



UK telematics apps

What 21,994 app-store reviews show

*A study of 19 UK operators, insurance and fleet.
We Are Systematic, 2026.*

Contents

	Foreword	04
	Executive summary	06
1	The two markets telematics serves	11
2	Method	13
3	The ratings say 4.5. The reviews say 2.	16
4	Everyday usage is the biggest problem	19
5	Consumer insurance: the customer impact of missed data	22
6	Fleet: the driver doesn't choose the app	27
7	AI has arrived, mostly between the customer and a person	30
8	When a driver disputes a score	32
9	Jobs-to-be-Done	34
10	What can be learned from By Miles	41
11	The strategic read: advantage rests on a record customers trust	44
12	Closing: ten questions to ask your product team	49
	About the author	51
	Appendix A1–A6: method, ratings, statistics, audits, sources, framework	52

■

"I like this app as it gives me the confidence that I can see how safely my daughter is driving and provides feedback for her to ensure she's driving safely."

Apple, insurer app, 5★, April 2022

■

"I followed all the instructions and after I check the trips are not recorded... I did everything right. And without any notice, you cancelled my policy... It isn't fair."

Google Play, insurer app, 1★, May 2024

■

"I had 5 drivers that were good quit on me after 3 months of using this."

Google Play, fleet app, 1★, August 2025

Foreword

We publish research like this because it's an example of the same evidence we use in our client work. When we help a product team decide what to build, what users actually say and do is fundamental.

Every month or two, we point that method at a whole category and publish what it finds. We've done it for [EV charging](#), [car companion apps](#) and [airline apps](#). This time it's telematics.

We do much of our work in automotive and mobility, and we're drawn to complex product ecosystems in particular, which made telematics a natural fit.

I want to be open and intentional about what this report is and isn't. Whilst the review data discussed can be very negative, this is not a teardown and our aim at all times was to write a fair assessment of the industry based on the user review data.

Telematics does something worthwhile, it offers insurance to younger or high premium drivers at a more affordable price point, it makes fleets safer and keeps them legal, and at its best it rewards good driving with money in people's pockets.

We studied this category because it matters, not to revel in bad reviews. The 21,994 reviews in this study include people who love these products, and you'll find them here too, including a whole section on one operator that got it particularly right.

I also have to declare an interest of sorts. Between January 2016 and March 2017 I was a product manager for SmartMiles at Hastings Direct, one of the operators in this study.

However, while the review record analysed here begins in 2011, every Hastings review in this study was written after I left.

I remember from that time a continuous tension between stakeholders; the product team wanted to build an engaging customer proposition that raised a grudge purchase to the level of smart-home tech, and saved some lives along the way through improved driving behaviour.

The underwriters wanted every hint of risk off their book, and saw the customer relationship as more of a 'nice-to-have'.

At the time, the unifying case study in every planning conversation was Vitality: proof that an insurance product (the thing you buy and hope never to use) can become a relationship people engage with, to the underwriter's benefit as much as the customer's.

That was the ambition for telematics then, and reading this report, it is interesting to reflect back on the gap between that ambition and the median customer experience we're now describing a decade later.

It is normal for us to caveat that review data tends to be polarised with a negative skew. That skew does not explain away the dissatisfaction these reviews describe.

This might be predictable, as issues here hit the customer's wallet harder and more directly than in other industries. But it is also true that the people it hits hardest are usually the ones with the fewest choices.

That's the really uncomfortable takeaway from reading the raw reviews. So many people describe doing everything right, only to be told they aren't.

If you want to check our working, the method, the audit trails and the named ratings table are in the appendix. We also flag where we've inferred something rather than measured it directly.

If you run one of these products, the ten questions at the end are a good place to start. Most should be answerable quickly from your own data, and will allow you to bring the report's findings into your own specific context.

Ed Roberts

We Are Systematic, 2026

Executive summary

Telematics products do not just serve their users, they monitor and judge them. Premiums, insurance cover, legal working hours and jobs ride on what these apps record and score.

We analysed 21,994 UK app-store reviews across 19 telematics operators, insurance and fleet. As far as we can establish, it is the first published study to mine organic end-user voice at scale in this category.

Our central finding is not that these products are disliked, though many are.

It is that the most consequential friction sits around the judgement itself, even though the largest volume of complaints is about getting in. The judgement is also where durable advantage in this market becomes available: a telematics business can cultivate an asset a competitor cannot simply copy, a driving record its customers believe and trust.

In our reading of the evidence, that trust rests on four pillars, each of which shows up in what customers write:

1. Access, whether the customer can get into the product at all.
2. Accuracy, whether the record matches what they actually did.
3. Transparency, whether they can see what was measured and challenge it.
4. Proportionality, whether they are penalised only for things they would accept as their own.

Where trust holds, we think it makes two advantages available:

- Switching costs, because a verified driving history is something a customer loses by leaving.
- Brand equity, because being known for judging people fairly is one of the ways to stand out on a price comparison site.

Section 11 makes the full case. The sections between show where each pillar holds and where it fails.

On this evidence that potential asset is under strain at nearly every operator we studied. That makes this a strategy problem before it is a software problem.

Telematics promises fairer prices for safe drivers, affordable cover for younger ones and safer and legal fleets. The technology is also evolving: policies have been shifting from black boxes to app-only measurement, and AI is rolling out in the form of customer service chatbots, all changes visible in the reviews themselves.

Yet the first half of 2026 is the worst period in the reviews we analysed, and the same problems repeat across eight years and nearly every operator studied.

Three facts frame the report as a whole:

1. Public measures of these products tend to read well while the recent written record reads badly. Store headlines for the major consumer apps run 4.5 to 4.7, company-level Trustpilot scores sit in a similarly high band, but recent written reviews for the same apps run 1.5 to 2.4.
2. The trend is not noticeably improving. The yearly mean has moved between 2.3 and 2.9 since 2018, and the first half of 2026 is the worst period in that record.
3. Many of the people inside that record cannot simply leave. Monitored cover is often the only affordable route for young drivers, and fleet drivers carry whatever app their employer chooses.

Those facts raise a question, the answer to which senior leadership at telematics operators should be interested in: why does sustained user dissatisfaction matter, if the category's basic bargain appears to be holding, with cheaper insurance for the customer and better loss ratios for the insurer?

Not mainly because of customer churn (in fact, in fleet scenarios it is *driver* churn that should be more concerning). Some dissatisfied customers will leave but many can't.

It matters because of where the dissatisfaction sits. When the product records wrongly, in insurance what weakens is the trusted-data asset telematics pricing rests on, in fleet it is the critical relationship with drivers.

When insurance users disengage, little may show inside the policy year, so the exit happens at renewal via a price comparison site, looking like price-shopping rather than product feedback.

Zoomed out, everything looks OK, but specific frictions are lost in the average.

This is not a homogeneous market. Section 1 maps the landscape: two worlds, consumer insurance and fleet, and within them apps that report on a self-contained device, apps tethered to a tag or box, and apps that are themselves the sensor. Each sub-category has its own findings but several reach across all.

At the major consumer insurers, the headline ratings say 4.5 and the written reviews say 2. Across operators with at least thirty recent written reviews to compare, the public store rating sits a median 2.2 stars above what recent written reviews describe.

For the major consumer operators, company-level Trustpilot scores sit in a similarly high band of 4.4 to 4.5, two to three stars above the recent written app store record.

The biggest complaint in the written record is getting into the product at all.

Login and reliability problems run at 42 mentions per 100 digital reviews, and at 29 on telematics-related reviews alone, the largest complaint on either count. Four in five of the reviews raising a login problem are one-star reviews, against just over half across the digital corpus.

Reviews describing these problems talk about deleting or uninstalling the app at roughly twice the baseline rate, but mention cancelling no more than other reviewers do. Section 4 sets out the hypothesis this points to: that customers keep the necessary policy while minimising use of the app.

Those exits won't necessarily show in the public ratings, and we'd be interested to know whether the signal is captured internally.

We extracted and scored nineteen user Jobs-to-be-Done, twelve of them clearing our evidence floor.

Telematics as a category seems to fit best at the edges. Every one of the twelve confirmed Jobs is underserved, three of them highly so.

Five Jobs reach served right or better, three of them on the borderline, and every one of them sits in the below-floor trends set rather than the confirmed set: a parent reassured about a young driver (two Jobs), paying only for miles driven, vehicle tracking, and car-health diagnostics.

None involves the customer's own score, and four of the five are each concentrated at one operator. We report them because the direction is consistent and interesting, not because the evidence is conclusive.

In consumer insurance, when the product misses data, customers describe getting the blame. Device pairing, journey recording, scoring and premium consequences behave as one coupled system in the reviews.

When the sensing layer drops a journey, the letters customers report receiving consider the gap to be theirs, warnings for "not providing data", up to threatened cancellation. Drivers describe being scored as if driving when they were the passenger, and marked down for phone use when the phone was plugged in playing music.

Across 10,442 reviews we found only two describing a successful challenge: a cancellation reversed after, the reviewer says, the insurer admitted its own systems were at fault, and a 93mph reading removed after a parent queried it as impossible. The correction route reviewers mention most often is narrower, the option to reassign who was driving.

In consumer insurance, AI has arrived almost entirely as a chatbot standing as a barrier between the customer and human assistance, at the moments the product has already let them down. Mentions of AI have risen roughly sevenfold since 2023, most of them from insurance customers, and almost none describe AI helping.

Meanwhile the app-only transition concentrates the sensing job in the customer's phone, and the phone-based policy already draws its own complaint genre, in which battery cost and background running are experienced as the product itself.

Both are significant technology choices with real trade-offs, and the reviews record how those trade-offs feel from the customer's side.

In fleet, app defects disrupt drivers' working hours. Professional drivers describe duty status flipping while they sleep and rest being redone because the app lost it.

In the UK the tachograph holds the legal record, but app defects could still expose the operator through drivers'-hours compliance and risk to its good reputation, the standing that determines whether it keeps its operating licence. One reviewer describes an update locking everyone in their company out of the app.

Issues are not unique to drivers. The manager-facing mobile apps are the worst-rated role in the study, on the only surface of their tools our data can see.

One operator showed the standard is achievable. By Miles matched a 4.7-star lived experience to its public reputation, on the same phones.

It is being wound down following its parent's acquisition, and section 10 takes the lessons of that case study further. A second, Marmalade, does score driving behaviour where By Miles only tracked miles driven, and among operators with at least 100 digital reviews it has the second-lowest reliability complaint rate and the second-highest share of four- and five-star reviews, after By Miles.

Its recent written sample is too small to compare on the same measures, so section 10 treats it as corroboration rather than a second case.

Most of this does not require new technology to improve. Section 9 scores the user needs by size of opportunity, section 11 sets out the strategic case for acting on them, and section 12 turns the findings into ten questions senior leadership can put to its product team.

Most should be answerable from data already held, and the diagnostic will help surface both the reason this gap has persisted and which of the two advantages might be available.

SECTION 1

The two markets telematics serves

Telematics turns driving into data. A sensor, a black box wired into the car, a tag on the windscreen, or increasingly the phone in the driver's pocket record journey data, typically including speed, braking, time and location.

Data and insights are reported back to the driver, insurer and/or business.

In the consumer world, telematics is mostly insurance. Insurers promise to price cover on how a person actually drives rather than on proxies such as age and postcode, which is why the category exists at all.

For young and high-premium drivers, monitored cover is often the only affordable route onto the road. Behind the insurer's brand there is often a specialist provider running the data platform.

