究的主要调查结果之一是，事实核查并非“一站式服务”，而是在报道研究和撰写的整个过程中反复出现的一项工作。我们得出的结论是，工具需要支持持续进行的事实核查工作，包括清楚呈现已经开展了哪些事实核查工作，以及由谁完成。

在开展这项研究时，我们所研究的新闻编辑部尚未成立专门的事实核查团队，而是依靠记者将事实核查融入日常工作。然而，随着近年来错误信息规模的不断扩大，事实核查已发展成为一种独立的实践（Graves and Cherubini 2016）。因此，PANACEA 项目的目标之一是了解事实核查团队成员如何开展这项工作、他们面临哪些挑战，以及事实核查支持工具随后可以在整个事实核查工作流程中发挥什么作用（Juneja and Mitra 2022）。

方法论

该项目的一个关键环节是与事实核查专家合作，深入了解处理错误信息所面临的挑战，以及在大流行病情境下错误信息的独特特征。我们的目标是尽可能全面地开展这项分析，因此项目邀请了在应对大规模错误信息挑战方面拥有截然不同专业知识和方法的合作伙伴。其中包括一家英国主流新闻机构和 Full Fact[2]，这一英国独立事实核查组织。

最初，我们的方法原本应当建立在一系列全新的事实核查工作民族志研究之上，旨在观察并详细了解这项工作如何在每时每刻被完成（Rouncefield and Tolmie 2016）。然而，新冠疫情使这一计划无法实施，因此，我们转而对英国一家主流新闻机构和 Full Fact 的事实核查团队成员开展了一系列半结构式访谈。2021 年底，我们共进行了八次访谈。受访者包括四名事实核查员，他们同时也是受过培训的记者；两名新闻编辑；以及两名数据科学家。访谈旨在确定事实核查流程中的关键阶段、每个阶段涉及的活动和人员，以及他们目前面临的主要挑战。最后，我们邀请受访者提出人工智能工具可能如何以及以何种方式帮助应对这些挑战。访谈问题概要见附录。

本研究已通过华威大学生物医学与科学研究伦理委员会的伦理审批。

访谈平均持续约 60 分钟。访谈内容经过录音和转录，随后由访谈者进行分析。根据先前的研究（（Tolmie et al. 2017））和文献综述，预先依据访谈问题确定了顶层主题。随后采用扎根理论（Birks et al. 2013）对受访者的回答进行归纳分析，以深入探讨这些主题。

[2]fullfact.org/

下一节将报告我们与事实核查员讨论的结果。

调查结果

本节报告访谈分析的结果。调查结果围绕事实核查工作流程的主要阶段展开，并通过访谈记录中的示例摘录加以说明。受访者以及受访者提及的个人和／或账号名称均已匿名处理。

挖掘主张

在开始事实核查之前，需要先找出候选主张，即“值得核查”的主张。搜索社交媒体通常是这一过程的重要环节，受访者解释了通常如何开展这项工作。

> “好，我们来聊聊推特。我们会先确定要关注的概念，有时则从偶然注意到的趋势入手。可以采取多种方式。热门话题本身或许就有用，也可以直接在推特上搜索。然后从搜索结果沿时间线往回追溯，看看其中是否提到了特定词语，再深入查看单条帖子……大多数时候，你都能找到某个爆红主张的首条帖子。有时发布者几乎没得到任何关注，后来有人注意到它，并据此提出了自己的主张。也可能有不同的人同时想到同一件事，所以有时很难判断最先发帖的人是谁。”

在新冠疫情这类事件期间，事实核查人员会意识到相关主题很可能出现假新闻文章，因此会更加集中注意力。

> “是的，我们会特别留意社交媒体上的某些频道，看看谁在推广什么。比如，几周内你会看到几篇关于有人向学校发送这些假的英国国家医疗服务体系疫苗信件的报道……所以，我们会关注它们在社交媒体上的传播情况……还有几篇关于疫苗信件的报道，目的是让家长和学生放心。然后也会持续关注那些反对它的示威活动。”

随着新冠疫情等事件不断发展，事实核查员可能会逐渐注意到一些特别活跃、传播虚假信息的账号，随后便会对这些账号进行更有针对性的监测。

选择要核查的主张

一旦确定候选论断，就必须决定优先对哪些候选论断进行事实核查。在两个组织中，这些决定最初都是在每日新闻例会开始时作出的。正如 Full Fact 的一名数据科学家所观察到的，这些决定源于协作努力，依赖编辑和事实核查人员的集体专业知识。

“早上，事实核查员会通过相互讨论来决定要核查哪些内容。具体情况差异很大。每项主张都各不相同，但通常需要弄清楚诸如究竟是谁说了什么、主张的具体内容是什么、受众是谁，以及相关背景是什么等问题？”

在以下摘录中，Full Fact 的编辑解释了这是如何完成的。

“团队会讨论提出的各种主张。我们会关注诸如这些主张的传播范围有多广？是谁提出了主张？这个人有多引人注目？他是普通公众成员，还是政治人物？这则内容是登上了全国性报纸的头版，还是第 36 页？此外，我们还会基本评估一项主张可能造成多大的危害：比如，‘注射漂白剂以避免感染冠状病毒’，这种情况下潜在危害可能非常大。如果主张不准确，我们就会综合所有这些不同因素，优先处理当天能够开展核查的内容……”

正如一位编辑所解释的，主流新闻机构也采用类似的流程。不过，一个显著的区别是，事实核查团队可以得到新闻机构记者的支持，因此，全天都会持续监测待核查的主张。

“我们大家会在各种不同的聊天应用上不断地互相闲聊，但每天也会开两次编辑会议……各个分社的人都会聚到一起……他们会说，正在发生这件事，这是我负责的一个主题团队。我还负责另一个主题团队，这意味着我们可以梳理出这些线索，然后说，‘好吧，这挺有意思的。’……有时也会遇到相反的叙事，我们就能把这些点串联起来。从监测的角度来说，事情大致就是这样运作的。”

