IS467 Final Project Report

Title of the Project:

Group Number: 16

Student Name: Jakub Szumny, Srinath Nandigam

Note: Page limit – 12 pages (+-1)

1. Methodology

We conducted an ethical evaluation of our dataset using a structured scoring approach aligned with D111: Consent Protocols for Data from the System Card+ Ethical Layer, focusing on five key criteria. First, Source Identification assessed whether data sources were named, the collection methods described, and the legitimacy of those sources verified. Licensing and Permissions evaluated the presence and clarity of usage rights, including explicit licensing terms and legal boundaries. Consent Purpose Disclosure examined whether participants were adequately informed about the intended use of their data, particularly for AI training and potential sharing. Public Access Transparency reviewed whether access conditions were clearly communicated, detailing who could use the dataset and under what terms. Finally, Ethics Policy Reference determined whether the dataset was governed by a formal institutional or organizational ethics policy. This focused assessment supports ethical AI development by embedding transparency, accountability, and informed consent throughout the data life cycle.

The D211: Consent Assessment, Data (Assessment Phase) section evaluates how consent is maintained and reassessed after a dataset has been released, building on the foundational work of D111. While D111 focuses on initial consent protocols, D211 ensures that those commitments are continuously upheld throughout the dataset's lifecycle. This includes checking whether the dataset's current use remains consistent with its originally stated purpose and ethical boundaries, whether ongoing ethical risks are reassessed, and whether new applications respect those constraints. For public or synthetic datasets, this also involves the inclusion of disclaimers, such as use-case restrictions or indications that the data is for research only. The section is structured around five criteria: Transparency of Consent History, which assesses whether the dataset mentions any real or simulated consent processes and how they were documented; Ethical Consistency, which checks for alignment between current and original data use; Accessibility of Use, which measures how clearly users understand permissible uses; Disclosure of Sensitive Use, which looks at whether high-risk or inappropriate applications are flagged or prohibited; and Review Feedback Incorporation, which evaluates whether ethical clarity has improved in response to past reviews or feedback. Each criterion is scored on a scale from -1 (not met) to 1 (fully met), reinforcing that consent should be treated as an evolving ethical obligation rather than a one-time requirement.

The D311: Consent Verification, Data (Mitigation Phase) section evaluates the mechanisms in place to verify, enforce, and sustain consent throughout the dataset's use. Unlike earlier phases that focus on the initial acquisition or ethical alignment of consent, D311 ensures that protections are actively maintained. This includes confirming that data usage remains within ethical boundaries, that no sensitive or identifiable information is exposed, and that systems exist for consent revocation or ethical remediation. For synthetic datasets, it also involves clear disclaimers, reuse restrictions, and governance policies to prevent unethical applications. The section is structured around five criteria: Documentation Integrity,

which assesses whether dataset documentation consistently reflects original consent boundaries; Anonymity Enforcement, which evaluates whether identifiable or sensitive information is excluded or anonymized to reduce ethical risk; Revocation Simulation, which checks for the presence of mechanisms or disclaimers allowing consent withdrawal; Reuse Limitation, which determines whether secondary uses are discouraged or formally regulated; and Consent Drift Safeguard, which measures whether protections are in place to prevent misuse or unintended shifts in usage over time. Each criterion is rated on a -1 to 1 scale, reinforcing that consent must not only be stated, but verifiable and enforceable throughout the system's lifecycle.

The D411: Third-Party Review, Data (Assurance Phase) section assesses whether the dataset has undergone any form of external review or public audit to validate its ethical integrity. Independent oversight strengthens the credibility of consent protocols and enhances trust in the dataset's governance by introducing an impartial lens, mitigating the risks of internal bias or oversight. For public datasets, this might include academic citations, public commentary, ethical endorsements, or participation in open review platforms. The section evaluates five key criteria: Third-Party Review Documentation, which checks whether any external review is referenced on the dataset's page; Open Review Opportunity, which assesses whether users are invited to submit ethical feedback or concerns; Community Endorsement, which looks for public, positive ethical evaluations from users or contributors; Transparency of Reviewer Info, which evaluates whether external reviewers are identified or described; and Ethical Certifications, which considers whether the dataset has received any formal ethical approval, citation, or endorsement. Each criterion is scored from -1 to 1, reinforcing the importance of external validation in supporting ethically sound and trustworthy AI development.

