One Feedback System Does Not Fit All: Localising Data-to-Text Driver Coaching for the United Kingdom and Nigeria

Iniakpokeikiye Peter Thompson[1,*] **Jawwad Baig**[1,*] **Ehud Reiter**[1] **Dewei Yi**[1]
[1]Department of Computing Science, University of Aberdeen,
King's College, Aberdeen AB24 3UE, United Kingdom
iniakpokeikiye.thompson1@abdn.ac.uk
{j.baig.18,e.reiter,dewei.yi}@abdn.ac.uk
*Equal contribution

Abstract

Data-to-text driver coaching is often presented as a generic pipeline from telematics events to advice. This paper argues that its content requires localisation because usefulness and credibility depend on drivers' knowledge, prevalent risks, regulation, infrastructure, and available data. Two independently developed systems in the United Kingdom and Nigeria are compared by tracing requirements through content selection, generation, and field evaluation. The UK system prioritises post-trip reflection, explanations tied to road and place context, and tone-sensitive wording. The Nigerian system combines legally grounded, once-daily Tips based on detected events with weekly persuasive Reports; it foregrounds safety education and alcohol-related risk in response to reported gaps in formal training and traffic-rule knowledge, as well as local road-safety priorities. Reliable speed-limit metadata supported speeding feedback in the UK, whereas its scarcity led the Nigerian evaluation to exclude speeding from its outcome metric. Both interventions were associated with reduced distance-normalised unsafe-event rates in their own field studies, although their designs and metrics preclude an effect-size comparison. The analysis yields a requirements-to-content design process for localising safety-critical NLG without treating a high-income deployment as the default.

1 Introduction

Natural Language Generation (NLG) has long been used to turn structured data into readable feedback for users who need to act on complex information (Reiter and Dale, 2000; Osuji et al., 2024; Sharma et al., 2024). In driver-feedback systems, this task is safety-critical: the text must be understandable and motivating, but it must also remain grounded in sensor observations, road context, and legal constraints. Reviews of driver-feedback systems and post-trip interventions show that sensing, presentation, timing, and behaviour-change mechanisms must be designed together (Rios-Torres and Malikopoulos, 2016; Michelaraki et al., 2021). The problem becomes harder when systems are deployed across different infrastructure and regulatory environments.

This paper compares two NLG-based driver-feedback systems described in recent doctoral system studies and related publications: DRIVING-BEACON, a United Kingdom (UK) system for post-trip telematics feedback (Baig, 2026; Baig et al., 2022), and Safe Drive Africa, a Nigerian mobile system for culturally attuned safety feedback (Thompson, 2025; Thompson et al., 2025). Both convert sensed behaviour into textual coaching, but they were designed from different local evidence rather than by translating a fixed intervention.

The central claim of this paper is that localising data-to-text coaching is fundamentally a problem of content selection, grounding, and framing rather than language translation. UK participants sought fair, contextual reflection that explained what happened, where, and why, without moralising. Nigerian stakeholders identified gaps in formal training and traffic-rule knowledge alongside alcohol-influenced driving, creating a distinct requirement for foundational education, explicit legal grounding, and alcohol-aware coaching. Architecture and sensor availability enable and bound these content strategies: reliable speed-limit metadata in the UK support speeding narratives, whereas data scarcity in Nigeria requires omitting ungrounded speed claims. Implementation differences alone are not the primary contribution; rather, we synthesize how contextual requirements dictate architectural tradeoffs.

Our contributions are:

- a requirements-to-content comparison show-

Proceedings of the 19th International Natural Language Generation Conference: Long Papers, pages 1041–1052
October 17-21, 2026 ©2026 Association for Computational Linguistics

ing how locally elicited risks, knowledge, and expectations shaped two NLG interventions;

- a synthesis of how road-data availability and safety priorities determine responsible content boundaries and evaluation targets;
- field evidence of behavioural effectiveness from both deployment contexts; and
- a five-stage design process, derived from the two deployments, providing a structured methodology for eliciting requirements, auditing evidence, defining messages, enriching content, and generating reports when localising safety-critical NLG systems.

2 Related Work

Early NLG systems for behaviour change used controlled templates and explicit tailoring to preserve traceability (Reiter et al., 2003; Williams and Reiter, 2008). Driver feedback followed a similar path: early data-to-text work generated explanations directly from telematics before SaferDrive integrated controlled content selection into a behaviour-change intervention (Braun et al., 2015, 2018). Reviews now distinguish real-time warnings from post-trip reflection and show wide variation in sensing, feedback modality, personalisation, and evaluation design (Rios-Torres and Malikopoulos, 2016; Michelaraki et al., 2021; Boylan et al., 2024; Kontaxi et al., 2025). Field studies nevertheless provide convergent evidence that telematics feedback can change measured behaviour, including randomised financial-incentive feedback, personalised programmes for young drivers, and safe-driving apps (Stevenson et al., 2021; Meuleners et al., 2023; Vankov et al., 2023; Warren et al., 2018).

Beyond research prototypes, usage-based insurance (UBI) systems turn telematics into scores, incentives, and post-trip advice. Surveys of telematics research cover risk modelling, safer transport, and insurance applications (Gao et al., 2022; Ghaffarpasand et al., 2022). User acceptance depends not only on usefulness but also on trust, social influence, and privacy (Quintero et al., 2023; Picco et al., 2023). Automotive-data studies accordingly emphasise user control, data minimisation, and governance when monitoring can affect premiums or other consequential decisions (Dowthwaite et al., 2024; Truby et al., 2024). Recent systems have generated actionable risk-mitigation suggestions from insurer data (Li et al., 2023), while a large randomised field experiment found that weekly UBI-style feedback and incentives reduced several measured risky behaviours (Ebert et al., 2026).