从本质上说，两位编辑描述的是一个分流筛选过程。他们指出，鉴于可用资源有限，所需投入的精力会限制每天选来进行事实核查的主张数量。一位新闻编辑介绍了用于评估某项主张是否值得核查的三个维度：传播、严重程度和放大效应。

“对我来说，这三项标准是传播、严重程度和放大效应，而我们在虚假信息领域的人基本上都得出了相同的结论。传播不仅是指这件事传播得有多广，还包括它是否跨越了不同语言，以及是否正在世界各地传播。……人们是否在参与其中？至于严重程度，这件事是否真的会造成现实世界的伤害，或者它是否来自一个可能对人们产生重大影响的人？显而易见的例子是唐纳德·特朗普注射漂白剂的说法；另一个很有参考价值的例子是诺瓦克·德约科维奇谈论反疫苗和反新冠肺炎的内容……这类事情影响重大，因为他并不是那种你会觉得会说出这种话的人 [，比如这样]。……而且他能够接触到一群受众和追随者，他们大概并非阴谋论圈子的一员。所以，当某件事被名人放大，尤其是被像他这样被认为正直、健康、堪称良好榜样的名人放大时，在我看来，这就符合严重程度这一项，也涉及整个放大效应。……。”

谈到传播，一家主流新闻组织的事实核查员谈及了他们目前用于在社交媒体上搜索的工具存在的一些不足。

“我的意思是，我们真正关注的是互动。通常，如果某件事被大量转发，却没有引发互动，那可能就没什么人讨论；但如果某件事获得了大量互动，有很多转推和评论，那就值得关注。所以我们必须看看情况。我的意思是，搜索工具有很多东西确实处理不了，其中之一就是讽刺和挖苦。因此，如果你想做情感分析，工具往往会完全漏掉这一点，因为它们不理解英国人用反话重复某件事的方式。这过去一直是个问题。”

正如一位新闻编辑所解释的，严重程度的问题在于试图估算可能造成的“现实后果”或危害。

“……观察事情何时开始从一堆胡说八道、只有少数人在谈论，转变为可能确实会造成现实后果的事情。到了那时，我们会在每天的例会上提出来，由我这样的人决定是否跟进。”

未被选中进行事实核查的主张可能会继续受到监测，并在日后重新纳入考虑，正如一家主流新闻机构的事实核查员所解释的那样。

”一篇关于伊维菌素、计划于今天发表的文章，已经筹备了好几个月。我的意思是，是啊，所以这件事可能很长一段时间都没有达到门槛，直到大型利益团体开始关注，随后它变成了反疫苗团体数百万美元的生意——这些团体确实从中赚了钱。人们开始争相购买时，这件事才越过了门槛。所以，是啊，潜在危害的门槛基本上就是在那时达到了。”

事实核查员举例说明了某个候选主张，其是否值得进行事实核查的决定，取决于对事实核查价值评估方法的一种相当细致入微的运用。事实证明，这是事实核查员经常需要处理的一种情况。在第一个例子中，对这位事实核查员而言，问题在于支持该主张的引述证据中存在错误，而非主张本身的真实性；同时，还需要权衡该主张可能的传播范围，以及若不予纠正可能造成的危害程度。还值得注意的是，这位事实核查员在判断这一新主张时，借鉴了此前的一项事实核查结果。

“我这周写的另一篇是关于 [...] 和她的电视节目，在节目中，她和 [. . .] 主张，因新冠肺炎住院的人中有 90% 未接种疫苗。这是一个很好的例子，说明我早已知道某件事不是真的，因为我最近写过这个话题。实际数字约为 36%，所以你完全可以说这个数字具有严重误导性。他们基本上是在说，你应该接种疫苗，因为这会大幅降低你住院的几率，然后他们用这个 90% 的数字来佐证这一说法。所以，他们的基本观点是对的……你应该接种疫苗，因为这确实会大幅降低住院的几率，但他们用来佐证这一点的数字是错的。这就是一个值得核查的重要例子，因为它是在全国电视台播出的。这种说法已经被多次提出，但造成伤害的可能性大概较低。”

在接下来的这段访谈摘录中，一位事实核查员举例说明了某项主张。这项主张需要经过一定的审慎讨论，以确定其具体所指的内容是否达到事实核查的门槛，而这一决策再次取决于对传播范围与危害程度之间平衡关系的权衡。

XXX [一位知名作者]在《The YYYY》 [一本英国周刊杂志]上发表了一篇文章，其中 [XXX]表示，疫苗无法阻止冠状病毒的传播。对此，你可以从两种不同的角度来理解。你可以说：“嗯，并不是每一个接种疫苗的人都能得到百分之百的保护，完全避免感染或传播病毒，因此，接种疫苗并不能彻底阻止所有传播情况。”但这可能并不是她想表达的意思。她也可能是想说：“疫苗能够阻止传播，因为在个体层面，有很多情况下，疫苗确实能够阻止病毒传播，例如，有些人从一开始就不会感染病毒。”那么，我们要核查的这句话究竟是什么意思呢？结合那篇文章的上下文，我认为她显然表达的是第二种意思。她是在说，接种疫苗对于减少传播完全没有任何作用，而这是不真实的，因为我们可以证明这一说法是错误的；相比之下，第一种可能的理解则是正确的。……因此，我们会进行很多类似这样深入细致的讨论……这类内容可能不会被那么多人看到……但其潜在危害却大得多，因为 [XXX] 正在精心构建一套基于数据的论证，以此解释接种疫苗对传播没有任何影响。如果人们相信这一说法并认真对待，就可能造成非常严重的危害。”

最后，放大效应是事实核查团队在商议是否开展某项具体事实核查时考虑的第三个因素。一位新闻编辑解释说，这取决于团队能否估计发布一项事实核查可能抬高某项主张的关注度，进而提高公众对它的认识；而这可能会适得其反，削弱事实核查本身的影响。