For the policyholder, the app is probably the whole relationship with the insurer. It is where journeys appear, where the driving score lives, and where any discount is earned or lost. The app's part in the measuring takes three main forms:

1. A self-contained device does the measuring and the app only reports back results.
2. The app is tethered to a tag or box that cannot record without the phone paired beside it.
3. In the newest app-only policies, the phone is the sensor and the app is the measurement.

The fleet world sells to businesses. A telematics platform gives an operator its vehicles' whereabouts, its fuel and efficiency numbers, its safety record, and compliance factors (drivers' hours, walkaround checks, the digital paperwork of keeping vehicles legal).

As is often the case in B2B, the software buyer is not the primary user. Transport managers or procurement teams choose the platform.

Drivers and operational teams carry the app because their employer requires it.

Priced fairly, good drivers save money. Monitored well, fleets run safer, cheaper and legally.

When these products work, insurers' loss ratios, drivers' premiums and everyone's road safety all move the right way.

In the consumer world there is another important dimension for the insurer:

Insurance is a product people buy hoping never to use. If a claim is never made the whole relationship is confined to the welcome and renewal invitation correspondence, separated by a year.

Telematics gives an insurer something it has never otherwise had, a legitimate, useful place in the customer's day.

Health insurance has already shown what that can become.

Vitality is an example of turning a grudge purchase into a relationship people engage with and value, to the underwriter's benefit as much as the member's; a RAND Europe study of more than 400,000 members found those on its Apple Watch benefit sustained a 34% increase in physical activity.

A telematics app, built well, is motor insurance's version of that chance.

But the promise of a valued customer relationship makes the app something more than a convenience. What it records sets the premium a young driver pays, decides whether cover continues, and feeds the record of a professional driver's working hours.

When a shopping app misbehaves, someone shops elsewhere. When a telematics app misses a journey, someone is judged on data that is wrong, and that can mean cancelled policies, lost discounts, redone rest, jobs put at risk.

And unlike shoppers, many of these customers cannot walk away.

SECTION 2

Method

We collected 22,003 app-store reviews (Apple App Store and Google Play, UK storefronts) across 19 UK-relevant telematics operators: seven insurance and telematics-insurance brands and twelve fleet platforms. Apple reviews were collected in two captures (13 and 20 July 2026).

Operators were selected by market coverage, the providers a UK driver is most likely to encounter with the most prominent app-first challengers included to capture where the underlying technology is heading.

After cleaning, 21,994 reviews remained. 10,442 of them (47.5%) are digitally relevant, and those form the analytical core. More than half of them (5,776) come from Admiral's and Hastings Direct's main customer apps, which carry each insurer's telematics product alongside its other cover. Most of those reviews do not mention telematics; where that changes a figure, we say so. Google Play supplies full review history.

Apple's feed caps at roughly the 500 most recent reviews per app, so long-run trends lean on Play.

Reviews were themed with keyword classifiers. Every classifier behind a published figure was precision-audited on random samples, and those that failed their thresholds were tightened, relabelled or withdrawn.

Audit records sit with the study data and are available on request, along with classifier definitions. We state clearly where a claim is inference rather than measurement.

Public headline figures were captured from store listings and Trustpilot on 13 July 2026.

The strategic analysis in section 11 applies Helmer's 7 Powers and a Porter-based read of demand structure. The full audit, including the powers that do not apply in this market and why, is summarised in appendix A6 and retained with the study data.

We did not approach the operators before publication. Everything here is drawn from public sources, the analysis is of what their customers wrote rather than of anything they told us, and we judged that a right-of-reply process would change what this report is.

Factual corrections go to hello@wearesystematic.com and we review and correct them promptly.

There are three honest limitations to flag upfront:

1. Written reviews skew negative everywhere, so this study's comparisons are between like and like. Written reviews against written reviews, cohort against cohort, year against year. The gap analysis compares public headline ratings with the recent written record.
2. The strongest objection to any study looking at online reviews is that the angriest reviewers are the ones the product has judged, the one-star review as revenge for a cancellation or a bad score. Three features of this data answer that:
 - i. The bulk of the complaints concern getting into the product at all, before any judgement has been made.
 - ii. Access complaints and judgement complaints barely appear in the same reviews, so one is not driving the other.
 - iii. One operator's book runs against the trend, showing that the reverse of the common pattern is clearly possible (section 10).

What remains true is that no individual reviewer's driving can be verified from their review, which is why section 5 rests on mechanisms and the industry's own published error rates, not on any single driver's word.

3. Nothing here contradicts aggregate success. A telematics programme can genuinely reduce claims across a book while individual customers inside it are mis-scored, locked out or wrongly blamed. Population averages and individual experiences answer different questions. This report is focussed on customer experience of the product, and we can grant at the same time that the aggregate benefits are real.

THE COHORT

The 19 telematics operators in this review

The UK's largest telematics providers by coverage, plus the most prominent app-first challengers.

INSURANCE

- 1 **Admiral**
- 2 **Hastings Direct**
- 3 **Zego** *
- 4 **By Miles**
- 5 **Marmalade**
- 6 **Carrot Insurance**
- 7 **insurethebox** †

FLEET

- 8 **Samsara**
- 9 **Geotab**
- 10 **Webfleet**
- 11 **Verizon Connect**
- 12 **Teletrac Navman**
- 13 **Quartix**
- 14 **RAM Tracking**
- 15 **ABAX**
- 16 **Radius (Kinesis)**
- 17 **Michelin Connected Fleet**
- 18 **Trakm8**
- 19 **Crystal Ball**

* Two brand lines in our data: general cover and Zego Delivery, its courier product, treated as its own persona in section 9.

† No app; Trustpilot presence only.

Selected by ranked UK market coverage, since no operator publishes UK-only installed-base or policy figures; full method in appendix A1.

SECTION 3

The ratings say 4.5. The reviews say 2.

Across the digital subset the mean review score is 2.4 out of 5. More than half of all digital reviews (54.6%) are one star. A quarter are five stars.

Only 11.6% sit at three or four.

The distribution is what we would consider 'normal' for online reviews: a large angry peak and a smaller delighted one, with almost nobody in between.

These products do work for some people.

Thirty per cent of the reviews we analysed are four or five star, and the satisfied reviewers tend to describe the same few things: the app shows them the journey it scored, it explains what it measured, the feedback arrives quickly enough to act on, and it does not mark them down for where or when they drove.

"This is genuinely the best black box I've ever heard of... There is no curfew, reports are available straight away after you finish driving and the areas that you're assessed on in your drive make sense... it does not punish you for things like where or when you drive!"

Apple, insurer app, 5★, April 2022

Reviews like that appear at more than one consumer insurer in this study, so the pattern is not the property of one operator or one architecture. It is what these products look like when they work.

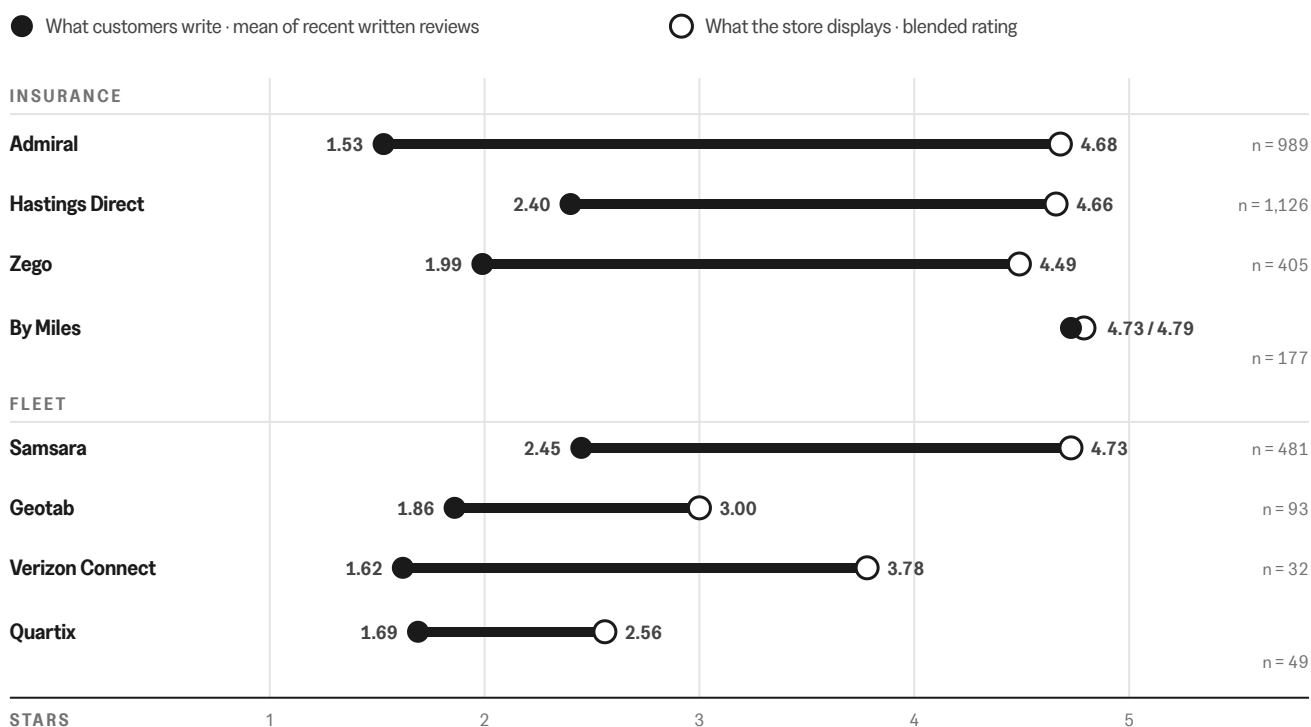
However, this report is about what it costs operators when they don't.

Set that against what the market displays: store headline ratings for the major consumer apps run between 4.5 and 4.7. The recent written record for the same apps runs between 1.5 and 2.4.

(By Miles scores 4.8 and 4.7 on the same two measures, see section 10).

The median gap, across operators with at least thirty recent written reviews to compare, is 2.2 stars. The largest gap in the cohort is around three stars.

What the store displays vs. what customers write



Blended store rating as displayed, against the mean of each operator's recent written reviews

UK telematics apps review study, We Are Systematic, 2026.

The obvious objection is that these are different populations. Headline ratings include millions of silent taps, written reviews are the self-selected angry.

That is exactly why the By Miles exception matters, where headline and written scores match to within a tenth of a star on the same constructs, which suggests the gap is not an artefact of how ratings work, though By Miles did not score driving (section 10). Written reviews can match a headline rating but most operators are nowhere near it.

"What's the point? You download the app, accept the terms and conditions just to try and login and get a message saying ERROR... 2026 and you can't even get an app that works."

Google Play, insurer app, 1★, February 2026

"As a new customer, I cannot yet review the insurance company, but I have now had a couple of weeks usage of their app. An app which crashes & freezes regularly."

Apple, insurer app, 3★, October 2025

For the major consumer operators, company-level Trustpilot scores run 4.4 to 4.5, a similarly high band sitting far above the written record.

The trend deepens the problem: the yearly mean has sat between 2.30 and 2.88 every year since 2018. Low and not recovering.

The first half of 2026, to 18 July, is the worst period in the record, at 2.04. A yearly average that looked stable for eight years has now dropped below its range.

STRATEGIC READ

A headline rating this far above the lived record offers little warning that anything is wrong. A healthy headline number is easy to read as reassurance rather than a reason to look closer, and none of the four pillars of trust is visible in it.

That matters most for brand equity: an operator cannot protect a reputation for fair treatment that it cannot see eroding.

SECTION 4

Everyday usage is the biggest problem

Before any product can be used, customers have to get in. Across business and consumer scenarios, under every architecture, being unable to do that is the complaint that towers over everything else.

