

从海洋到系统：探索以用户为中心的可解释人工智能在海事决策支持中的应用

Doreen Jirak¹ (✉)^[0009-0003-2839-3475], Pieter Maes², Armees Saroukanoff²,
and Dirk van Rooy¹

¹ University of Antwerp, Paardenmarkt 94, 2000 Antwerp, Belgium

{doreen.jirak,dirk.vanrooy}@uantwerpen.be

² Antwerp Maritime Academy (AMA), Noordkasteel Oost 6, 2030 Antwerp, Belgium

{pieter.maes, armeen.saroukanoff}@hzs.be

摘要 随着自主技术越来越多地影响海事操作，理解为什么人工智能系统做出某个决策变得和它决定什么一样重要。在复杂且动态的海洋环境中，对人工智能的信任不仅取决于其性能，还取决于透明度和可解释性。本文强调了在海事领域有效的人机团队合作中可解释的人工智能（xAI）作为基础的重要性，在这个领域内，知情监督和共享理解至关重要。为了支持以用户为中心的 xAI 集成，我们提出了一项特定领域的调查设计，旨在捕捉海事专业人士对信任、可用性和可解释性的感知。我们的目标是提高意识并指导开发符合海员和海事团队需求的用户中心 xAI 系统。

Keywords: 人机决策 · 可解释人工智能（xAI）· 信任 · 以用户为中心的设计 · 海事操作

1 介绍

海运业正在经历一场深刻的转型。随着人工智能、传感器融合和远程操作的进步，海上自主水面船舶（MASS）[8] 不再是投机性的概念，而是逐渐成为现实。诸如远程驾驶室操作 [2] 等海上任务要求重新定义船员和海事当局的角色与责任。然而，这种转变正在一个以保守性著称的领域内发生。与拥有悠久自动化历史和集中化监管的航空业不同，海运业高度分散且适应缓慢。海洋上的运营决策不仅受到技术限制的影响，还受到天气状况、团队动态及长期形成的航海惯例的影响。因此，将自主和混合系统整合到海上操作中提出了独特的挑战。可解释的人工智能（xAI）为弥合海事专业人士与 AI 系统之间的信任鸿沟提供了一个有前景的途径 [16]。通过使人工智能驱动决

策背后的理由透明且可理解 [13][11]，xAI 有可能促进用户信心，并支持人类操作员与智能系统的安全协作。在日常航海活动中，遵守 COLREGs（海上避碰规则）至关重要，在这种背景下，xAI 可以帮助澄清自主系统如何解读这些规则并在实时中作出调整。

本文介绍了旨在评估海事利益相关者对导航决策支持系统中可解释人工智能态度的经验研究设计。我们旨在调查可解释性特征如何影响用户信任、感知有用性和与 AI 辅助系统的互动意愿。为此，我们开发了一个基于调查的实验框架，其中包括基于场景的刺激物，以及在评估信任度、技术开放性 & 用户满意度的前后问卷和基于现实雷达任务的交互元素。鉴于数据收集正在进行中且存在参与者偏见的可能性，本文侧重于概念框架和设计理由，而不是提供详细的刺激材料或完整的调查项目。通过展示一个基于真实世界导航场景的调查框架，本研究旨在促进针对海事领域的可解释、值得信赖的人工智能系统的开发。结果将为人类与 AI 交互挑战提供早期见解，并指导未来关于海上合规自主系统的设计研究。以此方式，我们希望不仅提高对（团队）决策中涉及的人类因素的认识，而且还促进跨学科的研究以实现以用户为中心 xAI。

