数据匿名化攻防赛驱动的安全教学改革研究*

马瑞强**　　郭亚楠　孙赛格　蒋永生

内蒙古工业大学智能科学与技术学院
呼和浩特 010051
明治大学研究战略机构，日本东京 164-8525

内蒙古工业大学智能科学与技术学院
呼和浩特 010051

摘　要　生成式人工智能（AI）降低了文本与代码产出门槛，结果导向评价在数据安全与隐私保护教学中易失真。本文以国际数据匿名化攻防赛为真实任务载体，提出任务链与证据链耦合的过程证据框架，组织赛前两周基线复现、赛中四周匿名化与重识别验证迭代、赛后两周复盘与学术表达，并采用四维量规与抽样核验开展一致性检验，引入竞赛官方评分作外部对照。实践覆盖 60 名学生、15 组稳定团队、8 周周期，累计参与四轮攻防赛；代码提交率达 100%，人均有效提交次数增长超过 50%，迭代成绩均分由约 70.1 提升至 85.5。最终 15%的队伍进入国际竞赛前 10%并获奖。结果表明，该模式在教师工作量可控条件下显著提升了有效迭代密度与对照实验、失败复盘的证据质量，强化了匿名性与效用性权衡的机制解释能力，提升评价公平性与可解释性，为生成式 AI 时代数据安全实践教学提供可复制路径。

关键字　生成式人工智能，数据匿名化，攻防竞赛，过程证据，真实参与度，形成性评价

Data Anonymization Attack-Defense Competition-Driven Security Teaching Reform Research*

Ruiqiang Ma**　　Yanan Guo　Saige Sun　Yongsheng Jiang

College of Intelligent Science and Technology
Inner Mongolia University of Technology,
Hohhot 010051, China
Reasearch on Financial Strategic Institutions
Meiji University, Tokyo 164-8525, Japan.

College of Intelligent Science and Technology Inner Mongolia University of Technology,
Hohhot 010051, China

Abstract—Generative Artificial Intelligence lowers the barrier to producing text and code, making outcome-oriented assessment in data security and privacy education more prone to distortion. This study uses an international data anonymization attack–defense competition as an authentic task setting and proposes a process-evidence framework that couples a task chain with an evidence chain. The implementation spans an eight-week cycle: two weeks of baseline reproduction before the competition, four weeks of iterative anonymization and re-identification validation during the competition, and two weeks of post-competition reflection and scholarly communication. A four-dimension rubric and sampled checks are used for consistency verification, and official competition scores are incorporated as an external benchmark. The practice involved 60 students organized into 15 stable teams over eight weeks and completed four rounds of attack–defense tasks. The code submission rate reached 100%; the average number of effective submissions per student increased by more than 50%; and the mean iteration score rose from approximately 70.1 to 85.5. Ultimately, 15% of the teams ranked within the top 10% internationally and received awards. The results indicate that, with a manageable instructional workload, the proposed approach significantly increases effective iteration density and improves the evidence quality of controlled experiments and failure-based reflection, strengthens mechanism-level understanding of the privacy–utility trade-off, and enhances the fairness and interpretability of assessment, offering a replicable pathway for data security practice teaching in the era of generative AI.

Keywords—Generative AI, Data Anonymization, Adversarial Competition, Process Evidence, Authentic Engagement, Formative Assessment

1　引　言

生成式人工智能将信息检索、文本生成以及代码与推理输出整合到低门槛的对话式工具之中，使学生完成任务的策略逐渐由阅读理解、独立推导、形成答案转向提出问题、获取生成、再做润色。这种变化带来的挑战不止于学术诚信，更深层的问题在于它动摇了传统评价体系的隐含前提。长期以来，课程作业与阶段性项目多以最终成果为主要评价对象，默认结果质量能够反映学习质量。当 AI 能够稳定产出高质量文本与可运行代码时，结果导向评价更容易出现失真，表面同样完成得很好的作品背后，学生的学习投入、机制理解与实验能力可能差异巨大，教师也难以据此

*基金资助：本文得到国家人社部外国专家重点支撑计划（D20250092）和省部重点研发（2025YFSH0157）的资助。

**通讯作者：马瑞强 marq@imut.edu.cn。

判断真实掌握情况。

数据安全与隐私保护[1]课程尤其依赖过程性能力，包括威胁模型与问题界定、匿名化策略设计与权衡、对抗攻击与验证、指标选择与实验复现、结果解释与学术表达等。若仅以最终匿名化结果或最终报告评判学生学习成效，容易将生成式完成误判为机制性理解，使学生的推理过程与实验迭代不可见，从而削弱形成性反馈的有效性。基于此，本文引入国际数据匿名化攻防赛[2]作为真实任务载体，将竞赛的强约束、强对抗与强外部评价转化为教学改革的动力源，并提出过程证据驱动的评价与组织机制，旨在实现 AI 可用但学习必须可证，提升学生真实参与度与科研式能力发展。

