整合商业与社会健康决定因素：非临床健康决定因素的统一本体

Navya Martin Kollapally，理学硕士[1], Vipina Kuttichi Keloth，博士[2], Julia Xu，博士[3], James Geller，博士[1]1新泽西理工学院，美国新泽西州纽瓦克；[2]耶鲁大学，美国康涅狄格州纽黑文；[3]JX consulting，美国得克萨斯州

摘要

社会健康决定因素（*SDoH*）对人们健康与福祉的关键影响已得到广泛认可和研究。然而，健康的商业决定因素（*CDoH*）的影响直到最近才日益受到关注。构建 *CDoH* 本体有助于系统地识别和分类影响健康的商业因素。这些因素包括商品和服务的生产、分销与营销，可能对健康结局产生重大影响。本研究的目标是：*1*）利用 *PubMed* 文章和 *ChatGPT* 构建 *CDoH* 本体；*2*）将 *CDoH* 与现有的 *SDoH* 本体整合为统一结构，以促进本体复用；*3*）提出涵盖所有非临床健康决定因素的总体构想，并为其创建名为 *N-CODH* 的初步本体；*4*）验证 *ChatGPT* 提供的概念与现有 *SDoH* 本体之间的对应程度。

引言

根据世界卫生组织（WHO）的定义，健康的社会决定因素是指影响人们福祉的各种条件，包括出生、成长、生活和工作环境，以及获得医疗保健的机会[1]。这些条件受到经济、社会政策、政治和影响健康的商业因素等更广泛因素的影响。健康的商业决定因素是指商业实体影响个人和人群健康的情形、行动与不作为[2]。这些决定因素由追求利润的活动驱动，包括健康食品选择的可及性、营销与广告策略以及职场实践等因素。例如，企业采用的营销与广告策略会影响消费者行为和选择，可能导致不健康的行为与生活方式。因此，这些因素会影响可改变的风险行为，例如烟草使用、不健康饮食、缺乏身体活动以及有害的酒精摄入，进而导致超重和肥胖、血压升高、血糖水平升高、胆固醇升高，并最终引发心脏病、癌症、肝硬化、慢性呼吸系统疾病和糖尿病等危及生命的疾病。这些非传染性疾病可能包括生活方式疾病和心理健康问题。当非传染性疾病并非由个人所能控制，而是由商业活动造成时，可称为“工业流行病”或企业驱动型疾病[3]。在非传染性疾病导致的死亡中，心血管疾病占比最高（每年 1,790 万人），其次是癌症（930 万人）、慢性呼吸系统疾病（410 万人）和糖尿病（200 万人）。据估计，在美国，此类疾病每年造成的死亡占比为 88%，并导致 14% 的过早死亡（即年龄在 30 至 70 岁之间死亡）[4]。

文献中已提出多种定义和框架来描述商业健康决定因素。Kickbusch等人[2] 将商业健康决定因素定义为“私营部门为推广有损健康的产品和选择而采取的策略与方法。”作者将消费者与健康行为、个体化和选择列为微观层面的商业健康决定因素子类别；在宏观层面，则包括“全球风险社会”、“全球消费社会”和“全球化的政治经济学”。West等人[5] 将商业健康决定因素描述为“源自利润动机、影响健康的因素”。Lacy-Vawdon等人[6] 借鉴现有的商业健康决定因素和健康的社会决定因素定义，将商业健康决定因素定义为“最初围绕商业和/或企业权力体系形成的一系列系统”。作者承认，商业健康决定因素系统的影响可能是积极的，也可能是消极的，但主要重点应是预防和减少危害。Mialon[7] 以Kickbusch等人的研究为基础，构建了一个商业健康决定因素框架，其中列出以下因素：1）企业生产不健康商品；2）采取有害健康的商业、市场和政治实践；3）由企业实践塑造的全球性健康不良驱动因素。然而，作者指出，关于商业健康决定因素的研究有限，对全球性健康不良驱动因素关注不足，而且针对食品、酒精和烟草行业以外、同样促成商业健康决定因素的行业活动的研究也有限。总体而言，这些定义和框架凸显了商业活动与健康结局之间复杂的相互作用，也表明需要开展进一步研究并采取干预措施，以应对商业健康决定因素并促进健康公平。

如上所述，现有文献已广泛研究商业和企业利益对人群健康的影响。然而，将这些因素纳入商业健康决定因素框架仍是一个新兴研究领域[3,5,6]。现有的商业健康决定因素定义未能阐明商业健康决定因素与非传染性疾病相关风险行为之间的联系[3]。此外，这些定义也未同时考虑商业健康决定因素对人群健康的积极和消极影响。因此，有必要对商业健康决定因素的概念和分类进行标准化，以弥补术语方面的空白。为此，我们的目标之一是构建商业健康决定因素本体，以解决这些问题。根据Gruber[8], 的定义，本体是对特定目标领域共享概念化的正式、明确的规范。本体有望在模式层面整合多种信息来源，并可通过将关键词提升为本体概念及关系，从非结构化文本中实现信息检索。通过构建商业健康决定因素本体，我们旨在为这一特定领域提供一种途径，使知识管理工作能够基于共同的概念模型实现标准化[9]。