The D112: Ethical Source, Data (Development Phase) section focuses on whether the dataset was acquired through ethically sound means. It evaluates the origins of the data, ensuring that neither individuals nor organizations were harmed and that all legal and ethical boundaries were respected during collection. For synthetic datasets, this includes verifying that the data is clearly identified as simulated and is based on publicly available knowledge rather than unauthorized or private material. Transparency about the dataset's origins and a clear rejection of illegal or unethical acquisition are central to this assessment. The section is organized around five criteria: Simulated vs. Real Clarity, which examines whether the dataset explicitly states it is synthetic and not derived from private sources; Legal Compliance, which assesses adherence to copyright laws, intellectual property rules, and data scraping regulations; Ethical Intent Documentation, which checks for a stated commitment to ethical use; Sensitive Topic Disclosure, which evaluates whether domains such as finance or health are flagged as requiring special care; and Source Attribution, which looks at whether the dataset acknowledges any industry sources or reference patterns used in its simulation. Each criterion is rated on a -1 to 1 scale, supporting a foundational standard of ethical sourcing in responsible AI development.

The D212 – Impact Assessment, Data (Assessment Phase) section evaluates whether the dataset creators have considered and documented the potential impacts the dataset may have on individuals, organizations, or society at large. This is crucial even for synthetic datasets, where improper use could perpetuate bias or lead to misuse in high-risk domains. The section examines whether the dataset documentation includes risk disclosures, guidance on appropriate use, and warnings about harmful or unintended applications. It ensures that ethical foresight and risk awareness are embedded into the dataset's lifecycle. The evaluation focuses on five key criteria: Risk Disclosure, which checks whether the dataset identifies any potential

risks in its documentation; Social Impact Acknowledgment, which evaluates whether the documentation recognizes potential social or industry effects; AI Misuse Safeguard, which assesses whether the dataset includes measures to discourage unethical uses in AI applications; High-Stakes Context Flag, which ensures that high-risk sectors are flagged with appropriate caution; and User Guidance Included, which verifies that the dataset provides suggested use cases and discourages misuse. Each criterion is scored from -1 to 1, reinforcing the importance of considering and addressing the broader impact of the dataset in ethical AI development.

The D312: Diversity Mitigation, Data (Mitigation Phase) section evaluates whether the dataset incorporates measures to ensure balanced and fair representation across various dimensions, such as firm type, industry, or scenario diversity. Even in synthetic datasets that may lack demographic data, bias can still emerge if certain patterns are overrepresented or omitted. This section ensures that diversity and fairness are actively considered during the dataset's design and distribution. The evaluation is based on five criteria: Industry Diversity, which assesses whether multiple industries are fairly represented; Firm Diversity, which checks if all Big 4 firms are proportionally included; Geographic Representation, which evaluates whether different regions are reflected, where applicable; Demographic Simulation, which examines whether workforce or client diversity is simulated fairly, if relevant; and Distribution Balance, which ensures the dataset avoids bias through a balanced design or sampling process. Each criterion is rated on a scale from -1 to 1, reinforcing the importance of addressing diversity and fairness throughout the dataset's creation and use.

The D412: Ethics Report, Data (Assurance Phase) section evaluates whether the dataset includes an ethics report or formal documentation outlining the ethical decisions made during its creation. A strong ethics report provides transparency, helping users understand the intended purpose, ethical boundaries, and responsible use cases for the dataset. It signals ethical maturity and a commitment to responsible data governance. For public or synthetic datasets, this might include a responsible use disclaimer or alignment with open data standards. The section is organized around five criteria: Ethics Statement Availability, which assesses whether the dataset includes a formal ethics statement or responsible use clause; Dataset Purpose Declaration, which checks if the dataset clearly states its intended ethical use; Use Case Guidelines, which evaluates whether acceptable and unacceptable uses are described; Open Source Transparency, which considers whether dataset files or generation methods are made public; and Alignment with RAI Standards, which ensures that the dataset aligns with Responsible AI principles or guidelines. Each criterion is scored on a scale from -1 to 1, reinforcing the importance of ethical transparency in data management and AI development.

