Variational Autoencoders for Anomaly Detection in Choice Models

Lorenzo Munoz*[1], Stephane Hess[2], Thomas O. Hancock[3], and Geroges Sfeir[4]

[1]PhD Student, Institute for Transport Studies, University of Leeds, UK
[2]Professor of Choice Modelling, Institute for Transport Studies, University of Leeds, UK
[3]Senior Research Fellow, Institute for Transport Studies, University of Leeds, UK
[4]Research Fellow, Institute for Transport Studies, University of Leeds, UK

Short summary

Generative machine learning is increasingly gaining attention in the choice modelling field. By learning the underlying probability distribution of training data to synthesise new observations, these models offer a powerful framework for anomaly detection, i.e., identifying abnormal or atypical patterns in the input data. We introduce the use of variational autoencoders for anomaly detection. This is a generative model with two components, an encoder that compresses the input data into a low-dimensional latent representation, and a decoder that reconstructs the input from this latent representation to produce new data. By comparing the original data to reconstructions generated from samples of the latent space, reconstruction errors can be computed and used to flag atypical observations. We demonstrate that anomalies have a substantial impact on choice model outputs, such as values of time. This framework enables choice modellers to better detect outliers in high-dimensional datasets, thereby supporting data cleaning and model specification.
Keywords: Anomaly detection, generative machine learning, outliers, variational autoencoders.

1 Introduction

In choice modelling (CM), understanding and generalising the underlying structure of choice data is essential for behavioural analysis. Motivated by the substantial success of machine learning (ML) techniques across a wide range of disciplines, researchers have increasingly explored their application within CM. To date, most applications of ML in CM have focused on classification-oriented approaches, primarily aimed at improving predictive performance. Other applications include analysing large datasets, incorporating different types of data (e.g., text and images), assisting model specification, and capturing more complex patterns in explanatory variables, such as non-linearities and interactions (Sun et al., 2018; Han, 2019; Wong & Farooq, 2021; van Cranenburgh et al., 2022; S. Wang et al., 2020; Ortelli et al., 2021).

However, researchers in other fields—e.g., healthcare, software engineering, and finance—are paying increasing attention to generative models (Q. Wang et al., 2024). Generative ML refers to methods that learn the underlying probability distribution of training data to generate new, similar data samples (Harshvardhan et al., 2020; Sengar et al., 2024). Unlike classification algorithms, which focus on learning decision boundaries between classes, generative models aim to capture the latent structure of the data, enabling smooth interpolation within the latent space and supporting tasks that require flexible modelling of complex dependencies (Glorot & Bengio, 2010). By learning rich representations of behavioural patterns, generative models have the potential to broaden the analytical scope of CM and provide deeper insights into choice behaviour (Wong, 2019).

A widely used generative ML application is anomaly detection (AD), which refers to the task of identifying abnormal or atypical patterns within a given set of input data (Nguyen et al., 2024). Such deviations can yield valuable insights into the underlying mechanisms and structural irregularities (An & Cho, 2015). This process is challenging with marginal or distance-based methods (Munir et al., 2019), yet crucial because anomalies in data translate to significant, and often critical, actionable information in a wide variety of application domains (Chandola et al., 2009). AD has been successfully applied across diverse fields in ML, but remains largely unexplored within the CM literature.

AD identifies abnormal patterns through the reconstruction capacity of deep generative models, typically Artificial Neural Networks (ANNs). The aim is to reconstruct input data, and discrepancies between the original and reconstructed observations serve as anomaly scores (Nguyen et al.,

2024). Thus, the first step is to reconstruct the data using dimension reduction methods, where the input is reconstructed from a low-dimensional latent representation learned by the model. The difference between the original observation and its reconstruction is then computed as the reconstruction error, which can serve as an anomaly score. Among reconstruction-based approaches, variational autoencoders (VAEs) have gained particular prominence, as they learn compact latent representations of high-dimensional data and reconstruct inputs through deep generative networks, making them well suited for detecting unusual attribute configurations.
Anomalies can be inconsistent with economic theory Hultkrantz & Mortazavi (2001), cause substantial changes in model parameter estimates (McFadden, 1981), and mislead applications such as market simulations (Ku et al., 2017). This motivates the use of AD as a diagnostic tool in CM. This study contributes to the literature by introducing a generative ML framework for detecting anomalous observations in choice data, using reconstruction-based methods to flag atypical observations.

