

Using an Artificial intelligence chatbot to critically review the
scientific literature on the use of Artificial intelligence in
Environmental Impact Assessment (Second part)

使用人工智能聊天机器人对在环境影响评价中应用人工智能的
相关科学文献的评论性综述（下）

Alan Bond^{a, b}, Dirk Cilliers^b, Francois Retief^b, Reece Alberts^b, Claudine Roos^b, Jurie Moolman^b

^aSchool of Environmental Sciences, University of East Anglia, UK;

^bUnit for Environmental Sciences and Management, North-West University, South Africa

^a环境科学学院，东英吉利大学，英国；

^b环境科学与管理系，西北大学，南非

摘要：关于人工智能（AI）在环境影响评价（EIA）中可能发挥的作用，包括在研究中的作用，存在诸多不确定性。AI大型语言模型（LLM）聊天机器人有可能提高EIA研究的效率，但其输出结果可能引发担忧。本文研究了使用LLM聊天机器人对聚焦于AI在EIA中应用的文献进行评论性综述时可能节省的时间。通过结合使用ChatGPT和Elicit，对与AI在EIA中应用相关的文献进行了评析，识别出12个关键问题，并在3天半内实现从初步构思到完成本文。本文制定了一项协议，以帮助研究人员验证通过Elicit（或其他机器学习工具）提供的证据，这也是本文的一个新成果。通过3位同行评审员的评论，能够更客观地反思LLM聊天机器人输出的可信度、时间节省的合理性及未来在这种情况下应用LLM聊天机器人所需的研究。