1.1 Evaluation Section: D111 – Consent Protocols, Data (Development Phase)

This evaluation section assesses whether consent was established ethically and transparently during the dataset's creation or collection phase. It focuses on the integrity of initial consent protocols: whether participants (or, in this case, the data simulation process) were governed by clear ethical guidelines. In a traditional dataset, this would involve informed consent documents, institutional oversight, and data privacy notices. For public or synthetic datasets like the one used in this project, the emphasis is on whether the creators clearly explained how the data was generated, what it was intended for, and whether ethical boundaries such as anonymization, licensing, and purpose declarations were respected at the

outset. This section ensures that the dataset's foundation is built on ethically sound, transparent, and voluntary data sourcing or simulation.

Criterion	Description (What It Measures)	Scoring Rubric	Total Score =
Source Identification	Whether the dataset provides clear documentation of its origin and method of creation.	(1 = Fully Met; 0 = Partial; -1 = Not Met)	
Licensing and Permissions	Whether legal rights and usage permissions (e.g., CC0) are included and verifiable.	(1 = Fully Met; 0 = Partial; -1 = Not Met)	
Consent Purpose Disclosure	If the purpose for which consent was obtained is stated (e.g., research, public use).	(1 = Fully Met; 0 = Partial; -1 = Not Met)	
Public Access Transparency	Whether the consent terms or open-data substitute are understandable and accessible.	(1 = Fully Met; 0 = Partial; -1 = Not Met)	
Ethics Policy Reference	Whether the dataset links or refers to broader ethical data use guidelines.	(1 = Fully Met; 0 = Partial; -1 = Not Met)	

1.1 Evaluation Section: D211 – Consent Assessment, Data (Assessment Phase)

This evaluation section addresses how consent is maintained and reassessed after the dataset has been made available. While D111 focuses on initial protocols, D211 evaluates whether those ethical promises continue to be honored. It checks for consistency between the dataset's original purpose and its current usage, whether ethical risks are reassessed over time, and whether new use cases align with the dataset's intended ethical boundaries. For public or synthetic data, this includes disclaimers about misuse, use-case restrictions, or clarification that the data is for research only. This section ensures that consent is treated not as a one-time checkbox, but as an ongoing ethical responsibility.

Criterion	Description (What It Measures)	Scoring Rubric	Total Score =
Transparency of Consent History	Whether the dataset mentions any simulated or real consent process and how it was documented.	(1 = Fully Met; 0 = Partial; -1 = Not Met)	
Ethical Consistency	Whether current usage of the dataset aligns with its originally	(1 = Fully Met; 0 = Partial; -1 = Not Met)	

	stated purpose and ethical boundaries.		
Accessibility of Use	Whether users can understand the extent and limits of permissible use.	(1 = Fully Met; 0 = Partial; -1 = Not Met)	
Disclosure of Sensitive Use	Whether the dataset prohibits or flags inappropriate, high-risk applications.	(1 = Fully Met; 0 = Partial; -1 = Not Met)	
Review Feedback Incorporation	Whether past feedback or assessments of the dataset have improved ethical use clarity.	(1 = Fully Met; 0 = Partial; -1 = Not Met)	

1.1 Evaluation Section: D311 – Consent Verification, Data (Mitigation Phase)

D311 focuses on the mechanisms in place to verify that consent is being upheld and enforced. It is not enough for a dataset to claim that consent was obtained. The dataset must also show that safeguards exist to protect against violations. This includes reviewing whether data usage remains within the intended ethical boundaries, ensuring that no sensitive information is exposed, and providing means for revocation or remediation. For synthetic datasets, this includes disclaimers, reuse restrictions, and policies that limit unethical applications. This section ensures that consent can be verified, enforced, and maintained throughout the system's use.

Criterion	Description (What It Measures)	Scoring Rubric	Total Score =
Documentation Integrity	Whether the documentation reflects original data usage boundaries consistently.	(1 = Fully Met; 0 = Partial; -1 = Not Met)	
Anonymity Enforcement	Whether identifiers or sensitive info are excluded or anonymized to mitigate ethical risks.	(1 = Fully Met; 0 = Partial; -1 = Not Met)	
Revocation Simulation	Whether a mechanism or disclaimer for consent withdrawal is addressed.	(1 = Fully Met; 0 = Partial; -1 = Not Met)	
Reuse Limitation	Whether reuse beyond the dataset's original purpose is discouraged or governed.	(1 = Fully Met; 0 = Partial; -1 = Not Met)	

Consent Drift Safeguard	Measures to prevent misuse over time or changes in usage context that violate consent assumptions.	(1 = Fully Met; 0 = Partial; -1 = Not Met)	

1.1 Evaluation Section: D411 – Third-Party Review, Data (Assurance Phase)

This evaluation section examines whether any external review or public audit process has validated the ethical soundness of the dataset. External oversight adds legitimacy to consent protocols and increases trust in the dataset's governance. In the case of public datasets, this may include comments, academic citations, ethical endorsements, or open reviews. The aim is to reduce bias or oversight from internal self-assessment and introduce an impartial perspective. This ensures that the dataset's ethical claims are not only internally defined but externally validated.