本文的贡献体现在三个方面：第一，提出面向攻防竞赛场景的过程证据框架，将思路形成、AI 交互修正、实验对照与协作贡献等过程材料纳入评价核心；第二，构建适配大班或社团化培养的任务链组织方式，实现赛前准备、赛中迭代与赛后复盘的闭环；第三，提出外部竞赛成绩与内部过程数据联合的证据体系，为教学改革成效评估提供可复制的指标与方法。

2 竞赛情境与教学改革需求

2．1 国际数据匿名化攻防赛的任务特征与育人价值

国际数据匿名化攻防赛通常采用匿名者与攻击者双角色机制。参赛者既需要对数据进行匿名化处理以保护隐私[3]，又需要对其他队伍的匿名数据进行重识别攻击[4]。该机制天然强化了防护与攻击相互验证的科研范式，使学生在真实约束下理解隐私保护的边界条件[5]。竞赛评价通常同时包含匿名性与效用性两类指标，并通过惩罚机制抑制极端策略，从而要求参赛者围绕权衡开展迭代优化。这种真实任务结构使其具备明显的教学价值：它迫使学生明确研究问题与威胁模型，复现基线并进行对照实验，围绕指标解释结果并通过团队协作完成工程实现与答辩展示。

2．2 课程化改造的关键矛盾

将竞赛直接等同于训练拿奖容易造成两类偏差。一是只看最终排名，忽视多数学生的学习过程与能力成长，竞赛变为少数人的精英活动。二是只要求输出方案或代码，忽略验证与解释环节，使生成式 AI 更容易替代学生的真实思考。教学改革需要解决的关键矛盾是：如何在允许合理使用 AI 的同时，把学生的学习过程证据化，既能激励持续迭代，又能让教师在可控工作量下实施评价与反馈，并使竞赛成果转化为可迁移的科研能力。

3 过程证据框架与评价设计

3．1 总体结构：任务链与证据链耦合

本文将竞赛驱动的学习过程组织为贯穿赛前、赛中、赛后的任务链。赛前阶段侧重概念框架、工具链与基线复现，目标是让学生具备可复现的起点。赛中阶段侧重匿名化策略迭代与攻击验证[6]，目标是形成科研式循环，即提出假设、实施改动、开展对照、解释结果。赛后阶段侧重复盘与学术表达，目标是将过程证据沉淀为技术报告或海报展示，实现知识结构化。

与任务链并行的是证据链，包含四类证据。思路证据用于呈现问题界定、威胁模型、权衡假设与策略选择依据。交互证据用于呈现 AI 使用过程中的提示词演化、关键输出片段与纠错说明。验证证据用于呈现实验设计、对照与消融、指标变化与误差分析。参与证据用于呈现分工协作、代码贡献、组会讨论与互评反馈。每项任务要求提交最小充分证据包，即少量但关键的过程材料，以保证可核验性与低负担并存。

3．2 最小充分证据包与提交规范

最小充分证据包建议由一页思路卡、一份提示词演化表、一组关键对照实验记录与一次简短机制解释材料构成。思路卡要求学生写清研究对象、威胁模型、匿名目标与效用目标，并说明如何判断策略优劣。提示词演化表至少包含两轮提示词及其迭代理由，重点是展示学生如何发现 AI 输出的问题并通过实验验证纠偏。关键对照实验记录要求提供基线与改动方案在匿名性与效用性指标上的对比[7]，并说明指标选择依据。机制解释材料采用答辩要点，要求学生用自己的语言解释匿名性[8]与效用性权衡逻辑与失败原因。

3．3 量规评价与真实性约束

为避免评价复杂化，本文采用四维量规与抽样核验。量规包括证据完备度、机制解释力、验证与修正、反思与迁移四个维度。真实性约束不依赖对文本进行 AI 检测，而依赖一致性检验。第一，通过抽样验检验学生是否能解释自己的策略与实验结果。第二，通过随堂微测或阶段测验检验关键概念与机制理解是否稳定。第三，通过证据自洽性检验检查思路卡、提示词演化与最终方案是否逻辑一致。竞赛官方评分作为外部客观结果证据，与课堂过程证据共同构成可信的评价体系。