2 Methodology

VAEs (Kingma & Welling, 2013) are a class of generative models that combine deep learning with probabilistic graphical models to learn complex data distributions, while explicitly inferring structured latent representations from observed samples. This explicit probabilistic modelling enables meaningful interpretation of latent factors underlying decision-making processes (Haghi et al., 2026). VAEs consist of two components: an encoder that compresses the input data x (comprising k observations) into a low-dimensional latent representation z, and a decoder that reconstructs the input from this representation to produce synthetic data $\hat{x}$.
We used a fully connected ANN architecture, with multiple layers of artificial neurons. ANNs are highly flexible, non-linear, and complex approximations, which are the most common choice for approximating any mapping for both encoder and decoder (Borysov et al., 2019). The neuron is defined as the weighted sum of its inputs, followed by a non-linear activation function. Thus, the output layer l of an ANN is computed as $y_l = f(W_l y_{l-1} + b_l)$, where W_l and b_l are the weight matrix and the bias vector of layer l, and f is a non-linear activation function. We define the collection of all weights and biases as ϕ and θ for the encoder (p_ϕ) and the decoder (q_θ), respectively.
The VAE is trained by minimising the loss function in Equation 1. We differentiate between categorical and numerical variables based on the specification by Borysov et al. (2019). In practice, this means that numerical variables are reconstructed using a squared-error loss under a Gaussian assumption, while categorical variables are reconstructed using a cross-entropy loss based on a one-hot encoding, i.e., representing each category using a binary variable. The loss function combines a reconstruction term and a regularisation term enforcing proximity of the latent distribution to the prior $\mathcal{N}(0, I_{D_z})$, which is a standard multivariate normal distribution with zero mean and unit variance in each latent dimension, used to regularise the latent space. Hence, D_z is the dimension of the latent space z, and I_{D_z} is an identity matrix of size $D_z \times D_z$. This formulation allows the model to balance data fidelity and latent-space regularisation within a unified optimisation framework (Daoud et al., 2025).

$$\mathcal{L}_{\phi,\theta} = \sum_{k=1}^{N} ||x_k - \hat{x}_k||_{num} + ||x_k - \hat{x}_k||_{cat} + \beta D_{KL}[\mathcal{N}(\mu_k, \sigma_k)||\mathcal{N}(0, I_{D_z})], \tag{1}$$

where

$$||x_k - \hat{x}_k||_{num} = \frac{1}{2} \sum_{i \in num} (x_{ki} - \hat{x}_{ki})^2, \tag{2}$$

is the mean square loss for the reconstruction of numerical variables,

$$||x_k - \hat{x}_k||_{cat} = -\sum_{i \in cat} \sum_{j=1}^{D_i} x_{kij} log(\hat{x}_{kij}), \tag{3}$$

is the cross-entropy loss for the reconstruction of categorical variables, assuming a one-hot representation, and

$$D_{KL}[\mathcal{N}(\mu_k, \sigma_k)||\mathcal{N}(0, I_{D_z})] = -\frac{1}{2} \sum_{i=1}^{D_z} (1 + log(\sigma_{ki}) - \mu_{ki}^2 - \sigma_{ki}), \tag{4}$$

is the regularisation term represented by the Kullback-Leibler divergence (Kullback, 1951) between the D_z-dimensional latent variable distribution parametrised by the encoder output (with mean μ_k and standard deviation σ_k) and the standard Gaussian prior $\mathcal{N}(0, I_{D_z})$. Therefore, this term penalises the deviation from the predefined distribution. β is a hyperparameter representing a weighting coefficient between the reconstruction losses and the regularisation (Higgins et al., 2017). Thus, β is not estimated but selected through hyperparameter search. Figure 1 illustrates the VAE architecture.