为了构建本体，必须全面列出涵盖所研究领域的术语/概念[10]。为丰富领域本体，开发者通常依靠相关研究文章收集概念，以拓展本体的深度和广度。本研究使用 PubMed Central (PMC) 作为信息来源，查找该领域的相关研究，并最终从中提取概念。然而，即使进行了广泛搜索，收集所有相关概念并确保本体内容全面仍然颇具挑战。自然语言处理技术和聊天机器人已用于文本摘要任务。GPT（生成式预训练变换器）[11] 模型是一种经过大型文本数据集训练的语言模型。ChatGPT[12] 是由 OpenAI 公司开发的聊天机器人。它基于 OpenAI 的 GPT-3.5/4 系列大语言模型构建，并通过监督学习和强化学习技术进行微调。在本研究中，我们采用 ChatGPT 补充搜索策略；它能够针对自然语言提示生成类似人类的回复。我们提出一种新颖的人机协作概念收集方法，利用 ChatGPT 扩展概念集，以构建商业健康决定因素（CDoH）本体。

为全面整合商业化在 CDoH 框架中的影响，可能有必要拓宽现有范式的范围，以探讨这些范式影响健康的社会决定因素的方式。例如，与就业和工作条件相关的商业政策和实践会影响劳动者的社会经济地位，进而对健康结局产生影响。拓宽 CDoH 框架的范围，考察商业活动与健康的社会决定因素之间的相互作用，有助于我们更好地理解这些因素如何影响健康结局。因此，本研究借鉴了非临床健康决定因素的理念[13]（为此，我们提出 N-CODH 本体：非临床健康决定因素本体；读作 *en-code*），将健康的社会决定因素和商业健康决定因素对健康的影响整合起来。我们通过将现有的社会健康决定因素本体（SOHO； 参见论文末尾所列的所有本体术语表）与本文介绍的 CDoH 本体相结合，发布了 N-CODH 的初始版本[14] 。我们假设，N-CODH 本体有望变革研究、政策制定和公共卫生实践的方式，为塑造健康结局的复杂且相互关联的因素提供更全面、更细致的认识。

总而言之，我们正在利用 PMC 文章和 ChatGPT 开发商业健康决定因素本体。我们通过将商业健康决定因素与现有的社会健康决定因素本体（SOHO）整合，支持本体复用。我们提出了一个涵盖所有非临床健康决定因素的总体构想，并创建了一个名为 N-CODH 的初始本体。最后，我们正在验证 ChatGPT 提供的概念与 SOHO 本体之间的对应程度。

方法

文献综述与商业健康决定因素概念提取

我们依据 Noy[10] 提出的本体开发原则，构建了 CDoH 本体。其中一项本体设计原则是复用现有本体中的内容[15]。因此，我们首先检索了 NCBO BioPortal[16]。它是一个汇集多种本体和术语体系的平台，目前收录了 1,052 个本体和术语体系，包含 15,644,567 个类和 36,286 个属性。BioPortal 中的本体会复用其他本体的内容，以便对新类进行建模、覆盖某一主题领域、减少开发工作并支持应用程序。我们的本体旨在定义商业健康决定因素的健康影响，但我们未能在 BioPortal[17] 中找到此类本体。为得出这一结论，我们使用以下关键词进行了搜索：“健康的商业决定因素”“企业健康决定因素”“商业因素导致健康受损”“商业健康损害决定因素”“非传染性疾病的商业驱动因素”“非传染性设备的商业决定因素”，

“肥胖的商业决定因素”及其变体，并在 BioPortal 的查找本体和类搜索字段中进行检索。由于搜索未返回任何结果，我们因此从头开始开发 CDoH 本体。

我们采用了系统评价和荟萃分析优先报告条目框架（PRISMA 2020）[18]，如图 1 所示。为收集开发 CDoH 本体所需的相关文章，我们在 PubMed Central (PMC)[19] 中进行了范围检索，检索式为：*(commercial [所有字段] AND determinants [所有字段] AND ("health"[MeSH 术语] OR "health"[所有字段]) AND +framework [所有字段]) AND ("2018年1月17日"[出版日期] : "2023年1月15日"[出版日期])*. 检索共返回 23,342 篇全文文章。移除受禁发限制的文章后，23,094 篇全文文献进入下一轮筛选。在这一阶段，23,071 篇文章因符合以下排除标准而被剔除：“针对某一亚群体的研究，且缺乏更广泛的启示”以及“标题或摘要未讨论 CDoH 对健康／气候的影响”。我们确定了 23 篇全文文章不符合排除标准。我们进行了前向追溯（从已识别来源的参考文献目录中提取相关文章）和后向追溯（提取引用已识别文章的文献）。前向追溯帮助我们找到非学术文章，包括政府网站上的政策文件和人口统计数据，因此又从 PMC 以外补充了 14 篇文章。