4 教学实践设计与任务样例

4．1 赛前阶段：基线复现与问题界定

赛前阶段以数据集理解与基线方法复现为核心任务，要求学生明确字段含义、敏感属性与准标识属性，

形成初步威胁模型，并复现一个可运行的匿名化基线与一个重识别攻击基线[9]。教师在此阶段重点检查学生是否具备复现能力与记录习惯，避免后续迭代缺乏可信起点。

4．2　赛中阶段：匿名化策略迭代与攻击验证

赛中阶段围绕匿名性与效用性权衡展开，学生以小组为单位进行多轮迭代。每轮迭代必须提交最小充分证据包，尤其要呈现策略变更的理由、指标变化与攻击验证结果。匿名化侧可采用分组打乱、随机交换、按比例交换、关键字段再处理等策略组合[10]，并通过对照实验判断不同策略对匿名性与效用性的影响。攻击侧要求学生基于外部信息或相似性匹配开展重识别尝试，并记录攻击成功条件与失败原因，使防护策略在对抗检验中得到真实约束。

4．3　赛后阶段：复盘沉淀与学术表达

赛后阶段强调将竞赛过程转化为科研式成果表达。学生需完成技术报告或英文海报，围绕研究问题、方法路线、实验设计、结果分析与局限反思进行结构化呈现，并在班级或学院范围进行路演交流。教师在此阶段重点评价学生的论证逻辑、图表规范与证据链完整性，使过程证据真正沉淀为可迁移的研究能力。

5　成效评估与证据呈现

5.1　指标体系与数据来源

参与度指标包括上传代码驻留率、参赛迭代轮次、有效提交次数、BenchMark 系统完成率与展示次数，数据来源为提交记录与课堂观察。真实性与一致性指标包括抽样口头核验通过率与证据自洽率[11]，数据来源为核验记录与量规评分表。学习成效指标包括阶段测验得分、实验记录规范性评分与报告质量评分。学习体验指标通过问卷与访谈获取，关注作业负担、AI 使用感受、反馈有效性与投入感。外部证据包括竞赛官方评分与排名，用于校验学习结果的客观性。

5.2　结果呈现方式

本研究采用对比与趋势结合的方式呈现成效。对改革前数据，进行前后对比和行动研究式阶段对比，将前两周作为基线阶段，将后续稳定阶段作为对照。结果部分可报告参与度指标变化、抽样核验通过率、证据自洽率、报告质量提升情况，并结合竞赛成绩进行交叉佐证，过程如表 1 和表 2。

表 1 清晰地展示了项目的组织规模与周期，便于与日本 Group 进行横向对比；而表 2 直接回应用户“结合日本 Group 情况”的要求，进行结构化国际对照。通过引入固定周期的教学计划内竞赛机制，结合强制性的过程证据提交与核验体系，构建了一种强化真实参与度的结构化教学框架。与日本参照组的弹性课外项目制相比，本方法在确保同等竞赛产出的基础上，更强调学习过程的可追溯性与学术化沉淀，并通过规范 AI 工具的使用记录与反思，促进学生形成对技术辅助的批判性认知。这一模式为在有限教学周期内系统提升学生实践能力的真实性、可控性与教育价值提供了可复制的路径。

表 1　教改实践组织数据概览

表 2 中日小组实践模式对比简析

对比项	本研究（中国组）	日本明治大学（参照组）	对比启示
组织周期	固定8周（教学计划内）	弹性8-10（课外项目制）	证实8周是可行的集中攻关周期
核心驱动	过程证据（强制提交与核验）	结果导向与定期会议讨论	本研究引入了更结构化的过程管理框架
关键产出	竞赛成绩 + 过程证据包 + 技术报告	竞赛成绩 + 开源代码仓库	本研究强调过程的可追溯与学术化沉淀
AI工具使用	鼓励，但需记录交互过过程并解释	普遍使用，但文档记录较少	本研究更注重对AI辅助的元认知培养

5.3　典型案例

可保留两到三个典型案例。第一类案例表现为 AI 输出看似合理但缺乏验证，学生在抽样核验时无法解释指标变化原因，通过补充对照实验与修正说明后形

成完整证据链。第二类案例表现为匿名性提升导致效用显著下降，学生通过引入分层处理与参数搜索实现权衡改进。第三类案例表现为互评促进理解，学生在评价他组证据包时发现自身威胁模型遗漏并在下一轮迭代中修正。