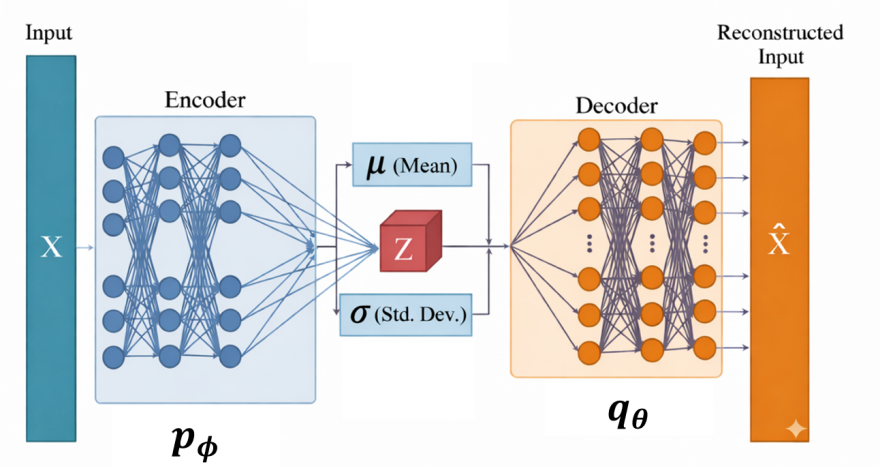


Figure 1: VAEs structure for anomaly detection.

During training, the encoder takes an observation x_k as input and maps it into two vectors μ_k (the mean of the latent variable z for the input data k) and σ_k (the corresponding standard deviation) of size D_z as shown in Equation 5. Together, these parameters define the multivariate normal distribution for the latent variable, with sampling implemented via the reparametrisation trick in Equation 6, where ε_k is a random noise vector drawn from a standard normal distribution and $\odot$ denotes element-wise multiplication. The reparametrisation trick expresses the latent variable as the sum of the mean and a scaled noise term, and enables efficient gradient-based optimisation.

$$\mu_k = f_\mu(x_k; \phi), \qquad \log \sigma_k = f_\sigma(x_k; \phi), \tag{5}$$

$$z_k = \mu_k + \sigma_k \odot \varepsilon_k, \qquad \varepsilon_k \sim \mathcal{N}(0, I_{D_z}), \tag{6}$$

The decoder reconstructs the input by mapping the latent space z back into the original feature space:

$$\hat{x}_k = g(z_k; \theta). \tag{7}$$

Because the decoder models the joint distribution of attributes, the VAE identifies anomalies that arise from subtle cross-attribute interactions that would be difficult to detect with traditional methods. The dimensionality of the latent space controls the level of compression: using a lower-dimensional latent space forces the model to learn only the most essential regularities in the data. Namely, VAE models the distribution parameters ϕ and θ, not the attributes $\hat{x}$ themselves. Training therefore proceeds until these parameters converge and the loss function is minimised.
The terms in Equation 2 and Equation 3 are denominated reconstruction errors. Large reconstruction errors indicate that the observation deviates substantially from the patterns captured by the VAE, and are therefore interpreted as anomalies. In a CM context, a high reconstruction error suggests that the attribute profile of the choice task is atypical relative to the rest of the dataset. This may correspond to data errors, rare or extreme attribute combinations, or inconsistencies in reported trip characteristics. Additionally, the values of the reconstruction error depend on the scale of the attributes; therefore, continuous variables are standardised prior to training to avoid scale biases in the reconstruction errors.
A key distinction between VAEs and standard autoencoders is that VAEs learn an explicit probabilistic representation of the data. Rather than mapping each observation to a single latent point, the encoder estimates a distribution over latent variables for each observation. As a result, the

model learns a full generative process that allows new latent samples to be drawn from the learned distributions and decoded back into the data space. This property enables sampling-based analysis after training, which is not possible with traditional autoencoders.
Thus, it is important to distinguish between reconstruction errors obtained directly from the training loss and those derived through sampling from the learned distributions. The reconstruction loss optimised during training corresponds to a single reconstruction per observation and, by itself, would not require a probabilistic model. This is also the case for traditional autoencoders, which map each observation to a single latent point and produce a single deterministic reconstruction, without modelling a distribution over latent variables. In contrast, because the VAE learns a distribution over latent variables, multiple latent samples can be drawn for the same observation and decoded to obtain multiple reconstructions. This allows reconstruction errors to be computed across samples, thereby reflecting uncertainty in the latent representation.

