A Appendix

This appendix complements the main paper with detailed materials supporting our framework for forecasting student engagement levels. Section A.1 lists the 28 non-cognitive questions and sample responses used to derive qualitative longitudinal features, showcasing the data's experiential richness. Section A.2 provides an extended discussion of related work, covering prior efforts in LLMs, time-series forecasting, and educational analytics. Section A.3 elaborates on limitations, addressing dataset constraints, imputation dependencies, and computational factors, enhancing transparency and reproducibility.

A.1 Non-Cognitive Questions and Response Options

Below is the complete list of 28 non-cognitive (NC) questions used to collect weekly student engagement data. Each question includes its prompting rule and response options.

- **Q1**: How much are you looking forward to your CS1 class lecture today?
  Rule: Prompted every Monday, Wednesday, and Friday at 12:01 PM (timeout 9240s)
  Options:
  1. I am really looking forward to it
  2. I am kind of looking forward to it
  3. I am not really looking forward to it
  4. I am not planning to attend today's lecture

- **Q2**: How well do you feel you understood the lecture material today?
  Rule: Prompted every Monday, Wednesday, and Friday at 3:25 PM on departure from the lecture hall (GPS-based, timeout 9240s)
  Options:
  1. I understood all of it well
  2. I understood most of it well
  3. There were some parts I didn't understand well
  4. There were many parts I couldn't understand well

- **Q3**: What are the (up to 2) most important reasons for your experience?
  Rule: If Q2 response is 1-4 (timeout 9240s)
  Options:
  1. The clarity (or lack of it) of the presentation
  2. The interestingness (or lack of it) of the content
  3. The amount that I prepared
  4. Something else

- **Q4**: You answered "Something else". Would you like to tell us more?
  Rule: If Q3 response is 4 (timeout 9240s)
  Options:
  1. FillText

- **Q5**: Reflecting on the CS1 class today, which statement best describes your feelings?
  Rule: Prompted every Monday, Wednesday, and Friday at 7:00 PM (timeout 9240s)
  Options:
  1. I thoroughly enjoyed it
  2. I mostly enjoyed it
  3. I enjoyed it for some parts of it
  4. I did not enjoy the lecture
  5. I was bored at the lecture
  6. I did not attend the lecture

- **Q6**: What are the (up to 2) most important reasons for your response?
  Rule: If Q5 response is 1-3 (timeout 9240s)
  Options:
  1. I love learning new things
  2. I am doing well in the class
  3. I like to be with my friends in the class
  4. I am not doing well but I like being with my friends
  5. I feel respected in the class

- **Q7**: What are the (up to 2) most important reasons for your response?
  Rule: If Q5 response is 4-5 (timeout 9240s)
  Options:
  1. I don't like learning new things
  2. I am not doing well in the class
  3. I don't like to be around my classmates
  4. I don't have any friends in the class
  5. My friends don't go
  6. I don't feel respected

- **Q8**: Select an answer that best describes your reflection on your CS1 lab.

Rule: Prompted every Monday on departure from lab (GPS-based, timeout 9240s)
Options:

1. I was able to complete all tasks
2. I was able to complete most tasks
3. I was unable to complete some tasks
4. I was unable to complete most tasks
5. I did not go to the lab today

- **Q9**: What are the (up to 2) most important reasons for your response?
  Rule: If Q8 response is 3-4 (timeout 9240s)
  Options:

  1. I did not study the relevant topics
  2. I studied but tasks were too difficult
  3. I did not seek help from lab assistants
  4. I did not get help from lab assistants
  5. I did not attend past lectures
  6. I don't have a partner

- **Q10**: What are the (up to 3) most important reasons for your response?
  Rule: If Q8 response is 5 (timeout 9240s)
  Options:

  1. Physically unwell
  2. Don't like being in lab
  3. Didn't study relevant topics
  4. Don't get help from assistants
  5. Did not attend past lectures
  6. No partners
  7. Can do tasks alone
  8. Attended another day

- **Q11**: Select up to 3 responses that best describe your experience with classmates in the last 2 days.
  Rule: Prompted every Tuesday and Thursday at 12:01 PM (timeout 9240s)
  Options:

  1. Learned something new
  2. Students near me work well together
  3. Learned something personal
  4. Comfortable asking for help
  5. Classmates respect my opinions
  6. Opinions not respected
  7. Didn't feel like talking
  8. Worked by myself