2 相关工作

虽然海运行业在采用自动化方面历来保守，但现在正面临由人工智能（AI）进步推动的技术变革浪潮。应用范围从导航任务中的决策支持到海上自主水面船舶（MASS[8]）上的完全自主控制。这些发展有望提高运营效率和经济效益，但也引发了关于如何合理将此类技术整合进海运作业复杂且不确定环境中的关键问题。例如，一个重要问题是海运安全。尽管技术取得了进步，人为错误仍然是海上事故的主要原因之一。一个常被引用的数字表明约 80% 的海运事故是由于人为错误造成的，这一数据源自 Berg 等人发布的一项研究 [1]。然而，根据调查的时间范围和海域的不同，这些数字波动在 60%[5] 到高达 96%[15] 之间。根本原因通常包括认知疲劳、沟通失误、情况过载或心理阻塞，特别是在压力之下或动态团队环境中。这些发现强调了不仅需要自动化任务，还需要补充人类表现并提高安全关键场景下的决策可靠性。因此，在海运环境中的 AI 系统整合必须超越技术稳健性。我们主张它应该是以用户为中心，认识到船员的操作常规、领域知识和生活经验。这要求采取一种整体的跨学科方法，考虑人与机器的能力。特别是，来自人类因素研究、认知科学、神经科学和社会心理学的见解对于理解船员如何解释、

适应并与这些新兴技术互动至关重要。向以用户为中心的人工智能交互和团队协作过渡的一个核心前提是信任 [12]。对自动化的信任不是一个单一的结构，而是包括了信任校准，既包含认知上的信任（例如，系统被感知到的能力、可靠性和可预测性）也包含情感上的信任（例如，安全感、与人类意图的一致性以及对系统行为的情感舒适度）[7]，还包括随着时间推移建立的信任。在像海事领域这样的高风险环境中，事故可能带来严重后果，设计以增强人工智能系统的信任至关重要。这引出了一个重要的方面，即如何在海事部门创建既可信又互动的系统。与受控模拟不同，现实世界中的信任是通过经验、机构培训和桥上的团队实践发展和塑造的。这不仅关乎人工智能系统或自主代理是否做出了“正确的”决策，更在于其决策是否可理解、可预测且符合用户的预期 [6]，特别是在系统行为偏离标准海事做法时。

在此背景下，可解释的人工智能（xAI）是建立可信和有效人机协作的关键。随着深度学习模型（通常简称为“AI”）在检测和识别任务中表现出超越人类的性能，它们决策过程缺乏透明度引起了人们的担忧。

xAI 通过开发方法来揭示所谓的 AI 决策“黑箱”，从而解决了这一问题，增强了透明性、可解释性和公平性。尽管近期以用户为中心的 xAI 研究集中在提高用户信任 [9] 和防止过度依赖 [3]，但航运业在这方面却受到的关注相对较少。

然而，Merwe 等人的最近一项研究 [14] 显示，代理透明度有可能改善海员的情境意识（SAW[4]），但这可能是交通复杂性和代理推理丰富程度的函数。同样地，Madsen 等人 [10] 的研究表明，代理透明度并没有均匀提高所有情境意识水平，并且信息显示需要校准到人类的认知资源上，进一步增强了对以用户为中心的 xAI 研究的需求。

为了解决这一缺口，我们提出了一项基于调查的研究，旨在捕捉海事专业人士如何看待人工智能驱动决策系统的可解释性、信任和易用性。我们的目标是为以用户为中心 xAI 交互设计奠定基础，以适应航运领域的独特认知和操作现实。

3 调查设计

作为我们持续研究的一部分，我们提出了一项调查，旨在捕捉用户对海事自主性中信任、易用性和可解释性的感知。我们的目标是提高意识，绘制信任维度（包括认知和情感方面），并确定未来以人类为中心的 xAI 系统的设计优先事项。这一努力不仅有助于 MASS 的发展，也有助于在安全关键且

操作要求高的领域中更广泛地整合可解释和可信的人工智能。我们的研究受到以下主要研究问题的指导：

1. RQ1: 海员对技术进步的态度及其信任倾向是什么? (预问卷)
2. RQ2: 可解释性特征 (xAI, “海上助手”) 的纳入在多大程度上影响了用户满意度、信任度以及最终采用此类系统的意愿? (问题后问卷)
3. RQ3: 海事专业人士如何感知海事助手的实用性和可靠性? 船员对将人工智能融入海事日常事务和团队工作流程有哪些担忧和期望? (问卷后问题, 开放性问题)