3 Results and discussion

We analysed the Swiss route dataset (Vrtic & Axhausen, 2002), available in *Apollo* (Hess & Palma, 2019). This dataset contains 388 individuals who participated in a stated choice survey regarding the introduction of a tilting train in Switzerland, yielding 3,492 observations. Each respondent was first presented with a complete description of their reported trip and then faced eight choice tasks between two route alternatives pivoted on their reported trip. Each alternative is characterised by travel time, cost, headway, and number of interchanges. We summarise the reconstructed variables in Table 1.

Table 1: Summary of the reconstructed variables.

Variable	Type	Description
tt1	Continuous	Travel time (mins) for route 1.
tc1	Continuous	Travel cost (CHF) for route 1.
hw1	Continuous	Headway (mins) for route 1.
ch1	Discrete	Number of interchanges for route 1.
tt2	Continuous	Travel time (mins) for route 2.
tc2	Continuous	Travel cost (CHF) for route 2.
hw2	Continuous	Headway (mins) for route 2.
ch2	Discrete	Number of interchanges for route 2.
hh_inc_abs	Continuous	Household income (CHF/year).
car_availability	Discrete	Car availability, 1 if yes, 0 if no.
commute	Discrete	Dummy variable for commute trips, 1 if yes, 0 if no.
shopping	Discrete	Dummy variable for shopping trips, 1 if yes, 0 if no.
business	Discrete	Dummy variable for business trips, 1 if yes, 0 if no.
leisure	Discrete	Dummy variable for leisure trips, 1 if yes, 0 if no.

We use 80% of the sample for training (with a 10% validation split), and 20% for testing. A hyperparameter search was conducted to optimise the autoencoder architecture, tuning the number of neurons in the hidden layer, D_z, the learning rate, β, batch size, and weight decay. We also fixed the anomaly threshold as the 95th percentile of the reconstruction errors, which is a standard practice. However, this is a modelling decision and can be adjusted depending on the application, the dataset, and the analyst's tolerance for false positives.
VAE performance was also evaluated in terms of convergence, reconstruction quality, and preservation of attribute relationships. Convergence was examined through the loss functions and their components; reconstruction quality through straightforward comparisons of original and reconstructed distributions; and correlations through Cramér's V (Cramér, 1999) for categorical variables, and through Pearson's correlation coefficient for continuous variables (Jackson et al., 2026). For each observation, we sampled 100 latent reconstructions as an illustrative first step. We experimented with larger numbers of samples and observed stable reconstruction-error distributions, indicating convergence of the results. Reconstruction errors were computed by comparing each reconstruction with the original data and then averaged across samples to obtain a single value per observation. As discussed earlier, the anomaly threshold was set at the 95th percentile of

the reconstruction-error distribution, resulting in the identification of 175 anomalous observations. The distribution of reconstruction errors is shown in Figure 2.