- **Q12**: Select up to 3 responses that best describe your experience with your instructor in the last 2 days.
  Rule: If Q11 response is 1-8 (timeout 9240s)
  Options:

  1. Instructor knows my name
  2. Instructor cares about me
  3. Acquainted with instructor
  4. Spoke informally
  5. Comfortable asking for help
  6. Not comfortable asking
  7. Instructor respects opinions
  8. Opinions not respected
  9. Didn't feel like talking

- **Q13**: How strongly do you feel you belong at UNL?
  Rule: Prompted every Tuesday and Thursday on departure from areas on campus where students usually gather outside of their classes or labs (GPS-based, timeout 9240s)
  Options:

  1. Really belong
  2. Bit like I belong
  3. Could belong
  4. Little out of place
  5. Don't belong

- **Q14**: How strongly do you feel you belong in the CS1 class?
  Rule: If Q13 response is 1-5 (timeout 9240s)
  Options:

  1. Really belong
  2. Bit like I belong
  3. Could belong
  4. Little out of place
  5. Don't belong

- **Q15**: What strategy do you typically use for solving assignments and lab problems? (Up to 3)
  Rule: Prompted every Tuesday and Thursday at 7:00 PM (timeout 9240s)
  Options:

  1. Use concepts from lectures/labs
  2. Categorize problems
  3. Solve without prior context
  4. Ask friends

5. Search online

- **Q16**: Which statement best describes your experience? (Up to 2)
  Rule: If Q15 response is 1-5 (timeout 9240s)
  Options:
  1. Attempt extra problems
  2. Ask instructor for more
  3. Only required problems
  4. Feel anxious
  5. Struggle with required problems

- **Q17**: What are the (up to 2) most important reasons?
  Rule: If Q16 response is 1-2 (timeout 9240s)
  Options:
  1. Love challenging problems
  2. Increase grade
  3. Be ahead
  4. Impress instructor
  5. Impress friends

- **Q18**: What grade do you think you might earn in CS1?
  Rule: Prompted every Saturday at 12:01 PM (timeout 9240s)
  Options:
  1. A
  2. B
  3. C
  4. D
  5. Not pass

- **Q19**: How confident are you in completing CS1 requirements?
  Rule: If Q18 response is 1-5 (timeout 9240s)
  Options:
  1. Very confident
  2. Confident
  3. Somewhat confident
  4. Little confident
  5. Not confident

- **Q20**: How confident are you in excelling in CS1?
  Rule: If Q19 response is 1-5 (timeout 9240s)
  Options:
  1. Very confident
  2. Confident
  3. Somewhat confident
  4. Little confident
  5. Not confident

- **Q21**: How satisfied are you with your performance in this class?
  Rule: Prompted every Saturday at 7:00 PM (timeout 9240s)
  Options:
  1. Very satisfied
  2. Satisfied
  3. Somewhat satisfied
  4. Little satisfied
  5. Not satisfied

- **Q22**: How do you think other students are performing compared to you?
  Rule: If Q21 response is 1-5 (timeout 9240s)
  Options:
  1. Much better
  2. Little better
  3. I'm a little better
  4. I'm much better

- **Q23**: How worried are you about your performance?
  Rule: If Q22 response is 1-4 (timeout 9240s)
  Options:
  1. Not at all
  2. Little
  3. Somewhat
  4. Worried
  5. Very worried

- **Q24**: How much do you see yourself as a future engineer or scientist?
  Rule: Prompted every Sunday at 12:01 PM (timeout 9240s)
  Options:
  1. Well suited
  2. Like but unsure
  3. Want to like but doubt
  4. Not for me

- **Q25**: How much do others see you as a future engineer/scientist?
  Rule: If Q24 response is 1-4 (timeout 9240s)
  Options:
  1. Very much

2. A lot
3. Somewhat
4. A little
5. Not at all

- **Q26**: How important is CS1 for your future career?
  Rule: If Q25 response is 1-5 (timeout 9240s)
  Options:
  1. Very important
  2. Important
  3. Somewhat important
  4. Little important
  5. Not important

- **Q27**: How important is doing well in college classes for a good life?
  Rule: If Q26 response is 1-5 (timeout 9240s)
  Options:
  1. Very important
  2. Important
  3. Somewhat important
  4. Little important
  5. Not important

- **Q28**: What type of on-campus extracurricular activities are you involved in?
  Rule: Prompted every Sunday at 7:00 PM (timeout 9240s)
  Options:
  1. Fraternity/sorority
  2. Social club
  3. Sports team
  4. None

A.2 Related Work

This research sits at the intersection of LLMs, time-series forecasting, and educational analytics, with a particular focus on handling missing data and feature selection in LE sequences. Below, we review prior work in these areas, highlighting gaps that our LLM-based framework addresses.