本体构建者经常面临的一个问题是，他们需要与“昂贵”的领域专家和本体专家合作。理想情况下，本体的贡献者应同时具备领域知识和本体方面的专业知识。为应对这类专家难以招募的问题，我们开展了一项试点研究，探索将 ChatGPT 用作“贡献者”。我们通过向 ChatGPT 提问，提取了商业健康决定因素对公共卫生的独特影响。示例提示包括：“商业健康决定因素对健康结果的影响”“商业健康决定因素对健康的影响有哪些子类别”“商业驱动因素和企业导致健康受影响的因素”“商业健康决定因素带来的气候灾害”“由企业造成、导致健康不良的 10 种气候变化影响”“列出私营部门中导致生活方式疾病的 20 类因素”等。我们提出了若干语义相近的问题，并从 ChatGPT 提取出 40 项独特影响。我们通过在 PMC 中检索相应文章，使用提取出的影响作为检索关键词，对每项影响进行了验证。分析后，我们新增了 72 篇先前综述中排除的文章。纳入阶段结束后，我们获得了 109 篇全文研究文章、报告和政策文件，用于概念提取。我们对这 109 份文献进行了人工审阅，以提取开发 CDoH 本体所需的全部概念。（未来，我们将使用最新的自然语言处理方法重新审视这一环节。）

商业健康决定因素本体的开发

在对概念收集阶段提取的所有概念进行全面分析后，我们将商业健康决定因素概念划分为五个主要类别：1）归因于商业因素的要素，2）归因于经济因素的要素，3）归因于环境因素的要素，4）归因于个体因素的要素，以及5）归因于社会因素的要素。我们使用 Protégé 5.5.0[20]，以网络本体语言（OWL）实现 CDoH 本体。Protégé 将“概念”称为“类”，并允许添加类之间的属性和关系。“Thing”类是 Protégé 中预定义的类，用作使用该工具创建的每个本体的根类。Protégé 支持用户编辑 OWL 本体，并使用推理机验证所开发本体的一致性和逻辑。使用 HermiT 1.4.3.456[21] 版推理机在 Protege 中进行了一致性检查。我们还为概念添加了对象属性和数据属性，以便描述归因于不同因素的要素之间的复杂关系。对象属性的示例包括“具有_教育_水平”，该属性将“个人”与“教育水平”关联起来；以及“具有_污染物，”该属性将“可用饮用水源”与“氡”“氟化物”等化学物质关联起来。

通过整合 *CDoH* 本体与另外三个本体构建 *N-CODH* 本体

考虑到商业健康决定因素与健康的社会决定因素之间存在重叠，商业活动可能影响社会因素，反之亦然。要有效应对这种复杂的相互作用，采取全面、整合的方法至关重要。在开发出 CDoH 本体后，我们介绍开发 N-CODH 本体的方法。我们导入了三个现有本体，这些本体包含影响非临床结局的因素，以提高 CDoH 本体的覆盖范围和灵活性[22]。我们将设计的本体与此前开发的、收录于 BioPortal[23] 的社会健康决定因素本体（SOHO[a]）进行了整合。SOHO 着重涵盖健康不公平在医院和从业人员中造成的医疗后果，并通过复用 HOME 本体[24]

[a]由于 SDoH 与 SOHO 历来容易混淆，我们使用 Agency FB 字体显示 SOHO 一词。为进一步减少混淆，我们没有为本文开发的 CDoH 本体命名。请参阅文末的术语表。

（少数群体公平医疗本体）。HOME 专门涉及医疗保健内外的隐性偏见所造成的医疗影响。通过在本体中标准化非传染性疾病的致病因素，我们可以应对其定义、分类和应用中存在的异质性问题。因此，通过将健康的社会决定因素概念加入 CDoH 本体，我们创建了一个全面涵盖非临床健康决定因素的本体。N-CODH 的开发是本研究的一项重要成果。此外，为表示事件随时间的演进，我们还从 BioPortal 导入了时间事件本体（TEO）至 N-CODH。我们向 N-CODH 添加了数据属性，例如“parts_per_million”，以表示饮用水中的最高化学污染物含量。我们为 N-CODH 添加了 CURIES 标识符注释，以确保互操作性，并使其更便于用作自然语言处理任务的金标准。