关键词：人工智能（AI）；环境影响评价（EIA）；未来挑战；伦理；问题

（续接2025年第4期IAPA文摘）

4 修改后的ChatGPT结果

ChatGPT生成了大量数据，如图1所示。尽管通过文献综述发现的证据验证了一级问题，但很快发现二级因素是存在问题的。这些问题可以分为以下3类：

（1）循环论证——一级问题作为二级因素重复出现。

（2）重复——在两个或多个一级问题下重复相同的二级因素。

（3）虚假信息[在Costello（2023年）之后]——ChatGPT识别的问题或因素本身并非真正的

问题或因素，而是事实陈述。

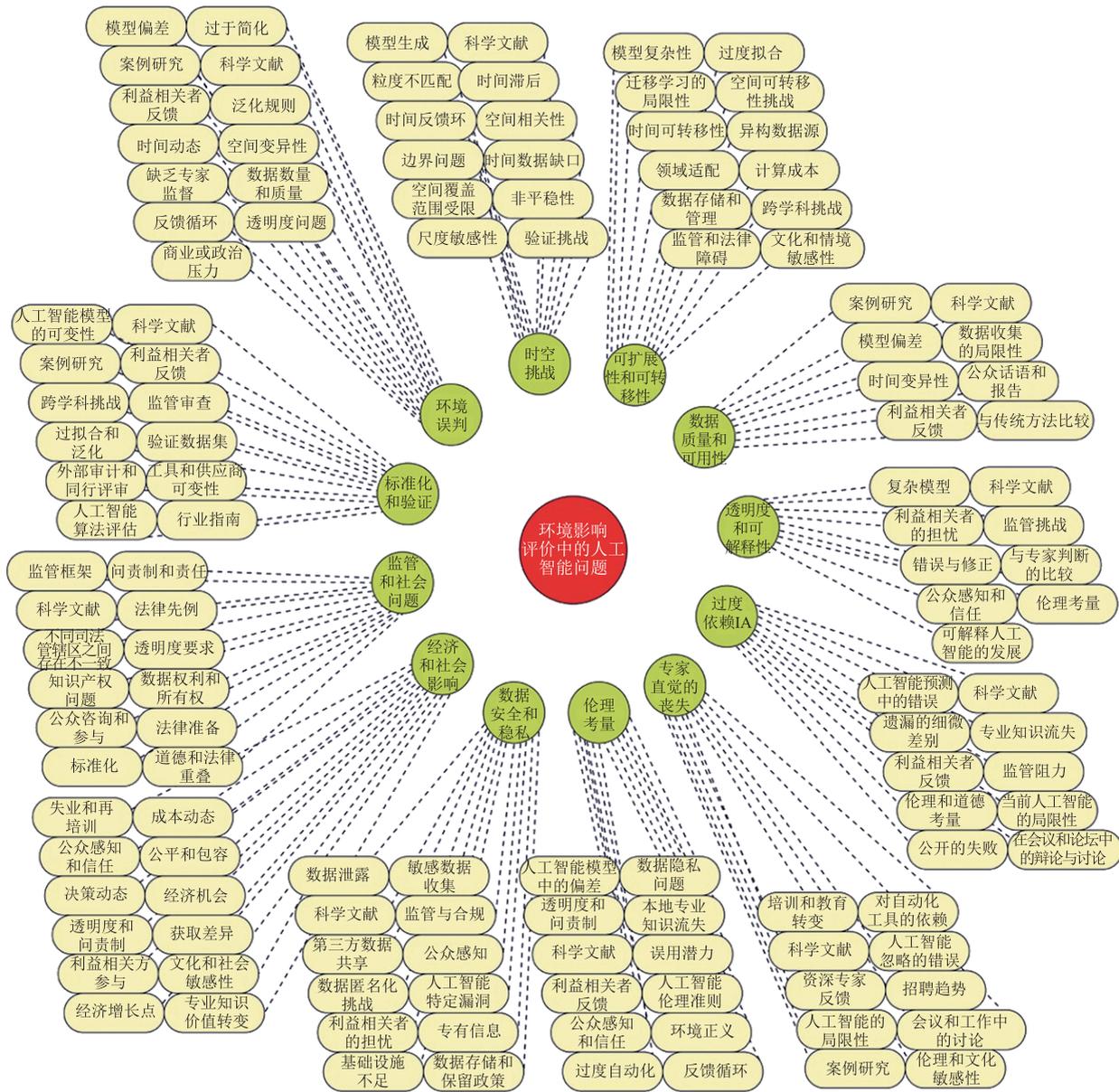
为了实现目标2，作者需要修改图1中总结的结果，以消除循环论证、重复和虚假信息。文本框3至文本框5通过作者团队的决策示例，依次解释了这3个类别是如何处理的。文本框3提供了两个关于如何处理循环论证的示例，文本框4提供了两个关于如何处理重复的示例，文本框5提供了两个关于如何处理虚假信息的示例。

这3类问题在每个类别中都得到了一致的处理，具体处理方法分别在文本框3至文本框5中详细说明，对每个类别的总结如下。

（1）循环论证——一级问题被重复作为二级因素。在这种情况下，删除多余的二级因素，取而代之的是在同名的一级问题和承载重复的一级问题名称作为二级因素的一级问题之间建立联系。在修改后的图中，这些循环论证替换都用非线性实线表示。

（2）重复——在两个或多个一级问题下重复相

环境影响评价 Environmental Impact Assessment



注：内部绿色圆圈表示一级问题，用虚线连接显示的所有二级因素（黄色椭圆形框）。

图1 ChatGPT识别EIA中AI的问题（未经作者解释）

同的二级因素。在这种情况下，所有重复的二级因素被聚合为一个二级因素，并通过线性实线将其与一级问题建立多个链接。

(3) 虚假信息——ChatGPT输出的问题本身不是问题或因素，而是对（通常未经证实的）事实的陈述。这些问题只是从图2中删除了。

人类研究人员编码的方法也会遇到这3类问题，这些方法在迭代的基础上进行——识别循环参数、重叠和明显不合适的代码。

图2清楚地表明，正如预期的那样，ChatGPT将

“问题”解释为挑战或潜在的不确定性。然而，重要的一点是，ChatGPT没有被要求确定使用AI的积极方面，这对所确定问题的负面框架产生了影响。

虽然图2的标题解释了如何解读这些关系，但关键的学习点是，第一级因素（内部绿色圆圈）是问题的类别。因此，它们仍然有些模糊，需要进一步解释。第二级因素（黄色椭圆）是更精确定义的问题，可能适用于未来的特定研究任务，例如，人们可能会问“在IA中使用LLM聊天机器人时，过度自动化会产生什么伦理问题？”。