图 1展示了调查的实验流程, 结构如下:

1. 简报: 欢迎并解释即将进行的演示。参与者表示同意 (GDPR), 并且可以在任何时候退出调查而不面临任何不利影响。此外, 不会从希望退出的参与者那里收集任何数据。
2. 调查: 具体的调查包含一些预问卷, 接着展示从船舶日常海事惯例中提取的情景, 例如, 避碰。信息通过雷达图像给出, 这些图像是基于关键时间步骤选择的, 比如当 COLREG 规则不适用时。在获得水手关于其行动建议 (基线条件) 的回答后, 我们展示了 “海事助手” 的输出 (参见图 1), 显示其决策以及参与决策过程中最重要的一项特征 (xAI 条件)。在展示情景之后, 后续问卷用于检查用户对自主系统的满意度和信任度。如所述, 可解释性必须与可用性并行发展。因此, 在此背景下以用户为中心的 xAI 设计应考虑系统性能以外的因素, 包括解释如何支持及时、安全且自信的人类决策的程度。虽然由于缺乏实际互动而无法应用标准的系统可用性量表, 但仍可以使用如 [6] 中建议的 xAI 指标, 并可将这些指标与针对个别场景定制的项目合并在一起。此外, 参与者还可以对展示的情景提出反馈。
3. 去 briefing: 收集包括参与者级别和海上经验年限的人口统计数据。我们决定在调查结束后再询问这些数据, 以避免任何由于年龄或经验相关问题引入的偏见。调查最后一个问题涉及数据库包含情况以用于进一步实验, 并解释了调查目标。

一旦数据收集完成, 未来的工作将报告这项研究的发现, 并提供一个经过验证的调查工具以供更广泛使用。我们还预计扩展这项研究, 探讨解释性偏好如何因角色 (例如, 导航员与工程师) 和人工智能熟悉程度的不同而有所不同。

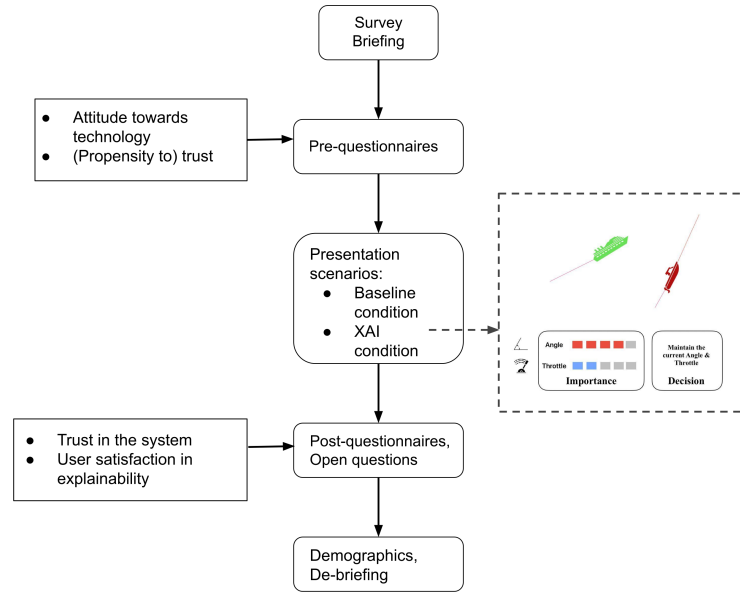


图 1. 前瞻性海洋 xAI 调查的实验流程图。

4 结论

在本文中，我们认为随着我们将智能系统引入海事领域，我们必须超越纯粹的技术叙事，需要将人为因素纳入决策过程，如认知资源和情感状态，以及信任和技术接受意愿。因此，我们介绍了正在进行的研究，以用户为中心的 xAI 为基础，并结合来自潜在用户的这些人为因素及重要领域的知识来改进海事人机决策。我们希望提高船员对即将来临的技术变化的认识，同时也说服 (x)AI 开发者构建以用户为中心接口，以创建有意义的人工智能集成，特别是在动态互动团队中，如海员。为了成功，混合学习和决策系统必须不仅与性能指标保持一致，而且还要与海上人员的实际经验和专业知识相契合。