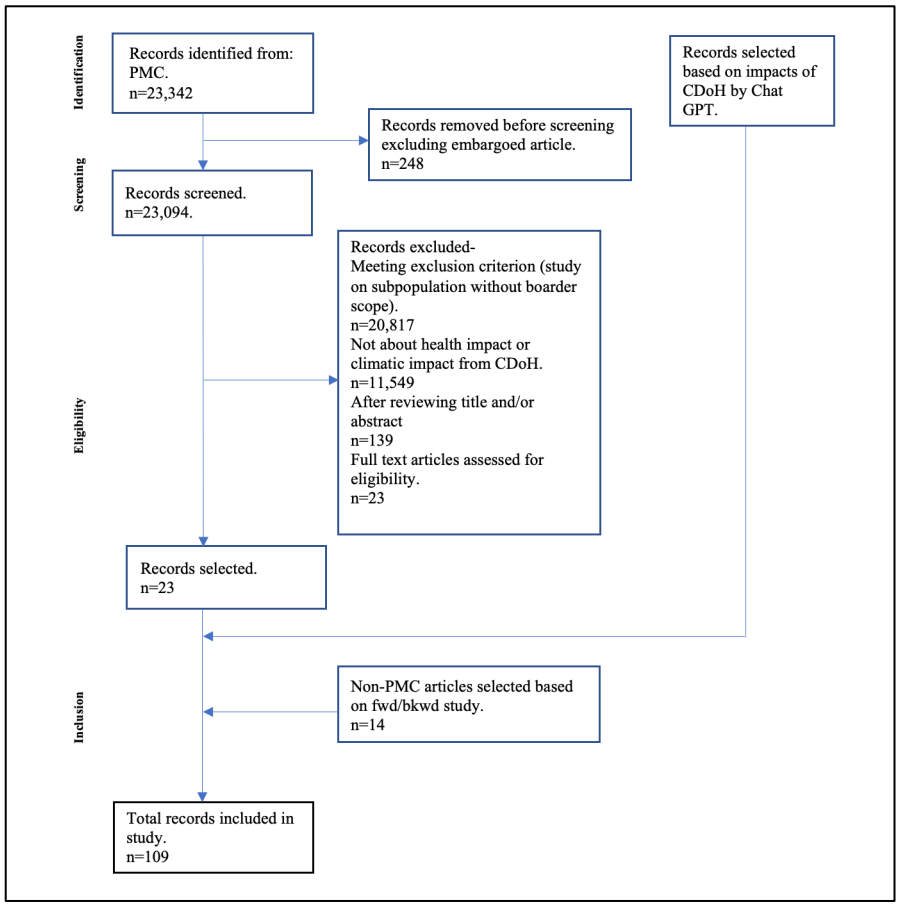


图 1。研究纳入的 PRISMA 流程图。

本体评估

本体评估是指根据一组评估标准判定本体质量的过程。本体评估的四种主要方法包括金标准比较、基于应

基于应用的评估：HermiT 推理机可用于判断本体是否一致且连贯。另一方面，OntoMetrics [26] 旨在评估本体的某些方面及其在知识表示方面的潜力。OntoMetrics 提供的指标描述了本体中与领域无关的特征，能够比 HermiT 提供更深入的见解。使用 Protégé 开发的 OWL 文件已作为

XML 文件上传至 OntoMetrics，以计算各项指标，尤其是模式度量指标。模式度量指标用于评估所设计本体的深度、宽度、丰富度和继承结构。关系丰富度反映了本体中关系的多样性及其分布。属性丰富度反映了为每个类定义的属性数量。它既能体现本体设计的质量，也能反映与实例数据相关的信息量。继承丰富度用于衡量信息在本体继承树不同层级间的分布，或父类的分支数量。这可以很好地表明本体中的知识是否得当地归入不同类别和子类别。类丰富度与实例在各个类之间的分布有关。

人工专家评估：在验证 N-CODH 本体的一致性、连贯性和语义正确性后，我们采用人工专家评估，考察所构建的本体是否正确涵盖了相关领域的各个方面。我们设计了一个电子表格，其中的概念对采用“父概念 ß IS-A- 子概念”的形式，以尽量减少歧义。父概念和子概念通过 IS-A 关系相连。表 1 展示了评估表中概念对的一个片段。

表 1。 提供给人工专家进行评估的概念对电子表格片段。

父概念	**关系（相同）**	**子概念**	**子概念?**	**更远 away**	**无关原因**
气候变化的影响 变化	ßIS-A-	不健康食品的营销 食品	No		子概念与 不健康食品推广 相关，与 父概念无关，而父概念 与气候变化相关。
饮食相关 精神病理学	ßIS-A-	暴食症	Yes		
化学风险涉及 饮用水	ßIS-A-	社交媒体影响了 健康结局	No		受影响的健康结局 社交媒体不能成为 饮用水中的化学风险的子概念 饮用水
贸易与 全球化 对健康的影响 差异	ßIS-A-	违反劳动标准		Yes	这些概念共享一种 祖孙 关系

两位人工评估者（VK 和 JX）收到的是同一份包含 100 对概念对的电子表格。在这 100 对中，我们提供了 10 对概念作为训练样本，用于展示本体的特点；另有 90 对需要评估。电子表格包含三类概念对：具有父概念与子概念关系的概念对、具有祖先_或¯祖父母与子概念关系的概念对，以及不具有层级关系的概念对。VK 和 JX 都知道电子表格中包含这些不同类型的概念对。电子表格中有三个空列，标题分别为“子概念”“更远”和“无关原因”。提供给评估者的 10 个样本中，有五个“子概念”字段填写了“否”，并在“无关原因”栏中提供了相应理由。三个“子概念”字段填写了“是”，两个“更远”字段填写了“是”。

对于每一对概念，如果评估者认为这些概念之间存在父子（IS-A）关系，则需在第四列（表 1 中的“子概念？”）填写“是”；否则填写“否”。如果答案为“否”，则需在“无关原因”列中填写原因。在评估阶段，这些原因为我们改进本体设计提供了方向。评估者认为这些概念之间存在祖孙关系或祖先关系（即由 IS-A 关系构成的链）时，还需在“更远”列中填写“是”。在这种情况下，评估者也需填写原因（在“无关原因”列中）。VK 和 JX 分别独立审阅了这些概念对，我们使用科恩 κ 系数[27] 检验来确定一致程度。科恩 κ 系数（κ）是一种表示两位评定者之间一致程度的统计系数。$\kappa > 0.4$ 被视为中度一致，$\kappa = 1$ 则表示完全一致。为评估各自结果的统计显著性，我们使用了 Fisher 精确检验[28]。