IAPA文摘

Abstracts of IAPA

文本框3 处理ChatGPT结果中的循环论证——作者干预的两个例子

补充数据文件提供了ChatGPT确定的二级因素。其中一些二级因素复制了一级问题（因此是循环的）的两个例子是：

(1) 一级问题“专家直觉的丧失”（见文本框2）与二级因素“伦理和文化敏感性”相关[文本框2（4），补充数据]。然而，已经存在一个一级问题“伦理考量”（见文本框2）。在这种情况下，删除了该二级因素，并修改了图1，在一级问题“专家直觉的丧失”和“伦理考量”之间建立联系。见图2。

(2) 一级问题“数据质量和可用性”（见文本框2）有一个相关的二级因素为“时间变异性”[见文本框2（1），补充数据]。在这种情况下，删除了该二级因素，并修改了图1，以在一级问题“数据质量和可用性”与“时空挑战”之间建立联系。见图2。

文本框4 处理ChatGPT结果中的重复——作者干预的两个例子

补充数据文件提供了ChatGPT确定的二级因素。其中一些二级因素在多个一级问题上重复的两个例子是：

(1) 一级问题“过度依赖AI”（见文本框2）与二级因素“公众感知和信任”有关[见文本框2（3），补充数据]；一级问题“伦理考量”（见文本框2）也与二级因素“公众感知和信任”有关[见文本框2（5），补充数据]。在这种情况下，从图1中删除了多余的二级因素，并添加了一个额外的链接，例如“过度依赖AI”和“伦理考量”都与一个二级因素“公众感知和信任”有关（见图2）（注：以同样的方式处理其他的相同二级因素的重复）。

(2) 一级问题“时空挑战”（见文本框2）与二级因素“跨学科挑战”[见文本框2（11），补充数据]有关；一级问题“标准化和验证”（见文本框2）也与二级因素“跨学科挑战”有关[见文本框2（9），补充数据]。在这种情况下，从图1中删除了多余的二级因素，并添加了一个额外的链接，例如“时空挑战”和“标准化和验证”都与一个二级因素“跨学科挑战”有关（见图2）。

文本框5 处理ChatGPT结果中的虚假信息——作者干预的两个例子

补充数据文件提供了ChatGPT确定的二级因素。其中一些二级因素本身不是问题或因素，而是对事实的陈述的两个例子是：

(1) 二级因素“案例研究”与4个独立的一级问题相关，描述如下：

a. 一级问题“数据质量和可用性”与二级因素：“**案例研究**：许多失败或未达到预期性能的AI项目都可以追溯到用于培训或验证的数据问题。例如，如果训练数据不能充分捕捉到一个地区的栖息地范围，或者数据存在许多不准确的地方，那么一个旨在基于卫星图像预测栖息地破坏的项目可能表现不佳”[见补充数据文本框2（1）]。没有具体说明案例研究，因此，无法确认证据基础。

b. 一级问题“专家直觉的丧失”与二级因素：“**案例研究**：具体案例研究，特别是那些在专家评审后对AI驱动的评价进行修订或纠正的案例研究，突出了AI和人类直觉的互补性。证明了虽然AI可以增强EIA过程，但无法取代人类专家的细微理解和判断”[见补充数据文本框2（4）]。没有具体说明案例研究，因此，无法确认证据基础。