6 讨论与改进

本研究表明，国际数据匿名化攻防赛具有天然的外部约束与真实评价指标，适合作为数据安全教学的高真实性任务载体。过程证据框架的关键价值在于把学习过程证据化，使 AI 的使用从替代完成转向验证与解释，并通过抽样核验与一致性检验实现真实性约束与工作量可控之间的平衡。推广应用时需注意分层与公平性，避免竞赛仅惠及少数学生；需通过模板化证据包、抽样复核与同伴互评控制教师负担；同时应明确 AI 使用规范与数据隐私边界，避免证据收集引发额外风险。后续可在多学期、多班级条件下开展更严格对照研究，进一步量化过程证据与学习成效之间的关联[12]。

持续改进已现实效，并获得前一年度一等奖，新参赛模式功不可没，图 1 所示为获奖证书，并得到学科组评估认可，结论如图 2 所示。


賞 状

PWS Cup 2024　総合優勝

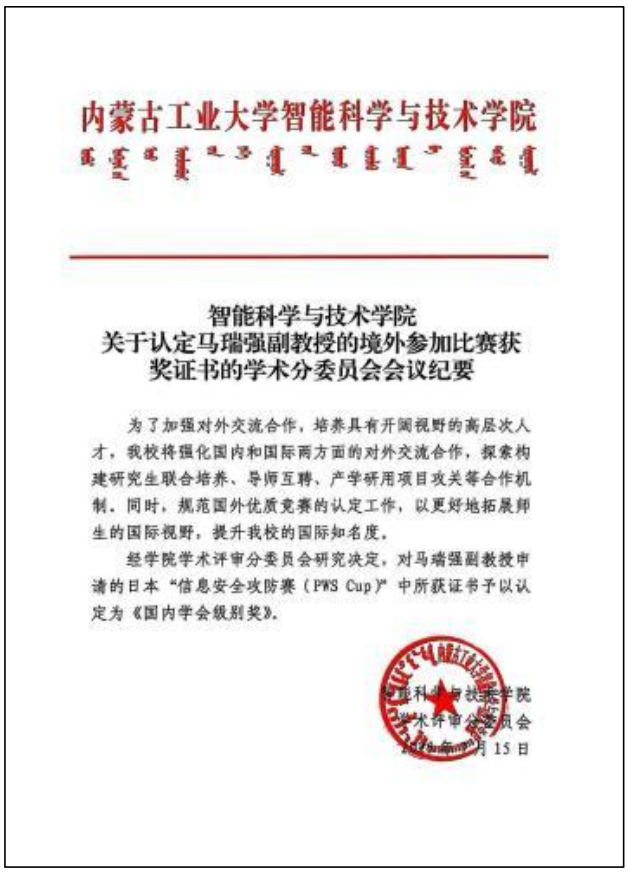
チーム名：Hots
馬 瑞強 殿(内蒙古工業大学)
郭 亜楠 殿(内蒙古工業大学)
崔 旭　殿(内蒙古工業大学)
楊 皓然 殿(内蒙古工業大学)

あなたがたは PWS Cup 2024 匿名化・属性推定コンテスト
において きわめて優秀な成績をおさめました
よってその功績をたたえ表彰します

2024年10月24日
一般社団法人 情報処理学会
コンピュータセキュリティ研究会
PWS組織委員会 PWS2024

実行委員長 千田浩司

図 1　优胜奖励证书


内蒙古工业大学智能科学与技术学院

智能科学与技术学院
关于认定马瑞强副教授的境外参加比赛获奖证书的学术分委员会会议纪要

为了加强对外交流合作，培养具有开阔视野的高层次人才，我校将强化国内和国际两方面的对外交流合作，探索构建研究生联合培养、导师互聘、产学研用项目攻关等合作机制。同时，规范国外优质竞赛的认定工作，以更好地拓展师生的国际视野，提升我校的国际知名度。

经学院学术评审分委员会研究决定，对马瑞强副教授申请的日本“信息安全攻防赛（PWS Cup）”中所获证书予以认定为《国内学会级别奖》。

15 日

图 2　优胜奖励证书认定

7 结束语

本研究表明，以国际数据匿名化攻防赛为真实任务载体、以过程证据链为核心评价框架的教学改革，在不禁用生成式人工智能（GAI）的前提下，有效解决了学习过程的不可见与能力的不可证难题，系统提升了学生的真实参与度与科研素养。实现了全覆盖的深度参与，代码提交率达到 100%，表明该模式成功激发了所有学生的全程投入。显著提升了过程质量与严谨性：迭代成绩均分从约 70.1 提升至 85.5，说明学生在机制解释、实验验证与反思修正等核心科研能力上取得了实质性进步。塑造了具有国际竞争力的成果，从无到有，15%的参赛队伍获得了国际竞赛前 10%的奖项，证明过程性训练能有效转化为顶尖的对外输出能力。固化了的有效学习行为：人均有效提交次数增长超过 50%，体现了学生从一次性交付到持续迭代研习的行为模式转变。综上所述，本研究验证了任务链与证据链的双链耦合模式的有效性，为生成式 AI 时代数据安全及相关领域的实践课程改革，提供了一条可操作、可验证且能兼顾公平与卓越的培养路径。后续研究可在更大范围与多学科背景下，进一步验证该框架的普适性与优化方向。