Login and access problems appear in 18.0% of all digital reviews (mean rating 1.37). Reliability and performance problems appear in 23.7% (mean 1.44). Together they run at 42 mentions per 100 digital reviews. On telematics-related reviews alone it is 29 per 100, still well ahead of any other complaint (see section 2).

More than a third of all digital reviews raise at least one, and the two often co-occur.

Over a third of login complaints also describe an app that will not run. Within these reviews the one-star share runs above three quarters, compared to the 54.6% baseline average.

"It randomly logs out of that app, refuses any attempt to log back in and when I successfully change password the app then tells me it doesn't recognise the password."

Google Play, insurer app, 1★, October 2024

Of 394 reviews describing opaque error states, unexplained codes, "an error occurred", or dead ends with no recovery route, 384 are insurance-side.

The reviews don't say why, but fleet apps tend to fail outright, crashing or freezing. The insurers' main apps tend to keep running but throw an unexplained error with no way forward.

"Even simple error messages are in terms of error codes that you cannot decipher. Looks like it has been designed by a bunch of amateurs."

Google Play, insurer app, 1★, January 2020

For some users, working around the app's problems has become just part of normal use. One reliability review in fourteen describes uninstalling and reinstalling the app, or clearing caches, as routine maintenance:

"I have to uninstall it and reinstall it probably twice a week at least."

Google Play, fleet app, 1★, September 2018

And at least in the reviews, the response from customer support follows a pattern. Users are advised to just keep trying, support conversations are closed before they're resolved, and users are left to start again.

"Spoke to a gentleman on customer service who advised to just keep trying."

Google Play, insurer app, 4★, October 2025

"Reported error to online agent and he was creating a case for me but during that the chat was closed and I can't access/open the chat anymore. I was asked to activate the device in 2 weeks but how can I do this now???"

Google Play, insurer app, 1★, May 2026

The most consequential finding here is the evidence pointing to disengagement.

Reviews describing reliability problems mention deleting or uninstalling the app at roughly twice the baseline rate, but mention cancelling the policy no more than other reviewers do.

Many customers opt for telematics policies because it is the only cover they can afford. The policy is necessary but the technology is a friction to be endured.

So when the product lets them down, they are no more likely to say they're leaving, but may already have lost considerable faith in the provider. If that is what's happening, the only daily engagement channel motor insurance has ever had is at risk.

There is a second type of disengagement, well documented but absent from our review data. Industry commentary reports heavy drop-off in telematics app engagement (Insurance Business UK, 2023).

Scores settle within a few weeks of ordinary driving, and once the number stops moving, there is little reason to open the app again. Someone who drifts for this reason is perhaps less inclined to write a review, which would explain why it doesn't show in our data.

An operator's own usage analytics might show the decline, though not the reason or its link to the renewal decision that follows.

Disengagement from the app doesn't necessarily mean interruption of data collection. On device-based policies the box records regardless, and what fades is the relationship.

In the case of phone-based policies where the app is the sensor, gaps in the record are likely to be detected, and customers describe being penalised for them (section 5).

We cover the data accuracy question in section 5 but the issue at stake here is the health of the relationship. The actual impact is deferred to renewal, largely not felt inside the policy year.

The book can look stable, and when a customer switches, it's hard to tell whether that's price-shopping or product feedback.

STRATEGIC READ

Access is the first of the four pillars, and the one the others depend on. A customer who can't use their app has nothing to trust and generates less data, so both switching costs and brand equity assume a working front door.

It is also a hard failure to see from inside, because on our reading of this evidence many of the people it affects keep the policy and stop opening the app. Inside the policy year that need not produce a cancellation or a formal complaint, only a fall in app use that an operator sees only if it measures it.

SECTION 5

Consumer insurance: the customer impact of missed data

Within the scoring chain itself, complaint dimensions move together in two tight clusters joined at the score.

The sensing dimensions, device pairing, journey recording, battery and background behaviour, co-occur at four to ten times the rate chance would predict.

Scoring and premium consequences co-occur at nearly thirteen times.

What doesn't couple directly is the premium and the sensing layer: customers connect the score to the price, and the sensing to the score, but they rarely name the sensing and the price in the same review.

The chain is joined in the middle, at the scoring. The industry's app-only transition is moving the sensing job into the customer's phone, and that architecture already has its own complaint genre (see the quotes below).

We were unable to robustly quantify the burden split between phone-sensing and device-sensing reviews, so this claim has to stay qualitative for now.

But whilst the causal direction is still an inference, the reviews do suggest a specific sequencing: the sensing layer drops, journeys record wrongly or not at all, scores go wrong, and this impacts pricing and cover decisions.

What makes this chain noteworthy is where the cost lands.

When the sensing layer misses journeys, customers describe the missing data being treated as their fault:

"Doesn't record 95% of my journeys. Then I receive emails and correspondence to say [the insurer] may cancel my insurance because I am not providing them with enough data!"

Google Play, insurer app, 1★, August 2022

Attribution problems follow the same pattern. The dominant fairness complaints in the review data concern being scored on driving that was not the reviewer's own:

"The app mark's me down for phone use when i am not even on my phone. my phone is either being plugged into the aux wire playing music, or being used for directions."

Apple, insurer app, 1★, June 2023

"Fundamentally there's a MASSIVE assumption that because my phone is moving, I must be driving my car."

Apple, insurer app, 2★, May 2023

"The app randomly decided that my braking was terrible out of the blue and halved my driving score to place me in the bottom 10% of drivers."

Google Play, insurer app, 1★, January 2025

The industry's own literature concedes some of this.

One vendor's published material describes distinguishing driver from passenger "in nearly 85% of cases" (2020).

A peer-reviewed validation study reports 91.2% specificity for non-driver trips, describing it as slightly lower and more variable than its sensitivity.

At the scale of a UK telematics book, even that small remaining percentage is not inconsequential to the people inside it. We couldn't find published data on how many UK policyholders are wrongly scored but the reviews suggest the number is not zero.

But the problem is perhaps more essential: even when the data is right, the product often cannot explain itself:

"Doesn't clearly explain exactly how the recording works, where in the evaluation process you are or what the discounts will be at time of renewal, if any."

Google Play, insurer app, 1★, December 2024

Searching the corpus for challenges that succeeded, we found only two. One reviewer says a major insurer moved to cancel their policy for a score too low while the app showed a perfect score, then admitted its systems were at fault and overturned the cancellation. They still gave one star, citing an unreliable app, a screenshot block that stopped them showing the insurer the problem, and the insurer's refusal to confirm the reversal in writing or to overturn a cancellation again. Separately, a parent at a young-driver insurer queried a 93mph reading they described as impossible, and was told it was inaccurate and had been removed.

A little way short of that sit two more:

- A fleet driver disputing the speed limits his app scored him against reports being told he can raise inaccuracies and have them corrected, and objects that the burden of auditing the measurement has been handed to the person being measured.
- An insurance customer describes a support team checking its own records, confirming his rating was above average, and finding in the process that the system had been calculating his mileage wrongly; he does not say whether anything followed.

The correction route reviewers mention most often is narrow, letting them reassign who was driving and nothing more.

One reviewer, at a young-driver specialist, Marmalade, describes an example of something better: an improvement system that lets users see where they went wrong and discuss it rather than have the policy terminated. A single review is not a pattern, and it describes a product feature rather than a successful appeal.

What it does show is the sequence that matters: a customer cannot challenge a score until they can see what it was based on.

Our data is limited to what customers describe, and they rarely describe contesting a score at all. We aren't claiming that no route exists anywhere in the apps studied, only that customers do not mention it.

That is the more useful observation, because it is what you would expect where the reasoning behind a score is not visible to them in the first place. Whether that absence is by design or by default is something an operator can settle for itself. Question 5 in section 12 is where to start.

"Have a dash cam so can prove otherwise yet no way to appeal just lose points for no reason."

Google Play, insurer app, 1★, June 2026

A missed journey, a wrong score and a warning letter arrive as one sequence for the customer. That sequence has reached the official record before, but only in fragments.

The Financial Ombudsman Service publishes guidance on telematics score and cancellation complaints, and in 2026 the Chartered Insurance Institute opened a consultation covering telematics use and how clearly it is explained to young drivers.

Our view is that these fragments belong together and should be assessed as a single customer journey or process chain.

A feeling of being watched

The emotional register running through these reviews is surveillance stress, closely linked to the physical evidence of battery-drain. Complaints of each co-locate at nearly nine times the baseline. Visible battery drain is a real-world reminder of being watched.

"This has made all driving a misery... The data collection is totally opaque... that is the source of the stress."

Apple, insurer app, 1★, April 2026

"I didn't consent to it recording my movements... got an Android notification saying 'trip recording suspended... low battery preventing [the app] from recording trips'. So it was running... Out. Of. My. Phone. You. Go."

Google Play, insurer app, 2★, February 2026

The app-only policy type is particularly susceptible to this problem. In one reviewer's characterisation:

"Avoid the spyware 'app based policy', drains battery quite quickly, constantly thinks your driving even when using public transport, trams and stuff."

Google Play, insurer app, 1★, July 2026

Accessibility

In 10,442 digitally relevant reviews our audited classifier found just five that discuss accessibility, and every one describes a barrier:

- An app "not designed for larger font sizes".
- Live chat unusable at small system fonts.
- Forced password re-entry called out as a burden on elderly drivers.
- A disabled customer citing the Equality Act's reasonable-adjustments duty.
- A visually impaired fleet app user, writing under US rules, reporting that the platform's "ADA compliance is nonexistent".

There is evidence that app-store reviewers under-represent disabled users. A 2025 study found doubts about whether feedback will be acted on among the barriers to reviewing at all.

Barriers also tend to stop use of the products before reviews get written, so absence of evidence here isn't evidence of absence. But for products whose users include older and disabled drivers on adapted vehicles, five mentions in 10,442 is an absence worth noting.

Accessibility is a legal duty as well as a product-quality question: one of those five reviewers invokes the Equality Act's reasonable-adjustments duty, and section 8 describes a UK mobility scheme that paused its telematics programme in 2026 after concerns were raised publicly that its scoring did not account for driving with hand controls.

STRATEGIC READ

Switching costs and brand equity both depend on trust in the record. This section covers two of the pillars that trust rests on, accuracy and transparency, and shows how they can fail together.

In our reading, accuracy is the part the industry talks about most publicly, and on this evidence that doesn't earn trust on its own: a customer who cannot see what was measured has no way to tell an accurate score from an inaccurate one, so investment in accuracy can earn less trust than it should. And the cost of a miss often falls on the customer, who is asked to supply data the product failed to collect. Both undercut the trust that either advantage depends on.

SECTION 6

Fleet: the driver doesn't choose the app

Fleet telematics carries a distinct kind of exposure: statutory compliance. For drivers in scope of the rules, mainly lorries and coaches, hours are a statutory record, and their rest is regulated to keep everyone safe.

In the UK that record lives in the tachograph, the device in the cab that logs driving time, and drivers'-hours rules are enforced by the DVSA and the Traffic Commissioners. The apps in our cohort work within that system as companions, handling duty logs, walkaround checks and hours alerts.

A quick note on vocabulary: some of the apps serve American drivers from the same UK app-store listing, so their reviews sometimes reference US rules: ELDs (electronic logging devices) and "hours of service".

We include them because these are the same products sold to UK fleets and the experience parallels are useful. The regulatory difference is that in the US, the app can itself hold the legal record, so a software defect might corrupt the record directly.