评估本体与 *ChatGPT* 的一致性

为探讨该本体与 ChatGPT 的一致性，我们采用了先前研究中为SOHO开发的评估表。社会健康决定因素本体（SOHO）可在 BioPortal[23] 上获取，曾由两位医学本体专家和一位医师进行评估。在本研究中，我们按照自然语言用户查询模式向 ChatGPT 提供概念对：“邻里与建成环境”ß*IS-A*—“邻近工业设施”，并询问：“这是否是有效的 IS-A 关系？”该箭头属于提供给 ChatGPT 的输入内容。ChatGPT 的回答要么是肯定的（类似于“是的，这是有效的父子关系”），要么是否定的（“不，这些概念之间不存在严格的 IS-A 关系”），并附上相应解释。对于 ChatGPT 给出否定回答的情况，我们会提出后续问题，以确定如何定义这种关系或如何修改子概念。表 2 展示了向 ChatGPT 提供的部分概念对。在SOHO验证研究中，60 对概念中有 20 对存在父子关系，20 对互不相关，其余 20 对存在祖孙关系（即这些概念彼此相关，但并非直接相关）。

表2. 提供给 ChatGPT 的概念对示例。

父概念	**关系**	**子概念**
粮食不安全的影响	ßis--a	营养不良导致的代谢紊乱
住房条件差	ßis--a	校园欺凌
经济不稳定	ßis--a	无法参加联邦援助
工作场所状况不佳	ßis--a	工作中团队成员搭配不佳

对于那些 ChatGPT 不认为通过 IS-A 链接相关联，而认为属于祖孙关系的概念对，我们尝试了一种新颖的评估方法：向 ChatGPT 提出一系列问题，并在图2中以图示方式进行说明。子图 a) 显示，我们向 ChatGPT 提出 B 是 A 的子概念。然而，ChatGPT 表示，它“认为”B 是 A 的孙概念。子图 b) 对此进行了图示说明。随后，我们要求 ChatGPT 列出 A 的子概念（子图 c)）。有趣的是，在某些情况下，它将 B 列为 A 的子概念（子图 d)）；而在其他情况下则没有。

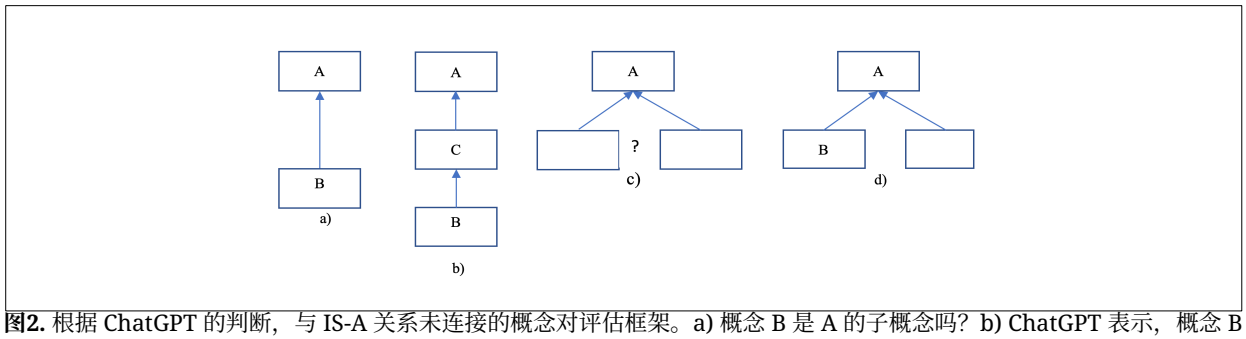


图2. 根据 ChatGPT 的判断，与 IS-A 关系未连接的概念对评估框架。a) 概念 B 是 A 的子概念吗？b) ChatGPT 表示，概念 B 是概念 A 的孙概念。c) 提示 ChatGPT 列出概念 A 的所有子概念。d) ChatGPT 列出 A 的所有子概念，其中包括 B，自相矛盾。

根据 SOHO，“住房条件差”ß是一个“虫害滋生的房屋”，但 ChatGPT 不同意这一关系，并指出：“住房条件差”和“虫害滋生的房屋”之间可能存在较远的层级关系。就 ChatGPT 的观点而言，住房条件差可以涵盖多种使住宅不符合标准的状况，其中一种可能就是虫害侵扰。接下来，我们提示 ChatGPT 返回 10 个与“住房条件差”存在 IS-A 关系的概念。ChatGPT 的回答包括昆虫或虫害侵扰，以及其他概念，例如住房拥挤、缺乏基本设施、暴露于环境危害、缺乏通风、无家可归等。在结果部分，我们将介绍这些案例的具体分解情况。我们总共使用了 276 个提示，以获取 ChatGPT 对这 60 对概念的评估结果。