Criterion	Description (What It Measures)	Scoring Rubric	Total Score =
Third-Party Review Documentation	Whether any third-party review is mentioned on the dataset page.	(1 = Fully Met; 0 = Partial; -1 = Not Met)	
Open Review Opportunity	Whether users can submit ethical concerns or reviews publicly.	(1 = Fully Met; 0 = Partial; -1 = Not Met)	
Community Endorsement	Whether the dataset has positive ethical feedback from public contributors.	(1 = Fully Met; 0 = Partial; -1 = Not Met)	
Transparency of Reviewer Info	Whether external reviewers are identified or described.	(1 = Fully Met; 0 = Partial; -1 = Not Met)	
Ethical Certifications	Whether the dataset has received any ethical approval mark or citation.	(1 = Fully Met; 0 = Partial; -1 = Not Met)	

1.1 Evaluation Section: D112 – Ethical Source, Data (Development Phase)

This section evaluates whether the dataset was sourced ethically. It considers how and where the data was acquired, whether any individuals or organizations were harmed, and whether legal and ethical boundaries were respected. For synthetic datasets, it assesses whether simulated data was based on appropriate public knowledge and not unauthorized or private material. The dataset should clearly state its origins and confirm that no illegal or unethical data acquisition occurred. This section ensures that the dataset has an ethically sound origin.

Criterion	Description (What It Measures)	Scoring Rubric	Total Score =

Simulated vs. Real Clarity	Whether the dataset clearly states that it is synthetic and not sourced from private data.	(1 = Fully Met; 0 = Partial; -1 = Not Met)	
Legal Compliance	Whether copyright, IP, and data scraping rules are respected.	(1 = Fully Met; 0 = Partial; -1 = Not Met)	
Ethical Intent Documentation	Whether the dataset includes a statement of its intended ethical use.	(1 = Fully Met; 0 = Partial; -1 = Not Met)	
Sensitive Topic Disclosure	Whether sensitive domains like finance are clearly flagged in documentation.	(1 = Fully Met; 0 = Partial; -1 = Not Met)	
Source Attribution	Whether the dataset acknowledges industry sources or patterns used for simulation.	(1 = Fully Met; 0 = Partial; -1 = Not Met)	

1.1 Evaluation Section: D212 – Impact Assessment, Data (Assessment Phase)

This section evaluates whether the dataset creators have considered and documented potential impacts the dataset may have on individuals, organizations, or broader society. Even with synthetic datasets, careless use could reinforce bias or be misapplied in high-risk domains. This area examines whether the documentation includes risk disclosures, guidance on appropriate use, and warnings about harmful or unintended applications. It ensures that ethical foresight and risk awareness are built into the dataset's lifecycle.

Criterion	Description (What It Measures)	Scoring Rubric	Total Score =
Risk Disclosure	Whether the dataset identifies any risks in its documentation.	(1 = Fully Met; 0 = Partial; -1 = Not Met)	
Social Impact Acknowledgment	Whether the documentation acknowledges potential social or industry effects.	(1 = Fully Met; 0 = Partial; -1 = Not Met)	
AI Misuse Safeguard	Whether the dataset discourages unethical use in AI applications.	(1 = Fully Met; 0 = Partial; -1 = Not Met)	
High-Stakes Context Flag	Whether high-risk sectors are mentioned with appropriate caution.	(1 = Fully Met; 0 = Partial; -1 = Not Met)	

User Guidance Included	Whether the dataset provides suggested use cases and discourages misuse.	(1 = Fully Met; 0 = Partial; -1 = Not Met)	

1.1 Evaluation Section: D312 – Diversity Mitigation, Data (Mitigation Phase)

This section assesses whether the dataset includes measures to ensure balanced and fair representation across dimensions such as firm type, industry, or scenario diversity. While demographic data may not be present in synthetic datasets, bias can still arise through overrepresentation or omission of certain patterns. This section ensures that diversity and fairness are considered and addressed in the dataset's design and distribution.