Recent Large Language Models (LLMs) make it easier to produce fluent and adaptive feedback, but they introduce risks when outputs must be factually defensible. Evaluation in high-stakes summarisation and feedback domains therefore encompasses factual consistency, task suitability, and human interpretation rather than only surface fluency (Novikova et al., 2017; van der Lee et al., 2019; Croxford et al., 2025). Hybrid architectures are a practical response: structured components decide what can be said, while neural components may adapt how it is said.

Persuasive technology research also matters because feedback is intended to influence future behaviour, not merely describe past events (Fogg, 2009). Meta-analytic road-safety evidence supports the relevance of attitudes, norms, and perceived behavioural control, while person-based intervention design argues for iterative grounding in target users' needs and contexts (Somoray et al., 2024; Yardley et al., 2015). Safe Drive Africa explicitly draws on the Theory of Planned Behaviour (TPB), using prompts that address attitudes, norms, motivation, and perceived behavioural control (Ajzen, 1991; Thompson, 2025). DRIVINGBEACON instead uses predefined user attributes to adapt tone while keeping factual claims under rule-based control (Baig, 2026).

Most work asks whether feedback is tailored to an individual. The present comparison adds a different level: the suitability of the *content strategy* for its socio-technical setting. Persuasive strategies are not received uniformly across cultures, as comparisons of Nigerian and Canadian users demonstrate, and broader driving research links behaviour to culture, income, and governance (Oyibo et al., 2018; Arslan and Özkan, 2024). Recent cultural-NLP research likewise distinguishes cultural adaptation from literal translation and warns that culture comprises multiple, context-dependent elements requiring explicit evaluation (Singh et al., 2024; Liu et al., 2025). Localisation here therefore extends beyond cosmetic translation or the addition of place names. It can change which behaviours are prioritised, whether advice provides rule education or supports reflec-

tion, which evidence establishes credibility, and which outcomes can be measured reliably. The two systems are examined as an analytic case comparison rather than as a controlled cross-country experiment.

3 From Local Context to Content

Table 1 summarises the evidence that motivated the two content strategies. The reported patterns describe the respective requirements samples and deployments; they do not characterise all UK or Nigerian drivers.

3.1 UK: reflection in a formal driving environment

Complementary methods addressed distinct requirements before implementation. Two structured surveys elicited report-format preferences and perceived influences on driving from the same 30 active UK drivers, recruited through personal and community networks; all held a DVLA-issued licence and had at least one year's experience. Thirteen respondents spanning ages, backgrounds, and driving environments joined two audio-recorded, semi-structured focus groups, using a documented guide, to examine privacy, tone, and context through interaction. Interviews with UK drivers ($n = 10$), police officers ($n = 4$), and instructors ($n = 2$) provided confidential individual and practitioner accounts. Survey responses were summarised descriptively. Qualitative recordings were transcribed verbatim and analysed inductively; the primary researcher iteratively grouped interview codes into higher-level themes. Triangulation across structured breadth, group interaction, and individual depth yielded requirements for content, timing, context, tone, and privacy (Baig, 2026).

The UK-focused evidence emphasised turning numerical events into explanations: what occurred, the relevant circumstances, why it was risky, and what to do next. Some participants also wanted educational reminders about legal thresholds, fines, and penalty points, but legal instruction was one requirement rather than the principal organising function of the reports. Participants considered real-time alerts potentially distracting and preferred daily or weekly post-trip summaries. Harsh or judgemental language was expected to reduce engagement; neutral, practical and educational wording was favoured (Baig, 2026).

Place changed the meaning of an event. Interviews highlighted schools, medical facilities, shopping areas, and other places with vulnerable road users. Consequently, speeding near a school was not merely another speed-limit breach: the report could connect it to children and time-specific local risk. Road type also shaped advice, including following distance on motorways, overtaking and unexpected hazards on rural roads, and crowded places in urban areas. Legal reminders remained a possible requirement, particularly when rules or penalties change, but they were not the principal content driver.

3.2 Nigeria: education, authority, and alcohol risk

To investigate Nigeria's local driving context, safety priorities, and feedback requirements, Safe Drive Africa conducted a mixed-methods stakeholder requirements study (Thompson, 2025). A mixed-methods design was chosen to capture both broad statistical patterns of driving behaviour and rich qualitative explanations of local road challenges. Stratified random sampling was used to recruit 80 key stakeholders across two complementary groups: 51 road drivers (commercial, private, and fleet) and 29 officials from the Federal Road Safety Corps (FRSC), Nigeria's federal road-safety regulator and enforcement agency. Data were collected over two months through face-to-face questionnaire distribution and a secure online form. Role-tailored parallel questionnaires contained 25 questions for drivers and 11 for FRSC officials, combining demographic, binary, multiple-choice, and open-ended items. For analysis, open-ended responses underwent inductive thematic analysis in NVivo to uncover key risk rationales, while quantitative survey items were analysed using descriptive statistics and Fisher's exact tests (Thompson, 2025).

The thematic analysis identified four principal unsafe-driving themes: non-compliant driving, aggressive driving, distracted driving, and driving in fatigue. Participants attributed these behaviours to driver habits, including alcohol/substance use and poor vehicle maintenance, limited formal training and traffic-rule knowledge, and external factors such as poor road conditions and traffic intensity (Thompson, 2025). These findings are consistent with Nigerian evidence on road-user attitudes, road-safety constraints, and locally salient speeding beliefs (Uhegbu and Tight, 2021;

Design question	UK evidence	Nigerian evidence	Consequence for generated content
What help is needed?	Drivers sought fair, actionable explanations of detected events and preferred post-trip reflection to distracting live alerts.	Stakeholders, specifically drivers and Federal Road Safety Corps (FRSC) officials, identified lack of training and knowledge of traffic rules as causes of unsafe driving, alongside demand for education and awareness.	UK reports explain behaviour in context; Nigerian Tips explicitly teach rules and consequences.
Which risks deserve emphasis?	Speeding, phone use, harsh manoeuvres, and risk near schools and other busy or sensitive places.	Non-compliant/aggressive driving, distraction, fatigue, and alcohol/substance use; the app could reliably instrument harsh braking, acceleration, swerving, and alcohol risk.	Each system selects locally relevant, credibly detectable behaviours rather than a universal inventory.
What makes feedback acceptable?	Supportive, non-judgemental language; clarity, fairness, and sensitivity to preferred feedback style.	Credible Nigerian legal references plus respectful, ability-building coaching that acknowledges local road realities.	Personalised tone in the UK; two complementary Nigerian surfaces for authority and persuasion.
What can data support?	Telematics supplied speed and speed-limit fields plus mapped/geofenced context.	Digital posted speed limits were too sparse and unreliable for consistent legal-speed classification.	UK content can discuss speeding and place risk; the Nigerian event rate and feedback avoid unsupported speeding claims.