LLMs for Time-Series and Sequential Data. Transformer-based LLMs have revolutionized NLP, excelling in tasks like text generation and classification (Bommasani et al., 2021). Recent efforts have extended their application to sequential data beyond text, such as time-series forecasting. Models like TimeGPT (Garza et al., 2024) and PromptCast (Xue and Salim, 2024) leverage LLMs' sequence modeling capabilities to predict numeric trends, often by verbalizing time-series into textual prompts. Research in this domain can be broadly categorized into model-centric and data-centric approaches (Sun et al., 2023).

Data-centric methods emphasize transforming time-series into representations suitable for pre-trained LMs, using embedding techniques to align time-series tokens with LM text spaces (Sun et al., 2023), augmenting embeddings with prompts containing dataset context or task instructions (Jin et al., 2024), two-stage fine-tuning (Chang et al., 2023), and zero-shot preprocessing of numeric data (Gruver et al., 2023). *Model-centric* approaches adapt LMs to time-series by fine-tuning specific layers (e.g., embedding, normalization) while freezing others (**?**), incorporating designs like time-series decomposition and soft prompts (Cao et al., 2023), framing forecasting as question-answering (Xue and Salim, 2024), or using prompt-tuning with few-shot learning (Liu et al., 2023).

While we adopt a model-centric approach by fine-tuning LLMs for forecasting, our work diverges by targeting experiential, qualitative LE data rather than numeric time-series. Unlike soft-prompt methods (Cao et al., 2023), we employ discrete prompts, and our focus on subjective engagement attributes in education addresses a domain where temporal dependencies and missingness remain underexplored by existing LLM-based time-series models.

Student Engagement Forecasting in Educational Analytics. Educational data mining has long explored student engagement through longitudinal data, often using cognitive metrics (e.g., grades) or behavioral logs (e.g., clickstreams) (Wang et al., 2014; Li et al., 2020). Machine learning methods like LSTMs and random forests have been applied to predict engagement or performance (Xu and Ouyang, 2022), but they typically rely on numeric features and struggle with the subjective, textual responses prevalent in LE data. Recent studies have incorporated non-cognitive (NC) factors—such as self-efficacy and motivation—using survey-based datasets (Fredricks, 2014; Sinatra et al., 2015), yet these efforts rarely address temporal dynamics or missingness systematically. Our approach differs by focusing on weekly NC trajectories, verbalizing them for LLM processing, and forecasting binary engagement shifts, offering a novel bridge between educational analytics and NLP.

Imputing Missing Data in LE Sequences. Missing data is a pervasive challenge in longitudinal studies, with implications for model accuracy and generalizability. Rubin's taxonomy classifies missingness as missing completely at random (MCAR), missing at random (MAR), or missing not at random (MNAR), with MNAR being particularly problematic due to its correlation with unobserved factors (e.g., disengagement) (Rubin, 1976). Traditional statistical methods, such as multiple imputation by chained equations (MICE) (van Buuren and Groothuis-Oudshoorn, 2011) and fully conditional specification (Van Buuren et al., 2006), estimate missing values based on observed data distributions. However, these approaches assume MCAR or MAR, require complete training sets, and struggle with LE data's qualitative heterogeneity and MNAR patterns, such as students skipping questions due to disinterest (Muzellec et al., 2020).