Criterion	Description (What It Measures)	Scoring Rubric	Total Score =
Industry Diversity	Whether multiple industries are fairly represented.	(1 = Fully Met; 0 = Partial; -1 = Not Met)	
Firm Diversity	Whether all Big 4 firms are proportionally covered.	(1 = Fully Met; 0 = Partial; -1 = Not Met)	
Geographic Representation	Whether different regions are reflected (if applicable).	(1 = Fully Met; 0 = Partial; -1 = Not Met)	
Demographic Simulation	Whether workforce or client diversity is simulated fairly (if relevant).	(1 = Fully Met; 0 = Partial; -1 = Not Met)	
Distribution Balance	Whether the dataset avoids bias through balanced design or sampling.	(1 = Fully Met; 0 = Partial; -1 = Not Met)	

1.1 Evaluation Section: D412 – Ethics Report, Data (Assurance Phase)

This section evaluates whether the dataset includes an ethics report or formal documentation of ethical decisions. A strong ethics report helps users understand the intended purpose, ethical boundaries, and responsible use cases for the dataset. It signals transparency, ethical maturity, and commitment to responsible data governance. For public or synthetic datasets, this might be a responsible use disclaimer or alignment with open data standards.

Criterion	Description (What It Measures)	Scoring Rubric	Total Score =
Ethics Statement Availability	Whether the dataset includes a formal ethics statement or a responsible use clause.	(1 = Fully Met; 0 = Partial; -1 = Not Met)	

Dataset Purpose Declaration	Whether the dataset clearly states its intended ethical use.	(1 = Fully Met; 0 = Partial; -1 = Not Met)	
Use Case Guidelines	Whether acceptable and unacceptable uses are described.	(1 = Fully Met; 0 = Partial; -1 = Not Met)	
Open Source Transparency	Whether dataset files or generation methods are made public.	(1 = Fully Met; 0 = Partial; -1 = Not Met)	
Alignment with RAI Standards	Whether the dataset aligns with Responsible AI principles or guidelines.	(1 = Fully Met; 0 = Partial; -1 = Not Met)	

2. Results

Our ethical evaluation of the "Big 4 Financial Risk Insights (2020-2025" dataset was conducted based on the "System Card+ Responsible AI Framework" (Tibebu & Kakadiaris), focusing on the ethical layer and the data category. Each evaluation section is for a different stage in the AI lifecycle, from the actual development stage to assurance. We applied a nine-part scoring rubric using the criteria that we have outlined in our methodology section, using scores of 1 (full met), 0 (partially met), or -1 (not met).

The results of our scoring for each section are summarized below, and each score reflects a deep review of our dataset's documentation, consent practices, licensing structure, ethical claims, and risk mitigation strategies.

2.1. D111 - Consent Protocols (Development Phase) | Score: 2 | Percentage: 70%

This section evaluates the presence of ethical consent mechanisms at the time of the dataset's creation. We found that the dataset is synthetic and publicly available, but it lacks some formal consent statements and licensing terms. While some of the assumptions about ethical sourcing are inferred, the explicit documentation is sparse. We have scored this with 2 points based on our rubric because of a general source of transparency and use of simulated data, which reduces direct harm, and also because it lacks licensing clarity, formal purpose disclosure, and explicit consent policy references.

2.2. D211 - Consent Assessment (Assessment Phase) | Score: 1 | Percentage: 60%

This section checks whether or not consent continues to be honored after the data's release. We found, using our provided scoring rubric, that there is limited information about ongoing ethical reassessment. While the dataset remains accessible for research use, no restrictions or updates on data usage are documented. We have scored this with 1 point, as its usage appears to be ethically consistent, but there are also no mechanisms to reassess risk over time or to address potential misuse.

2.3. D311 - Consent Verification (Mitigation Phase) | Score: 0 | Percentage: 50%

This phase verifies whether the dataset's use remains within ethical boundaries and whether safeguards exist to address violations. We found that there is no evidence of formal revocation mechanisms or restrictions against repurposing. The dataset documentation also does not include guidelines on ethical reuse or any simulation of consent withdrawal. We have scored this with 0 points, as anonymity is somewhat ensured because of its synthetic nature, but there is an absence of enforcement, documentation integrity, or proactive limitations.