Table 1: Local evidence and its consequences for content. The contrast concerns study populations and deployment evidence, not national stereotypes.

Uzondu et al., 2022; Etika et al., 2021). Crucially for content design, "lack of training" covered limited formal instruction, limited traffic-rule knowledge, and difficulty interpreting signs. Stakeholders called for continuing education and stricter enforcement; 78 of 80 (97.5%) supported a monitoring-and-feedback app (Thompson, 2025). This requirements-led approach also responds to reviews calling for road-safety mobile systems to cover the full information cycle rather than detection alone (Aghayari et al., 2024).

Those findings motivated two content functions. Event-triggered Tips address foundational knowledge by citing applicable Nigerian rules, fines, and penalties. Weekly Reports address habits and attitudes through supportive reflection. Alcohol-influenced driving received dedicated treatment because stakeholders repeatedly identified it as a major local risk and requested detection of alcohol consumption. The project subsequently collected locally labelled smartphone-sensor data and trained a post-trip alcohol classifier on 152 trips (reported nested-cross-validation area under the receiver operating characteristic curve (AUC) 0.855). Alcohol status can therefore be acknowledged empathetically in a Report, with plan-ahead alternatives when risk is detected and positive reinforcement otherwise.

3.3 Comparative boundary

The contrast does not imply that UK drivers know every rule, that all Nigerian drivers lack licences, or that alcohol-impaired driving is unique to Nigeria. Rather, the local studies supplied different evidence about the content that each intervention needed to foreground.

A separate exploratory strand provides further requirements evidence without extending the comparison to a third system. Fifteen semi-structured interviews with Pakistan-based Uber drivers, conducted in Urdu or English and analysed inductively, highlighted dense traffic, inconsistent signage, unpredictable pedestrian activity, weak enforcement, and a preference for daily or weekly reports with reminders linked to markets, schools, mosques, and busy junctions (Baig, 2026). No feedback system was implemented or evaluated in Pakistan, so these observations illustrate another potential context for adaptation rather than evidence of system effectiveness. The analysis is consequently limited to study-specific findings instead of attributing uniform characteristics to national populations.

4 Two Localised Content Strategies

4.1 DRIVINGBEACON: contextual post-trip reflection

DRIVINGBEACON produces daily or weekly in-app narrative reports from UK trip summaries. Its content planner selects detected speeding, mobile-phone use, harsh braking and acceleration, trip time, road type, and proximity to geofenced sensitive locations. It then combines events into a readable account rather than presenting a dashboard score alone (see Figure 1(a)) (Baig, 2026).

Three content properties follow from the requirements. First, explanations are *situated*: a report can identify the event and connect it to a familiar location or road condition. Second, advice is *actionable*: the report explains why the behaviour matters and suggests an achievable change. Third, wording is *socially acceptable*: a lightweight profile records age group, predominant driving environment, sensitivity/stress level, preferred feedback style, and report frequency. These attributes condition tone and emphasis, not the measured facts. A user can receive friendly or direct language, but an event cannot disappear because of a profile.

This strategy treats feedback as supported reflection for drivers already operating in a comparatively formal licensing and enforcement setting. It can include brief legal information, but its distinctive value is turning telematics into a fair, contextually specific narrative. Pilot participants preferred enhanced reports containing phone-use and geofenced context to basic event summaries, although the eight-person, four-week pilot was not powered to show behavioural efficacy.

4.2 Safe Drive Africa: education plus persuasion

Safe Drive Africa produces two deliberately separate forms of feedback for Nigerian drivers (Thompson, 2025). **Tips** (Figure 1(c)) are short messages generated and displayed once per day from unsafe behaviours detected that day, such as harsh braking, rapid acceleration, or swerving; they are not presented continuously while the driver is driving. Behaviour keywords retrieve entries from a curated knowledge base built from the Nigerian Highway Code and National Road Traffic Regulations. The prompt requires applicable law sections, penalties, and fine values to be copied from the retrieved context, while adding a plain-language explanation and practical action. Tips therefore identify the unsafe behaviour and provide the corresponding rule and consequence.

Content layer	Locally motivated configuration
Core purpose	**UK:** explain post-trip events in context. **Nigeria:** teach rules and support behaviour change.
Evidence	**UK:** event, time, and road/place. **Nigeria:** event statistics plus retrieved law, fine, and penalty.
Risk focus	**UK:** speed, phone use, manoeuvres, and sensitive places. **Nigeria:** manoeuvres and alcohol influence.
Adaptation	**UK:** profile-sensitive tone and road advice. **Nigeria:** legal retrieval, local realities, and TPB coaching.
Surface	**UK:** integrated daily/weekly Report. **Nigeria:** event Tips plus weekly Report.

Table 2: The systems implement different locally motivated content strategies.

Reports (Figure 1(b)) are weekly, driver-specific narratives over structured statistics: total and most frequent unsafe behaviours, an example date/time and road, trip duration and distance, locations, and alcohol status. A two-stage generation-and-reflection procedure applies TPB principles implicitly. It praises safe habits, acknowledges local realities such as congestion, poor roads, or rain, gives two or three ability-building actions, and uses a supportive tone. Alcohol-positive reports directly but empathetically recommend not driving after drinking and planning alternatives; alcohol-negative reports reinforce the safe pattern. The Report is therefore not a longer Tip: the former supports motivation and perceived control, while the latter supplies precise legal education.