In the UK, as the tachograph holds the record, issues could instead risk drivers' hours exposure and the operator's good repute (the "licence-to-operate" standing the Traffic Commissioner assesses). Whichever regime is in play, we found defects tend to reach the driver with a similar shape:

"Changed me from sleeper berth to on duty WHILE I WAS SLEEPING AND THE TRUCK WAS OFF!! It's not robots driving these trucks, you know that right?!!... I had no choice but to redo the 10 hours. You don't know what you cost me. I missed something so damn important."

Google Play, fleet app, 1★, US market, June 2023

"Every time I log out of this app it is keeping me logged on duty and connected to the vehicle."

Google Play, fleet app, 1★, June 2025

"It'll notify me that I'm running out of HOS when I've been logged out for hours, but not that I'm about to go over my break time."

Google Play, fleet app, 1★, US market, May 2021

In our Jobs analysis (section 9), keeping working hours legal and duty status right is one of the two largest opportunities for fleet and delivery drivers, scoring 14 out of a possible 19.

Errors cost drivers rest, pay and standing. Satisfaction could hardly be lower.

It is worth noting that a single large platform contributes 61% of our fleet digital subset. We checked every fleet aggregate with and without it: the patterns hold, but some fleet averages are notably worse without it, so that platform lifts rather than drags the fleet figures we quote.

Fleet drivers are required to use these apps by their employers, and software updates arrive whether they want them or not. Of 75 complaints about app updates in our data, 55 are from fleet, and the widest-reaching case we found describes a whole company locked out:

"Ever since the update everyone in our company has never been able to log in."

Google Play, fleet app, 2★, January 2025

The cost is not abstract. In the words of one reviewer who manages drivers:

"I had 5 drivers that were good quit on me after 3 months of using this."

Google Play, fleet app, 1★, August 2025

And the buyers are not compensated with better tools themselves from what we can see: manager-facing apps are the worst-rated role in the study, and within fleet they sit at 2.17 against the 2.64 of the driver apps their staff are required to use.

As far back as 2018 a fleet manager was writing *"knowing where my vehicles were 3 hours ago with no ability to refresh is useless"*. In 2021 another was locked out of the live map on mobile while the desktop still worked.

The people who choose these products are not the people who use them all day, a misalignment of incentives that may weaken the pressure to fix it, though as complaints accumulate a product unloved by the workforce is on less sure footing at the next procurement cycle.

For UK operators the direction of travel raises the stakes. Compliance data that was once read off the device in the cab is increasingly collected and submitted digitally, under obligations phasing in from 2024 through 2026.

The more of that record passes through software, the more an operator's standing rests on software it did not write, and must trust implicitly. When drivers describe duty status they did not set and rest they had to take twice, trust is once again the asset at stake.

STRATEGIC READ

This section is about access and accuracy in a setting where the person bearing the failure is not the person who bought the product. Fleet's switching costs are real and held by the platforms. The evidence here shows how costs can fall largely on the workforce.

For the fleet buyer the exposure is in driver retention and compliance confidence. Section 11 gives fleet its own reading for exactly this reason.

SECTION 7

AI has arrived, mostly between the customer and a person

Mentions of AI in the data have risen roughly sevenfold since 2023 on a rate basis, from about five reviews across the whole of 2023 to around thirty in the first half of 2026, reaching 1.29% of 2026 reviews. Comparing 2023-24 against 2025-26 (raw counts 12 and 48), that's a 3.5x rate increase.

Of those 48 recent mentions, 37 are insurance customers and 11 are fleet.

Read the insurance mentions and they describe almost entirely one phenomenon: the "chatbot wall", AI deployed between customers and humans. At exactly the moments the product has already let them down, they describe the experience as deflection not help.

"I have been trapped in their live chat queue for over an hour trying to speak with a human agent to process a formal UK GDPR Data Erasure Request... The automated chatbot keeps cycling me through endless loops just to avoid connecting me to a real person."

Google Play, insurer app, 1★, June 2026

"The responses felt scripted, almost as if I was talking to an AI rather than a real person. I was getting responses similar to what we get when chat with Gemini."

Apple, insurer app, 1★, June 2026

"Keep trying to drive you to ineffective and inefficient chatbots. When they can't help you have to phone up and the first message you get is to tell you to go online."

Google Play, insurer app, 2★, June 2026

Distrust of AI in insurance is well documented in survey research (Insurance Business UK, 2026). A recent European survey of over 4,000 policyholders found more than a quarter say nothing would improve their confidence in insurers' use of AI.

Strong-anger vocabulary has also risen, to 1.12% of 2026 reviews against 0.45% in 2023-24.

The absolute numbers are still small (tens, not hundreds), which is why we can read it as a corroborating trend rather than a full finding.

For the organisations spending on AI in this category, the customer record offers a blunt audit. So far, the spend is visible to customers only as an unwelcome wall between them and a person.

STRATEGIC READ

This section is about transparency, and specifically about the half of it that decides whether a customer can act on what they have been told. Switching costs and brand equity both rest on trust, and trust is decided at the moments a customer needs something from an operator.

This is one of those moments, and on this evidence it is being increasingly staffed by a chatbot. The problem is not automation, which answers a real cost problem, but placement and effectiveness.

Many of the customers who reach it have already been let down by the product, and so need particular care, not deflection.

SECTION 8

When a driver disputes a score

Beyond the misattribution cases in section 5, we found an interesting thread that deserved its own discussion. There is a small but real cluster of reviewers who accept that the app measured their car's movement, but dispute what it concluded, pointing at a cause other than their own driving.

Among the various reasons offered, cornering (especially at roundabouts) is the most consistent single dispute in the pool. Thirty-five reviews across five operators describe ordinary cornering being scored as harsh.

Perhaps surprisingly, eight of them come from four- and five-star reviews, satisfied customers reporting the same experience as the unsatisfied, which makes it hard to dismiss as revenge.

"...after a couple months it has been fine, slightly annoying when going around a roundabout at 15mph is classed as harsh cornering however this has never resulted in a red journey so it doesn't really matter..."

Apple, insurer app, 4★, July 2022

The other causes vary and have small samples for each, and so they are best analysed as a thematic cluster:

- Braking hard for hazards, including a learner penalised for practising emergency stops and a reviewer asking for *"an option to include whether you had to break hard to avoid an accident"*
- Electric vehicles' greater torque and regenerative braking
- Road-data artefacts, *"speeding when it hasn't factored a new flyover (flyover 50mph limit, old road below 40mph limit)"*

What makes this more than a complaints list is that in many of these disputes, the algorithm may well be right.

- Greater torque produces faster acceleration, and an operator might well argue it is the driver who must adapt to the vehicle, not the measurement.
- An emergency stop is real deceleration, and the correct response to the hazard, but also could be the result of a moment of inattention.
- A 15mph roundabout may or may not exceed a sensible lateral threshold.

We cannot tell from review data who is right in any given case, and that is the point, because on this evidence neither can the customer. Section 5 found only two reviewers describing a successful challenge, and correction routes that otherwise mostly reassign who was driving.

The reviewer asking for "an option to include whether you had to break hard" is requesting a way to put context against a measurement. This may or may not be feasible, but zooming even further out, this is a question of clarity.

Greater transparency about why the scoring works the way it does might make all these complaints and requests redundant.

Each time the customer sees a data point that contradicts their lived experience, that data point becomes a grievance. We also see signs this increases with the changing vehicle parc: EVs, adapted vehicles and assisted driving all produce driving signatures likely to generate more of these disputes.

One recent example: in 2026 a UK mobility scheme paused its telematics programme after concerns were raised publicly that its scoring did not account for driving with hand controls.

STRATEGIC READ

This section is about two of the pillars, proportionality and transparency, and shows how closely they are connected: a problem in one becomes a problem in the other. Trust requires customers to understand and accept how their product's measurement corresponds to their own lived experience.

The advantage is not open to whoever scores customers most leniently, but to whoever can both measure accurately and show their working in a way the customer can understand and accept, if not agree with.

SECTION 9

Jobs-to-be-Done

Sections 3 to 8 describe what goes wrong in these products. This section asks what customers are trying to get done with them, and how well each of those Jobs is served. From 10,442 reviews we identified nineteen Jobs and scored each with the Ulwick opportunity method.

Twelve clear our evidence floor of fifty mentions and are reported as confirmed. The remaining seven sit below it, and four of them are also concentrated at one operator; they are scored on the same basis but reported separately as trends.

They are interesting enough to discuss and worth further research, but not a basis to build a strategy on.

We have further split the Jobs into two sections for consumer insurance and commercial (fleet and delivery), since the buyer and the stakes differ.

The Ulwick opportunity calculation

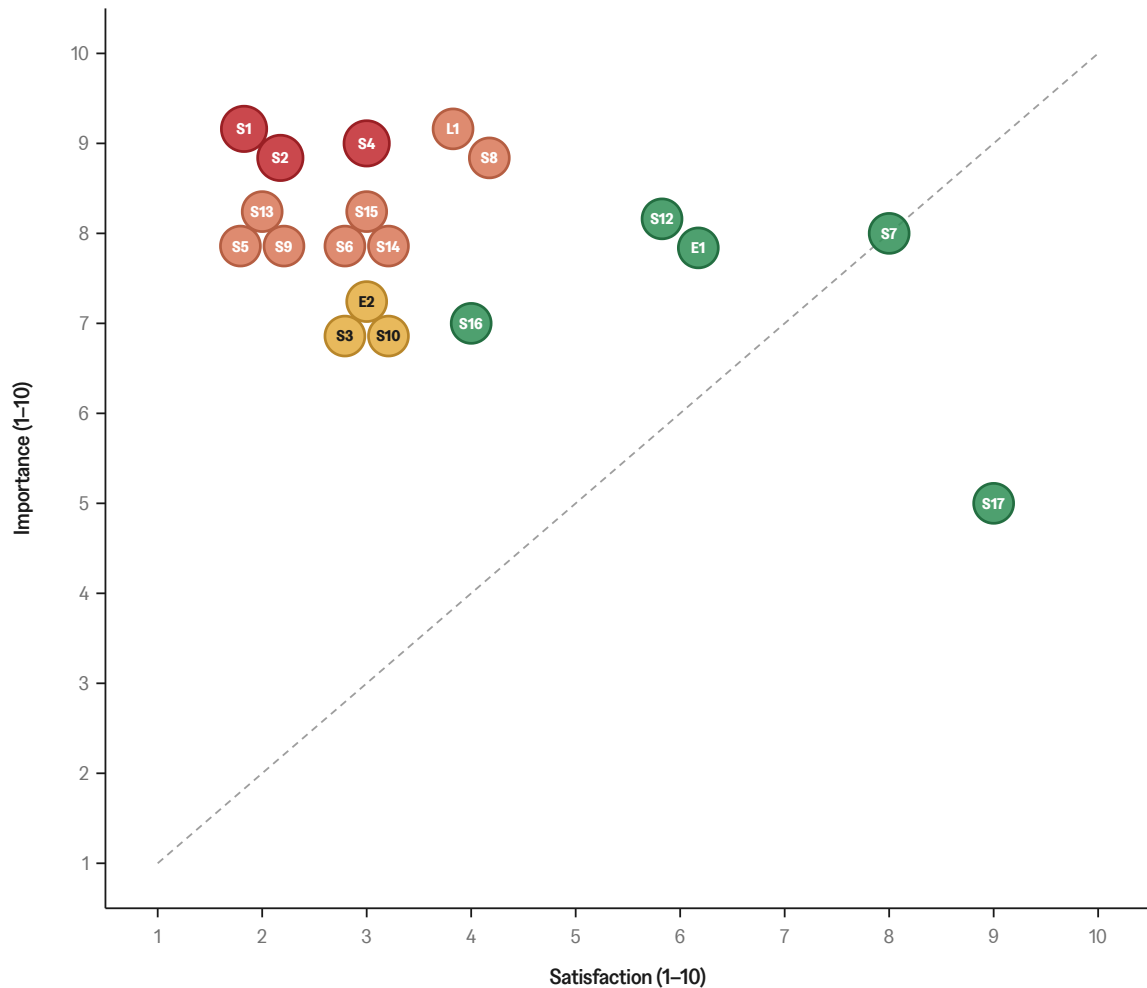
Opportunity = Importance + the shortfall, where the shortfall is Importance minus Satisfaction, never counted below zero.