“……放大效应是我们一直在努力应对的问题，而且你知道，由于拥有 [主流新闻机构的] 平台，你知道，如果我们说，‘哦，这正在发生，我们必须对此非常谨慎’，尤其是涉及 [XXX] 和其他阴谋论时……我们希望在不让它显得具有吸引力的情况下报道它。”

由此可见，在撰写事实核查内容以供发布时，团队还会重新考量放大效应这一因素。

寻找证据

一旦确定哪些主张达到核查价值阈值，实际的事实核查工作便随即开始。正如我们将在下文摘录中看到的，访谈揭示了所用方法存在显著差异。在 Full Fact，通常采用的方法是汇集相关文献资料来源，再进行审阅，以确认收集到的证据是否与该主张一致。这个过程从在线搜索来源开始。一位事实核查组织的事实核查员解释说，工作量正是在这一步骤中大幅增加。

“有时，即使你知道大致的主题领域，要筛选该主题下的信息来源仍然有很多工作要做。若有工具能简化这一过程，或者至少列出你可能想查看的内容，无疑会让事情容易一些。同样，对于可以联系的专家，也可以提供一些值得考虑的人选或组织之类的信息……我们首先会核查的是：他们以前接受过事实核查吗？这个人是否传播过错误信息？这是我们需要注意的吗？还有，他们在某个领域具备怎样的专业知识，诸如此类。”

如上文摘录所揭示的，科学论文或报告本身可能并不总是被认为足以确定某一主张的真相。有时，事实核查员会发现有必要与该领域公认的专家交谈。

“在很多情况下，可能需要阅读已发表的报告……在更专业的情况下，也可能需要咨询专家。许多与大流行病相关的报道中，他们会采访医生或医学研究人员：‘你怎么知道 mRNA 疫苗确实安全？’……征求专家意见，或请专家解释。”

如下文摘录所示，决定采用何种方法对一项主张进行事实核查，需要权衡多种因素，其中包括对相关主题的熟悉程度。

“我们的文章会在不同程度上呈现专家观点。有些文章可能只是直接报道有人对贫困统计数据提出了主张，然后说明这些贫困统计数据反映了什么，以及读者需要了解这些数据的哪些注意事项……我们不一定需要去采访专家。如果涉及更广泛的主张，我们通常会咨询专家……如果主张不那么具体，或者

如果涉及的话题我们不太熟悉。随着事实核查工作的推进，你确实会逐渐积累相关领域的专业知识。如果某个话题我们不太熟悉，我们通常会咨询专家，甚至只是问问“我们应该去哪里查找这方面的信息？这听起来对吗？”在咨询过专家之后，无论他们是为我们指明了信息来源，还是提供了他们对该话题的看法，我们都会呈现这些内容；而且我们已经尝试评估此人在该领域的权威性，以及其观点在多大程度上符合该话题的共识（如果存在共识的话）。”

相比之下，在主流新闻机构中，出于组织层面的原因，与专家交谈是默认的事实核查方式，因为人们认为这是维护其公正声誉的途径。

“我们总会设法去采访一位专家……如果你想查明是否有人说过美国有这么多人死于新冠肺炎，你可以查约翰斯·霍普金斯大学的统计数据之类的。但如果我们要做的事情涉及更详细的内容，我们就不能说这些话是我们自己说的……公正性非常重要……我们总会想办法采访某位专家……[这样我们就可以说]我们采访了X专家或Y专家，他们是这么说的。’……始终设法采访一位专家，如果可能的话，还要采访多位专家。我们很少会认为仅凭一个来源就足够可信，可以单独作为依据。”

在寻找可信专家时，事实核查员必须警惕某些虚假信息来源为掩盖其对当今议题的真实立场而采取的各种手段。

“有很多这样的团体。他们把自己包装得仿佛既支持气候变化，又反对气候变化，因此，一名需要快速行动的记者看到‘气候信心联盟’之类的名称时，就会想，哦，这很有意思……[但]后来才发现，实际上那是一个煤炭生产商联盟……”

受访者强调，事实核查的过程必须非常迅速，而急于得出结果的压力可能导致人们采取走捷径等做法，使这一过程不够彻底。

“如果你一直反复查阅相同的来源，就会担心产生某种固有的潜意识偏见。因为作为记者，你总是在以每小时一千英里的速度做事，而且你知道某个网站更容易阅读。它可能更容易证实你想做的事情，或者类似的想法，于是你实际上会变得有些懒惰，不再真正开始更广泛地核查信息。”

在某些情况下，找出该联系哪些专家相当直接。正如这位新闻编辑所解释的，事实核查员会从以往的事实核查工作中积累相关知识，知道该找谁交谈。

“我是说，现在又是这样，因为我们做新冠肺炎报道已经很久了。我们知道可以找哪些人，也知道他们值得信赖。”

然而，当一个新主张出现时，挑战在于找到符合新闻组织专家标准的人。一家主流新闻组织的事实核查员描述了这样一个事实核查的例子。

“当出现一些我们其实并不太了解的事情时。你知道，我们都是研究人员，也都是以记者为职业的，我们知道如何处理这类事情，但尤其是当出现的事情不一定属于你的专业知识领域时。5G 相关内容开始出现时，你确实会发现自己在 Google 上搜索，试图弄清楚 5G 运营商所属的全国性组织是哪一个，以及诸如此类的信息，而他们可能总部设在柏林，……于是你会花很多时间试图找到合适的人。”

Full Fact 的编辑更详细地描述了他们识别可信专家的具体方法，揭示了这一过程可能会耗费多长时间。

“我们最终进行事实核查的许多事情，都已经在新闻中或由其他事实核查员进行了相当多的报道……也有人曾与其他事实核查员交谈过。这至少能让你了解当事人是否愿意或有时间接受媒体组织采访。接下来，你必须评估我们是否认为他们是可靠的信息来源。这通常需要深入调查他们的资历。他们是学者吗？是什么让他们有资格就这个话题发言，并让我们乐于称他们为该领域的专家……只要能查到，我们就会了解他们的职业背景。我们还会看看他们此前对这个话题说过什么：一方面，他们可能已经就我们正在寻找的确切内容发表过看法；另一方面，我们也非常清楚，我们希望确保自己接触的是独立组织。还要留意我们推荐的专家是否与此事有利益牵连，因此我们会努力找出任何可能的危险信号。”