The content inventory is bounded by measurement. The app uses smartphone accelerometer, gyroscope, and Global Positioning System (GPS) streams for motion and unsafe-event detection. Its post-trip alcohol model was trained and evaluated using voluntary daily alcohol self-reports as ground-truth labels; those self-reports are not operational input features to the classifier. Fatigue and phone distraction were raised by stakeholders but omitted because they could not be measured credibly without extra hardware. Speeding was also omitted from the principal evaluation measure because reliable digital speed-limit metadata were not consistently obtainable.

5 How the Pipelines Support the Content

Architecture is secondary to the content argument, but it explains how each strategy is made dependable (Table 3). Both systems separate sensed data from user-facing language and use a large language model (LLM) only after structured evidence has been assembled. They differ in where content authority resides.

Factual control in DRIVINGBEACON. Cloud telematics supplies trip and waypoint data, including measured speed and available speed limits; mapping and geofences add road and place context. Deterministic rules decide which facts and advice enter the draft. The LLM is limited to paraphrasing the resulting report under instructions to preserve numbers and behavioural meaning; it does not select incidents or infer new facts. This yields an audit path from each report statement back to a trip variable or rule while still permitting tone variation.

Dual grounding in Safe Drive Africa. Deployment context also influenced component placement. DRIVINGBEACON could rely on cloud telematics in its UK deployment, whereas intermittent mobile connectivity and data cost in Nigeria motivated local persistence and on-device processing in Safe Drive Africa. The content-relevant distinction is between two evidence packages. A Tip is grounded by the detected event and retrieved Nigerian regulation. Its prompt prohibits invented laws, fines, or penalties and returns structured fields. A Report is grounded by aggregated driver statistics and passes through a second reflective prompt that checks coherence and implicit TPB coverage. Thus legal accuracy and motivational narrative use different controls instead of forcing both into one prompt.

Availability of speed-limit information. In the UK deployment, the Damoov telematics service supplied waypoint-level vehicle-speed and corresponding speed-limit fields, allowing evidence-based speeding feedback. No equivalent sufficiently reliable digital source of posted speed limits was available for the Nigerian deployment. The source study reports that roadside speed-limit signs were missing, inconsistent, or undocumented (Thompson, 2025). GPS speed alone therefore could not establish a legal breach reliably. The Nigerian system excluded speeding from its unsafe-events-per-kilometre measure rather than generating an unsupported claim.

Models and inference settings.
UK system: DRIVINGBEACON used OpenAI's Chat Completions application programming interface (API) with `gpt-3.5-turbo`, sending the assembled evidence and profile instructions as one user-role message. Its source study does not report decoding settings for this call.
Nigerian system: Safe Drive Africa sent the same retrieved legal context independently to `GPT-5.1` and `Gemini Nano Banana Pro` for Tips and displayed both outputs. Its OpenAI Tip call used temperature 0.5 and `max_tokens=750`; its Gemini Tip call used Google Generative AI Android Software Development Kit (SDK) 0.9.0 with no explicitly configured decoding parameters. Reports used `gpt-4-turbo` for initial generation and `gpt-4o-mini` for mandatory revision, both with temperature 0.0 and `max_tokens=350`. Safe Drive Africa's OpenAI requests used the Chat Completions endpoint, `https://api.openai.com/v1/chat/completions`. For its generation calls, top-p, seed, safety settings, and dated model snapshots were not specified, so provider or SDK defaults applied.

Illustrative feedback surfaces. Figure 1(a) shows how the UK surface integrates incidents and contextual reflection in a report. The Nigerian surfaces separate a longer weekly coaching narrative (Figure 1(b)) from a legally grounded Tip (Figure 1(c)).

6 Evidence of Effectiveness

The two field evaluations examine whether locally designed feedback is associated with safer behaviour, but they do not provide a head-to-head test. Table 4 therefore reports within-study evidence and makes non-equivalence explicit.

6.1 UK deployment

The 30-driver study recorded four weeks without reports (pre-intervention), five weeks with personalised contextual reports (intervention), and three weeks after reports were withdrawn (post-intervention). Baig (2026) reports means of 0.6881, 0.3654, and 0.2403 incidents per mile; dividing by 1.609344 gives *unsafe events per kilometre*: 0.4275 pre-intervention, 0.2270 intervention, and 0.1493 post-intervention. Lower rates indicate fewer detected unsafe events: these values

System	Local evidence	Content selection	Generation control	Driver-facing output
UK	Telematics events, speed-limit fields, maps, and geofences	Rules select incidents and road/place advice	Constrained paraphrase of the rule-selected draft	Contextual daily or weekly Report
Nigeria	Smartphone events, alcohol classification, and legal knowledge base	Keyword legal retrieval for Tips; weekly statistics for Reports	Structured Tip output; two-step Report generation and revision	Once-daily legal Tips; weekly persuasive Reports

Table 3: How locally available evidence determines content selection, generation control, and feedback delivery in the UK and Nigerian systems.

Dimension	DRIVINGBEACON: UK	Safe Drive Africa: Nigeria
Design	12-week within-subject deployment: 4-week pre-intervention, 5-week intervention, and 3-week post-intervention with feedback withdrawn.	Within-subject pre-intervention/intervention deployment over approximately 25 weeks, excluding transition trips.
Data	30 drivers; daily or weekly Reports; phase-level driver event rates.	54 analysed drivers (27 commercial, 14 private, 13 company); 802 trips; once-daily Tips and weekly Reports.
Outcome	Unsafe events/km from speeding, phone use, and harsh braking/acceleration; geofenced events remain in the total and place context supports interpretation.	Unsafe events/km from harsh braking, rapid acceleration, and swerving; speeding excluded.
Result	0.4275 pre-intervention, 0.2270 intervention, and 0.1493 post-intervention; significant phase effect and all pairwise differences.	1.44 pre-intervention and 1.16 intervention, a 19.4% reduction ($p < .001$); 43/54 drivers (79.6%) improved.
Limitations	No parallel control group; all drivers received feedback during intervention; the personalisation effect was not isolated; post-intervention observation lasted three weeks.	No parallel control group; intervention trip count did not correlate significantly with improvement; alcohol-classifier outputs were excluded from the outcome.

