

Behavioural Synthetic Traffic Generation for 5G and 6G Network Validation: A Hybrid GAN-VAE Framework Extended by a Behavioural Codebook Hybrid

Faris Muhammad

VIAMI Solutions UK, CTO Office, Stevenage, United Kingdom

Corresponding author: Faris Muhammad

Abstract—Validation of adaptive 5G and 6G networks requires synthetic traffic that reproduces realistic subscriber behaviour, telecom protocol validity, and target operational key performance indicators (KPIs) while preserving privacy. Conventional rule-based generators and aggregate statistical models capture protocol exchanges or workload averages, but fail to model how users and devices adapt under context, mobility, congestion, and policy. We present BSTG, a behaviour-conditioned, constraint-aware synthetic traffic framework that learns a latent behavioural persona manifold from anonymised traces using a hybrid GAN-VAE, projects generated traces onto a telecom-valid set via a constraint engine, and closes generation against target KPIs through a closed-loop digital twin. We further introduce a next-generation Behavioural Codebook Hybrid (BCH) variant that replaces the monolithic generator with a four-stage pipeline: per-persona vector-quantised autoencoders (VQ-VAE) learning discrete behavioural codebooks, an autoregressive transformer over codewords, a Wasserstein GAN closure, and constraint-aware token rejection during decoding. On a 5G quasi-real corpus, BCH raises aggregate corpus fidelity from 0.79 (GAN-VAE baseline) to 0.96 and resolves minority-class collapse on machine-type personas without ad-hoc normalisation. On a 6G IMT-2030 preview corpus, BCH remains in the realistic band on all five persona classes.

Index Terms—synthetic traffic generation, 5G, 6G, behavioural modelling, hybrid GAN-VAE, vector quantisation, autoregressive transformer, network digital twin, constraint-aware generation, KPI fidelity.

I. INTRODUCTION

Modern 5G cores and emerging 6G architectures are increasingly data-driven and adaptive. Functions such as the Network Data Analytics Function (NWDAF) [1], the Policy Control Function (PCF), and slice management consume rich behavioural inputs. Validating such functions and their AI components demands synthetic traffic that goes beyond protocol-level fidelity, capturing how users and machines adapt under congestion, mobility, and policy.

Conventional approaches each fall short. Commercial rule-based platforms [2], [3] expose parameterised call models with limited behavioural adaptability. Statistical models such as Markov-modulated Poisson processes [4] and ON-OFF models [5] capture aggregate volumes but treat sessions as memoryless. Trace replay preserves real protocol state but raises privacy concerns and is anchored to a particular capture window. Recent generative work on time-series traffic [6], [7] focuses on packet headers or marginal distributions rather than behaviourally coherent multi-session traces.

We present BSTG, an end-to-end framework for behaviour-conditioned, constraint-aware synthetic traffic. The framework learns a latent behavioural manifold using a hybrid VAE-

GAN, projects generated traces onto a telecom-valid set via a constraint engine, and aligns generated traffic with target operational KPIs through a closed-loop digital twin. We further introduce a Behavioural Codebook Hybrid (BCH) variant that replaces the monolithic generator with a four-stage pipeline organised around per-persona behavioural vocabularies learned by vector-quantised autoencoders [8]. The main contributions are:

- A unified hybrid GAN-VAE framework that learns a controllable behavioural persona manifold and synthesises scenario-conditioned, multi-step telecom traffic.
- A constraint validation and repair layer that projects generated traces into a telecom-feasible set (protocol ordering, 5QI admissibility, handover feasibility, slice consistency).
- A KPI fidelity loop that closes generation against target outcomes via a simulator, surrogate model, or digital twin.
- A next-generation BCH variant with per-persona VQ-VAE codebooks, an autoregressive transformer over learned behavioural primitives, a Wasserstein GAN closure, and constraint-aware token rejection during decoding.
- Empirical evaluation on 5G and 6G IMT-2030 corpora showing that BCH raises aggregate corpus fidelity from 0.79 to 0.96 and remains in the realistic band on all five 6G persona classes.

II. RELATED WORK

A. Rule-based and scripted generators

Commercial test traffic platforms expose parameterised call models and rule-based scripts that reproduce protocol message exchanges with tunable rates and mixes [2], [3]. These platforms are aligned to 3GPP UE conformance specifications and are essential for radio and core interoperability, but their behavioural adaptability is limited and their scenario libraries do not generalise to context-dependent or policy-driven behavioural shifts.