结果

构建的 *N-CODH* 本体

使用 Protégé 开发的 CDoH 本体包含 317 个类和 675 条公理，以及 27 个对象属性和 19 个数据属性。图3展示

它将 CDoH 本体与现有的健康的社会决定因素本体 SDHO、医疗公平本体（HOME）和时间事件本体（TEO）整合：

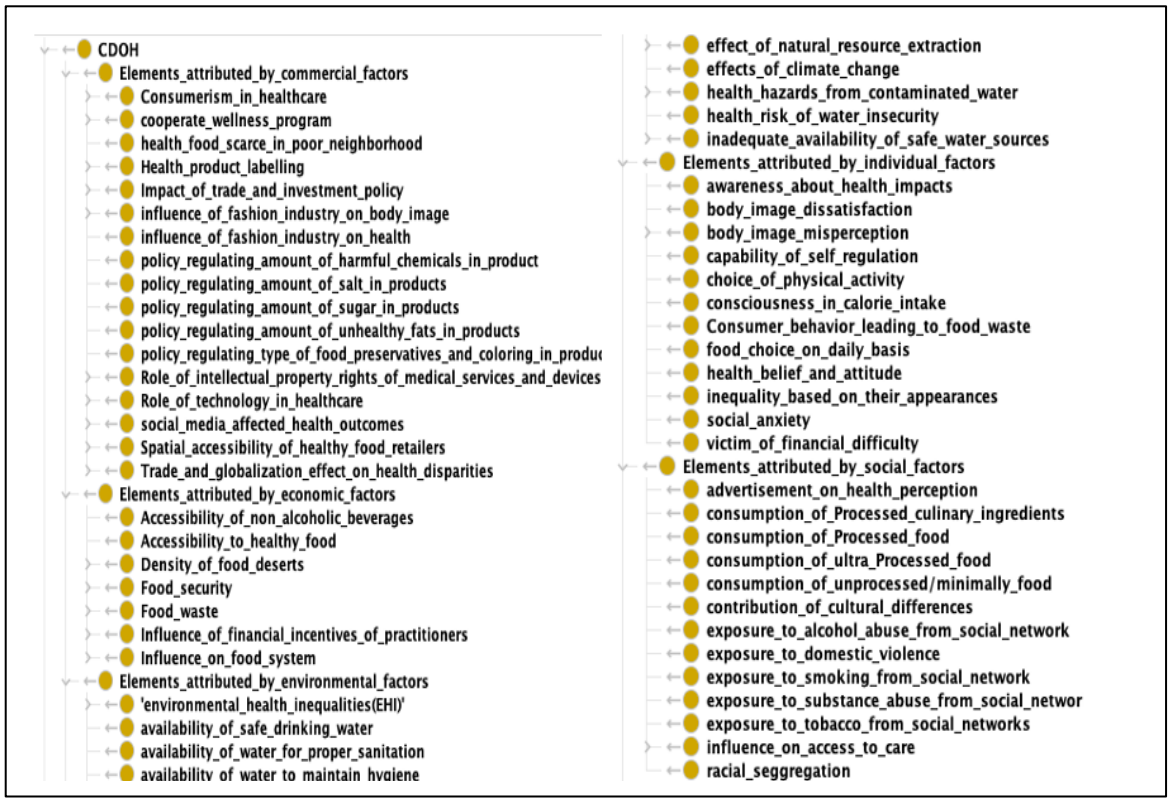


图3。Protégé 中 CDoH 本体的主要类及直接子类。

N-CODH 的指标质量

根据 Protégé 中运行的 HermiT 推理机，N-CODH 是一个连贯且一致的本体。我们使用 OntoMetrics 对该本体进行了分析，以获取模式度量指标，结果如表3所示。N-CODH 本体旨在全面表示健康的商业决定因素**和**健康的社会决定因素的影响。其特点是属性丰富度较低，而继承丰富度较高。继承丰富度代表了本体的横向性质，表明继承层级较少，且每个类对应的子类数量较多。N-CODH 主要由类-子类关系构成（而非语义关系），因此语义关系丰富度较低；语义关系丰富度体现了关系的多样性及其在本体中的分布。

N-CODH 的人工评估结果

人类专家进行了评估，以确保 N-CODH 所表示的领域知识准确反映人类的直觉。VK 和 JX 分别独立评估了 90 对随机概念，其中包括 32 对 IS-A 关系对、 14 对祖父母 - 子代关系对，以及 44 对错误地以 IS-A 关系连接的无关概念对。表4展示了用于计算科恩 κ 系数的输入值。我们得到 $\kappa=0.50502$，这表明一致率为 74.44%，进而说明两位评估者对本体的看法具有一般程度的一致性。表5和表6分别列出了与每位评估者对应的费希尔精确检验混淆矩阵。在该指标中，具有层级关系的概念对包括 IS-A 关系和祖先¯孙辈关系。两位评估者的 p 值均为 <0.0001，小于 $p=0.05$。这表明评估结果具有统计显著性[28]。根据专家反馈，我们为提高清晰度，重命名了 N-CODH 中两个父概念。“农贸市场交通可达性”改为“前往农贸市场的交通可达性”。“被驱逐出境的恐惧”改为“从事危险工作的非法劳工被驱逐出境的恐惧”。