Machine learning has advanced imputation with generative models. GAIN (Yoon et al., 2018) uses Generative Adversarial Networks (GANs) to impute numeric values, while MIWAE (Mattei and Frellsen, 2019) extends importance-weighted autoencoders for MAR data. Transformed Distribution Matching (TDM) (Zhao et al., 2023) aligns incomplete batches distributionally, excelling across missingness types. These methods, however, falter with LE sequences' textual NC features and MNAR missingness, where context-aware solutions are needed. Techniques like LOCF (Liu, 2016) are inadequate, ignoring behavioral causes of missingness. Transformer-based approaches like TabMT (Gulati and Roysdon, 2023) and LLM pre-training on tables (Yang et al., 2024) show promise but overlook LE-specific patterns. Our LLM-informed imputation uses GPT-4o to generate textual descriptors, capturing MNAR context without numeric estimation.

Feature Selection for Qualitative High-Dimensional Data. Feature selection—identifying the most relevant features from high-dimensional datasets—is critical for enhancing model performance, reducing computational complexity, and improving interpretability (Guyon and Elisseeff, 2003). In domains like LE data, with its rich, qualitative NC attributes, effective feature selection is paramount yet challenging. Traditional methods include statistical techniques like variance thresholding and correlation-based selection (Jain et al., 2000), alongside machine learning approaches such as feature importance from tree-based models (e.g., random forests) and regularization (e.g., LASSO) (Hastie et al., 2009). Recently, deep learning has introduced automated feature selection via attention mechanisms and feature masking, learning relevance within neural architectures (Ying et al., 2024; Cherepanova et al., 2023).

These methods, however, often rely on statistical or linear assumptions, which may fail to capture the nuanced, non-linear, and semantically driven relationships in qualitative LE data (e.g., self-reported engagement). For instance, correlation-based selection might overlook features with subtle contextual importance, while deep learning approaches typically require large, labeled datasets—scarce in educational settings. We propose a novel zero-shot feature selection approach using GPT-4o, leveraging its advanced reasoning and world knowledge to assess the semantic relevance of NC features for predicting student engagement. Unlike traditional and deep learning methods, our LLM-based strategy excels in high-dimensional, textual data, offering a scalable, context-aware alternative that aligns with LE data's subjective nature and enhances downstream forecasting.

While prior studies apply LLMs to time-series, impute missing values, or select features in structured data, none address the combined challenges of qualitative LE sequences, MNAR missingness, and engagement forecasting in education. Our three-tier framework—imputation, zero-shot feature selection, and fine-tuned forecasting—extends NLP techniques to this domain, emphasizing LLM-based feature selection as a key innovation, and outperforms traditional and generative baselines by embracing LE data's textual richness.

A.3 Limitations

While our LLM-based framework demonstrates the promise of LLMs in forecasting student engagement levels from qualitative longitudinal data, several limitations warrant consideration. First, our dataset, comprising 960 trajectories from students within a single university's introductory programming courses, is modest in size compared to typical NLP corpora. This scale might limit the robustness and the generalizability of our findings to diverse academic disciplines or educational settings. Furthermore, the domain-specific nature of student engagement and the verbalization of non-cognitive features might mean that the observed performance advantages, such as those of encoder-only LLMs (e.g., RoBERTa) over numeric baselines, may weaken with different distributions of non-cognitive features or variations in verbalization styles. This sensitivity to feature quantity and modality was also suggested by our ablation studies. The limited dataset size could also impact the model's ability to generalize to non-academic longitudinal experiential data, such as workplace engagement.

Second, our approach relies on GPT-4o for imputing missing data exhibiting MNAR patterns. While this zero-shot strategy effectively leverages the model's contextual understanding, it introduces a dependency on an external, proprietary model, potentially raising concerns about reproducibility and cost. Moreover, there is a risk that the textual patterns generated by GPT-4o could introduce a bias, potentially skewing downstream forecasting if these patterns do not perfectly align with the true underlying engagement signals. For our baseline comparisons, we employed zero-imputation for missing values in the numeric data. While straightforward, this method might undervalue the potential of traditional ML models, as more sophisticated imputation techniques like MICE (van Buuren and Groothuis-Oudshoorn, 2011) could potentially narrow the performance gap.

Finally, the computational demands of fine-tuning LLMs like Gemma2 9B and Mixtral 8x7B are substantial, requiring significant resources (e.g., 8× A40 GPUs for our experiments). This resource-intensive process could pose a barrier to scalability for broader datasets or limit the accessibility of our approach for researchers with constrained computational resources.