B. Statistical and stochastic models

Classical traffic models use Poisson processes, Markov-modulated Poisson processes [4], ON-OFF models [5], and self-similar or heavy-tailed extensions to match aggregate volumes and inter-arrival distributions. These models are analytically tractable but treat sessions as memoryless or weakly stateful and do not represent goal-directed multi-session user behaviour, multi-application interleaving, or context-conditioned shifts.

C. Trace capture and replay

Replay tools deterministically reproduce captured packet streams or session logs. They preserve real protocol state but are anchored to a particular capture window, raise privacy and re-identification concerns, and provide no behavioural extrapolation outside the captured distribution.

D. Generative models for traffic and time series

Generative adversarial networks (GANs), variational autoencoders (VAEs), and more recently diffusion models have been applied to network telemetry and packet headers [6], [7]. Most published work targets marginal distribution fidelity at the packet or flow level and does not jointly model behavioural state, protocol validity, and operational KPIs. Pure-GAN setups exhibit mode collapse on minority subscriber classes, while pure-VAE setups over-smooth burst structure.

E. Digital twins and privacy-preserving synthetic data

Digital twins of radio access and core networks provide simulation environments for what-if analysis [13]. Most published twins emphasise physical-layer fidelity and use simplified or replayed workloads rather than learned behavioural traces. Privacy-preserving training via differentially private stochastic gradient descent (DP-SGD) [10] and membership-inference auditing [15] is necessary when learning from subscriber data.

III. FRAMEWORK ARCHITECTURE

Fig. 1 shows the BSTG architecture, organised into six functional modules: data ingestion and behaviour abstraction, representation learning, conditioned generation, constraint validation and repair, KPI fidelity evaluation, and executable scenario compilation. The ingestion module receives anonymised network logs, charging records, mobility events, and KPI observations, converting raw telemetry into structured event sequences, dwell-time series, and contextual windows under pseudonymisation and k-anonymity controls. The representation learner maps traces and context into a latent behavioural space using a conditional VAE.

The conditioned generator produces synthetic traces given a sampled latent vector, a scenario context, and an optional noise input; the preferred embodiment is a hybrid VAE-GAN whose recurrent generator carries a temporal-coherence objective that preserves event ordering, dwell times, and state-dependent transitions. The constraint engine evaluates seven protocol-validity constraints (throughput ceiling, handover feasibility, 5QI admissibility, protocol ordering, slice consistency, dwell-time bounds, and policy trigger rate); repairable violations are corrected by a projection operator and unrepairable ones trigger regeneration. The KPI fidelity module compares induced KPIs against a target envelope and feeds residual mismatch back to the generator parameters. Validated traces are emitted as executable scenarios for downstream test platforms.