For example "get credit for every journey I make" scores importance 9 and satisfaction 2. Opportunity: $9 + (9-2) = 16$.

A Job scores highly underserved at 15 or more, underserved from 12 to just under 15, moderately underserved from just above 10 to just under 12, and served right at 10 or below. Over-served is reserved for satisfaction of 8 or more running at least two points beyond importance. Three Jobs in the trends table, S12, E1 and S16, sit at exactly 10, the boundary between served right and moderately underserved, and should be read as boundary calls, not comfortable verdicts.

JOBS-TO-BE-DONE · VISUALISATION

Fourteen of nineteen Jobs are underserved



- Opportunity 15 and above · highly underserved
- Opportunity 12 to just under 15 · underserved
- Opportunity just above 10 to just under 12 · moderately underserved
- Opportunity 10 and below · served right or better

was

Each Job plotted by importance against satisfaction, coloured by Ulwick opportunity score.

Consumer insurance Jobs

Ref	The Job	Type	Scope	Imp	Sat	Opp	Servedness
L1	Get and keep affordable cover as a driver the standard market prices out	Functional	Large	9	4	14	Underserved
S1	Get credit for every journey I make	Functional	Small	9	2	16	Highly underserved
S2	Keep my record accurate when the car is shared	Functional	Small	9	2	16	Highly underserved
S5	Check my score, cover and status whenever I need to	Functional	Small	8	2	14	Underserved
S10	Know what is recorded about my driving and what it means for my policy	Functional	Small	7	3	11	Underserved (moderate)
S3	Know what to change in my driving to lower my premium	Functional	Small	7	3	11	Underserved (moderate)
S4	Keep my cover valid and my premium fair as my miles and circumstances change	Functional	Small	9	3	15	Highly underserved
S6	Get my cover documents and driving record whenever I need them	Functional	Small	8	3	13	Underserved
S9	Get unstuck quickly when the app can't help	Functional	Small	8	2	14	Underserved

Every Job above is underserved, none reaches served right or better.

Commercial Jobs

Jobs for fleet and delivery drivers are a different context, so we have separated these out from the consumer insurance set above.

Ref	The Job	Type	Scope	Imp	Sat	Opp	Servedness
S8	Get covered for work the moment I start driving, without paying for time off the road (delivery)	Functional	Small	9	4	14	Underserved
S14	Get my mandatory vehicle checks and paperwork done quickly and right (fleet)	Functional	Small	8	3	13	Underserved
S13	Keep my driving hours legal and my duty status right (fleet)	Functional	Small	8	2	14	Underserved

Two confirmed commercial Jobs rest on concentrated evidence: S8 on one operator's courier product, and S13 on two fleet platforms. We score S13 for UK drivers, for whom the tachograph rather than the app holds the legal record.

Below are the trend Jobs from both sections: Jobs that appeared in the data but sat under 50 mentions each, several of them concentrated at one operator. We include them because the trends are interesting enough to discuss and worth further research, but they are not confirmed on the data available. First, for consumer insurance:

Ref	Signal	Type	Scope	Imp	Sat	Opp	Servedness	Why it's here, not above
E2	Drive normally with the monitoring out of mind	Emotional	Small	7	3	11	Underserved (moderate)	41 mentions
S7	Pay only for the driving I actually do (low-mileage)	Functional	Small	8	8	8	Served right	42 mentions, 39 at one operator
S12	See how safely my young driver is driving when I'm not in the car (parent)	Functional	Small	8	6	10	Served right	37 parent-voice mentions, 26 at one operator
E1	Feel reassured that my child is safe on the road without me beside them (parent)	Emotional	Small	8	6	10	Served right	same cluster as above
S16	Track my vehicle's location	Functional	Small	7	4	10	Served right	40 mentions
S17	Catch car problems before they leave me stranded	Functional	Small	5	9	5	Over-served	37 mentions, 35 at one operator

And for fleet:

Ref	Signal	Type	Scope	Imp	Sat	Opp	Servedness	Why it's here, not above
S15	Know where my vehicles and people are right now, from wherever I am (fleet manager)	Functional	Small	8	3	13	Underserved	33 mentions, below floor; scored from the mobile surface this corpus can see

Analysis and commentary

We identified several interesting takeaways from both the summaries above and the process of putting them together.

Firstly, every Job we have enough evidence to score with confidence across consumer, fleet and delivery is underserved, three of them highly so, and the two highest opportunity scores in the study both concern the accuracy of the consumer driving record: getting credit for every journey, and keeping the record right when the car is shared. There is friction at the core of this product category.

The initial read of this situation is twofold: firstly there is opportunity to improve every customer interaction point compared to competitors by focussing on helping the customer 'get the job done' better. Secondly, an underserved Job is always a sign of disruption risk.

If an alternative product category emerges that serves the Job better, customers will quickly move to adopt it. Therefore improvements in this area have both strategic growth and risk reduction dimensions.

Notable for marketing teams, we searched for 'social Jobs', how a driver wants to be seen by others, and found almost nothing.

Across 10,442 reviews we could identify four clear expressions: a driver wanting to share a trip score and blocked by a screenshot restriction, a parent describing a teenager competing with friends on weekly scores, and two drivers complaining that a leaderboard was broken or not believable.

Thirty-six reviews mention a score and another person in any connection at all. Against an evidence floor of fifty mentions, that is far too thin to score as a Job.

Social awareness is not missing entirely, since emotional and reassurance language appears in ninety-four reviews and reviewers name family members, colleagues and employers in a hundred more, so the near-silence is about social motivation specifically rather than about what reviewers write. And three of the four expressions are people reaching for a social feature the product already has and finding it broken or blocked.

Nobody in this corpus describes **driving well simply to be seen driving well**.

The other trend we can't help but notice is the hints of strong fit that can be seen at the edges, and this is one reason we wanted to include the trends alongside the confirmed Jobs. Within the lower-confidence trends we can see examples of this technology really serving customers' needs:

- Parents reassured about a young driver's safety
- Drivers paying only for the miles driven
- Useful vehicle location and car-health diagnostics

The warmest material in the whole evidence base is found here, including parents and grandparents describing exactly the reassurance the product promised them.

But every one of these clusters sits below our evidence floor or is concentrated at a single operator (one of which is being wound down, see section 10), so against an industry-wide brief they remain points of interest rather than conclusive findings.

It could be that this is simply the shape product-market fit takes in this category. The measurement is imperfect and probably always will be. Nobody enjoys being scored regardless of accuracy, so the sensible play is to focus on the broader value the category brings and accept some poor sentiment.

However, we think there is evidence that points the other way. Section 8 argued that greater transparency about how scoring works might resolve disputes it currently can't.

It's notable that the same style of monitoring reads as coaching at some operators and surveillance at others. 'Acceptance' then looks like something that can be built.

We noted earlier that satisfaction skews harsh in review data, so full marks are hard to earn from this source. Yet the same method that found every underserved score above also found fit in the trends tables.

In most categories we study, the map shows a product doing some things well and some badly. Here we have a category that fits in the 'macro' (lower premiums, safer young drivers on the UK casualty record, reassured parents), but contains friction in the specific (the moments of judgement). Every confirmed Job is underserved, and none of the Jobs that reach served right concerns the reviewer's own score: they are about a young driver's safety, miles paid for, a vehicle's location and its health.

This is compatible with a healthy book despite being unloved by a sizeable section of its user base.

Looking at the data as a whole, the raw record is more encouraging than the Jobs table might suggest, and we want to make sure this is acknowledged.

Engaging with the judgement at all still predicts dissatisfaction: reviews that mention the score are satisfied 15.5% of the time against a 30.4% baseline, and the judgement comes up roughly twice as often in complaints as in praise.

But when satisfied reviewers do mention the score, they are specific. We analysed the 115 four- and five-star reviews that refer to scoring, plus ten more from a wider search, and 63 of the 125 praise the judgement itself rather than the app around it, and they give consistent reasons.

Reviewers specifically praise:

1. Seeing each journey scored on a map
2. Being told what was measured
3. Getting the result as soon as they finish driving
4. Being scored less strictly than by other black boxes

One notes that the app does not count music played through the car, the situation a section 5 reviewer describes being penalised for.

These reviews are a minority of the satisfied minority, but they show two of the four pillars, transparency and proportionality, working in products already on the market.

Where the scoring is shown and proportionate, some customers accept it and can say why, which suggests the shortfall in the Jobs table is one of execution.

In section 10 we look at an operator that seemed to be winning beside the judgement rather than inside it, but is being wound down before we could learn whether that advantage compounds into more.

STRATEGIC READ

Our thesis is that durable position in this market is built on drivers at minimum accepting the judgement as legitimate, which is what the four pillars add up to. Not necessarily liking it, but trusting it enough to stop disputing it as arbitrary.

Our Jobs-to-be-Done analysis suggests the market hasn't cracked this yet. No Job inside the judgement reaches served right.

SECTION 10

What can be learned from By Miles

We set out in this report to produce a fair assessment of the UK telematics market, despite the clear dissatisfaction present in the review data.

Every claim we've made so far invites the same defence: that this is simply a hard category, where phones are occasionally unreliable and telematics drivers are unhappy regardless of who holds their policy. However, the example of one operator shows this can't be the whole story.

By Miles, a pay-by-mile insurer, ran the same kind of product as everyone else in this study (phone app, mileage tracking, data-driven pricing) with the same phones and (we must assume) comparable users.

Unlike the rest of the market, its recent written reviews average 4.73 stars and its reliability complaint rate, 9.3% of its digital reviews, is the lowest of any operator with at least 100 digital reviews. We do note that pay-by-mile also attracts lower-mileage drivers, so the comparison is not exactly like for like.

Its early reviews included familiar complaints ("Great insurance. Terrible app.", 2024). But unlike its peers, its later reviews show the complaints answered.

"Best insurance I've had, 4th year now... loved being able to track my mileage, also makes you think twice about unnecessary journeys. User friendly app and best of all very economical..."

Apple, By Miles app, 5★, March 2026

There are some key differences between By Miles and the rest of the cohort which may at least partially explain the discrepancy.

Firstly, the architecture helps. By Miles measured miles with a self-contained plug-in device, so its app had the reporting job rather than the sensing one.

But this is not the whole story. Several rivals' apps sit on the same architecture with far worse records.

Secondly, By Miles measured miles driven but it didn't score driving behaviour, which reinforces the pattern we've seen throughout, that the telematics category appears to fit best outside of scoring and judgement.

A clear problem for an industry built around just that.

The corresponding user Job (paying only for the driving you actually do) is one of the few in this study that scores as served right, though with most of its mentions at a single operator it remains a trend not a firm conclusion.

By Miles stopped taking new customers on 26 November 2025 and stopped offering renewals on 6 January 2026. Following Aviva's acquisition of Direct Line Group, trade press reported that investment was concentrating on core brands; an Aviva spokesperson said the book would be wound down over the next eighteen months.

Popularity and loss ratio are of course different measures, and we can't see By Miles's underwriting result from outside. A long-term view of whether the counter-positioning candidate we describe in appendix A6 would have paid off will probably remain unknown.