Table 4: Evaluation summary. Rates use a common distance unit, but different event definitions, durations, and designs make effect sizes non-comparable.

correspond to approximately 43, 23, and 15 events per 100 kilometres. The intervention mean was 46.9% below the pre-intervention mean, while the post-intervention mean was a further 34.2% lower, or 65.1% below pre-intervention. A repeated-measures one-way analysis of variance (ANOVA) on the original per-mile rates found a significant phase effect, $F(2, 58) = 85.2$, $p < .0001$. Tukey honestly significant difference comparisons were significant for pre-intervention versus intervention ($p < .0001$), intervention versus post-intervention ($p = .0018$), and pre-intervention versus post-intervention ($p < .0001$). This pattern is consistent with improvement during the intervention and short-term maintenance, but the three-week post-intervention phase is too short to establish durable behaviour change. Because all participants received the complete intervention, the study cannot isolate tone personalisation, contextual content, or observation effects.

Qualitative responses illuminate content reception. Participants described the reports as helpful and non-blaming, valued friendly or profile-appropriate tone, and reported that references to speeding near schools made an abstract event feel consequential. Suggested improvements included visual or map summaries, adjustable report frequency and detail, audio summaries, and less repetitive wording. These observations support acceptability of the UK strategy, though they are not independent evidence of efficacy.

6.2 Nigerian deployment

The Nigerian study analysed 54 drivers and 802 trips. Driver-level unsafe behaviours per kilometre (UBPK) fell from 1.44 to 1.16, a 19.4% mean reduction; 43 drivers (79.6%) improved (Thompson, 2025). A Wilcoxon signed-rank test found the pre/intervention difference significant ($p < .001$). The study also introduced *bad-day frequency*: the proportion of intervention trips whose trip-level UBPK exceeded the 75th percentile of that driver's pre-intervention trip-level UBPK distribution. This driver-specific threshold represents the riskiest quarter of pre-intervention trips. Drivers reaching similar final event rates sometimes followed very different trajectories, from steady improvement with no bad days to volatile improvement with repeated relapses. This suggests an NLG opportunity beyond changing

(a) DRIVINGBEACON Report

This Wednesday, you drove 12 miles along the M1 and M25 motorways, and while every journey is a chance to build your driving confidence, there are some important areas that need your attention.
Your maximum speed reached 86 mph, which is well above the 70 mph motorway limit. As a young man on the roads, it is especially important to build safe habits early; exceeding the speed limit by such a margin can result in hefty fines, penalty points, and even a driving ban that could affect your independence and career prospects.
You also used your mobile phone while driving for approximately 2 miles, which is not only illegal but puts you and others at serious risk. Research shows you are four times more likely to be involved in a crash when using your phone behind the wheel, and fines can reach up to £1,000.
On a more positive note, you kept harsh braking to just one incident, which shows some awareness of the road ahead. However, the two harsh acceleration events suggest a rushed driving style; try leaving a few minutes earlier to avoid the temptation to speed up. Remember, the motorway demands full concentration and safe following distances, especially at higher speeds. Take it steady, build your experience with patience, and you will become a confident, responsible driver.

(b) Safe Drive Africa weekly Report

Hello,
I want to start by commending you on maintaining a zero alcohol level during your trips. This is a significant achievement, especially given the challenges on the roads in West Africa. Your commitment to sober driving is commendable and sets a high standard for road safety.
Navigating through the bustling streets, dealing with rough patches, unpredictable weather, and the occasional encounters with area boys can be quite challenging. These realities make your adherence to safe driving practices even more crucial.
I've noticed from your recent driving data that harsh acceleration has been a frequent issue, particularly noted six times, with a notable instance on February 11, 2025, at 07:42 on Ikorodu Road. This behavior can be risky, especially on busy roads where sudden speed changes can lead to accidents. To help manage this, try to maintain a steady pace, accelerating gradually. This will not only enhance your safety but also improve your vehicle's efficiency.
Additionally, your trips in the afternoon seem to have the most incidences. It might be helpful to plan your routes ahead of time during these hours to avoid high traffic areas. Taking a few minutes before your trip to relax and mentally prepare can also reduce the likelihood of rushed decisions while driving.
Keep up the good work, and remember, every trip is a new opportunity to enhance your safety and that of others on the road. Your responsible choices can inspire those around you to prioritize safety as well.
Your Driving Safety Specialist Agent.

(c) Safe Drive Africa legal Tip

Title: Keep Your Lane Steady to Stay Safe.
Meaning: Swerving happens when your car makes sudden side movements, often because of quick reactions to traffic or road obstacles. These sharp shifts can make other road users uncomfortable and increase the risk of a crash.
Penalty: Suspension or bar from holding a driver's licence.
Fine: N50,000 or imprisonment for up to 2 years.
Law: Regulation 167(1) of National Road Traffic Regulations.
Summary Tip: Focus on keeping a steady lane position. Look ahead, ease off the accelerator when traffic gets tight, and guide the car smoothly instead of making sudden moves. This keeps you in control and protects everyone around you.

Figure 1: Source-study examples: (a) a DRIVINGBEACON Report; (b) a Safe Drive Africa weekly Report; and (c) a Safe Drive Africa legal Tip.

average risk: content cadence and emphasis could respond to stability.