加快这一流程的一种方法，是与那些以可信、可靠著称的人士和组织建立人脉网络。

“我们可以联系网络中的组织。如果他们不合适，就说，好吧，我们正在寻找一位研究这一主题的人。你能想到有人可能对这项具体的事实核查有所帮助吗？如果这些办法都不奏效，就只能再次进行一些研究，尝试寻找合适的人，比如查看各个大学的院系之类的。然后直接联系陌生人，和大学新闻官交谈，询问是否有人可以由你介绍给我们，使得你能够把这段话接下去”

就这一主题发表演讲，或联系在我们可能开展研究的特定领域从事工作的智库。

随着事实核查员积累了对虚假信息宣传活动中常见主题的经验，他们会不断拓展自己的可信专家网络，从而更容易找到适合特定事实核查的专家。

“随着我们这个组织的发展，我们如今较少遇到这种情况，因为我们已经成功建立了更大的专家网络。但可以肯定的是，当我们遇到从零开始的情况时，这会相当耗时。”

当同一主题或相关主题出现新的主张时，先前对该主题进行的事实核查也会有所帮助；尤其是当负责核查的事实核查员对该主题并不熟悉时。

“如果我们无法确认网上读到的内容是否是可靠的信息来源，就很难加以利用。对于我们所有人来说，获取陌生主题信息时，一个非常常见的来源是先前的事实核查，也就是我们自己做过的核查……一旦 [某项主张] 经过了这一极其繁琐的编辑流程并最终发表，它就相当可靠；如果有人此前写过一篇关于同一主题的文章，其中附有你可能需要的所有链接，能带你找到相关文件，那么这就是熟悉该主题的绝佳途径。”

显然，除了能够加快新一轮事实核查之外，以往的事实核查记录还有助于在团队内部分享知识并积累专业知识。

审查事实核查

随后，团队成员会审查事实核查的结果，以确保没有出现任何错误，正如 Full Fact 的一位事实核查员所解释的那样。

“审稿人实际上所做的工作量与撰写者几乎相同，因为他们需要重复整个建立证据基础的过程……而这个过程可能非常、非常复杂，有时还会持续很长时间……协作确实非常、非常重要。其主要目的是检查撰写者是否正确完成了所有工作，同时也能让不止一个人理解相关主题。”

Full Fact 的数据科学家强调了审查事实核查结果的重要性。

“每篇文章至少有一位审核人员负责审阅，基本上就是逐字逐句、仔仔细细地检查。查看是否有任何数据或分析，并确认我们是否正确解读了所有内容。我们的计算是否正确？是否链接了所有信息来源？语法是否正确……编辑流程、审核阶段都是一个高度协作的过程。”

Full Fact 的编辑对此作了进一步阐述。

“这算是我们的标准做法，必须有三双眼睛把关。如果内容比较复杂，后续工作或媒体工作量很大，参与人数就可能增加。对于特别复杂、我们希望投入大量工作处理的内容，有时甚至可能需要八到十个人参与。但我们现在规定，事实核查至少要有三个人参与……内容越敏感，团队中第三位把关者的资历就需要越高。”

对于这家主流新闻机构而言，事实核查审核的重点是确认在选择专家意见时已进行了尽职调查。

“我们认识一些人，知道可以向他们求助，也知道他们值得信任，诸如此类。但没错，你知道，我们总会核查、再核查、反复核查，绝不能在这些事情上出错，因为如果你说别人散布虚假信息，结果自己却弄错了，那就等于给对方送上绝佳机会……如果 [一位事实核查员] 来找我说‘这是我写的一篇文章’，我会问：你采访了谁？……这基本上会是我首先提出的问题。”

显然，对这两个组织而言，若未能进行全面审核，可能会严重损害其值得信赖的声誉。

撰写事实核查报告

最后一个阶段是撰写事实核查结果并予以发布。在这一阶段，一些组织规范会发挥作用。有些类型的主张很容易处理，正如下文 Full Fact 的一位数据科学家所举的例子。

“基本上，是指那些定量的主张，并且有明确、可信的原始来源可供我们核实。比如宏观经济学方面的官方数据。像当前的通货膨胀率之类。每个月都会公布三种不同的通货膨胀率指标。如果有人说零售价格通胀下降了5%，那要么是对的，要么就是错的……”

撰写其他类型的主张可能需要采取更细致入微的方法。两个组织都强调，必须以尊重平衡原则的方式呈现调查结果，正如下列四段摘录所强调的那样。第一段谈的是平衡问题，这是新闻组织十分重视的议题。

“我认为，对公众而言，一个挑战在于，如何判断一个故事的双方是否平等、平衡或存在偏见。例如，有人一直在抱怨 [. . .] 对一些其实已经有定论的辩论是如何呈现双方观点的。比如气候变化。你可能会有一个否认气候变化的人，但你是否应该给予他与那个认为气候变化确实存在的人平等播出时间。

而且，如果你给予这种虚假等同、虚假平衡，公众最终可能会认为，谁也不真正清楚情况，但实际上，证据表明大多数人的看法是压倒性的……

接下来的三段摘录强调，事实核查的调查结果必须以一种让读者能够自行判断的方式呈现。Full Fact 的编辑解释说，他们的目标并非影响读者。

“我们并不是想让任何人在文章中对某个话题形成定见，我们的想法是把信息呈现给他们，让他们基本上能够自行判断主张、话题、主张相关的一切。所以，你知道，除了对这个做法稍作引导，告诉他们这是我们所写内容的摘要之外，我们并不想影响任何人。”