"So pleased with this company. A reasonable starting cost. Then pay each month for the miles you've travelled. They inform you how much each journey costs..."

Apple, By Miles app, 5★, June 2026 (written more than six months into the wind-down)

By Miles is not the only operator whose customers are largely satisfied.

Marmalade, a young-driver and learner specialist, has the second-lowest reliability complaint rate of the operators with at least 100 digital reviews, at 10.9% of its digital reviews, and 56% of its digital reviews are four or five star against 30.4% across the corpus.

It is the more interesting of the two for our purposes, because unlike By Miles it does score driving behaviour, so its satisfied customers are describing a judgement they accept rather than the absence of one.

Its recent written sample is 24 reviews, below the threshold we use in appendix A2, so we draw no comparison of written means and treat it as corroboration rather than a second case study. Its four- and five-star reviews that discuss the score date almost entirely from 2024 or earlier, so it shows what was achieved, and its recent sample is too small to say whether that has continued.

The challenge to an operator running a fuller scoring mechanism is to make that judgement itself trustworthy enough to earn similar results, and section 9 suggests that is possible rather than hypothetical.

STRATEGIC READ

By Miles shows that a reliable app is achievable on the same phones, and on the same architecture as rivals with far worse records, but it had the advantage of never having to ask its customers to accept a score. Marmalade, on a smaller sample, suggests customers can accept the score if the execution is right. Together they place the gap where we highlighted it in section 9, the judgement, and suggest an operator can close it. Whether closing it pays back is the question By Miles's wind-down leaves open, and section 11 sets out why we think it would.

SECTION 11

The strategic read: advantage rests on a record customers trust

This section applies Helmer's 7 Powers and a Porter-based read of demand structure to the evidence above; the full audit is in appendix A6. This is an external view built on the review record, so we make no claims about operators' internal systems, data or economics.

A telematics insurer's ability to create a durable structural advantage in the market relies on one asset: a driving record the customer trusts.

Data is not in itself enough, every operator holds its own book of behavioural data. What matters is the trust, and from what we've found this trust is under strain. Across operators with a reportable written sample, the public store rating sits a median 2.2 stars above what recent written reviews describe, so the signal an operator watches and the signal its customers give are not the same signal.

What trust rests on

Trust in the record breaks down into four thematic areas we saw recurring throughout this report:

1. Access

A customer who cannot get into the product has nothing to trust. Section 4 finds this the largest complaint in the written record, and it sits upstream of everything else.

2. Accuracy

Does the record match what the driver actually did? Section 5 finds device pairing, journey recording, scoring and premium consequences behaving as one coupled system, with the cost of a miss landing on the customer.

3. Transparency

Can the customer see what was measured and why, and can they do anything about it? Section 5 finds products that often cannot explain themselves; the satisfied reviewers in section 3 describe the opposite, an assessment whose criteria make sense to them. Seeing the reasoning is also what makes contest possible, because a customer cannot argue with reasoning they cannot see. One of section 5's successful challenges came from a reviewer who says their app showed a perfect score when the insurer moved to cancel, so a visible record and a route to dispute it are two different things and both are needed to achieve true transparency.

4. Proportionality

Is the driver penalised only for things they would accept as their own? Section 8's disputes are almost entirely about this, and the praise makes the same point in reverse when reviewers say a product does not penalise where or when they drive.

Accuracy is the pillar most often mistaken for the whole, and in our reading it is the one the industry talks about most publicly. The four do not appear equally in the data. **Access** and **accuracy** dominate the complaint record.

Proportionality, and the visible half of **transparency**, appear mostly the other way round, in the praise, where satisfied reviewers single out products that show them where on a map, explain what was measured, and do not penalise things they consider outside their control.

The other half of **transparency**, the ability to contest, appears mostly in the complaint record, which is what you would expect of something customers only reach for when they are already unhappy.

That asymmetry does not mean the two matter less. It means a customer rarely complains about an explanation they never knew was available, which is why an operator reading only its complaint volumes would reasonably conclude that accuracy is the whole problem.

That distinction matters commercially, because accuracy is also the pillar an operator is least likely to own outright. Specialist providers sit in the stack behind many of the brands in this study, running the data platform.

Where the sensing and analysis is bought in rather than built, accuracy may be improving for the provider as well as (or instead of) the operator.

We don't know the agreements individual operators have with their technology providers, so this is a question only an operator can answer about itself. Question 8 in section 12 is where to start.

Trust does not divide the same way. The customer has no relationship with the provider and, we can probably assume, in most cases doesn't know it exists.

Whoever runs the measurement, it is the operator's name on the policy, the app and the letter, so the operator carries the trust and the loss of it either way. An operator may therefore be bearing the reputational consequence of measurement it does not control.

Two advantages trust makes available

Trust in the record is not itself the advantage. It is the condition on which two advantages become available, and we found no evidence that either is held by any consumer insurer in the cohort.

Advantage 1: Switching costs

A customer's own verified driving history, richer than a no-claims bonus, plausibly gives them a reason not to switch, since starting again elsewhere means losing it.

A record you believe is wrong stops being an asset you protect and becomes a burden to escape. We discussed what this looks like in section 5.

Insurers already know how to build switching costs deliberately, as multi-product bundling shows.

Data-portability rights narrow this advantage but don't remove it. They give customers their raw data, not the insurer's verification of it, which is where the value is actually realised.

Advantage 2: Brand equity

Where price is often the only thing separating one insurer from another on a comparison site, being known for treating customers fairly becomes one of the ways to stand out.

While brand equity is hard earned and quickly lost, true brand power takes a long time to realise and the impact of poor experiences may not be immediately felt.

Each new cohort of business doesn't necessarily check the review record before purchasing, being driven largely by price, but a bad experience during that first policy term can lose that cohort for a generation and reduce gains for years to come.

Both advantages rest on the same trust foundation, which is why erosion of that trust undercuts both at once. That erosion could come from any of the four pillars: a record the customer believes is wrong, a score they cannot see the basis of or argue with, a penalty they think unfair, or simply not being able to get in.

It could equally come from the customer disengaging without generating a signal the operator can see, so that when renewal comes it looks the same as price-shopping. In every case, it is the trust that is damaged first.

Why are these advantages unclaimed?

The probable answer to this is that the UK insurance market doesn't push operators to claim them. Personal lines insurance largely runs through price-ranked comparison sites, where the quality signals available describe the cover rather than the lived product (for example, Defaqto stars).

Retention is tied to the policy year, so assuming customers can't or don't cancel, any loss isn't felt until the end of the term.

Why they are still worth claiming

The best possible outcome for any business in any market is a durable advantage that a competitor cannot copy.

The two advantages compound together, and both start from the same place. A record customers believe is one they stay with, so retention of good risks improves in the current book.

That shows up in pricing at new business and attracts the next cohort. Over time, the switching costs and the brand power earned through fair treatment make good customers more willing to stay even when a rival quotes less, so the book keeps its best risks without discounting to win them back.

A competitor can see the prices and match them. What it cannot copy quickly is a reputation for judging people fairly, or the verified history a customer would have to abandon to leave.

What this means for fleet

Fleet is a different model and the consumer insurance learnings above don't transfer directly. In fleet, the buyer is not the user.

Product quality is therefore only weakly rewarded at acquisition and only weakly punished in-term. The benefits of investment in product are felt at less frequent, procurement-driven events.

Benefits and losses are slow building but with a longer tail when they do hit.

Fleet switching costs already exist and are held by the platforms, organisational rather than trust-based, and the evidence shows the cost of product issues falling largely on the workforce (section 6), not those making the purchase decision.

The buyer's strategic exposure is largely internal and operational, in driver retention and compliance confidence. If fleet operators value drivers as a scarce resource, it is in their interests to ensure those drivers are not pushed to a competitor by poor tooling.

Why now

Durable advantages last as long as the current technology cycle, and moments of transition offer a chance for players to reset the board and claim new advantages against the leader. Telematics could be seen to be in such a transition.

The sensing layer is moving into the customer's phone, re-baselining the trust relationship for a generation of policies.

Three recent attempts on or near this ground have ended in withdrawal: insurethebox ceased renewals in January 2025; VitalityCar, a Vitality and Covéa venture in UK motor, closed in 2023 citing market-wide claims inflation; and By Miles is winding down as discussed in section 10.

We're not saying these facts are necessarily linked, and the stated reasons for withdrawal differ where they are stated at all; By Miles's, as reported, was its new parent's decision to concentrate on core brands. What the three do suggest is that this ground is difficult and thinly occupied, which is also why both advantages appear to remain uncontested.

SECTION 12

Closing: ten questions to ask your product team

To close, we've put together a set of ten questions for operators in the study (or indeed any organisation in a comparable industry). They should be easily answerable from data you already hold.

Which opportunities to prioritise or which advantages are most available will differ by operator. This quick diagnostic is intended to help understand your starting position, which is the first essential step in any transformation.



If you run a product that scores or judges the people who use it, in insurance, financial services, gig platforms, employee software or elsewhere, the patterns in this report likely apply to you in some form.

If you're interested in your own product's review record, we can run a sample of this report's method against it and send you the results, with no charge or obligation. Write to hello@wearesystematic.com with the app or industry you want us to look at and we'll be in touch.

SECTION 12 · DIAGNOSTIC

Ten questions to ask your product team

1

What percentage of our recent written reviews describe not getting into the product at all, and who monitors that number? (Section 4)

2

How many of our users have stopped using the app while keeping the policy, and what is that doing to our data coverage per policy? (Section 4)

3

When our sensing layer misses a journey, what does the customer receive from us, and does it blame them? (Section 5)

4

What is our measured driver/passenger misattribution rate, and what happens to the customers it catches? (Section 5)

5

Can a customer challenge a measurement itself, rather than just reassign a trip, and how many tried last year? (Section 5)

6

If a customer exercised their right to meaningful human intervention in a solely automated decision, or a regulator asked how our scoring meets the FCA's price-and-value and consumer-understanding outcomes, what would we show them? (Section 5)

7

As the vehicles our customers drive change, from EVs to adapted and assisted-driving cars, how confident are we that our scoring keeps up, and when did we last test it? (Section 8)

8

Where our sensing and scoring is provided by a third party, what can we show a customer who challenges a score, and how quickly, and is any of that ours to decide? (Sections 5 and 11)

AND FOR FLEET OPERATORS IN PARTICULAR:

9

If a release stopped every driver at one of our fleet customers logging in, how long would it take us to know? (Section 6)

10

Do we know how much of our driver turnover comes down to poor tooling, rather than pay or conditions? (Section 11)

About the author



Ed Roberts is a partner and chief strategy officer at [We Are Systematic](#), leading the firm's strategic direction across regulated and operationally complex sectors. His focus is how product decisions carry commercial, and increasingly legal, consequences for the businesses that ship them. He worked as a product manager on a UK telematics proposition in 2016 and 2017; the full disclosure is in the foreword. ed@wearesystematic.com · [linkedin.com/in/ed-roberts-49272b33](https://www.linkedin.com/in/ed-roberts-49272b33)



was

We Are Systematic is a UK product strategy consultancy. We work out what is worth building and how it pays back, then design it on the evidence. Because we work upstream of development, we've no incentive to recommend building something that isn't worth it. Our work is in automotive and mobility, financial services, and AI and IoT products.

This is the seventh report in our published research series, following *European airline apps: State of UX 2026* and *UK public EV charging apps: State of UX 2025*.