The ancillary post-trip alcohol classifier was a bagged decision-tree model evaluated on 152 labelled trips using nested 5×5 cross-validation. Beyond AUC 0.855 (95% confidence interval (CI) 0.794–0.909), it achieved mean F1 0.589 (95% CI 0.435–0.711) and mean recall 0.615 (95% CI 0.433–0.773). Aggregated out-of-fold predictions yielded accuracy 0.829, precision 0.576, recall 0.613, F1 0.594, and specificity 0.884. These results describe trip classification, not the behavioural-intervention outcome (Thompson, 2025).

There was no significant correlation between the number of intervention trips and the reduction in unsafe events per kilometre (Pearson $r = -0.071$, $p = .610$; Spearman $\rho = -0.043$, $p = .757$). The evidence therefore does not support a simple "more app use causes more improvement" claim. Four follow-up interviews instead help explain perceived usefulness: participants valued legal references and fines as credible, locally relevant support, while supportive wording reduced the sense of punishment. This aligns with the dual content rationale but remains a small qualitative sample.

6.3 Scope of the comparison

Both studies show statistically supported reductions in unsafe events per kilometre while locally adapted feedback was deployed, and both report favourable reception of the content features that

distinguish them. Neither includes a contemporaneous randomised control, so "associated with" is more defensible than "caused". The common unit does not make the rates equivalent: the two measures contain different event sets, most notably because speeding is excluded in Nigeria. Ranking the systems by rate or percentage change would therefore be invalid. The comparative result is instead a replication at the design-principle level: locally derived content can be made operational and evaluated in two materially different settings.

7 A Design Process for Localised NLG

The comparison yields a five-stage design process, consistent with person-based and mobile road-safety design guidance (Yardley et al., 2015; Aghayari et al., 2024): (1) *requirements elicitation* identifies stakeholder needs, risks, languages, and feedback timing; (2) *evidence audit* determines which behaviours, contextual variables, and authorities are available and credible; (3) *message definition* maps input data and evidence to information that the system may communicate; (4) *message enrichment* adds supported information such as legal consequences, practical advice, or encouragement; and (5) *report generation* realises the selected messages and additional information in the language and form required by stakeholders. Following standard NLG terminology, a message is a domain-level unit of information selected for communication before its wording is chosen (Reiter and Dale, 2000). This process is derived from two deployments, but crossover or component-level experiments are needed to establish which choices are genuinely locale-specific.

Stage 1: Requirements elicitation. In both systems, localisation began with eliciting local stakeholder needs, prevalent risks, and existing knowledge rather than adopting default assumptions. DRIVINGBEACON's interviews revealed that UK drivers wanted contextual post-trip reflection, sensitivity to personal feedback styles, and explanations tied to sensitive places like schools. Safe Drive Africa's stakeholder survey identified gaps in formal training and traffic-rule knowledge, alongside alcohol-influenced driving, creating a clear demand for educational rules, authoritative legal backing, and alcohol-aware coaching. Importantly, requirements elicitation must characterize the specific study population and deployment context without generating sweeping national stereotypes: the findings reflect elicited needs from the respective cohorts rather than uniform traits of all UK or Nigerian drivers.

Stage 2: Evidence audit. A safety-critical NLG system must systematically audit local data infrastructure to determine which behaviours, contextual variables, and authorities can be verified reliably. DRIVINGBEACON could discuss speeding because cloud telematics supplied waypoint-level vehicle speed and reliable posted speed limits. Conversely, because digital speed-limit metadata were sparse and unverified in the Nigerian deployment, Safe Drive Africa excluded speeding from both driver feedback and its evaluation metric. The core principle of evidence auditing is strict: when a required contextual variable cannot be grounded in local data, the system should omit the claim entirely rather than infer or hallucinate ungrounded assertions.

Stage 3: Message definition. Message definition maps audited sensor observations to the core communicative units that the system intends to convey. Rather than deploying a universal inventory of driving events, each system defined messages reflecting its audited risks: DRIVINGBEACON focused on speeding, phone distraction, and manoeuvring near sensitive places, whereas Safe Drive Africa defined messages around harsh manoeuvres, trip timing, and classifier-detected alcohol status. In both systems, a message represents a structured domain-level entity established before lexical realisation.

Stage 4: Message enrichment. Once core messages are defined, message enrichment attaches supported domain knowledge to make the feedback persuasive, actionable, and socially acceptable. In DRIVINGBEACON, enrichment incorporated driver profile attributes (age group, feedback sensitivity) to condition narrative tone, alongside place-specific risk information (such as risks to pedestrians near schools). In Safe Drive Africa, enrichment retrieved applicable laws, regulations, and statutory penalties via a curated legal knowledge base for daily Tips, and structured motivational advice using TPB principles (acknowledging local road conditions, offering ability-building tips, and praising sober trips) for weekly Reports.

Stage 5: Report generation and delivery. The final stage realises the enriched messages into user-facing text, selecting generation controls

and delivery cadences matched to content function. DRIVINGBEACON generates integrated daily or weekly Reports using rule-based selection paired with constrained LLM paraphrasing to maintain factual integrity. Safe Drive Africa employs dual generation pathways: deterministic legal retrieval with structured LLM constraints for once-daily Tips, and a two-stage LLM generation-and-reflection pipeline for weekly persuasive Reports. Both deployments delivered feedback in English. For deployments in multilingual environments, report generation would also require establishing stakeholder language requirements and validating outputs in target languages rather than treating machine translation as sufficient localisation.

8 Conclusion

The comparison is not primarily between two software stacks, but between two locally motivated content strategies. UK feedback prioritises contextual post-trip reflection and tone-sensitive explanation. Nigerian feedback places greater emphasis on legal and safety education, locally identified alcohol risk, and persuasive weekly coaching. Data availability further determines the scope of evidentially supported content: UK speed-limit context supports speeding narratives, while Nigerian metadata scarcity requires omission of that claim and metric. Both deployments provide qualified evidence of behavioural improvement in context. The findings connect local evidence to content choice, grounding, generation control, and evaluation, without treating either country's system as a universal default.