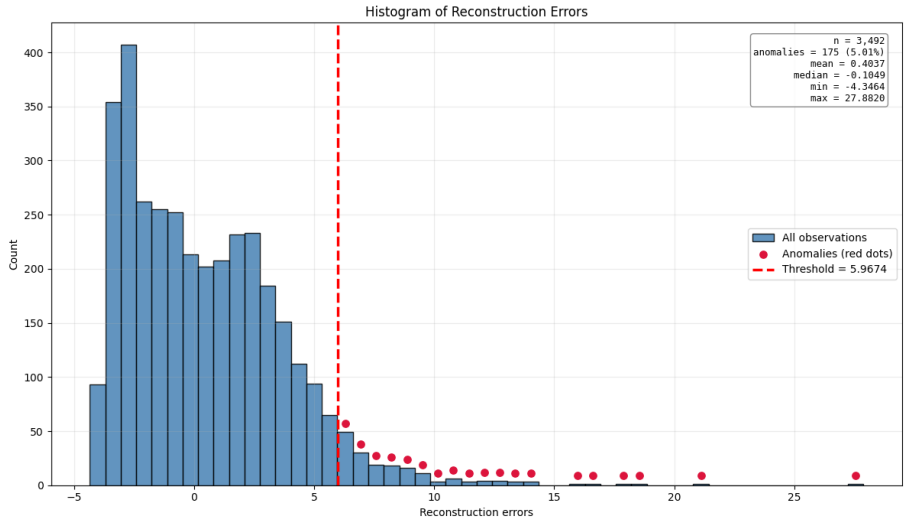


Figure 2: Histogram of reconstruction errors.

The ten observations with the largest reconstruction errors are presented in Table 2, along with the corresponding trip attributes. These observations exhibit substantially higher reconstruction errors than the rest of the sample, suggesting that they may exert disproportionate influence in CM applications.

Table 2: Observations with the highest reconstruction errors and associated attributes.

User ID	**Trip ID**	**tt1**	**tc1**	**tt2**	**tc2**	**hw1**	**hw2**	**ch1**	**ch2**	**Rec. error**
16117	1183	144	201	169	155	15	60	0	0	27.88
23321	3060	21	34	30	26	15	30	1	2	21.20
16117	1188	145	206	101	268	30	15	2	1	18.86
19691	2088	13	27	19	20	15	30	1	2	18.21
19691	2087	20	18	15	20	60	30	2	2	16.86
20352	2286	15	24	14	31	15	15	1	2	16.15
14074	453	199	88	284	77	15	60	0	0	14.30
23321	3053	32	30	43	23	60	15	1	0	14.15
14787	692	117	118	131	87	15	60	0	1	13.93
16617	1350	269	72	385	62	30	60	2	0	13.65

We first estimate a multinomial logit model using the full dataset (hereafter Baseline Model), in which continuous attributes enter the utility function as level-of-service variables with associated coefficients β, while discrete attributes are captured through alternative-specific constants with coefficients α. We then re-estimate the model after excluding the ten observations with the highest reconstruction errors identified in Table 2 (hereafter Filtered Model). Although the exclusion of a fixed number of observations is inherently arbitrary, this step serves as a benchmark exercise to illustrate the potential impact of anomalous observations on model estimates and behavioural interpretation. Table 3 reports the parameter estimates and corresponding robust t-ratios for both models, along with the percentage differences between estimates. The comparison shows that excluding the most anomalous observations leads to substantial changes in the magnitude of parameter estimates, specifically in the marginal utility of travel cost, travel time and in the shopping and leisure specific constants.

We compare the models in terms of key economic indicators used for transport project appraisal, such as the value of travel time (VTT) and value of headway time (VHT), in Table 4. Excluding these observations results in a significant reduction in the value of travel time (5.6%) and in the value of headway time (8.4%). These results illustrate how a small number of atypical observations

Table 3: Parameter estimates and robust t-ratio for Baseline and Filtered models.

Parameter	**Estimate (Bas.)**	**t-ratio (Bas.)**	**Estimate (Filt.)**	**t-ratio (Filt.)**	**Diff. (%)**
β_{tt}	-0.061	-8.94	-0.062	-9.59	+3.5
β_{tc}	-0.133	-5.66	-0.142	-6.65	+7.0
β_{hw}	-0.038	-16.16	-0.038	-16.17	+0.4
β_{ch}	-1.155	-18.76	-1.159	-18.69	+0.4
α_{car_av}	-0.246	-2.54	-0.247	-2.57	+0.6
$\alpha_{commute}$	0.115	1.38	0.116	1.39	+0.9
$\alpha_{shopping}$	0.144	0.99	0.164	1.16	+13.5
$\alpha_{business}$	-0.132	-0.85	-0.136	-0.88	+3.2
$\alpha_{leisure}$	0.084	1.14	0.093	1.27	+10.5

can materially affect key model outputs. Importantly, this does not suggest that such observations should be removed mechanically; rather, it highlights the value of anomaly detection as a diagnostic tool for identifying influential cases that require closer inspection in CM.