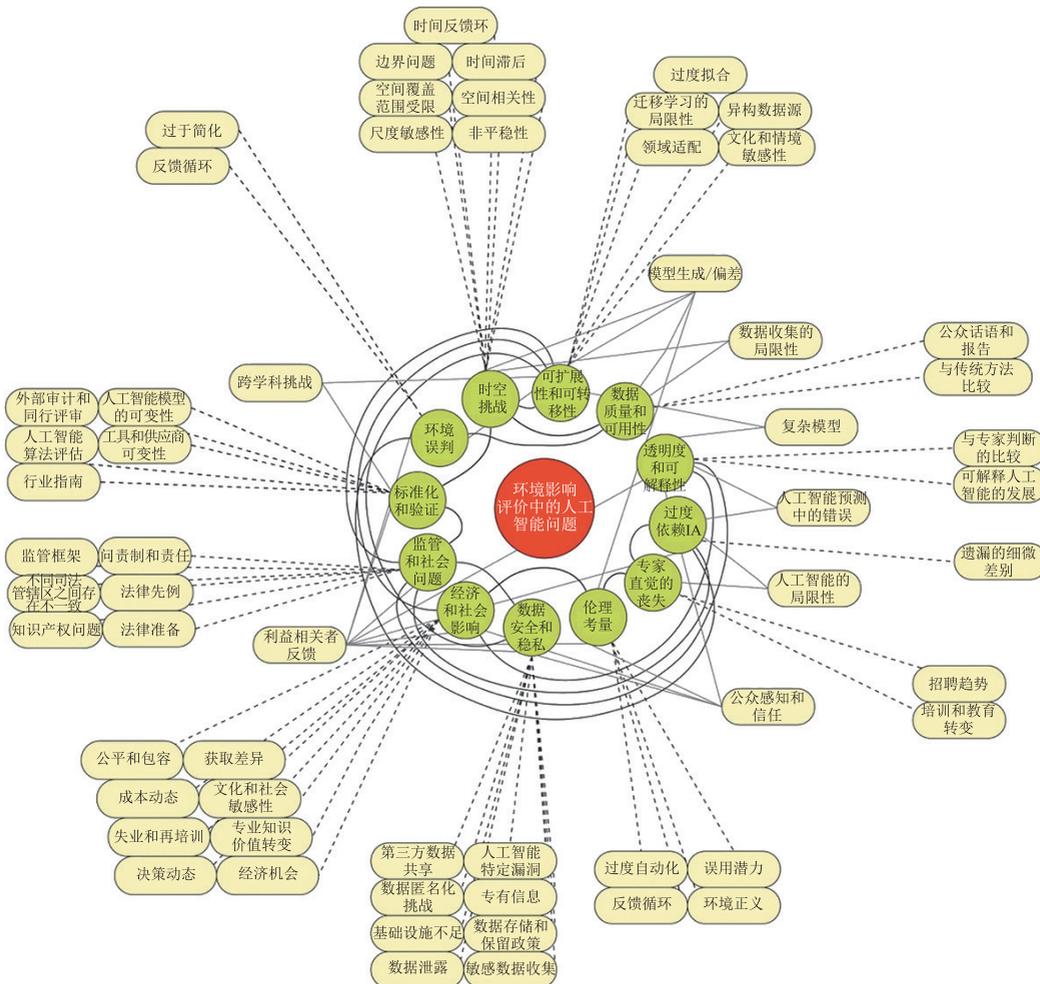
c. 一级问题“标准化和验证”与二级因素：“**案例研究**：在实际案例中，AI驱动的评价或模拟产生了相互冲突或不准确结果，凸显了缺乏标准化和有效性带来的挑战”[见补充数据文本框2（9）]。同样，没有具体说明案例研究，因此，无法确认证据基础。

d. 一级问题“环境误判”与二级因素：“**案例研究**：存在一些案例，AI驱动的评价或模拟由于数据问题、模型限制或参数设置不当等因素提供了误导性结果”[见补充数据文本框2（10）]。这里的案例研究不是一个具体的因素，它只是一个事实陈述，即与AI相关事项导致的问题可以通过案例研究来识别。

环境影响评价 Environmental Impact Assessment

(2) 二级因素“科学文献”与10个独立的一级问题相关，描述如下：

- a. 一级问题“数据质量和可用性”与二级因素：“**科学文献**： 学术论文和出版物通常强调高质量数据的重要性。许多研究表明，在有限或低质量数据上训练的模型会导致误解或预测失误” [见补充数据文本框2（1）]。这里未指明的科学文献仅指出了与AI相关的数据质量问题，而不是具体的证据。
- b. 一级问题“透明度和可解释性”与二级因素：“**科学文献**： 许多学术论文和文章指出了某些AI模型的不透明性。例如，关于使用AI进行物种识别或栖息地评价的研究表明，虽然模型可以准确，但准确理解它们为什么做出具体预测可能具有挑战性” [见补充数据文本框2（2）]。这里同样没有具体说明科学文献，因此，无法验证证据。
- c. 一级问题“过度依赖AI”与二级因素：“**科学文献**： 讨论AI在各个学科中的实施的研究和文章经常强调人类监督的重要性。许多研究论文的共识是，虽然AI可以增强决策过程，但它不应该完全取代人类的判断” [见补充数据文本框2（3）]。这里同样没有具体说明科学文献，因此，无法验证证据。
- d. 这一问题在其他7个一级问题中重复出现，所有这些问题都未完全将科学文献确定为一个因素 [见文本框2（4） - （6）和（8） - （11）]。