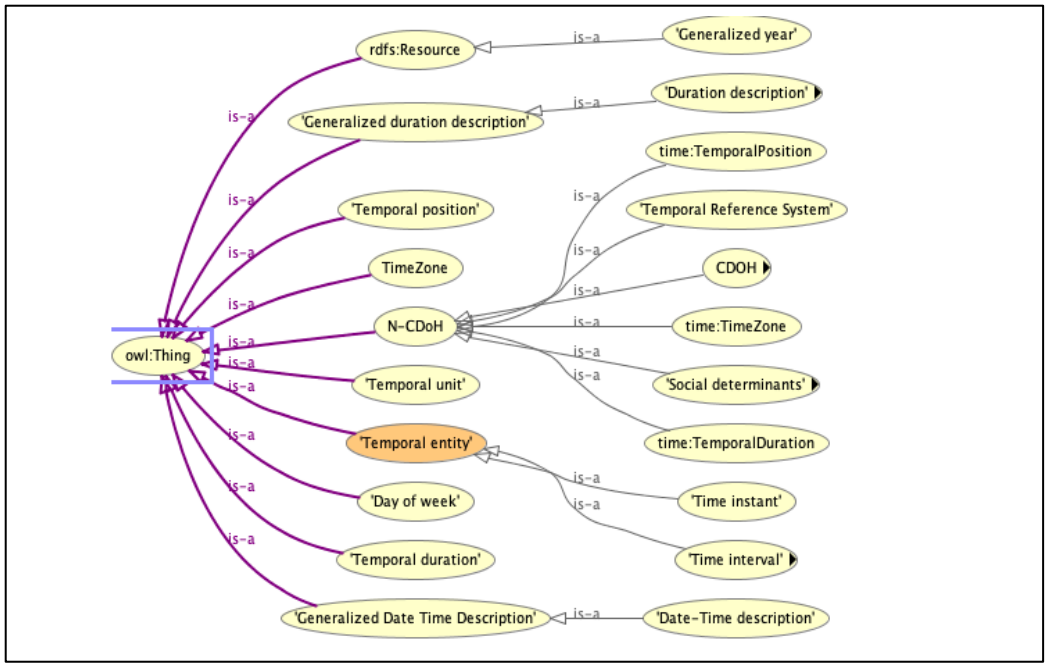


图 4。 N-CODH 的顶层层级类（图中称为 N-CDoH）。

表 3。 OntoMetrics 返回的模式度量指标。

指标	**值**
属性丰富度	0.008876
继承丰富度	0.98816
关系丰富度	0.12336
公理/类比率	4.49905
类/关系比率	0.88713

表 4。 科恩卡帕系数输入指标

描述	**数量**
两位评估者均同意纳入	31
两位评估者均同意排除	36
第一位评估者希望纳入	3
第二位评估者希望纳入	20

ChatGPT 的验证研究结果： ChatGPT 认定，从 SOHO 中选取的 20 对无关概念不应通过 IS-A 关系连接。它还正确识别了 20 对祖孙关系。父子关系方面的结果则不够理想。对于父子概念对，初始一致数=9，不一致数为 11。根据图2.c)，我们尝试为其中 11 对里的 7 对建立父子关系。对于其中 5 对，ChatGPT 在第二步将子概念识别为子概念，对应图2.d)。其余两对则未能如此识别。在剩余的 4 对（=11–5–2）概念对中，ChatGPT 认为 3 对由“部分”关系连接，另 1 对由“类型”关系连接。就我们的用途而言，我们认为类型关系与父子（IS-A）关系足够相似。

表5。 评估者1的混淆矩阵。

混淆矩阵	层次相关概念对	不相关概念对
评估为层次相关概念对	39	0
评估为不相关概念对	7	44

表 6。 评估者 2 的混淆矩阵。

混淆矩阵	层次相关概念对	不相关概念对
评估为层次相关概念对	42	11
评估为不相关概念对	4	33

结论

在本项目中，我们创建了一个本体，用于阐述商业健康决定因素（CDoH）对健康的影响，其中包括由商业行为引发的气候变化所造成的健康危害等概念。我们使用 Protégé 5.5.0 开发了 CDoH 本体，其中包含 675 条公理、317 个类、27 个对象属性和 19 个数据属性。关于 CDoH 的研究表明，我们需要将其与此前开发的社会健康决定因素本体（SOHO）、少数群体公平医疗保健本体（HOME）以及时间事件本体（TEO）整合，从而开发出 N-CODH。初始版 N-CODH 本体包含 611 个类和 2603 条公理。

为了评估 N-CODH 本体，我们使用了 HermiT 推理机和 OntoMetrics 工具，并请两位人类专家评估其领域覆盖范围。我们还开展了一项验证研究，以确定 ChatGPT 是否能够辅助本体开发。 通过将 ChatGPT 作为“贡献者”，我们得以补充文献发表和概念收集工作，并扩大本体的覆盖范围。这种人机协作方法有望降低构建本体所需的成本和时间，同时保持较高的准确性和严谨性。在验证研究期间，ChatGPT 提示我们，60 对概念中有 11 对并非严格的 IS-A 关系。 因此，通常而言，本体开发人员可以借助 *ChatGPT* 重新审视并检查父子概念对，并作出必要调整。 换言之，ChatGPT 可用作一项重要工具，用于验证其他相关概念对，从而丰富本体，使其达到所需的粒度。

局限性与未来工作

本研究的一项局限是，概念由人工审阅确定。第二项局限是，仅纳入了来自 PMC 的研究文章。第三项局限是，我们采用的排除标准可能导致遗漏相关概念。为弥补这些不足，未来将利用自然语言处理技术，从政策文件、人口调查、死亡率调查、临床记录、科学出版物等来源提取相关概念。排除标准也将放宽。最后一项值得注意的局限是，人工审阅仅由两位专家完成。条件允许时，将增加第三位专家加入团队。

术语表

SOHO：健康的社会决定因素本体：收录于 BioPortal，由本团队此前开发。CDoH 本体：健康的商业决定因素本体（无缩写）：收录于 BioPortal，在本文中开发 HOME：少数群体公平健康本体：收录于 BioPortal，由本团队此前开发。N-CODH（*en-code*）：非临床健康决定因素本体：收录于 BioPortal/GitHub，在本文中开发 TEO：时间事件本体：收录于 BioPortal。已导入 N-CODH。