两个事实核查团队都投入了相当多的精力，以确保文章的中立性。在以下引文中，一位团队成员描述了 Full Fact 如何做到这一点。

“这项工作中有很大一部分是将调查发现撰写成文。文章的写作必须非常中立、清晰，易于任何人阅读，并以一种不带评判色彩的方式呈现所有证据，让读者能够自行判断是否相信这一主张。而撰写、编辑和发布的整个过程具有很强的协作性。我认为，一篇文章在发表之前会由三位事实核查员审阅，以确保内容保持中立，并确认其中没有任何需要进一步仔细核实的主张。”

在这第三段摘录中，一位 Full Fact 数据科学家指出，虽然人工智能确实可以协助为事实核查寻找证据，但如何呈现证据仍需要人类的技能和专业知识。

“我们的受众……不希望有人说这是真的，相信我们就好。他们希望我们替他们进行研究……然后让他们自己做出判断。你能让机器找到证据，再由事实核查员筛选并呈现给公众吗？这是一个可行的目标。但要自动判断某件事是真是假，极大程度上取决于对现实世界的理解。这对任何人工智能来说都非常、非常困难。”

主流新闻机构的一位编辑解释说，对某项主张完成事实核查后，在撰写该主张之前会重新评估传播放大测试；如果随后决定发布，该测试将决定如何撰写。

“……这则报道是否涉及公共利益？我们是否在放大某种行为，或者可能会助长不良行为？因此，在报道之前，我们必须对内容进行编辑审查……我们希望报道这件事，但不让它看起来具有吸引力。所以，我们会采取各种措施，比如不添加实时链接。我们会在内容上叠加‘虚假’或‘误导性’之类的标注，而且我们对此非常谨慎。”

不要留下太多数字痕迹，以免人们找到你——你知道，我们最不希望看到的就是人们说：“哦，真有意思”，然后掉进那个兔子洞。是的，所以我们必须非常负责任。我认为我们对此有着非常敏锐的意识。”

总之，事实核查结果的撰写方式，在两个事实核查团队看来，与确保事实核查准确一样，对打击虚假信息至关重要。

计算工具支持

事实核查员一致认为，有些类型的主张可以自动核查。不过，他们也认为，这种方式只适用于他们需要处理的全部主张中的很小一部分。

“有些类型的主张可以由机器核实……那些具有量化性质、且有明确可信的原始来源可供核对的主张。如果有人说零售价格通胀下降了 5%，那要么是对的，要么是错的；这类主张有可能由机器核查……但即便运行得再完美，也只能覆盖人们所提出主张中极小的一部分。”

因此，受访者很愿意提出其他计算工具可以支持他们工作的方式。一家主流新闻机构的新闻编辑解释了这些工具如何帮助发现候选论断。

“理想情况下，某件事突然走红，而我们知道它正在走红，是因为我们那套了不起的设备提醒我们：‘有个主张出现了，说接种疫苗会让你更可能生双胞胎之类的。’然后还能使用事实核查工具，它可以说：‘好的，这里有六个不同的来源，看看它们对此怎么说。’”

这位新闻编辑接着更详细地说明了这样做的一些益处。

“对我来说，提前获知哪些事情正在迅速走红，简直就是一座金矿，因为我们只能等着别人给我们通风报信。要是某些事情只在特定社群或特定地理区域内传播，我们却没能发现，这会让我非常沮丧；我认为，这也会使我们的内容难以像我们希望的那样多元化。因此，如果有工具能够指出哪些内容……我想，[……]正在着手研究这方面的东西，他们开始关注自己所说的类似主张的陈述，以及这些陈述如何浮现出来。但这对我们来说将会是一个真正的变革性突破……”

这家主流新闻机构的一位事实核查员指出，现有搜索工具存在不足。

”……可靠的情感分析。因为我们过去用过的那些工具，嗯，虽然已经尽力了，但似乎无法应对人类情感。还有撒谎之类的事情，这些往往会让整个社群感到沮丧，因为大家都知道，能够……

浏览大量数据来识别情感，并判断某人是否真正投入其中，而不是只是敷衍了事。”

识别哪些主张值得核查所面临的挑战，促使 Full Fact 开发计算工具来协助这一流程。一位来自该事实核查组织的数据科学家对此作出了解释，同时强调，自动化整个事实核查流程并非可行目标。

“与事实核查员交流后发现，决定核查什么的决策过程非常、非常复杂，很大程度上依赖他们的专业知识和直觉，所以我一直基本上拒绝尝试开发一种算法，来预测某项内容是否值得核查。相反，我们尝试开发一些工具，以便呈现主张，让他们随后判断是否值得核查。因此，我们所做的就是向事实核查员展示各种各样的主张……我从未编写过一个系统来预测哪些主张值得核查，原因在于一位事实核查员说，你知道，这是我们的流程。每次核查都要经过数小时的讨论。我们不能就这么把它自动化。”

他随后更详细地介绍了用于协助识别值得核查的主张的方法，包括哪些类型的新闻报道会被忽略，哪些会被优先处理。

“其中一个 [工具]本质上是一个索引，涵盖英国媒体过去24小时的内容。比如，取出报纸中的每个句子，然后判断它是否提出了主张？……这个算法基本上会识别体育赛事结果、名人八卦、婚讯等内容，并将它们全部排除在考量范围之外。接着，它会把剩下的内容大致按类别分组，比如健康、教育、犯罪率等宽泛类别。它只是在寻找报纸上最常被讨论的报道。如果一则报道被多位不同的记者报道，就会被列在我们的榜单前面。事实核查员早上会查看这个榜单，问自己：在我感兴趣的宽泛领域里，有没有什么重要报道是我可能遗漏的？阅读只需几分钟，这一点很重要。我们并不是要求他们花几个小时学习或使用这个工具，只需花五分钟浏览一下早间内容……我想，过去几个月里，他们已经根据使用该工具发现的内容发布了两三篇事实核查报道。”