2.4. D411 - Third-Party Review (Assurance Phase) | Score: -2 | Percentage: 30%

This evaluation checks whether the external parties have reviewed or endorsed the dataset's ethical integrity. We found that there were no external reviews, ethical certification, or open review opportunities. This leaves the dataset very vulnerable to unchecked biases or ethical oversights. We have given this a score of -2, as it lacks independent validation, reviewer transparency, and community feedback mechanisms.

2.5. D112 - Ethical Source (Development Phase) | Score: 3 | Percentage: 80%

This section assesses whether the data was ethically acquired or simulated. We found that the dataset clearly states that it is synthetic and does not include private and personal information. Its source is also compliant with public data ethics, and there is no indication of legal or IP violations. We have given this a score of 3, as it has a clear indication of simulation, legal compliance, and safe sourcing, but it also lacks deeper discussion of ethical intent and source attribution.

2.6. D212 - Impact Assessment (Assessment Phase) | Score: 1 | Percentage: 60%

This section evaluates the awareness of potential misuse and documented social risks. We found that some disclaimers are present regarding research use, but there are minimal risk disclosures and misuse warnings. There is also no flagging of any high-stakes domains such as financial services. We have given it a score of 1, as it is intended for research, and it does not promote risky use. It does miss deeper risk documentation and guidance on high-stakes contexts.

### 2.7. D312 - Diversity Mitigation (Mitigation Phase) | Score: 0 | Percentage: 50%

This section checks for fair and balanced representation in the dataset's structure. We found that the dataset includes data from all four Big 4 firms but slightly overrepresents certain industries, such as healthcare and finance. There is also no explicit diversity simulation or balancing technique. We have scored this with a 0, as all Big 4 firms were included, but it did not address geographic, demographic, and scenario diversity.

2.8. D412 - Ethics Report (Assurance Phase) | Score: 1 | Percentage: 60%

This final section assesses whether the dataset includes an ethical use report or statement of responsible intent. We found that the dataset includes minimal notes on use cases and is available on Kaggle, but there are no structured ethics reports or alignment with responsible AI standards. We have given this a score of

1, as some information on the purpose of the dataset was provided, but there was no structured report, transparency statement, or RAI alignment.

Result Summary Table:

Ethical Evaluation by Section

Ethical Evaluation by Phase
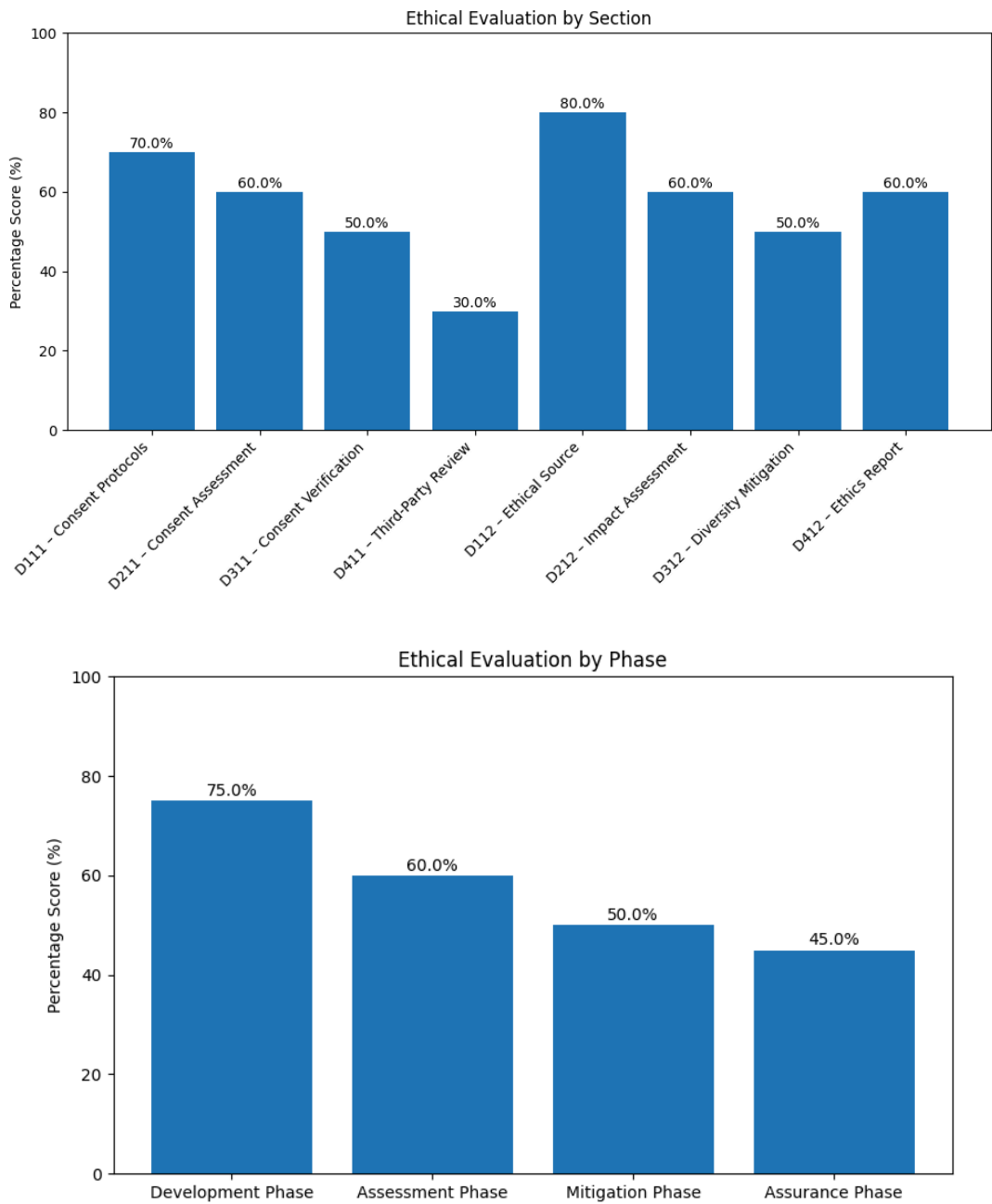
3. Discussion