参考文献

1. 世界卫生组织。健康的商业决定因素。[在线]。；2021 [引用于2023年2月10日。网址：https://www.who.int/health-topics/commercial-determinants-of-health#tab=tab_1。2. Kickbusch I、Allen L 和 Franz C，《健康的商业决定因素。》，《柳叶刀·全球健康》，第4卷，第e期，第895-896页，2016年。3. Lee K 和 Freudenberg N，《应对健康的商业决定因素，始于更明确的定义和测量。》，《全球健康促进》，第27卷，第2期，第3-5页，2020年6月。4. 世界卫生组织（WHO）。《非传染性疾病进展监测报告》。；2022年。5. West R 和 Marteau T，《关于 Casswell（2013）的评论：健康的商业决定因素》，《成瘾》，第108卷：686，2013年。6. Lacy-Vawdon Cd、B.Vandenberg、Livingstone CH。《正视房间里的大象：健康的商业决定因素。》《BMJ Global Health》。2022年；7. 7. Mialon M，《健康的商业决定因素概述》，《全球健康》，第16卷，第74期，2020年。8. Gruber TR。《迈向用于知识共享的本体设计原则》。《国际人机交互研究》。1993年：第907-928页。

9. Stephan G、Pascal H 和 Andreas A，《知识表示与本体》，柏林：语义网服务，2007年。10. McGuinness D 和 Noy NF，《本体开发 101：创建你的第一个本体指南。》，2001年。11. Radford A、Wu J、Child R、Luan D、Amodei D 和 Sutskever I，《语言模型是无监督多任务学习器。》，计算机科学，2019年。12. OpenAI，《ChatGPT》，[在线]。网址：https://openai.com/blog/chatgpt/。[访问日期：2月2023]。13. Golembiewski E、Allen KS、Blackmon AM、Hinrichs RJ、Vest JR。《结合非临床健康决定因素与临床数据开展研究与评估：快速综述。》JMIR 公共卫生监测。2019年10月7日；5(4)：e12846。doi：10.2196/12846。PMID：31593550；PMCID：PMC6803891。14. Kollapally NM、Geller J。NCBO BioPortal。[在线]。；2023年。网址：https://bioportal.bioontology.org/ontologies/N-CDOH?p=summary。15. Ochs C、Perl Y、Geller J、Arabandi S、Tudorache T 和 Musen MA，《BioPortal 中本体复用的实证分析》，生物医学信息学杂志，第71卷，第165-177期，2017年7月。16. Whetzel P、Noy NF、Shah N、Alexander P、Nyulas C、Tudorache T 和 Musen MA，《BioPortal：国家生物医学本体中心提供全新 Web 服务，增强访问和使用软件应用中本体的功能。》，*Nucleic Acids Res*，第W541-5页，2011年7月。17. Noy NF、Shah N、Whetzel PL、Dai B、Dorf M、Griffith N、Jonquet C、Rubin D、Storey MA、Chute CG 和 Musen MA，《 BioPortal：轻点鼠标即可访问本体和集成数据资源》，*Nucleic Acids Res*，第 37 卷，第W170-3页，2009年7月。18. 英国医学杂志。《PRISMA 2020声明：系统综述报告指南的更新。》；2021年。报告编号：n71。19. 美国国家医学图书馆。PubMed Central（PMC）。[在线]。[引用于2023年2月。网址：https://www.ncbi.nlm.nih.gov/pmc/。20. Musen MA。《Protégé项目：回顾过去，展望未来。》《《AI Matters》。计算机协会人工智能特别兴趣小组。2015年6月；1。21. Glimm B、Horrocks I、Motik B、Stoilos G、Wang Z。HermiT：一个 OWL 2推理器。《J. Autom. Reason.》2014年；15(3)：245-269。22. Kollapally NM，GitHub。[在线]。网址：https://github.com/navya777/N-CDoH。23. Kollapally NM、Chen Y、Xu J 和 Geller J，《健康的社会决定因素领域本体。》，电气与电子工程师协会 - BIBM 2022，第2403-2410页。24. Kollapally NM、Chen Y 和 Geller J，《少数群体公平健康本体（HOME）》，知识工程与本体开发国际会议，第1727页，2021年 - 。25. Raad J 和 Cruz C，《本体评估方法综述。》，第七届国际知识发现、知识工程与知识管理联合会议论文集，第10.5220/0005591001790186页，2015年。26. Tello、Lozano A 和 Pérez AG，《ONTOMETRIC：选择适当本体的方法》，《数据库管理杂志》，第15卷，第1-18页，2004年。27. Idostatistics。Cohen κ系数计算器。[在线]。[引用于2023年2月。网址：https://idostatistics.com/cohen-kappa-free-calculator/。28. Fisher精确检验 [在线]。；2023年。网址：https://www.socscistatistics.com/tests/fisher/default2.aspx。29. language Be。标识符。[在线]。[引用于2023年1月。网址：https://biological-expression- language.github.io/identifiers/。30. 罗斯托克大学。Ontometrics维基。[在线]。网址：https://ontometrics.informatik.uni-rostock.de/wiki/index.php/Schema_指标。