这位数据科学家接着描述了如何帮助事实核查员估算核查某项具体主张所需的工作量。

“提供一份关于该主张相关证据的易于获取的摘要，其中包括来源和专家、可信的意见或更多信息来源，使事实核查员能够非常迅速地判断这一项是否值得再深入挖掘一些。他们仍然需要亲自完成这项工作，而这仍然需要这类专业知识，也就是研究和讲述故事。可是如果你可以一开始就给他们足够的证据，这样一来，某件事即便单看似乎没什么希望，结合一些数据之后，他们也可能会想：其实值得再深入了解一下。我认为，这类工具确实可能产生影响。

这家主流新闻机构的一位事实核查员提出，在咨询他们的人脉网络却未能找到一位或多位合适的专家进行交流时，工具可以发挥作用。

“你可以让算法学会识别某个来源在上个月的文章中被用作可信来源达 10 次，这会很有帮助，因为新闻业与学术界截然相反。我们讲求速度，尽力而为……有时你会寻找经过验证的可靠来源，因此，任何能让你走捷径的东西都会大有裨益。”

Full Fact 的数据科学家以寻找有关某项主张的证据为例，谈到了如何恰当地平衡工具的能力与人类专业知识这一挑战。

“我想，事实核查员本身就是经验丰富的记者，已经深入这个领域，因此，最能帮助他们的方式，是比他们自己查找证据更快地提供证据链接。或者提供一些他们原本可能会遗漏的内容。但这颇具挑战，因为这基本上就是他们技能组合的核心部分，而我们却试图将其自动化……例如，我认为，工具面临的挑战不仅是找出故事的双方并呈现出来，还要以某种方式展示证据的整体倾向；我认为事实核查员擅长这一点，如何实现自动化将会很有意思。”

Full Fact 的数据科学家也承认，将用户可能难以理解其行为方式的复杂计算工具引入事实核查工作流程，本身就会引发信任方面的问题。

“存在这样的风险：算法是一个黑箱，除非积累大量经验，否则没人会真正信任它。对事实核查员来说，可解释性往往意味着提供摘录、引文、片段以及指向其他来源的链接。一手来源。比如，提供大学主页的链接，说明此人受雇于这所大学，并附上其主页，这可以作为一种证据。如果其他记者也报道过某个人，可以附上他们文章中的链接和引文；如果此人还被《纽约时报》和其他二十来份报纸引用，你就知道他们在媒体上广受谈论。你想展示证据，但不想把它简化成一个数字或列表；你想说明他们是如何被引用的。”

Full Fact 的一位新闻编辑解释说，他们改进事实核查的方法不仅是引入计算工具，还包括让这一流程更具协作性。

“未来几年，我们努力的方向之一未必是自动化这一流程本身，而是让更多人参与整个事实核查流程，让支持者加入进来，并吸引更多人参与其中。基本上，就是不再局限于团队中参与这一流程的成员。我们很清楚，外面有大量错误信息。我们目前的团队规模仍然相对较小，无法仅靠自己核查所有信息。我们希望让更多人参与这一流程，同时寻找改进方法，让流程实现更高程度的自动化。具体要怎么做，我们还在摸索中，但这就是我们最终的目标。”

鉴于团队协作已经在保障流程质量方面发挥着关键作用，因此强调让这一过程更具协作性或许并不令人意外。

讨论

对事实核查团队成员的访谈揭示了构成事实核查工作流程各个阶段的细节，我们的调查结果总体上印证了 Micallef 等（2022）的研究。流程始于发掘候选论断，随后对这些论断进行审核，以决定哪些值得核查。在这一过程中，会运用一种基于传播、严重程度和放大效应的非正式评估方法，为决策过程提供一定程度的可重复性和质量保证，尽管这些因素的具体应用往往不可避免地相当微妙。因此，尽管已经开发出可自动识别值得核查论断的工具，但显然，这些工具无法取代人类的专业知识（Demartini、Mizzaro 和 Spina（2020）），也无法取代对遵守组织政策意味着什么的理解（Bittner（1965））。

重要的是，这项研究揭示了整个事实核查过程如何依赖协作。协作不仅对于确保流程的稳健性至关重要，还有助于团队内部知识和专业知识的积累。因此，事实核查员对将人类专业知识完全从这一过程中剔除的可行性持怀疑态度，也就不足为奇了。

在识别出值得核查的论断后，实际的事实核查工作便开始了。在这一过程中，我们观察到，由于两个组织在“新闻与媒体生态系统”中所处的位置不同，其工作流程也存在差异。这使得它们以不同的方式界定最终目标，而这些差异乍看之下可能十分细微，但随后却会沿着事实核查工作流程逐步向前传导。

事实核查组织通常会在科学论文以及可信、可靠组织发布的文件等在线来源中寻找证据，以支持或反驳主张。对于主流新闻机构而言，这一流程依赖于寻找愿意接受引述的可信专家。与前一种方法相比，后一种方法还有两点优势。首先，它能让事实核查员获得针对主张性质量身定制的答案。其次，证据搜索可能无法解决以下主张：由于其本质，这些主张从未有过经过检验的记录；因此，咨询专家将是唯一的选择，尤其是在迫切需要尽快完成核查时。例如，新冠肺炎疫情期间提出的一些主张（如“注射漂白剂可以治愈新冠肺炎”）可能属于这一类。

无论采用何种方法，这两个组织的目标都是让读者自行判断一项主张是真是假。对 Full Fact 而言，这意味着以不带评判的方式呈现可信证据的摘要；而对主流新闻机构而言，则意味着呈现可信专家的观点。

显然，事实核查方法上的差异需要体现在工具支持的设计中。考虑到这一点，我们目前正在开发工具，协助识别专家，供事实核查员随后联系并采访、引用其观点（Zhang、He 和 Procter，2023）。不过，无论事实核查员遵循何种实践，他们很可能都记得自己是否核查过类似的主张；通过查阅自己或事实核查团队其他成员保存的记录，他们便能更快地处理新的主张。

