Review of Previous Trimester Methods:

SDV-GAN (Synthetic Data Vault - Generative Adversarial Networks):

- Overview: SDV-GAN is a synthetic data generation method that utilizes GANs to produce realistic data that mirrors the distribution of real data. It's commonly used for generating tabular data in various domains.

- Strengths: High fidelity in data replication, particularly effective in capturing complex relationships in the data.

- Weaknesses: Computationally intensive, may require significant tuning to avoid mode collapse or overfitting.

Mimesis:

- Overview: Mimesis is a tool for generating fake data with a focus on structured and realistic outputs. It's commonly used for generating personal data (e.g., names, addresses) for testing purposes.

- Strengths: Easy to use with a variety of data types, high customizability.

- Weaknesses: May not be suitable for generating highly complex or interdependent data structures.

Mockaroo:

- Overview: Mockaroo is a cloud-based platform that provides an easy interface for generating structured synthetic data across multiple domains. It allows users to define schemas and generate data that matches these schemas.

- Strengths: User-friendly interface, supports a wide range of data types and schemas.

- Weaknesses: Limited in handling complex data dependencies, reliant on predefined data patterns.

Faker:

- Overview: Faker is a Python library that generates fake data for testing and development. It's versatile and can be used to create various types of random data, from names and addresses to more complex structures.

- Strengths: Lightweight, easy to integrate with Python projects, highly customizable.

- Weaknesses: Not specifically designed for generating complex, interdependent datasets.

Research on Other Synthetic Data Methods:

TableGAN:

- Overview: TableGAN is a GAN-based method specifically designed for generating synthetic tabular data. It uses a generator to create data and a discriminator to evaluate its realism compared to the real data.

- Applications: Particularly useful for generating synthetic financial data, where relational data integrity is crucial.

- Challenges: Balancing data diversity with privacy guarantees, as the generated data should not inadvertently reveal information about real individuals.

GANBLR:

- Overview: GANBLR is a variant of GANs that focuses on balancing between learning the distribution of real data and ensuring that the generated data maintains privacy guarantees. It uses techniques such as differential privacy to enhance the privacy of the synthetic data.

- Applications: Ideal for scenarios where privacy is paramount, such as generating synthetic data for financial transactions.

- Challenges: Managing the trade-off between data utility and privacy, as increasing privacy often reduces the utility of the data.

Other Techniques:

- SMOTE (Synthetic Minority Over-sampling Technique): Although not a GAN-based method, SMOTE is commonly used to address class imbalance by generating synthetic samples of the minority class.

- CTGAN (Conditional Tabular GAN): Similar to TableGAN, CTGAN focuses on generating synthetic tabular data but adds conditional generation capabilities, allowing more control over the generated data.

References:

- Patki, N., Wedge, R., & Veeramachaneni, K. (2016). *The synthetic data vault*. 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA), pp. 399-410. DOI:10.1109/DSAA.2016.49
- Xu, L., Skoularidou, M., Cuesta-Infante, A., & Veeramachaneni, K. (2019). *Modeling Tabular data using Conditional GAN*. Advances in Neural Information Processing Systems, 32, 7335-7345. arXiv:1907.00503
- Mimesis Documentation. *Mimesis: Fake Data Generator*. GitHub Repository
- PyPi. (2023). *mimesis 5.5.0 documentation*. PyPi
- Mockaroo. (n.d.). *Mockaroo: Realistic Data Generator*. Website
- Mockaroo. *API Documentation*. API Docs
- PyPi. (2023). *Faker 18.9.0 documentation*. PyPi
- Harris, W. (2020). *Generating Test Data with Python's Faker Library*. Real Python. Real Python
- Park, N., & Shin, J. (2019). *Data Synthesis based on Generative Adversarial Networks*. Proceedings of the VLDB Endowment, 12(11), 1399-1410. DOI:10.14778/3342263.3342631
- Xu, L., Bower, A., Gowda, S., Skoularidou, M., & Veeramachaneni, K. (2020). *Learning Fair Representations for GANs*. In Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI). arXiv:2012.06667