APPENDIX

A1. Method

Population. 19 operators selected by UK market coverage. The providers a UK driver is most likely to encounter, plus the most prominent app-first challengers as a stated structural inclusion (a study of the category's hardware-to-app transition must contain both generations). Insurance cohort: Admiral, Hastings Direct, insurethebox (no app; Trustpilot presence only), Marmalade, Carrot Insurance, Zego, By Miles. Fleet cohort: Samsara, Geotab, Webfleet, Verizon Connect, Teletrac Navman, Quartix, RAM Tracking, ABAX, Radius (Kinesis), Michelin Connected Fleet, Trakm8, Crystal Ball. No operator publishes UK-only installed-base or policy figures, so selection is defended as ranked coverage with scale bands rather than share percentages. The cohort includes the European fleet leader, the global subscription leader, the UK independents (one still AIM-listed; one acquired and delisted in 2025) and the identifiable leadership of the UK telematics insurance market. The coverage evidence behind the selection is retained with the study data. Two market exits during the study window (insurethebox ceased renewals January 2025; By Miles wind-down) are treated as evidence the cohort captured a consolidating market. Of the 19, insurethebox has no app, and ABAX and Trakm8 contribute 26 and 8 digital reviews respectively, one each in the recent window. Per-operator claims are made only where samples support them. Zego's UK app-store presence carries two internal brand tags, Zego Delivery (694 reviews) and Zego Sense (494 reviews). Section 9 treats the delivery brand as a distinct persona from mainstream consumer motor cover, since it insures gig courier work rather than private driving. Admiral and Hastings Direct run their telematics products (LittleBox Pod and YouDrive) inside each insurer's main customer app, which also serves its non-telematics customers. In the 24-month window, 97 of Admiral's 989 written reviews and 195 of Hastings Direct's 1,126 use a telematics term; on those reviews alone the recent means are 1.91 and 2.10, against 1.53 and 2.40 overall. The gap between headline and written record holds on either basis, but these two operators' figures, including their reliability rates, describe the whole app. Together the two apps supply 5,776 of the 10,442 digital reviews, and 5,017 of those use no telematics term. Without those 5,017, login and reliability run at 29 rather than 42 mentions per 100 (insurance 22, fleet 35), the 2026 reliability rate is 21.9% rather than 27.1% and no longer the highest year, and the largest headline gap is 2.8 rather than 3.15 stars. The median gap and the 2026 low in the yearly mean both hold. The telematics-term flag is a keyword list and misses some telematics reviews; a wider list moves these figures by under two points.

Data. 22,003 reviews from the Apple App Store (UK storefront, capped at ~500 most recent per app; two captures, 13 and 20 July 2026, combined and deduplicated) and Google Play (full history, UK, 13 July 2026), deduplicated and language-filtered to 21,994. Digital-relevance filtering produced the 10,442-review analytical subset (47.5%). Trustpilot company scores were captured from public pages on 13 July 2026. No Trustpilot review text was collected. Reddit was excluded (platform access restrictions).

Analysis. Twelve keyword themes and nine pain categories, each precision-audited on random samples. Audits were run as independent AI review passes with human spot-checks, and six themes and five pain categories were tightened and rerun. Categories failing a 10% false-positive threshold (15% for colloquial pain language) were either fixed or demoted to qualitative treatment. Audits measured precision (false positives) rather than recall (missed mentions). Our classifiers were built to catch complaints, so they tend to under-count praise. We tested one theme for what it missed and found the miss rate varied sharply by operator, from under 2% to 25%, because satisfied reviewers write shorter and more casually. Our theme percentages are therefore minimums rather than measurements, and comparing two operators on one of them only means something when the gap between them is wide. Co-location measured as lift against independence (observed co-occurrence over the product of the two marginal rates), computed on all digital reviews except where a figure is marked insurance-only.

Cohort and review-length controls were run as sensitivity checks rather than applied to the published figures: restricting to reviews of 50 words or more reduces the lifts in this cluster to a range of 3.8 to 4.7 while leaving every one of them positive and significant. The published figures are the uncontrolled ones and should be read as upper bounds. Jobs extracted ground-up from the corpus with every count preserved as a runnable probe and the four- and five-star strata mined deliberately alongside the complaint record, scored with Ulwick ODI, and passed through an independent faithfulness audit. Importance and Satisfaction are 1-10 judgements made against the corpus itself, not against an external benchmark. Because review data is a harsh instrument, with 54.6% of this corpus at one star and only 30.4% at four or five, satisfaction is anchored relatively: a Job whose evidence cluster runs far below the 30.4% four-and-five-star baseline scores 2-3, a cluster at 80-95% scores 8-9, and clusters near baseline score mid-range. Importance is argued per Job from the stakes evidenced in the reviews. Scoring was a ground-up re-extraction on 17 August 2026 with every quotation re-traced to source, then audited for faithfulness and proofed by the study author on 18 and 24 August, which withdrew one Job and restated four. The catalogue, probes and audit records are retained with the study data. Signals computed as rates per window (2023-24 vs 2025-26; sample growth ratio 1.16x). With roughly twenty classifiers and probes run, borderline results are treated as colour, not claims, and labelled so where they appear. Bot-detection on Google Play flagged 5.8% of rows (maximum for an operator with a stable sample 11.8%; one operator with only 15 Play rows flagged at 13.3%). The detector flags reviews under fifteen characters or matching a list of generic one-phrase verdicts, so it is a low-information filter rather than a test of authenticity, and it falls disproportionately on five-star reviews (18.4% of five-star Play rows against 1.6% of one-star), which are systematically shorter. Flagged rows do not alter any finding's direction, and no review quoted in this report is flagged. Classifier definitions, audit samples and the capture record are available on request.

Limits. Written reviews over-represent the dissatisfied. Raw review volume partly measures storefront capture: Apple's feed caps at roughly the 500 most recent reviews per app, so a large count can mean a product reached that ceiling rather than that a theme is widespread, and long-run trends lean on Play. Where a figure depends on volume we lead with breadth, the number of operators and markets showing a pattern, rather than the count. All comparisons are like-for-like and the headline-ratings gap is itself the finding (the fuller answer to the negativity objection is in section 2). Samsara contributes 61% of the fleet digital subset. Fleet aggregates were checked with and without it. Several small fleet operators fall below per-operator claim thresholds and are treated qualitatively. Some fleet quotes use US regulatory vocabulary from mixed-market apps, marked where quoted. UK framing follows the tachograph and drivers' hours regime. Accessibility and microcopy categories fell below quantitative floors (5 and 4 genuine reviews) and are reported qualitatively. Fleet quotes without a country marker cannot be located. Mixed-market apps serve non-UK drivers through the same UK storefront. A quantitative split of battery burden between phone-sensing and device-sensing reviews was attempted and withdrawn as too sensitive to segment definition. The architecture point rests on the audited battery-surveillance co-location and the app-only complaint genre, and several operators run device and app-based policies simultaneously, so no per-operator architecture label is used in any quantitative comparison; where section 10 describes By Miles's plug-in device and rivals' box-based policies, it relies on what those operators' customers describe. Manager-role figures reflect mobile apps only. Desktop consoles do not appear in app-store data. The corpus cannot see silent disengagement (users who stop opening the app write no reviews), and it cannot see inside any operator. Claims about what operators' internal systems or data show are made nowhere in this report, and where we raise the question we label it as one. Where we reason about what signal does or does not reach an operator, we are reasoning from the review record and say so. The foreword's account of the author's own time inside one operator is personal recollection, offered as disclosure rather than as evidence, and no finding rests on it. We flag where these blind spots matter.

APPENDIX

A2. Headline ratings

Store headline and Trustpilot figures were captured 13 July 2026; written-review figures are computed on the digital subset (analysis run 20 July 2026). Store headlines are each operator's own public aggregate rating, blended across both stores by ratings volume.

The comparison table carries the eight operators with at least 30 recent written reviews, the threshold below which we do not report a written mean. The remaining eleven cohort members follow it, with their public figures listed for completeness and no comparison drawn. Full cohort selection is in A1.

Operators with a reportable written sample (n≥30)

The store and Trustpilot figures are public; the blended rating and the written mean are ours, computed as A1 describes. This is not a ranking by us. A wide gap means an operator's written record and its store headline are describing different things; it is not a judgement on the product. Admiral's and Hastings Direct's figures are for each insurer's main app, which serves all its products (see A1).

Operator	Cohort	Operator's own store rating (blended)	Mean of recent written reviews (24 mo)	n	Trustpilot (company-level)
Admiral	Insurance	4.68	1.53	989	4.5 (156k)
Hastings Direct	Insurance	4.66	2.40	1,126	4.4 (194k)
Zego	Insurance	4.49	1.99	405	4.4 (19k)
By Miles	Insurance	4.79	4.73	177	4.6 (11.7k)
Samsara	Fleet	4.73	2.45	481	3.2 (525)
Geotab	Fleet	3.00	1.86	93	—
Verizon Connect	Fleet	3.78	1.62	32	3.5 (883)
Quartix	Fleet	2.56	1.69	49	4.8 (1.3k)

Also in the cohort: sample too small to report

Radius (Kinesis) has no written reviews inside the recent 24-month window, so its recent-n is zero rather than merely small. These eleven operators had fewer than 30 recent written reviews, so no written mean is reportable and no headline-to-written comparison is drawn for them. Their public figures are listed for completeness only. A low store rating in this table is a public aggregate, not a finding of this study.

Operator	Cohort	Operator's own store rating (blended)	Recent written reviews (n)	Trustpilot (company-level)
Marmalade	Insurance	4.22	24	3.9 (12.5k)
Carrot Insurance	Insurance	1.82	13	4.6 (7.5k)

Operator	Cohort	Operator's own store rating (blended)	Recent written reviews (n)	Trustpilot (company-level)
insurethebox	Insurance	no app	—	4.2 (5.1k)
Teletrac Navman	Fleet	2.44	18	3.6 (3.1k)
Webfleet	Fleet	4.75	10	—
RAM Tracking	Fleet	4.58	7	4.7 (2.3k)
ABAX	Fleet	4.04	1	4.0 (325)
Michelin Connected Fleet	Fleet	3.30	2	none
Trakm8	Fleet	3.38	1	3.9 (56)
Radius (Kinesis)	Fleet	3.36	0	—
Crystal Ball	Fleet	—	—	4.8 (103)