Table 4: Value of travel time and value of headway time for Baseline and Filtered models.

Metric	**Baseline**	**Filtered**	**Diff (%)**
VTT (CHF/min)	0.451	0.426	-5.6
VHT (CHF/min)	0.282	0.258	-8.4

4 Conclusions

We detected anomalies in the stated preference dataset, and the extension of this framework to revealed preference data is ongoing work. The appropriate treatment of anomalies depends heavily on the case study and on the nature of the anomalous observations, which may range from data errors to anomalous attribute combinations. Our findings illustrate how AD can serve as a diagnostic tool to flag observations or individuals that exert disproportionate influence on model estimates. The proposed framework provides a scalable benchmark for rapidly identifying such cases in high-dimensional datasets, such as revealed preference surveys and GPS trajectories.
Future work will be using conditional VAEs for AD. In this setting, the model is trained by conditioning explicitly on the observed attributes, while the choice is treated as the variable to be reconstructed. This allows the model to assess whether an observed choice is plausible given the attribute configuration, rather than merely identifying unusual combinations of attributes. This provides a framework to directly flag observations in which the choice appears inconsistent with the explanatory variables. In addition, an important extension of this framework concerns the use of reconstruction probabilities as a likelihood-based anomaly score. While the present analysis focuses on reconstruction errors, reconstruction probabilities exploit the full probabilistic structure of variational autoencoders by assessing how likely an observation is under the learned data-generating process.

Acknowledgements

The authors acknowledge the financial support by the European Research Council through the advanced grant 101020940-SYNERGY.

References

An, J., & Cho, S. (2015). Variational autoencoder based anomaly detection using reconstruction probability. *Special lecture on IE*, *2*(1), 1–18.

Borysov, S. S., Rich, J., & Pereira, F. C. (2019). How to generate micro-agents? a deep generative modeling approach to population synthesis. *Transportation Research Part C: Emerging Technologies*, *106*, 73–97.

Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM computing surveys (CSUR)*, *41*(3), 1–58.

Cramér, H. (1999). *Mathematical methods of statistics* (Vol. 9). Princeton university press.

Daoud, A. M., Elkomy, O. M., Khedr, W. I., & Hosny, K. M. (2025). A new hybrid model for improving outlier detection using combined autoencoder and variational autoencoder. *Scientific Reports*.

Glorot, X., & Bengio, Y. (2010, 13–15 May). Understanding the difficulty of training deep feedforward neural networks. In Y. W. Teh & M. Titterington (Eds.), *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (Vol. 9, pp. 249–256). Chia Laguna Resort, Sardinia, Italy: PMLR. Retrieved from `https://proceedings.mlr.press/v9/glorot10a.html`

Haghi, H., Patterson, Z., & Farooq, B. (2026). Choice set generation in work destination choice modeling with variational autoencoders. *Transportation Research Record*, 03611981251409202.

Han, Y. (2019). *Neural-embedded discrete choice models* (Unpublished doctoral dissertation). Massachusetts Institute of Technology.

Harshvardhan, G., Mahendra, K. G., Manjusha, P., & Siddharth, S. R. (2020). A comprehensive survey and analysis of generative models in machine learning. *Computer Science Review*, *38*, 100285. Retrieved from `https://www.sciencedirect.com/science/article/pii/S1574013720303853` doi: https://doi.org/10.1016/j.cosrev.2020.100285

Hess, S., & Palma, D. (2019). Apollo: A flexible, powerful and customisable freeware package for choice model estimation and application. *Journal of choice modelling*, *32*, 100170.

Higgins, I., Matthey, L., Pal, A., Burgess, C., Glorot, X., Botvinick, M., ... Lerchner, A. (2017). January 24-26). beta-vae: Learning basic visual concepts with a constrained variational framework. proceedings of the international conference on learning representations, toulon, france.