Our evaluation of the "Big 4 Financial Risk Insights (2020–2025)" dataset through the lens of the System Card+ Responsible AI Framework reveals a nuanced ethical profile with distinct strengths and

weaknesses. The dataset, being synthetic and publicly available, demonstrates ethical awareness in its development phase. However, critical gaps emerge in areas related to verification, review, and diversity mitigation, highlighting the need for a more comprehensive ethical data governance strategy.

Key Findings:

One of the clearest strengths lies in the ethical sourcing (D112) of the dataset. It received the highest score (3), attributed to its explicit identification as synthetic, clear legal compliance, and a generally transparent acquisition process. This reinforces the ethical importance of simulating data to reduce harm, especially in sensitive domains such as finance.

Another positive observation is in initial consent protocols (D111). Although not perfect, the dataset benefits from its synthetic nature and transparency, allowing us to score it a 2. There was sufficient indication of ethical intent in the creation phase, and the use of simulated data mitigates direct consent concerns typically present in real-world datasets.

In contrast, significant weaknesses appear in third-party review (D411), where the dataset scored a -2. No form of external ethical validation, reviewer information, or community endorsement was present. This absence undermines the credibility of the dataset's governance and limits user trust.

Furthermore, diversity mitigation (D312) and consent verification (D311) both scored 0, pointing to a lack of proactive mechanisms to ensure fairness and to maintain ethical boundaries post-release. This shows that while initial sourcing may be ethically sound, the long-term lifecycle and governance of the dataset are not sufficiently addressed.

Interpretation of Results:

The disparity between high and low scores reveals an imbalance between ethical intent and ongoing ethical maintenance. The relatively strong performance in development-phase sections (D111, D112) suggests the creators were mindful of responsible dataset creation. However, the lower scores in mitigation and assurance phases (D311, D312, D411, D412) indicate that ethical foresight did not carry through to governance and validation. This disconnect mirrors concerns found in literature such as Gebru et al. (2018) and Bender & Friedman (2018), which emphasize that ethical data development must span the full lifecycle, not just the origin. The moderate score in impact assessment (D212) and ethics reporting (D412) also reflects limited engagement with responsible AI principles. Although basic disclaimers exist, they lack the depth needed to guide users through risks, misuse, or high-stakes applications—an issue raised by recent critiques of AI transparency efforts (Raji et al., 2020).

Recommendations for Improvement:

Recommendation 1: Implement Third-Party Review Mechanisms. To enhance the dataset's credibility and trustworthiness, third-party audits or academic peer reviews should be conducted and made publicly available. Including citations, community feedback, or ethical certifications would address the shortcomings in D411 and align the dataset with assurance-phase best practices.