Data availability. Baig (2026) reports aggregate results and an anonymised phase-level extract for 10 drivers. Participant-level UK data are not publicly released because of privacy constraints. Safe Drive Africa participant-level and trip-level data are likewise not publicly released because location traces, driving-behaviour records, and alcohol self-reports create privacy constraints; aggregate results are reported in the source study (Thompson, 2025).

9 Limitations

This is a comparative synthesis of two source studies, not a preregistered cross-country experiment. The systems differ in recruitment, duration, sensors, feedback exposure, event definitions, and statistical analysis. The requirements samples cannot represent all drivers in either country, and the four Nigerian follow-up interviews particularly limit claims about reception. The alcohol classifier predicts trip-level patterns associated with self-reported consumption; it is not a breathalyser or legal determination of impairment. The paper does not independently evaluate generated-text factuality, cultural resonance, or whether each content strategy would outperform the other in the same country.

10 Ethical Considerations

Driver monitoring raises privacy, surveillance, and fairness concerns. Both projects obtained institutional ethical approval and informed consent. Location traces, behavioural profiles, and alcohol self-reports are sensitive; deployment beyond research would require data minimisation, transparent access and retention policies, meaningful opt-out, and protection against punitive secondary use, consistent with empirical privacy preferences and regulatory analyses of automotive telematics (Dowthwaite et al., 2024; Truby et al., 2024). Legal content also requires versioning and review when regulations change. The analysis consequently reports study-specific evidence and uncertainty without associating observed knowledge or infrastructure gaps with nationality itself.

Acknowledgments

Peter Thompson's doctoral research was funded by the Tertiary Education Trust Fund (TETFund), Nigeria, with institutional support from Niger Delta University, Nigeria.

References

Hossein Aghayari, Leila R. Kalankesh, Homayoun Sadeghi-Bazargani, and Mohammad-Reza Feizi-Derakhshi. 2024. Functionalities of mobile solutions for preventing road traffic health and safety issues in the context of information management cycle. *Transportation Research Record: Journal of the Transportation Research Board*, 2678(4):649–658.

Icek Ajzen. 1991. The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2):179–211.

Burcu Arslan and Türker Özkan. 2024. Role of culture, income level and governance quality on driver behaviours. *Journal of Road Safety*, 35(3):10–23.

Jawwad Baig. 2026. *Personalised Data-to-Text Feedback for Safer Driving*. Ph.D. thesis, University of Aberdeen. Doctoral thesis manuscript.

Jawwad Baig, Guanyi Chen, Chenghua Lin, and Ehud Reiter. 2022. DrivingBeacon: Driving behaviour change support system considering mobile use and geo-information. In *Proceedings of the First Workshop on Natural Language Generation in Healthcare*, pages 1–8. Association for Computational Linguistics.

James Boylan, Denny Meyer, and Won Sun Chen. 2024. A systematic review of the use of in-vehicle telematics in monitoring driving behaviours. *Accident Analysis & Prevention*, 199:107519.

Daniel Braun, Ehud Reiter, and Advaith Siddharthan. 2015. Creating textual driver feedback from telemetric data. In *Proceedings of the 15th European Workshop on Natural Language Generation (ENLG)*, pages 156–165. Association for Computational Linguistics.

Daniel Braun, Ehud Reiter, and Advaith Siddharthan. 2018. Saferdrive: An nlg-based behaviour change support system for drivers. *Natural Language Engineering*, 24(4):551–588.

Emma Croxford, Yanjun Gao, Nicholas Pellegrino, Karen Wong, Graham Wills, Elliot First, Frank Liao, Cherodeep Goswami, Brian Patterson, and Majid Afshar. 2025. Current and future state of evaluation of large language models for medical summarization tasks. *npj Health Systems*, 2(1):6.

Anna Dowthwaite, Dave Cook, and Anna L. Cox. 2024. Privacy preferences in automotive data collection. *Transportation Research Interdisciplinary Perspectives*, 24:101022.

Jeffrey P. Ebert, Catherine C. McDonald, Ruiying A. Xiong, Dina Abdel-Rahman, Neda Khan, Maya Irving, Athena Lee, Arjun Patel, Sadie Friday, Brian Tefft, William J. Horrey, Subhash Aryal, Michael O. Harhay, and M. Kit Delgado. 2026. A national randomized field experiment testing the impact of simulated usage-based insurance programs on driver safety. *Accident Analysis & Prevention*, 225:108331.

Anderson Etika, Natasha Merat, and Oliver Carsten. 2021. Identifying salient beliefs underlying speeding behaviour: An elicitation study of nigerian drivers. *Transportation Research Interdisciplinary Perspectives*, 9:100279.

BJ Fogg. 2009. A behavior model for persuasive design. In *Proceedings of the 4th International Conference on Persuasive Technology*, pages 1–7. ACM.

Guangyuan Gao, Shengwang Meng, and Mario V. Wüthrich. 2022. What can we learn from telematics car driving data: A survey. *Insurance: Mathematics and Economics*, 104:185–199.

Omid Ghaffarpasand, Mark Burke, Louisa K. Osei, Helen Ursell, Sam Chapman, and Francis D. Pope. 2022. Vehicle telematics for safer, cleaner and more sustainable urban transport: A review. *Sustainability*, 14(24):16386.

Armira Kontaxi, Apostolos Ziakopoulos, and George Yannis. 2025. Exploring the impact of driver feedback on safety: A systematic review of studies in real-world driving conditions. *Transportation Research Part F: Traffic Psychology and Behaviour*, 114:118–140.

Hong-Jie Li, Xing-Gang Luo, Zhong-Liang Zhang, Wei Jiang, and Shen-Wei Huang. 2023. Driving risk prevention in usage-based insurance services based on interpretable machine learning and telematics data. *Decision Support Systems*, 172:113985.