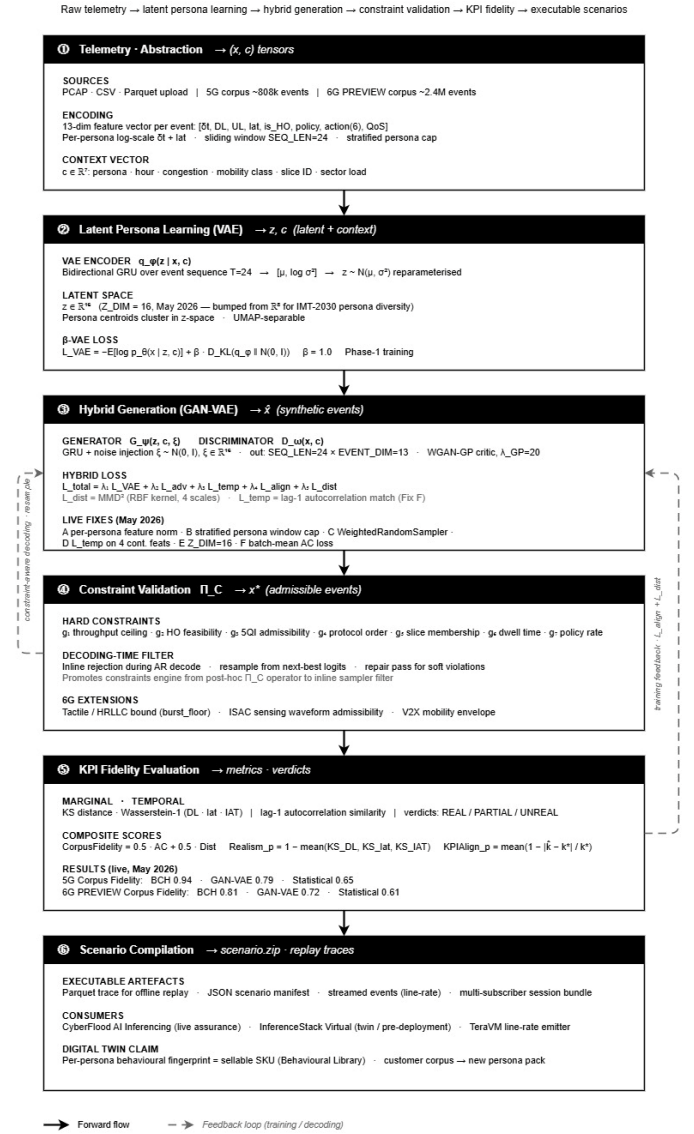


Fig. 1. BSTG framework architecture: from raw telemetry, through latent persona learning and hybrid generation, to constraint validation, KPI fidelity evaluation, and executable scenario compilation.

IV. MATHEMATICAL FORMULATION

Let the training corpus be $D = \{(x_i, c_i, k_i)\}$, where x_i is an event trace, c_i a context vector (time, location, congestion), and k_i an observed KPI vector. A conditional VAE provides a smooth, controllable latent persona space:

$$q_\phi(z | x, c) = N(\mu_\phi(x, c), \Sigma_\phi(x, c)) \quad (1)$$

$$\mathcal{L}_{VAE} = -\mathbb{E}_q[\log p_\theta(x | z, c)] + \beta D_{KL}(q_\phi || p(z)) \quad (2)$$

A conditional adversarial component sharpens higher-order structure that pure reconstruction objectives tend to blur. With generator G_ψ and critic D_ω , the Wasserstein form with gradient penalty [9] reads:

$$\mathcal{L}_{WGAN-GP} = \mathbb{E}[D_\omega(x, c)] - \mathbb{E}[D_\omega(G_\psi(z, c, \xi), c)] - \lambda \mathbb{E}[\|\nabla D_\omega(\hat{x}, c)\|_2 - 1]^2 \quad (3)$$

Generated traces must be telecom-valid, not merely realistic. Letting $g_j(x) \leq 0$ denote $J = 7$ validity constraints

(throughput ceiling, handover feasibility, 5QI admissibility, protocol ordering, slice consistency, dwell-time bounds, policy trigger rate), the constraint loss and projection-based repair operator are:

$$L_{\text{valid}} = \sum_j \max(0, g_j(\tilde{x}))^2 \quad (4)$$

$$\tilde{x}_{\text{feasible}} = \Pi_C(\tilde{x}) = \arg \min_{\{x' \in C\}} \|x' - \tilde{x}\|_2^2 \quad (5)$$

Operational realism is measured by the network outcomes induced by generated traffic. An outcome model F estimates KPIs; in the reference configuration evaluated here, KPI alignment is reported as the mean per-KPI relative agreement to a target envelope:

$$A_{\text{KPI}} = (1/J) \sum_j (1 - |F_j(\tilde{x}_{\text{feasible}}) - k_j^*| / k_j^*) \quad (6)$$

The combined generator-side objective sums weighted contributions from VAE reconstruction L_{VAE} , adversarial realism L_{adv} , temporal coherence L_{temp} (lag-1 autocorrelation match across continuous channels), latent-space alignment $L_{\text{align}} = \|\mathbf{R}(\tilde{x}, c) - \mathbf{z}\|^2$ (which ties the generator output back to the encoder's latent space and stabilises hybrid training), and a distribution-matching term L_{MMD} based on maximum mean discrepancy on inter-arrival time, downlink throughput, and latency channels: $L_{\text{total}} = \lambda_1 L_{\text{VAE}} + \lambda_2 L_{\text{adv}} + \lambda_3 L_{\text{temp}} + \lambda_4 L_{\text{align}} + \lambda_5 L_{\text{MMD}}$, with $(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5) = (1.0, 0.5, 0.3, 0.2, 0.3)$ in the reference configuration. Validity (Eq. 4) and KPI alignment (Eq. 6) are evaluated at decoding / evaluation time rather than as training-time penalties; a Wasserstein-2 KPI alignment term $W_2(P_{\text{F}}(\tilde{x}_{\text{feasible}}), P_k)$ and differentially private training via DP-SGD [10] are planned extensions and are not enabled in the configuration evaluated here.