事实核查员得出有关某项主张的结论后，团队的第二位成员有时甚至第三位成员会对其进行复核，以确保整个过程足够周全、没有遗漏，也不存在错误。这凸显了团队合作对于确保事实核查质量至关重要。完成复核后，事实核查结果便会整理成文并发布。两个组织都不希望被视为判断主张真伪的权威。相反，他们强调让受众自行判断，这通常意味着以平衡的方式呈现所收集的证据；尤其是因为，受众是否认可事实核查的价值，取决于事实核查员能否赢得读者的信任（Nakov 等，2021）。

受访者强调，事实核查可能极具挑战性，并指出主张的数量不断增加、可支配的资源有限，以及尽可能快速完成整个流程的压力。这表明，事实核查工具若能维护一个可搜索的核查结果存储库，将大有裨益；这样一来，当出现涉及类似主题的新主张时，便可以重新利用此前的事实核查结果（Das 等，2023）。

总而言之，事实核查团队成员指出，工具可以在事实核查工作流程的所有阶段以多种方式为他们提供帮助。总而言之，这些方式包括：发现主张、确定主张的优先级、搜索证据和/或专家，以及评估证据和/或专家。此外，事实核查工作中还有一个在这方面未被提及、但值得研究的方面，那就是协助撰写一份平衡的说明，阐述支持或反对某一特定主张的证据。最后，在我们之前的工作中，我们认为，为事实核查提供支持应集中在单一工具或仪表板中会是一个错误（Tolmie 等，2017），而我们在此报告的专门事实核查团队之间实践的多样性进一步印证了这一观点。

结论与未来工作

本文报告了对从事事实核查工作的新闻从业者开展一系列半结构化访谈所得的调查结果。我们发现，事实核查工作流程由一系列阶段构成。每个阶段对事实核查员而言都有其独特的挑战：他们需要在压力之下迅速判断某项主张的真实性并发布结果。显然，事实核查实践形式多样，而工作流程的稳健性取决于团队成员协作且持续的努力，这印证了我们早期研究的调查结果（Tolmie 等，2017）。

受访者指出，计算工具可以在事实核查工作流程的各个环节提供帮助：发掘主张、判断哪些主张值得核查，以及搜集证据。受访者对自动化这一流程的前景持怀疑态度。他们认为，事实核查员是参与其中的人类，理所当然必须由他们在评估主张时做最终裁定。正如其他领域引入先进决策工具时一样，事实核查工作中也有重要的专业和组织层面的原因，决定了这一点应当保持不变（Procter, Tolmie, and Rouncefield 2022）。因此，与近期的若干研究（例如 Demartini, Mizzaro, and Spina (2020)、Rubin (2022) 和 Das et al. (2023)）一样，我们强调，不应试图将事实核查自动化，而应确保工具建立在对事实核查工作的深入理解之上，包括对其协作层面的理解。

当然，任何新工具都必须易于使用；这也意味着，若要让专业人士信任这些工具，并认可它们适用于预定用途，就必须解决可解释性方面的挑战（Das 等，2023；Nakov 等，2021）。通过采用民族志研究方法（Rouncefield 和 Tolmie，2016），以及让事实核查员积极参与设计与开发（Wolf，2020），可以实现这些目标。前者对于深入理解工作及其对工具设计的影响至关重要；后者则有助于确保用户对计算工具能力的预期切合实际（Slota，2020；Arana-Catania、Van Lier 和 Procter，2021）。

最后，我们认为，要应对事实核查方法的多样性以及自然语言处理技术持续快速发展的局面，关键在于提供一套事实核查工具箱，即一组基于开放标准、彼此互操作的工具。这套工具可以配置，以适应特定事实核查团队及其成员的工作实践；也可以随着这些实践的变化重新配置。

致谢

我们谨此感谢各位事实核查团队成员，他们慷慨地拨出时间支持本项目。

本研究得到英国工程与自然科学研究委员会的资助（资助编号：EP/V048597/1）。此外，YH 和 ML 分别获得由英国研究与创新署资助的图灵人工智能奖学金（资助编号：EP/V020579/1、EP/V030302/1）。

参考文献

Abumansour, A. S.; and Zubiaga, A. 2021. QMUL-SDS 在 CheckThat! 2021：增强预训练语言模型，以估算阿拉伯语推文的核查价值。载于 CLEF（工作论文）。Arana-Catania, M.; Kochkina, E.; Zubiaga, A.; Liakata, M.; Procter, R.; and He, Y. 2022. 运用自注意力的自然语言推理：疫情主张的真实性评估。载于《2022年北美计算语言学协会分会年会：人类语言技术》论文集，1496–1511。Arana-Catania, M.; Van Lier, F.-A.; and Procter, R. 2021. 用于武装冲突调解的机器学习。arXiv 预印本 arXiv:2108.11942。B´elair-Gagnon, V.; Larsen, R.; Graves, L.; and Westlund, O. 2023. 平台事实核查合作中的知识工作。《国际传播期刊》，17：21。Birks, D. F.; Fernandez, W.; Levina, N.; and Nasirin, S. 2013. 信息系统研究中的扎根理论方法：其性质、多样性与机遇。《欧洲信息系统期刊》，22(1)：1–8。Bittner, E. 1965. 组织概念。《社会研究》，239–255。Bontcheva, K.; Liakata, M.; Procter, R.; and Scharl, A. 2015. 社交媒体中的谣言与欺骗研讨会：检测、跟踪与可视化。载于第24届万维网国际会议论文集。Brandtzaeg, P. B.; L¨uders, M.; Spangenberg, J.; Rath-Wiggins, L.; and Følstad, A. 2016. 关于社交媒体的新兴新闻核查实践。《新闻实践》，10(3)：323–342。Das, A.; Liu, H.; Kovatchev, V.; and Lease, M. 2023. 面向事实核查的以人为中心的自然语言处理技术现状。《信息处理与管理》，60(2)：103219。Demartini, G.; Mizzaro, S.; and Spina, D. 2020. 人机协同人工智能：应对在线错误信息的挑战与机遇。《IEEE数据工程公报》，43(3)：65–74。Diseases, T. L. I. 2020. COVID-19信息疫情。《柳叶刀·传染病》，20(8)：875。Dougrez-Lewis, J.; Kochkina, E.; Arana-Catania, M.; Liakata, M.; and He, Y. 2022. PHEMEPlus：利用外部证据增强社交媒体谣言核查。载于第五届事实提取与验证研讨会（FEVER）论文集，49–58。Dougrez-Lewis, J.; Liakata, M.; Kochkina, E.; and He, Y. 2021. 学习用于推特谣言真实性分类的解耦潜在主题。载于计算语言学协会论文集：ACL-IJCNLP 2021 调查结果，3902–3908。线上：计算语言学协会。Graves, L.; and Cherubini, F. 2016. 欧洲事实核查网站的兴起。技术报告，路透社新闻研究所。