Chen Cecilia Liu, Iryna Gurevych, and Anna Korhonen. 2025. Culturally aware and adapted NLP: A taxonomy and a survey of the state of the art. *Transactions of the Association for Computational Linguistics*, 13:652–689.

Lynn Meuleners, Michelle Fraser, Mark Stevenson, and Paul Roberts. 2023. Personalized driving safety: Using telematics to reduce risky driving behavior among young drivers. *Journal of Safety Research*, 86:164–173.

Eva Michelaraki, Christos Katrakazas, George Yannis, Ashleigh Filtness, Rachel Talbot, Graham Hancox, Fran Pilkington-Cheney, Kris Brijs, Veerle Ross, Hélène Dirix, An Neven, Roeland Paul, Tom Brijs, Petros Fortsakis, Eleni Konstantina Frantzola, and Rodrigo Taveira. 2021. Post-trip safety interventions: State-of-the-art, challenges, and practical implications. *Journal of Safety Research*, 77:67–85.

Jekaterina Novikova, Ondřej Dušek, Amanda Cercas Curry, and Verena Rieser. 2017. Why we need new evaluation metrics for NLG. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2241–2252. Association for Computational Linguistics.

Chinonso Cynthia Osuji, Thiago Castro Ferreira, and Brian Davis. 2024. A systematic review of data-to-text NLG. *arXiv preprint arXiv:2402.08496*.

Kiemute Oyibo, Ifeoma Adaji, Rita Orji, Babatunde Olabenjo, and Julita Vassileva. 2018. Susceptibility to persuasive strategies: A comparative analysis of nigerians vs. canadians. In *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization*, pages 229–238. ACM.

Angèle Picco, Arjan Stuiver, Joost de Winter, and Dick de Waard. 2023. The use of monitoring and feedback devices in driving: An assessment of acceptability and its key determinants. *Transportation Research Part F: Traffic Psychology and Behaviour*, 92:1–14.

Juan Quintero, Victoria Karaseva, Freya Gassmann, and Zinaida Benenson. 2023. User acceptance factors of usage-based insurance. In *Advances in Intelligent Traffic and Transportation Systems*.

Ehud Reiter and Robert Dale. 2000. *Building Natural Language Generation Systems*. Cambridge University Press.

Ehud Reiter, Roma Robertson, and Liesl M. Osman. 2003. Lessons from a failure: Generating tailored smoking cessation letters. *Artificial Intelligence*, 144(1–2):41–58.

Jackeline Rios-Torres and Andreas A. Malikopoulos. 2016. An overview of driver feedback systems for efficiency and safety. In *2016 IEEE 19th International Conference on Intelligent Transportation Systems*, pages 667–674. IEEE.

Mandar Sharma, Ajay Kumar Gogineni, and Naren Ramakrishnan. 2024. Neural methods for data-to-text generation. *ACM Transactions on Intelligent Systems and Technology*, 15(5):1–46.

Pushpdeep Singh, Mayur Patidar, and Lovekesh Vig. 2024. Translating across cultures: LLMs for intralingual cultural adaptation. In *Proceedings of the 28th Conference on Computational Natural Language Learning*, pages 400–418. Association for Computational Linguistics.

Klaire Somoray, Katherine M. White, Barry Watson, and Ioni Lewis. 2024. Predicting risky driving behaviours using the theory of planned behaviour: A meta-analysis. *Accident Analysis & Prevention*, 208:107797.

Mark Stevenson, Anthony Harris, Jasper S. Wijnands, and Duncan Mortimer. 2021. The effect of telematic based feedback and financial incentives on driving behaviour: A randomised trial. *Accident Analysis & Prevention*, 159:106278.

Iniakpokeikiye Peter Thompson. 2025. *Using an AI App to Encourage Safer Driving in Non-Western Countries: A Nigerian Context*. Ph.D. thesis, University of Aberdeen.

Iniakpokeikiye Peter Thompson, Dewei Yi, and Ehud Reiter. 2025. An end-to-end system for culturally-attuned driving feedback using a dual-component NLG engine. *arXiv preprint arXiv:2509.04478*.

Jon Truby, Rafael Dean Brown, and Imad Antoine Ibrahim. 2024. Regulatory options for vehicle telematics devices: Balancing driver safety, data privacy and data security. *International Review of Law, Computers & Technology*, 38(1):86–110.

Uchenna Nnabuihe Uhegbu and Miles R. Tight. 2021. Road user attitudes and their reported behaviours in abuja, nigeria. *Sustainability*, 13(8):4222.

Chinebuli Uzondu, Samantha Jamson, and Greg Marsden. 2022. Road safety in nigeria: Unravelling the challenges, measures, and strategies for improvement. *International Journal of Injury Control and Safety Promotion*, 29(4):522–532.

Chris van der Lee, Albert Gatt, Emiel van Miltenburg, Sander Wubben, and Emiel Krahmer. 2019. Best practices for the human evaluation of automatically generated text. In *Proceedings of the 12th International Conference on Natural Language Generation*, pages 355–368. Association for Computational Linguistics.

Daniel Vankov, Ronald Schroeter, and Andry Rakotonirainy. 2023. Provisional drivers intend to speed less: The positive outcome for young drivers of a safe-driving app randomised trial. *Transportation Research Interdisciplinary Perspectives*, 21:100877.

Ian Warren, Andrew Meads, Robyn Whittaker, Rosie Dobson, and Shanthi Ameratunga. 2018. Behavior change for youth drivers: Design and development of a smartphone-based app (backpocketdriver). *JMIR Formative Research*, 2(2):e25.

Sandra Williams and Ehud Reiter. 2008. Generating basic skills reports for low-skilled readers. *Natural Language Engineering*, 14(4):495–525.

Lucy Yardley, Leanne Morrison, Katherine Bradbury, and Ingrid Muller. 2015. The person-based approach to intervention development: Application to digital health-related behavior change interventions. *Journal of Medical Internet Research*, 17(1):e30.