Future work could address these limitations by expanding the dataset to include larger, multi-domain samples, exploring the transferability of our framework to different types of longitudinal experiential data, investigating the use of lightweight LLMs to reduce computational costs, and developing or evaluating alternative imputation strategies that are less reliant on proprietary models.

References

Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, Erik Brynjolfsson, Shyamal Buch, Dallas Card, Rodrigo Castellon, Niladri Chatterji, Annie Chen, Kathleen Creel, Jared Quincy Davis, Dora Demszky, Chris Donahue, Moussa Doumbouya, Esin Durmus, Stefano Ermon, John Etchemendy, Kawin Ethayarajh, Li Fei-Fei, Chelsea Finn, Trevor Gale, Lauren Gillespie, Karan Goel, Noah Goodman, Shelby Grossman, Neel Guha, Tatsunori Hashimoto, Peter Henderson, John Hewitt, Daniel E. Ho, Jenny Hong, Kyle Hsu, Jing Huang, Thomas Icard, Saahil Jain, Dan Jurafsky, Pratyusha Kalluri, Siddharth Karamcheti, Geoff Keeling, Fereshte Khani, Omar Khattab, Pang Wei Koh, Mark Krass, Ranjay Krishna, Rohith Kuditipudi, Ananya Kumar, Faisal Ladhak, Mina Lee, Tony Lee, Jure Leskovec, Isabelle Levent, Xiang Lisa Li, Xuechen Li, Tengyu Ma, Ali Malik, Christopher D. Manning, Suvir Mirchandani, Eric Mitchell, Zanele Munyikwa, Suraj Nair, Avanika Narayan, Deepak Narayanan, Ben Newman, Allen Nie, Juan Carlos Niebles, Hamed Nilforoshan, Julian Nyarko, Giray Ogut, Laurel Orr, Isabel Papadimitriou, Joon Sung Park, Chris Piech, Eva Portelance, Christopher Potts, Aditi Raghunathan, Rob Reich, Hongyu Ren, Frieda Rong, Yusuf Roohani, Camilo Ruiz, Jack Ryan, Christopher Ré, Dorsa Sadigh, Shiori Sagawa, Keshav Santhanam, Andy Shih, Krishnan Srinivasan, Alex Tamkin, Rohan Taori, Armin W. Thomas, Florian Tramèr, Rose E. Wang, William Wang, Bohan Wu, Jiajun Wu, Yuhuai Wu, Sang Michael Xie, Michihiro Yasunaga, Jiaxuan You, Matei Zaharia, Michael Zhang, Tianyi Zhang, Xikun Zhang, Yuhui Zhang, Lucia Zheng, Kaitlyn Zhou, and Percy Liang. 2021. On the opportunities and risks of foundation models. *ArXiv*.

Defu Cao, Furong Jia, Sercan O Arik, Tomas Pfister, Yixiang Zheng, Wen Ye, and Yan Liu. 2023. Tempo: Prompt-based generative pre-trained transformer for time series forecasting. *Preprint*, arXiv:2310.04948.

Ching Chang, Wen-Chih Peng, and Tien-Fu Chen. 2023. Llm4ts: Two-stage fine-tuning for time-series forecasting with pre-trained llms. *Preprint*, arXiv:2308.08469.

Valeriia Cherepanova, Roman Levin, Gowthami Somepalli, Jonas Geiping, C. Bayan Bruss, Andrew Gordon Wilson, Tom Goldstein, and Micah Goldblum. 2023. A performance-driven benchmark for feature selection in tabular deep learning. *Preprint*, arXiv:2311.05877.

Jennifer Fredricks. 2014. *Eight Myths of Student Disengagement: Creating Classrooms of Deep Learning*. Corwin Press, Thousand Oaks, California.

Azul Garza, Cristian Challu, and Max Mergenthaler-Canseco. 2024. Timegpt-1. *Preprint*, arXiv:2310.03589.

Nate Gruver, Marc Finzi, Shikai Qiu, and Andrew Gordon Wilson. 2023. Large language models are zero-shot time series forecasters. *Preprint*, arXiv:2310.07820.

Manbir S Gulati and Paul F Roysdon. 2023. Tabmt: Generating tabular data with masked transformers. *Preprint*, arXiv:2312.06089.

Isabelle Guyon and André Elisseeff. 2003. An introduction to variable and feature selection. *J. Mach. Learn. Res.*, 3(null):1157–1182.