V. BEHAVIOURAL CODEBOOK HYBRID (BCH)

Empirical evaluation of the GAN-VAE configuration surfaced three structural limitations: persistent minority-class collapse on machine-type personas, capacity trade-offs where gains on downlink throughput regressed latency, and a reliance on per-persona normalisation that functioned as a patch rather than a fix. To address these we extend the framework with a Behavioural Codebook Hybrid (BCH) variant. BCH retains the persona library, constraint engine, and closed-loop KPI alignment but replaces the monolithic generator with a four-stage pipeline organised around a learned per-persona behavioural vocabulary.

A. Stage 1: Per-persona VQ-VAE

For each persona p in P , a separate VQ-VAE is trained on event windows of length T . The encoder maps a window to a latent sequence that is quantised to a persona-specific codebook of K codewords of dimension D ($K = 256$, $D = 64$ in our configuration). The decoder reconstructs the window from the codeword sequence. Each codeword encodes a behavioural primitive such as an IoT-keepalive ping, a video-segment burst, or a handover with buffer fill. Codebook entries are updated with an exponential moving average and dead-code revival; acceptance requires reconstruction loss ≤ 0.01 and utilisation $\geq 60\%$.

B. Stage 2: Autoregressive transformer over codewords

A decoder-only transformer ($\sim 3M$ parameters per persona) is trained by cross-entropy on Stage 1 codeword indices, conditioned on a scenario context token (hour of day, congestion class). Operating on a learned discrete vocabulary lets the transformer capture long-range causal structure between behavioural primitives, including the staging of preparatory control-plane events before a data-plane burst.

C. Stage 3: Wasserstein GAN closure

The discrete sampler is composed with the Stage 1 decoder to produce a continuous trace. A discriminator is trained on full decoded traces against real traces using a WGAN-GP loss; the gradient back-propagates only through the decoder, while the encoder and codebook stay frozen. This corrects inter-primitive seams that the autoregressive model alone cannot smooth.

D. Stage 4: Constraint-aware decoding

Hard physical constraints (bounded throughput, bounded inter-arrival time) are evaluated inline during autoregressive sampling. At each step the decoder candidate is materialised, constraints are checked, and any violating token is rejected; the next-most-probable token is sampled. The constraint engine is therefore promoted from a post-hoc projection operator to a decoding-time filter.

VI. EMPIRICAL RESULTS

We evaluate the framework on an anonymised behavioural trace corpus of $\sim 20,000$ - $25,000$ events per backend across 80 synthetic subscribers and five canonical personas (high-mobility streaming, low-mobility browsing, IoT or machine-type periodic reporting, policy-intensive entertainment or vehicular, and roaming), plus a separate IMT-2030 (International Mobile Telecommunications 2030) preview corpus covering five 6G persona classes: XR or Hologram, Integrated Sensing and Communication (ISAC), Massive IoT, Tactile or High-Reliability Low-Latency Communication (HRLLC), and Vehicle-to-Everything (V2X) or Autonomous Driving. Anonymisation uses pseudonymisation and k -anonymity controls.

A. Aggregate fidelity

Table I reports aggregate corpus fidelity on the 5G in-domain corpus, defined as a weighted combination of marginal Kolmogorov-Smirnov agreement, lag-one autocorrelation agreement, and per-persona distribution agreement. BCH raises aggregate fidelity from 0.79 (GAN-VAE) to 0.96 and outperforms both baselines on every component metric.

TABLE I. Aggregate corpus fidelity (5G quasi-real corpus).