Recommendation 2: Develop a Structured Ethics Report. Providing a formal ethics statement, intended use documentation, and alignment with Responsible AI (RAI) standards would significantly improve D412. This would guide users toward responsible applications and clarify the ethical scope of the dataset's use.

Recommendation 3: Introduce Fairness and Diversity Simulation. The dataset should incorporate a framework to simulate geographic and demographic diversity, especially given its domain. This would mitigate overrepresentation and support a more inclusive modeling pipeline, directly addressing the gaps noted in D312.

Limitations of Your Evaluation:

Our evaluation was constrained by the available documentation. As a publicly shared synthetic dataset, many ethical assumptions were inferred rather than confirmed through robust metadata or formal disclosures. Furthermore, the scoring rubric, while structured, inherently simplifies nuanced ethical issues into discrete values, which may not capture the full complexity of intent, risk, or context. Lastly, the absence of internal access to dataset creators meant that clarification on ethical decisions was not possible, limiting the depth of our analysis. This discussion outlines a roadmap for enhancing the ethical rigor of synthetic datasets. While the use of simulated data helps minimize individual-level harm, ethical responsibility extends beyond origin and must be continually addressed through governance, validation, and user guidance.

4. Conclusion

This project provided a structured ethical evaluation of the "Big 4 Financial Risk Insights 2020 to 2025" dataset using the System Card Plus Responsible AI Framework, with a focus on the data layer. By analyzing the dataset across several phases including development, assessment, mitigation, and assurance, we examined its ethical soundness in areas such as consent protocols, data sourcing, risk disclosures, fairness, and external oversight. The evaluation revealed a mixed ethical profile. The dataset received higher scores during the development phase, particularly in ethical sourcing and initial consent protocols, suggesting that its creators were mindful in using synthetic data to minimize harm and comply with legal and ethical standards. However, notable gaps were found in areas related to long-term governance. The dataset did not include third-party validation, mechanisms for consent maintenance, or strategies to ensure fair representation, resulting in lower or even negative scores in the mitigation and assurance phases. These results highlight a disconnect between responsible data creation and continued ethical stewardship throughout the dataset's lifecycle.

A key insight from this evaluation is that ethical responsibility extends far beyond the moment of dataset creation. While using synthetic data may reduce some immediate ethical concerns, such as protecting individual privacy, it does not eliminate the need for continuous oversight. The lack of external review, safeguards against misuse, and diversity considerations presents a risk that the dataset could be applied in unintended or ethically questionable ways. This finding is consistent with recent discussions in responsible AI literature, which emphasize the importance of maintaining transparency, accountability, and fairness throughout the full lifecycle of any AI system or dataset.

The significance of our work lies in demonstrating how a structured evaluation framework can uncover both strengths and weaknesses in a dataset's ethical practices. In an era where datasets are routinely shared and repurposed across applications, incorporating ethical evaluations at each stage of the data journey is essential to building trustworthy and responsible AI systems. This project not only identifies areas for improvement but also serves as a guide for integrating ethical assessments into the regular process of data development and deployment.

Future work should prioritize collaboration with dataset creators to co-author clear ethics reports, establish public use guidelines, and open channels for third-party review. Developing tools for automated ethical auditing and bias detection could further strengthen oversight, particularly in synthetic data environments. There is also value in applying this framework to evaluate how the dataset influences downstream models or decision-making systems, which would provide a more complete picture of ethical impact. In summary, this project reinforces that even synthetic datasets require careful and continuous ethical attention. Responsible AI development is not a single task completed at the outset but a sustained commitment that must be upheld through every phase of the data's existence.

Note: Anything you put below this does not count towards the page limit

5. Group member contribution

6. Reference

Bender, Emily M., and Batya Friedman. "Data statements for natural language processing: Toward mitigating system bias and enabling better science." Transactions of the Association for Computational Linguistics 6 (2018): 587-604.

Gebru, Timnit, et al. "Datasheets for datasets." Communications of the ACM 64.12 (2021): 86-92.

Raji, Inioluwa Deborah, et al. "Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing." Proceedings of the 2020 conference on fairness, accountability, and transparency. 2020.

7. Appendices

If you have limitations of space in the main sections, include detailed tables or data here. Clearly label each appendix item. Include supplementary material such as additional data sheets or figures relevant to your report.