注：有3个同心圆：1) 内部绿色圆圈代表一级问题，用线条显示与所有二级因素（黄色椭圆形框）的连接，原本在二级因素中列出的相同问题的循环论证已被替换为与其他一级问题的连接；2) 中间黄色椭圆形框是二级因素，通过实线与多个一级问题连接；3) 外黄色椭圆形框是二级因素，通过虚线与单一的一级问题连接。

图2 ChatGPT在作者解释后确定的EIA中的AI问题和因素

IAPA文摘

Abstracts of IAPA

5 结论

研究旨在实现两个目标：

(1) 缩短对文献进行评价所需的时间；

(2) 综合当前知识，就研究人员和从业者在使用AI应用于EIA时需要考虑的问题展开讨论。

对于第一个目标，尝试使用AI工具（ChatGPT 4.0和Elicit）在尽可能短的时间内为IA社区提供有用的产出。本文在研究开始后的3天半后提交（尽管承认在后续的审稿过程中额外花费了一些时间，而且尚不清楚如果没有使用AI，修订所需的时间是否会有所不同）。在使用AI提高研究效率的同时，非常谨慎，没有简单地接受结果，而是为ChatGPT的输出寻找证据。使用ChatGPT和Elicit这些AI工具所花费的时间只占据了一小部分。大部分的人力投入都用于识别证据以支持ChatGPT的主张，尽管Elicit还存在缺陷，但其在许多情况下仍然是识别相关文献的有用手段。为协助研究人员验证通过Elicit（或其他机器学习工具）提供的证据而开发的协议是本研究的一个新成果，这可以协助未来的研究人员。另一个重要的时间投入是绘制网络图（使用Miro，<https://miro.com/app/dashboard/>），然后解释结果网络图以去除不一致性和循环引用。该解释步骤复制了人类研究员对学术文献的标准编码，涉及检查和改进代码的迭代过程。尽管没有对照实验来作为衡量所需时间的基准，但仍然是任何作者从想法到产出所花费的最短时间。研究表明，只要结果真实可靠，像ChatGPT这样的AI工具可以显著提高研究任务的效率。一个尚未回答的问题是，如果使用基于学术数据库的系统文献综述会产出什么结果。沿着这些思路进行的未来研究可能有助于明确可能节省的时间范围，并且更重要的是比较“传统”与使用AI辅助文献综述方法的研究结果。

因此，鉴于ChatGPT所采取的方法存在不确定性，使用AI的主要缺点在于用户对其调查结果的信任。这指出了有趣的未来研究领域，即比较AI与人类研究员进行文献综述的结果。LLM在未知的数据集上使用未知的算法进行初始编码。然而，人类

研究员既容易出错，又受到资源限制，同时还面临数据库访问问题和综合分析时的认知局限。确定哪种方法更好本身就是一项任务，确定哪种方法更能赢得读者的信任也是如此。此外，研究还提出了其他问题：

(1) 研究人员需要对文献有多深入的了解，才能够使得通过LLM聊天机器人提供的帮助变得可信（或者是否需要了解文献）；

(2) 研究人员需要具备什么样的知识才能评价LLM聊天机器人的输出结果。

初稿审稿人对这种快速撰写论文的方式表示了一些不满。特别是提出了以下担忧。

追求快速输出可能会损害研究的质量，从而影响其稳健性。

- 自动化分析节省的时间，削弱了人类思维的价值，从而威胁到研究的深度和知识进步。

- 查找参考文献节省的时间可能会使某些出版机构享有特权（无法知道这种情况是否正在发生）。

本文承认这些担忧是合理的。通过向聊天机器人提出不同的问题，有些问题或许能够得到解决。但关于缺乏洞察力以及对出版来源的无意识偏袒的担忧似乎更难以轻易避免。需要研究人员在进行研究时积极考虑这些问题。