Configuration	Fidelity	Autocorr.	Distribution
Statistical baseline	0.671	0.620	0.723
GAN-VAE (Sec. IV)	0.792	0.812	0.771
BCH (Sec. V)	0.962	0.945	0.979

B. Per-persona realism

Fig. 2 reports the per-persona realism score on the 5G corpus, defined as one minus the mean of three KS distances (downlink throughput, latency, inter-arrival time). The realistic band is set at ≥ 0.85 on all three metrics simultaneously. BCH lands in this band on all five 5G personas (0.99-1.00); the GAN-VAE baseline falls below threshold on three personas, most visibly on Browse or Indoor (0.75) and Entertainment or Vehicle (0.55).

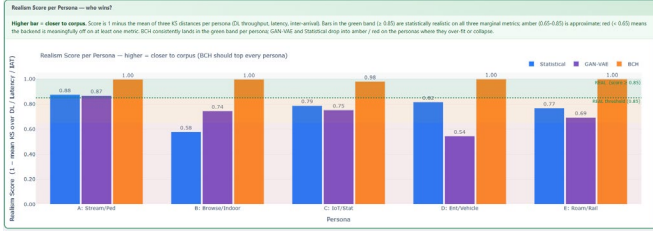


Fig. 2. Per-persona realism on the 5G corpus. BCH (orange) clears the 0.85 realistic band on all five personas; GAN-VAE (purple) falls below threshold on three and the statistical sampler (blue) on four.

Repeated on the IMT-2030 6G preview corpus (Fig. 3), the gap widens substantially. BCH stays at or near 1.00 on every persona, while GAN-VAE collapses into 0.42-0.57 and the statistical sampler into 0.35-0.66. The more heterogeneous 6G service classes place greater stress on per-class fidelity, which the per-persona behavioural vocabulary handles directly.



Fig. 3. Per-persona realism on the IMT-2030 (6G) preview corpus. BCH remains at or near 1.00 on every persona while both baselines drop into the unrealistic band.

C. Codebook separation

Fig. 4 plots the five IMT-2030 personas in the (latency criticality, data-rate intensity) plane. The dispersion ellipses do not overlap and centroids are well separated: Massive-IoT and XR / Hologram occupy opposite corners; Tactile or HRLLC sits in a narrow band on the right. This confirms that the per-persona VQ-VAE codebooks learn behaviourally distinct primitives rather than collapsing to a shared representation, supporting the deployment of each codebook as a stand-alone behavioural module.

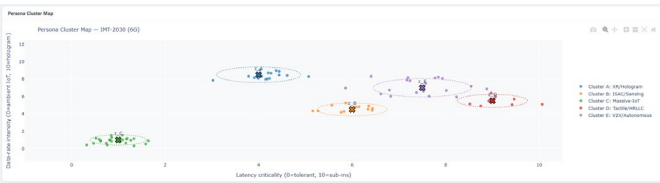


Fig. 4. Persona cluster map for the IMT-2030 (6G) corpus, plotting five synthetic-subscriber clusters in the (latency criticality, data-rate intensity) plane. Cluster centroids (cross markers) are well separated and dispersion ellipses do not overlap.

D. Applications and deployment

The framework targets a number of concrete validation workloads. First, it supports testing and benchmarking of NWDAF analytics workflows [14], including anomaly detection and load prediction, by generating behaviourally diverse subscriber populations with controllable persona mixes. Second, it supports PCF and NEF policy validation under context-conditioned subscriber behaviour, including realistic responses to congestion and slice steering. Third, the per-persona codebook of BCH is directly productisable as a behavioural library module: an enterprise corpus can be used to train a new persona codebook that extends the library without retraining the autoregressive backbone, providing a customer-extensible deployment model. Fourth, repeatable regression libraries built from named behavioural scenarios and controlled persona mixes enable continuous-integration testing of adaptive 5G and 6G functions.

VII. CONCLUSION

We have presented BSTG, a behaviour-conditioned, constraint-aware framework for synthetic 5G/6G traffic that combines a hybrid GAN-VAE generator with telecom-validity constraints and a closed-loop KPI fidelity objective. We have further introduced the Behavioural Codebook Hybrid variant, replacing the monolithic generator with a four-stage pipeline of per-persona VQ-VAE codebooks, an autoregressive transformer over codewords, a Wasserstein GAN closure, and constraint-aware decoding. On a 5G quasi-real corpus, BCH raises aggregate corpus fidelity from 0.79 to 0.96 and resolves minority-class collapse on machine-type personas. On a 6G IMT-2030 preview corpus, BCH remains in the realistic band on all five persona classes, while the baselines drop sharply. Future work covers dedicated kernel-MMD and membership-inference attack evaluations and pure-GAN / pure-VAE ablation studies.