以国际数据匿名化攻防赛为真实任务载体，构建过程证据驱动的数据安全教学改革方案，通过任务链

组织赛前准备、赛中迭代与赛后复盘，通过证据链要求提交最小充分证据包，并以四维量规与抽样核验实现可操作评价。实践表明，该模式在不禁用生成式 AI 的前提下提升了学生真实参与度与实验验证质量，使学习过程可见、可核验、可追溯，为生成式 AI 时代数据安全课程的评价改革与科研能力培养提供了可复制路径。

综上，研究结果表明，以国际数据匿名化攻防赛为牵引、以过程证据为核心的教学改革能够同时提升真实性、参与度与科研式能力，并且在教师工作量可控的条件下可落地运行。其关键有效性来自三点：一是任务链把赛前准备、赛中迭代与赛后复盘串成连续闭环，使学生必须经历提出假设、实施改动、对照验证、解释结果的科研式循环，而不是停留在一次性提交；二是证据链与最小充分证据包将思路形成、AI 交互演化、实验对照与协作贡献强制纳入提交与评分，使“生成式完成”难以冒充“机制性理解”，从根本上降低结果导向评价在生成式 AI 环境下的失真；三是四维量规与一致性检验形成稳定的真实性约束，配合竞赛官方评分这一外部客观标准，实现过程证据与结果成绩的双重校验，提升评价公平性与可解释性。在可推广性方面，本研究已在 60 名学生、15 组、8 周周期的组织框架下验证其可执行性，证明该模式不仅适用于少数精英参赛，也可通过稳定小组、模板化证据包与抽样核验扩展到更大范围的育人场景。后续研究建议在多学期、多班级条件下进一步引入对照设计与更细粒度的过程数据分析，以量化过程证据质量对学习成效的预测作用，并持续优化证据包的成本与收益边界。

参 考 文 献

[1] Akash T R, Lessard N D J, Reza N R, et al. Investigating Methods to Enhance Data Privacy in Business, Especially in sectors like Analytics and Finance[J]. Journal of Computer Science and Technology Studies, 2024, 6(5): 143-151.

[2] https://www.iwsec.org/pws/2024/cup24.html

[3] Liu B, Ding M, Shaham S, et al. When machine learning meets privacy: A survey and outlook[J]. ACM Computing Surveys (CSUR), 2021, 54(2): 1-36.

[4] Chen M, Cang L S, Chang Z, et al. Data anonymization evaluation against re-identification attacks in edge storage[J]. Wireless Networks, 2024, 30(6): 5263-5277.

[5] Abd Razak S, Nazari N H M, Al-Dhaqm A. Data anonymization using pseudonym system to preserve data privacy[J]. Ieee Access, 2020, 8: 43256-43264.

[6] Yan Y, Herman E A, Mahmood A, et al. A weighted k-member clustering algorithm for k-anonymization[J]. Computing, 2021: 1-23.

[7] Madan S, Goswami P. A technique for securing big data using k-anonymization with a hybrid optimization algorithm[J]. International Journal of Operations Research and Information Systems (IJORIS), 2021, 12(4): 1-21.

[8] Vasa J, Thakkar A. Deep learning: Differential privacy preservation in the era of big data[J]. Journal of Computer Information Systems, 2023, 63(3): 608-631.

[9] Ye M, Shen W, Zhang J, et al. Securereid: Privacy-preserving anonymization for person re-identification[J]. IEEE Transactions on Information Forensics and Security, 2024, 19: 2840-2853.

[10] Sei Y, Okumura H, Ohsuga A. Re-identification in differentially private incomplete datasets[J]. IEEE Open Journal of the Computer Society, 2022, 3: 62-72.

[11] Manzanares-Salor B, Sánchez D, Lison P. Evaluating the disclosure risk of anonymized documents via a machine learning-based re-identification attack[J]. Data Mining and Knowledge Discovery, 2024, 38(6): 4040-4075.

[12] 马瑞强,崔旭,郭亚楠,杨皓然，通过参加PWSCup攻防赛以提升学生科研能力之探索[J]，计算机技术与教育学报(2325-0208)，2024年12月 第12卷 第6期，P78-84.