对于第二个目标，鉴于无法确定更传统的方法是否会得出相同（或更全面或更不全面）的结果，本文认为图2对于希望在EIA研究和实践中参与或管理AI使用的人来说，是一个有用的起点。本文对一级问题真实性进行了独立的证据检查，相信这些问题可信且被广泛认可。尽管证据存在重复和循环，但这些问题已在图2中得到解决。这代表了当前以这种方式使用LLM聊天机器人的局限性，并且LLM在未来很可能会有所改进，仅需要更少的调整。因此，EIA研究人员可以开始着手解决这些问题。图2提供了一个研究路线图，列出了需要被管理或解决的问题。

然而，审稿人对输出提出了一些担忧。具体如下：

- 仅搜索验证问题的参考文献会遗漏可能会破坏

环境影 响 评 价

Environmental Impact Assessment

调查结果的反证。

- 审稿人强调，作者不得不修改ChatGPT的初始输出表明了初始输出是没有帮助的。

- 有批评指出，对于已确定的问题而言，其输出的结果在实践意义方面没有得到进一步阐释。

- 在识别“问题”与这些可能对IA的规划和委托过程产生的后果之间存在脱节。

- 从ChatGPT学习是有价值的这一观点被质疑是肤浅的。特别是目前根本没有更深入地了解对于使用AI管理环境要素（如物种丰富度、分布和连通性）的空间数据的局限性和后果。

- 基于AI的输出没有产生任何新的见解。

本文承认这些观点有一定的合理性。虽然本文最初只是进行文献综述以识别AI在EIA中应用的问题，而不是进一步考虑更广泛的影响，但这些评论确实促使作者反思了对结果的某种不适感，因为很难感受到有显著的所有权。如果没有这种所有权的感受，就不愿意进一步探索其意义。如果AI在EIA研究中的使用范围扩大，这可能会成为问题，具体取决于每个案例中的目标。本文认为，尽管目前可用的聊天机器人不会产生任何新的见解，因为它们只是搜索现有知识，但文献综述也是如此。在这种情况下，在一份出版物中列出的问题可能会被认为是新颖且有用的，可以呈现以前没有发表过的知识的综合——但读者同意的程度是主观的。本文不清楚ChatGPT是如何解释“AI”这一术语的，因此也不清楚进行搜索时的边界。关于在IA中AI使用相关问题的识别仍然存在相当大的不确定性，因为尚不清楚

ChatGPT将什么解释为AI。

最后谈谈本文所做事情的伦理问题。本文认为，在研究中越来越多地使用AI是不可避免的，如果通过使用AI可以提升研究的效率，那么忽视它是愚蠢的。本文确实相信，在适当的透明度下就可以合理地使用AI，且本文已经尽力做到这一点。真正的挑战在于知道AI在何处被使用但没有被承认——这才是跨越伦理界限的地方。一位审稿人还评论说，项目提议者已经在EIA中产生了大量的不会对决策产生不利影响的虚假信息，他询问是否在看待AI输出时使用了一种更具批判性的视角，而不是针对过程中的一些利益相关者——这也许是另一个潜在的未来研究领域。

在评析过程中还确定了在以EIA为重点的研究中使用AI的其他挑战。也许随着LLM变得越来越复杂，关于节省时间的评论可能会变得不那么重要，LLM可以为人类思维创造空间，这些空间在其他情况下是无法获得的。审稿人提出的其他问题似乎与良好的研究实践有关，并确保对LLM的依赖不会导致不良实践，如不寻找反证（本文就犯了这一错误）。最终，本研究表明需要意识到AI是如何提供帮助的，以及它如何构成威胁。在没有审稿人提出的这些真实的威胁的情况下，肯定有一些中间地带可以感受到益处。

参考文献：请见原文

译者：三捷环境工程咨询（杭州）有限公司
吴成志 马焯怡