Hultkrantz, L., & Mortazavi, R. (2001). Anomalies in the value of travel-time changes. *Journal of Transport Economics and Policy (JTEP)*, *35*(2), 285–299.

Jackson, N. J., Espinosa-Dice, N., Yan, C., & Malin, B. A. (2026). Reward-guided generation improves the scientific utility of synthetic biomedical data. *medRxiv*, 2026–03.

Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.

Ku, Y.-C., Chiang, T.-F., & Chang, S.-M. (2017, March). Is what you choose what you want?—outlier detection in choice-based conjoint analysis. *Marketing Letters*, *28*(1), 29–42. doi: 10.1007/s11002-015-9389-3

Kullback, S. (1951). Kullback-leibler divergence. *Encyclopedia of Machine Learning*, 581–583.

McFadden, D. (1981). Econometric models of probabilistic choice. *Structural analysis of discrete data with econometric applications*, *198272*.

Munir, M., Chattha, M. A., Dengel, A., & Ahmed, S. (2019). A comparative analysis of traditional and deep learning-based anomaly detection methods for streaming data. In *2019 18th ieee international conference on machine learning and applications (icmla)* (pp. 561–566).

Nguyen, H. H., Nguyen, C. N., Dao, X. T., Duong, Q. T., Kim, D. P. T., & Pham, M.-T. (2024). *Variational autoencoder for anomaly detection: A comparative study.* Retrieved from `https://arxiv.org/abs/2408.13561`

Ortelli, N., Hillel, T., Pereira, F. C., de Lapparent, M., & Bierlaire, M. (2021). Assisted specification of discrete choice models. *Journal of Choice Modelling*, *39*, 100285. Retrieved from `https://www.sciencedirect.com/science/article/pii/S175553452100018X` doi: https://doi.org/10.1016/j.jocm.2021.100285

Sengar, S. S., et al. (2024). Generative artificial intelligence: A systematic review and future directions. *Multimedia Tools and Applications*. ("These models acquire an understanding of the underlying patterns ... generating novel data with akin characteristics.")

Sun, Y., Jiang, Z., Gu, J., Zhou, M., Li, Y., & Zhang, L. (2018). Analyzing high speed rail passengers' train choices based on new online booking data in china. *Transportation Research Part C: Emerging Technologies*, *97*, 96-113. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0968090X18301694` doi: https://doi.org/10.1016/j.trc.2018.10.015

van Cranenburgh, S., Wang, S., Vij, A., Pereira, F., & Walker, J. (2022). Choice modelling in the age of machine learning - discussion paper. *Journal of Choice Modelling*, *42*, 100340. Retrieved from `https://www.sciencedirect.com/science/article/pii/S1755534521000725` doi: https://doi.org/10.1016/j.jocm.2021.100340

Vrtic, M., & Axhausen, K. W. (2002). The impact of tilting trains in switzerland: A route choice model of regional-and long distance public transport trips. *Arbeitsberichte Verkehrs-und Raumplanung*, *128*.

Wang, Q., Wang, S., Zheng, Y., Lin, H., Zhang, X., Zhao, J., & Walker, J. (2024). Deep hybrid model with satellite imagery: How to combine demand modeling and computer vision for travel behavior analysis? *Transportation Research Part B: Methodological*, *179*, 102869. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0191261523001947` doi: https://doi.org/10.1016/j.trb.2023.102869

Wang, S., Wang, Q., & Zhao, J. (2020). Deep neural networks for choice analysis: Extracting complete economic information for interpretation. *Transportation Research Part C: Emerging Technologies*, *118*, 102701. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0968090X20306161` doi: https://doi.org/10.1016/j.trc.2020.102701

Wong, M. (2019). *Generative modelling and machine learning methods for travel behaviour analysis* (Unpublished doctoral dissertation). Toronto Metropolitan University.

Wong, M., & Farooq, B. (2021). Reslogit: A residual neural network logit model for data-driven choice modelling. *Transportation Research Part C: Emerging Technologies*, *126*, 103050. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0968090X21000802` doi: https://doi.org/10.1016/j.trc.2021.103050