Trevor Hastie, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd edition. Springer.

A.K. Jain, R.P.W. Duin, and Jianchang Mao. 2000. Statistical pattern recognition: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):4–37.

Ming Jin, Shiyu Wang, Lintao Ma, Zhixuan Chu, James Y. Zhang, Xiaoming Shi, Pin-Yu Chen, Yuxuan Liang, Yuan-Fang Li, Shirui Pan, and Qingsong Wen. 2024. Time-LLM: Time series forecasting by reprogramming large language models. In *Proceedings of the International Conference on Learning Representations (ICLR)*.

Xiang Li, Xinning Zhu, Xiaoying Zhu, Yang Ji, and Xiaosheng Tang. 2020. Student Academic Performance Prediction Using Deep Multi-source Behavior Sequential Network. In *Advances in Knowledge Discovery and Data Mining*, Lecture Notes in Computer Science, pages 567–579, Cham. Springer International Publishing.

Xian Liu. 2016. Methods for handling missing data. In Xian Liu, editor, *Methods and Applications of Longitudinal Data Analysis*, chapter 14, pages 441–473. Academic Press.

Xin Liu, Daniel McDuff, Geza Kovacs, Isaac Galatzer-Levy, Jacob Sunshine, Jiening Zhan, Ming-Zher Poh, Shun Liao, Paolo Di Achille, and Shwetak Patel. 2023. Large language models are few-shot health learners. *Preprint*, arXiv:2305.15525.

Pierre-Alexandre Mattei and Jes Frellsen. 2019. Miwae: Deep generative modelling and imputation of incomplete data sets. In *Proceedings of the International Conference on Machine Learning*, pages 4413–4423. PMLR.

Boris Muzellec, Julie Josse, Claire Boyer, and Marco Cuturi. 2020. Missing data imputation using optimal transport. In *Proceedings of the 37th International Conference on Machine Learning*, ICML'20. JMLR.org.

D. B. Rubin. 1976. Inference and missing data. *Biometrika*, 63:581–592.

Gale M. Sinatra, Benjamin C. Heddy, and Doug Lombardi. 2015. The challenges of defining and measuring student engagement in science. *Educational Psychologist*, 50(1):1–13.

Chenxi Sun, Yaliang Li, Hongyan Li, and Shenda Hong. 2023. Test: Text prototype aligned embedding to activate llm's ability for time series. *Preprint*, arXiv:2308.08241.

S. van Buuren and K. Groothuis-Oudshoorn. 2011. mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software*, 45(3):1–67.

Stef Van Buuren, Jean-Paul Brand, Karin Groothuis-Oudshoorn, and Donald B. Rubin. 2006. Fully conditional specification in multivariate imputation. *Journal of Statistical Computation and Simulation*, 76(12):1049–1064.

Rui Wang, Fanglin Chen, Zhenyu Chen, Tianxing Li, Gabriella Harari, Stefanie Tignor, Xia Zhou, Dror Ben-Zeev, and Andrew T. Campbell. 2014. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '14, pages 3–14, New York, NY, USA. Association for Computing Machinery.

Weiqi Xu and Fan Ouyang. 2022. The application of AI technologies in STEM education: a systematic review from 2011 to 2021. *International Journal of STEM Education*, 9(1):59.

Hao Xue and Flora D. Salim. 2024. Promptcast: A new prompt-based learning paradigm for time series forecasting. *IEEE Trans. on Knowl. and Data Eng.*, 36(11):6851–6864.

Yazheng Yang, Yuqi Wang, Sankalok Sen, Lei Li, and Qi Liu. 2024. Unleashing the potential of large language models for predictive tabular tasks in data science. *Preprint*, arXiv:2403.20208.

Wangyang Ying, Dongjie Wang, Haifeng Chen, and Yanjie Fu. 2024. Feature selection as deep sequential generative learning. *Preprint*, arXiv:2403.03838.

Jinsung Yoon, James Jordon, and Mihaela van der Schaar. 2018. Gain: Missing data imputation using generative adversarial nets. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80, pages 5689–5698. PMLR.

He Zhao, Ke Sun, Amir Dezfouli, and Edwin V Bonilla. 2023. Transformed distribution matching for missing value imputation. In *International Conference on Machine Learning*, pages 42159–42186. PMLR.