The framework relies on the availability of representative anonymised behavioural traces and on a suitable outcome model for KPI alignment. Privacy is addressed at the data-abstraction layer through pseudonymisation and k-anonymity controls. Differentially private training via DP-SGD [10] is a planned extension and is not enabled in the reference configuration evaluated here; membership-inference and linkage risk should therefore be audited per deployment. Future work covers dedicated kernel-MMD and membership-inference attack evaluations, pure-GAN and pure-VAE ablation studies that isolate the contribution of each component, and extension of the per-persona codebook approach to multi-domain corpora covering radio, core, and transport layers in a single learned vocabulary.

REFERENCES

- [1] 3GPP TS 23.288, "Architecture enhancements for 5G System (5GS) to support network data analytics services."
- [2] VIAVI Solutions, "TM500 Network Tester," product page, 2026.
- [3] VIAVI Solutions, "TeraVM 5G SA Core Emulator," data sheet, 2024.
- [4] H. Heffes and D. M. Lucantoni, "A Markov Modulated Characterization of Packetized Voice and Data Traffic," IEEE J. Sel. Areas Commun., vol. 4, no. 6, pp. 856-868, Sep. 1986.

- [5] D. Anick, D. Mitra, and M. M. Sondhi, "Stochastic Theory of a Data-Handling System with Multiple Sources," *Bell Syst. Tech. J.*, vol. 61, no. 8, pp. 1871-1894, Oct. 1982.
- [6] J. Yoon, D. Jarrett, and M. van der Schaar, "Time-series Generative Adversarial Networks," in *Proc. NeurIPS*, 2019.
- [7] N. Sivaroopan et al., "NetDiffus: Network Traffic Generation by Diffusion Models through Time-Series Imaging," *Computer Networks*, vol. 251, art. 110616, 2024.
- [8] A. van den Oord, O. Vinyals, and K. Kavukcuoglu, "Neural Discrete Representation Learning (VQ-VAE)," in *Proc. NeurIPS*, 2017.
- [9] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved Training of Wasserstein GANs," in *Proc. NeurIPS*, 2017.
- [10] M. Abadi et al., "Deep Learning with Differential Privacy," in *Proc. ACM CCS*, 2016.
- [11] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," in *Proc. ICLR*, 2014.
- [12] A. Vaswani et al., "Attention is All You Need," in *Proc. NeurIPS*, 2017.
- [13] L. U. Khan, W. Saad, D. Niyato, Z. Han, and C. S. Hong, "Digital Twin of Wireless Systems: Overview, Taxonomy, Challenges, and Opportunities," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 4, pp. 2230-2254, 2022.
- [14] A. Sevgican et al., "Intelligent network data analytics function in 5G cellular networks using machine learning," *J. Commun. Networks*, vol. 22, no. 3, pp. 269-280, June 2020.
- [15] R. Shokri, M. Stronati, C. Song, and V. Shmatikov, "Membership Inference Attacks Against Machine Learning Models," in *Proc. IEEE S&P*, 2017.

AUTHOR BIOGRAPHIES



FARIS MUHAMMAD is Senior Director, Enterprise AI Architecture at VIAVI Solutions (CTO Office), with corporation-wide remit across VIAVI. He is a distinguished engineering and technology leader with over twenty-five years delivering large-scale software, AI/ML, and telecommunications products, including more than a decade of hands-on AI/ML technical leadership. Since 2015 he has driven significant AI/ML innovation, beginning with his first patent filing that year and now holding nine granted patents, alongside one book, four theses, and more than thirty-five technical publications. His work spans AI/ML for network test, telecommunications data analytics, intelligent automation, generative models for synthetic traffic, network digital twins, and AI-driven service assurance. He holds an Executive MBA from Imperial College Business School, London, and is a Visiting Professor at Brunel University of London. He is frequently engaged as a strategic advisor to senior executives and universities on technology strategy, AI/ML product direction, and organisational design.