Guo, Z.; Schlichtkrull, M.; and Vlachos, A. 2022. 自动化事实核查综述。《计算语言学协会学报》，10：178–206。Juneja, P.; and Mitra, T. 2022. 事实核查的人力与技术基础设施。《ACM人机交互汇刊》，6(CSCW2)：1–36。Kochkina, E.; Hossain, T.; Logan IV, R. L.; Arana-Catania, M.; Procter, R.; Zubiaga, A.; Singh, S.; He, Y.; and Liakata, M. 2023. 评估神经网络谣言核查模型的泛化能力。《信息处理与管理》，60(1)：103116。Krishna, A.; Riedel, S.; and Vlachos, A. 2022. ProoFVer：用于事实核查的自然逻辑定理证明。TACL，10：1013–1030。Lazarski, E.; Al-Khassaweneh, M.; and Howard, C. 2021. 使用自然语言处理进行事实核查：综述。《设计》，5(3)：42。Micallef, N.; Armacost, V.; Memon, N.; and Patil, S. 2022. 真还是假：研究专业事实核查员的工作实践。《ACM人机交互汇刊》，6(CSCW1)：1–44。Nakov, P.; Corney, D.; Hasanain, M.; Alam, F.; Elsayed, T.; Barr´on-Cede˜no, A.; Papotti, P.; Shaar, S.; and Martino, G. D. S. 2021. 自动化事实核查：协助人工事实核查员。arXiv 预印本 arXiv:2103.07769。Procter, R.; Tolmie, P.; and Rouncefield, M. 2022. 让人工智能担责：在医疗保健领域提供可信人工智能面临的挑战。《ACM人机交互汇刊》。Rouncefield, M.; and Tolmie, P. 2016. 工作中的民族方法学。Routledge。Rubin, V. L. 2022. 社交媒体内容核验：从欺骗检测到自动化事实核查。《SAGE社交媒体研究方法手册》。Si, J.; Zhou, D.; Li, T.; Shi, X.; and He, Y. 2021. 主题感知的证据推理与立场感知的聚合，用于事实核查。载于计算语言学协会第59届年会暨自然语言处理国际联合会议第11届会议论文集（第1卷：长篇论文），1612–1622。Slota, S. C. 2020. 跨分布式能动性进行设计：价值观、参与式设计与构建具有社会责任感的人工智能。Good Systems 出版研究。Tolmie, P.; Procter, R.; Randall, D. W.; Rouncefield, M.; Burger, C.; Wong Sak Hoi, G.; Zubiaga, A.; and Liakata, M. 2017. 支持在新闻实践中使用用户生成内容。载于2017年ACM CHI计算系统中的人为因素会议论文集，3632–3644。Wolf, C. T. 2020. 人工智能模型及其世界：通过工作实践视角研究数据驱动的人工智能/机器学习生态系统。载于《可持续数字社区：第15届国际会议 iConference 2020，瑞典布罗斯，2020年3月23—26日，论文集15》，651–664。Springer。Zeng, X.; Abumansour, A. S.; and Zubiaga, A. 2021. 自动化事实核查：综述。《语言与语言学指南》，15(10)：e12438。Zhang, W.; Gui, L.; and He, Y. 2021. 面向新冠肺炎大流行病中多模态不实新闻检测的监督式对比学习。载于第30届 ACM 国际信息与知识管理大会论文集，3637–3641。Zhang, W.; He, Y.; and Procter, R. 2023. NewsQuote：一个基于引文抽取与归因、用于专家推荐的数据集。载于国际 AAAI 网络与社交媒体大会论文集。AAAI Press。Zhao, R.; Arana-Catania, M.; Zhu, L.; Kochkina, E.; Gui, L.; Zubiaga, A.; Procter, R.; Liakata, M.; and He, Y. 2023. PANACEA：一个针对 COVID-19 的自动化错误信息检测系统。arXiv 预印本 arXiv:2303.01241。Zhu, L.; Fang, Z.; Pergola, G.; Procter, R.; and He, Y. 2022. 在社交媒体中用于疫苗态度检测的立场与方面主题解耦学习。载于2022年计算语言学协会北美分会年会：人类语言技术论文集，1566–1580。

附录：半结构化访谈问题

关于您和您的团队/组织

- 您的背景是什么?
- 您的团队/组织的目标是什么?
- 您的职责是什么？您担任该职务多久了?

事实核查工作

- 事实核查流程的主要阶段有哪些?
- 事实核查工作是如何组织的：– 如何发现主张？– 如何决定某项主张是否值得核查？– 哪些类型的证据对于事实核查最重要？– 目前使用哪些类型的工具，例如用于查找证据的工具？– 采取哪些措施来确保流程可靠且及时？– 事实核查工作面临的主要挑战是什么，原因又是什么?

- 事实核查流程的结果如何使用?
- 团队成员是否各司其职?
- 团队成员会参加培训吗?

支持事实核查的新工具

- 您目前使用的工具有哪些局限?

- 在事实核查流程的哪个阶段（或哪些阶段），新工具最能发挥作用?
- 新工具可以发挥哪些作用?
- 要让这些工具有效且值得信赖，会面临哪些挑战?

- 您认为事实核查可以自动化吗？自动化是否可取?