Acknowledgments. 作者感谢 Arian Sabaghi Khameneh (imec/IDLab) 提供的 xAI 图像。本研究在 DEFRA AHOI 项目下进行，该项目由比利时皇家高等国防学院资助，合同编号为 23DEFRA002。

Disclosure of Interests. 作者声明与本文内容相关的不存在任何利益冲突。

参考文献

1. Berg, N., Storgård, J., Lappalainen, J.: The impact of ship crews on maritime safety. *Publications of the Centre for Maritime Studies, University of Turku A* **64**, 1–48 (2013)
2. Bhuiyan, Z.: Bridge simulator as a training platform for future remotely controlled mass navigators. In: *Maritime Autonomous Surface Ships (MASS)-Regulation, Technology, and Policy: Three Dimensions of Effective Implementation*, pp. 301–316. Springer (2024)
3. Buçinca, Z., Malaya, M.B., Gajos, K.Z.: To trust or to think: cognitive forcing functions can reduce overreliance on ai in ai-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction* **5**(CSCW1), 1–21 (2021)
4. Endsley, M.R.: Supporting human-ai teams: Transparency, explainability, and situation awareness. *Computers in Human Behavior* **140**, 107574 (2023)
5. Erol, S., and, E.B.: The analysis of ship accident occurred in turkish search and rescue area by using decision tree. *Maritime Policy & Management* **42**(4), 377–388 (2015). <https://doi.org/10.1080/03088839.2013.870357>
6. Hoffman, R.R., Mueller, S.T., Klein, G., Litman, J.: Measures for explainable ai: Explanation goodness, user satisfaction, mental models, curiosity, trust, and human-ai performance. *Frontiers in Computer Science* **5**, 1096257 (2023)
7. Legood, A., van der Werff, L., Lee, A., den Hartog, D., van Knippenberg, D.: A critical review of the conceptualization, operationalization, and empirical literature on cognition-based and affect-based trust. *Journal of Management Studies* **60**(2), 495–537 (2023)
8. Li, S., Fung, K.: Maritime autonomous surface ships (mass): implementation and legal issues. *Maritime Business Review* **4**(4), 330–339 (2019)
9. Liao, Q.V., Varshney, K.R.: Human-centered explainable ai (xai): From algorithms to user experiences. *arXiv preprint arXiv:2110.10790* (2022)
10. Madsen, A.N., Brandsæter, A., van de Merwe, K., Park, J.: Improving decision transparency in autonomous maritime collision avoidance. *Journal of Marine Science and Technology* pp. 1–19 (2025)
11. Madsen, A.N., Kim, T.E.: A state-of-the-art review of ai decision transparency for autonomous shipping. *Journal of International Maritime Safety, Environmental Affairs, and Shipping* **8**(1-2), 2336751 (2024)
12. Mayer, R.: An integrative model of organizational trust. *Academy of Management Review* (1995)

13. van de Merwe, K., Mallam, S., Engelhardtzen, Ø., Nazir, S.: Towards an approach to define transparency requirements for maritime collision avoidance. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. vol. 67, pp. 483–488. SAGE Publications Sage CA: Los Angeles, CA (2023)
14. van de Merwe, K., Mallam, S., Nazir, S., Engelhardtzen, Ø.: The influence of agent transparency and complexity on situation awareness, mental workload, and task performance. *Journal of Cognitive Engineering and Decision Making* **18**(2), 156–184 (2024)
15. Sánchez-Beaskoetxea, J., Basterretxea-Iribar, I., Sotés, I., Machado, M.: Human error in marine accidents: Is the crew normally to blame? *maritime transport research*, 2, 100016 (2021)
16. Veitch, E., Alsos, O.A.: Human-centered explainable artificial intelligence for marine autonomous surface vehicles. *Journal of Marine Science and Engineering* **9**(11), 1227 (2021)