APPENDIX

A3. Key statistics referenced in the text

Digital subset mean 2.38 (n=10,442). Distribution: 54.6% 1★ / 8.6% 2★ / 6.3% 3★ / 5.3% 4★ / 25.1% 5★. Access and reliability: login 1,880 mentions (18.0%, mean 1.37, four- and five-star share 5.2% against a 30.4% corpus baseline, one-star share 81.1% against 54.6%); reliability 2,477 (23.7%, 1.44, 5.7% against 30.4%, one-star share 76.5% against 54.6%); overlap 680; combined rate 42 mentions per 100 digital reviews (41.7 exact; insurance 44, fleet 35). Reliability decomposition: hard failure 488; opaque errors 394 (97% insurance, mean 1.18); slow 283; freeze 255; reinstall ritual 174 (7.0%); chronic 127 (mean 1.19). Reliability rate by year (digital subset, n in brackets): 2018 25.2% (282), 2019 19.5% (533), 2020 21.8% (1,031), 2021 24.7% (1,480), 2022 24.5% (1,460), 2023 24.4% (1,359), 2024 22.2% (1,152), 2025 23.1% (1,741), 2026 to 18 July 27.1% (1,232). The 2018-2025 series is flat (homogeneity $p = 0.165$); 2026 sits above it ($p = 0.004$ against pooled 2019-2025); not on the narrower basis in A1 (21.9%), so the report does not treat it as a finding. 2018 falls below the 300-row per-year floor used elsewhere and is reported here for completeness; it is also 77% fleet against 61% insurance in 2019, so the earliest comparison carries a cohort-mix confound. Yearly mean, digital subset: 2.33 (2018) rising to 2.88 (2019) then holding between 2.30 and 2.65 through 2025; 2.04 in 2026 to 18 July, the only year below the band. The equivalent series on the full cleaned corpus runs 2.70 to 3.58 with 2.78 in 2026; the report's yearly means are digital-subset figures throughout. Consequence coupling (all classifiers audited; records with study data): deletion and uninstall talk 1.9% vs 0.93% baseline, rest of digital subset excluding the reliability set (lift 2.0, $z = +3.8$); cancellation talk is no higher in reliability reviews than elsewhere on either basis (0.73 on the full corpus, 1.01 on the narrower basis in A1), so the report claims only that; a stricter clean-abandonment variant is directionally consistent but not statistically significant and carries no claim. Duty logs 437 mentions (fleet only), mean 2.07. Co-location lifts: score x premium 12.7; device x battery 10.2; trip x device 7.3; battery x privacy (insurance) 8.9; login x score 0.29 (0.31 on long reviews only). AI mentions: 0.18% (2023) → 1.29% (2026); window rates 0.23% → 0.80% (3.5x, sample growth 1.16x); window raw counts 12 → 48; Play-only series 0.21% → 1.18%; recent-window cohort split 37 insurance / 11 fleet. Strong anger: 1.12% (2026) vs 0.45% (2023-24). Score-dispute attribution scan (§8; classifier and audit retained with the study data): 45 reviews across 6 operators disputing a scoring outcome with an external attributed cause, passenger and phone-placement cases excluded; roundabout/cornering sub-group 35 across 5 operators (8 at 4-5★, mean 2.46); hazard-forced braking 8, EV dynamics 5 and road/GPS artefacts 3, reported qualitatively. Legacy phase-6 EV probe (13 mentions, 7 in 2025-26) retained in the signal trail, superseded in the report by the scan. Accessibility: 5 genuine mentions in 10,442. Manager-role mean 2.17 (fleet); driver 2.57 pooled, being fleet drivers 2.64 (n=1,948) and insurance/gig-courier drivers 2.39 (n=694); consumer 2.33. Per-operator reliability mention rate ranges from 9.3% to 38.0% of an operator's digital reviews (operators with $n \geq 100$ digital reviews; the top figure is Admiral's main app, 20.2% on its telematics-related reviews); the pattern appears at every operator size. All figures computed on the combined two-capture corpus (analysis run 20 July 2026).

Study headline and cross-referenced figures. Reviews collected: 22,003; after cleaning, 21,994 reviews remained (§2). Cohort: 19 UK-relevant telematics operators, seven insurance and telematics-insurance brands and twelve fleet platforms (§2).

Digitally relevant subset: 10,442 of them (47.5%) (§2). Collection dates: Apple App Store, two captures, 13 and 20 July 2026 (§2); Google Play, 13 July 2026 (A1); Trustpilot, 13 July 2026 (§2). Store headline ratings for the major consumer apps run between 4.5 and 4.7; the recent written record for the same apps runs between 1.5 and 2.4 (§3). Median gap, across operators with at least thirty recent written reviews to compare: 2.2 stars (§3). The largest gap in the cohort is around three stars (§3). By Miles: recent written reviews average 4.73 (§10); store rating (blended) 4.79 (A2); reliability complaint rate 9.3% of its digital reviews, the lowest among operators with at least 100 digital reviews (§10). Jobs to be Done: Nineteen Jobs were scorable (§9); five of the nineteen sitting at served right or better (Executive summary); every Job we have enough evidence to score with confidence across consumer, fleet and delivery is underserved, three of them highly so (§9); the highest opportunity score in the study is 16 out of a possible 19, shared by "Get credit for every journey I make" and "Keep my record accurate when the car is shared" (§9).

APPENDIX

A4. Classifier audit trail (summary)

Every theme and pain classifier was screened on random samples (20 per theme, 15 per pain category, match-context re-checks after tightening), with each sample judged individually. These are screening samples, not powered tests. A twenty-item sample is well suited to catching a badly wrong classifier, since a classifier misfiring a quarter of the time would return a clean sheet less than one time in three hundred, and it caught eleven, one of which was cutting a mention count by 84%. It cannot put a tight bound on what remains: a sample returning two false positives is consistent with a true rate anywhere between about 1% and 32%. The thresholds below (10% false positives, 15% for colloquial pain language) were the trigger for tightening and rerunning a classifier, not a certification of the ones that passed. Residual precision on the surviving classifiers is unmeasured, and prevalence figures should be read with that latitude as well as the recall floor noted in A1. Six themes and five pain categories exceeded false-positive thresholds at first pass and were tightened and rerun; the largest corrections cut premium-mechanics mentions from 369 to 59 (69 on the final two-capture corpus) and accessibility mentions from 28 to 5. No pre-tightening figure appears in this report. The consequence-coupling and signal classifiers behind the abandonment, cancellation, AI, anger and EV figures were audited in a further pass (13 July 2026, re-verified on the final two-capture corpus 20 July): two were tightened, one construct was relabelled, and two constructs were withdrawn as unreliable (a signal multiplier, and a sensing-architecture segment split). A further classifier (score-dispute attribution, section 8) was built and audited on 20 July: 24 samples judged across four sub-groups, one negation false-positive pattern tightened out and the group rerun; definitions and assignments retained with the study data. Audits measure precision, not recall. Full audit records are retained with the study data and available on request.

APPENDIX

A5. External sources

1. Movingdots (Swiss Re), "Transport mode recognition and the driver-passenger dilemma", 20 November 2020 (driver/passenger distinction "in nearly 85% of cases"). movingdots.com/news-insights/transport-mode-recognition-and-the-driver-passenger-dilemma
2. Ebert et al., "Validation of a smartphone telematics algorithm for classifying driver trips", *Transportation Research Interdisciplinary Perspectives* 25, May 2024; University of Pennsylvania (96.5% overall accuracy; 91.2% specificity for non-driver trips, described in the abstract as slightly lower and more variable than sensitivity). Cited from the published abstract. [pmc.ncbi.nlm.nih.gov/articles/PMC12416331/](https://pubmed.ncbi.nlm.nih.gov/articles/PMC12416331/)
3. Financial Ombudsman Service, guidance on telematics ("black box") insurance complaints, updated 6 September 2022. financial-ombudsman.org.uk/businesses/complaints-deal/insurance/motor-insurance/telematics-black-box-vehicle-insurance
4. Chartered Insurance Institute road-safety consultation following a coroner's report on young-driver deaths, Insurance Business UK, 28 May 2026.
5. Motability "Drive Smart" telematics scheme paused, announced c.14 May 2026; Disability News Service, 21 May 2026; Carwow, 15 May 2026.
6. By Miles wind-down: quotes to new customers ceased 26 November 2025, renewals 6 January 2026, By Miles company update (help.bymiles.co.uk); Insurance Post, December 2025.
7. insurethebox ceased offering renewals: Tom Luckham, "Insurethebox ceases offering renewals or policies", Insurance Post, 10 January 2025.
8. Smart tachograph obligations and drivers' hours enforcement context, DVSA and Traffic Commissioners' published guidance and annual reporting, 2024-2026.
9. RAND Europe, "Incentives and physical activity" (RR2870), the Vitality Active Rewards with Apple Watch assessment: 400,000+ members, UK/US/South Africa; commissioned by Vitality's parent, independently conducted. rand.org/pubs/research_reports/RR2870.html
10. "Why there are huge drop-off rates among telematics app users", Insurance Business UK, 2023 (industry commentary on score plateau and engagement decay).
11. We Are Systematic, *UK public EV charging apps: State of UX 2025*. wearesystematic.com/research/uk-public-ev-charging-apps-state-of-ux-2025
12. Hamilton Helmer, *7 Powers: The Foundations of Business Strategy*, 2016.
13. Michael E. Porter, "The Five Competitive Forces That Shape Strategy", Harvard Business Review, January 2008.
14. VitalityCar (Covéa Insurance and Vitality, UK motor): launched 16 June 2021, Covéa Insurance press release, covered by Insurance Times, Insurance Business UK and Claims Media; withdrawal citing market-wide claims inflation, Insurance Times, 2023 (insurancetimes.co.uk/news/vitality-announces-it-will-leave-car-insurance-market/1445620.article).

15. Santos, Oliveira & Eler, "Why Are Accessibility App Reviews Rare? Exploring Motivations, Barriers and Opportunities Among Users with Disabilities," INTERACT 2025 (Springer, 2026). Survey of 104 users with disabilities on motivations and barriers to app-store review participation; the abstract lists doubts about feedback effectiveness, technical obstacles and lack of interest without ranking them. doi.org/10.1007/978-3-032-04999-5_11
16. Guidewire, 2026 European Insurance Consumer Survey (4,000+ policyholders, UK/France/Germany/Spain): 26% say nothing would increase confidence in insurers' AI use, up from 24% in 2025; EIOPA's 2025 Eurobarometer survey found only 30% of European consumers would trust AI-generated recommendations for insurance products. Reported in Insurance Business UK, 26 May 2026. insurancebusinessmag.com/uk/news/technology/one-in-four-consumers-unconvinced-by-insurers-use-of-ai--survey-576516.aspx
17. LexisNexis Risk Solutions, analysis of UK Department for Transport road casualty data (Stats19), 2011-2017: a 35.32% reduction in collision rates among 17-19 year-old drivers with telematics insurance. Company-commissioned research on government data, published 22 February 2019; not independently peer-reviewed. risk.lexisnexis.co.uk/about-us/press-room/press-release/20190222-black-box

Store and Trustpilot headline figures were captured from the operators' public listings on 13 July 2026; the capture record is retained with the study data.

APPENDIX

A6. Strategic framework audit (summary)

The strategic analysis applies Hamilton Helmer's 7 Powers as its audit instrument, testing each power for both of Helmer's conditions (a benefit to the holder and a barrier to rivals), supplemented by a Porter-based read of demand structure. Section 11 publishes the argument that survived that audit. The table below summarises all seven verdicts, including the powers that do not apply in this market and why. The full working audit is retained with the study data. No claim is made about any operator's internal economics, in either direction.

Power	Verdict in this market	Basis
Scale economies	Unclaimed at the operator layer on the evidence available; whether it could accrue at the vendor layer is an open question invisible to this study	Accuracy compounds only on a clean pipeline; the same reliability complaints repeat at every operator size
Network economies	Not applicable in consumer; live at the vendor layer in fleet, unassessed here	No customer-to-customer value link in monitored insurance; cross-fleet benchmarking exists as a marketed vendor proposition, outside this corpus and not power-tested here
Counter-positioning	A candidate was observed, never proven; the test can no longer be run	Pay-by-mile's incumbent-cannibalisation logic is inference; the wind-down rationale is stated as portfolio concentration
Switching costs	Consumer: unclaimed, and at risk of forfeit where trust in the record fails. Fleet: held by the platforms, with costs falling largely on the workforce	Sections 5 and 6
Branding	Unbuilt; any damage latent in the written record	Section 3's measured gap; renewal consequences are inference
Cornered resource	No; reframed as the foundational asset the claimable advantages share	Every operator holds a book; none is exclusive
Process power	No evidence of process power anywhere in the cohort; the market structure described in section 11 rewards it weakly	Eight flat years; one small team achieving cohort-best execution suggests the capability is imitable

Corrections policy: factual corrections to hello@wearesystematic.com; we review and correct promptly.

was

What 21,994 app-store reviews show A study of 19 UK operators, insurance and fleet.

CORRECTIONS AND ENQUIRIES

hello@wearesystematic.com

We Are Systematic, 2026