Open Access

Journal of Artificial Intelligence & Robotics

Article type: Research Article

ParkIQ for predictive parking guidance and accountable kerbside management

Syed Darda Rahman*

School of Computer Science and Technology, University of Bedfordshire, Luton, United Kingdom.

Corresponding Author: Syed Darda Rahman

School of Computer Science and Technology, University of Bedfordshire, Luton, United Kingdom.

Email: darda.rahman52@gmail.com

Received: Sep 11, 2026

Accepted: Oct 21, 2026

Published Online: Oct 28, 2026

Website: www.joaiar.org

License: © **Rahman SD** (2026). *This Article is distributed under the terms of Creative Commons Attribution 4.0 International Licens.*

Volume 3 [2026] Issue 2

Abstract

Deploying artificial intelligence in urban parking requires a distinction between forecasting an uncertain state and establishing evidence for an operational decision. This paper introduces ParkIQ, a practical architecture for predictive parking guidance and accountable kerbside management. The framework combines a temporal convolutional network with gradient-boosted contextual models to estimate bay occupancy at six horizons from 15 to 90 minutes. A shared, versioned bay representation connects these estimates to regulatory eligibility, data freshness and evidence provenance. The design separates learned predictions from observed events and rule-based eligibility checks: forecasts inform guidance, while enforcement support requires relevant observations, applicable restrictions and meaningful human review. Signed audit logs provide the reference integrity mechanism, with permissioned ledger anchoring treated as an optional experimental comparator. The paper contributes an integration specification that can be tested in practice. It does not claim a new learning algorithm. The evaluation plan addresses missing observations, chronological leakage, calibration, model selection, transfer and service degradation. Five hypotheses define comparisons against simpler forecasting models, alternative guidance policies and conventional evidence handling. A staged protocol combines independent occupancy audits, offline temporal and spatial holdouts, fault-injection tests and a controlled guidance trial. UK kerbside operations, initially English civil parking enforcement, provide the application context. The manuscript reports no implementation, experimental dataset, measured performance or deployment. The paper establishes the conditions under which future experiments could determine whether the integrated architecture improves decision quality sufficiently to justify its complexity.

Keywords: Artificial intelligence; Parking occupancy prediction; Temporal convolutional networks; Probabilistic calibration; Human oversight; Kerbside management.

Introduction

AI systems that support physical-world decisions must manage both predictive uncertainty and the authority to act. Parking provides a concrete setting in which these requirements intersect. Learned occupancy estimates can help drivers choose where to search, but they cannot establish whether a particular vehicle has breached a restriction. ParkIQ addresses this separation through a proposed combination of probabilistic forecasting, explicit rule evaluation and accountable human decision support. Its focus is applied AI in a cyber-physical information system; it does not claim robotic actuation or autonomous vehicle control.

The operational value of parking information depends on more than whether a space is empty. A driver needs a space that is likely to be available at arrival, permits the intended use and duration, and is reachable without excessive search. An authority needs evidence about restrictions and actual behaviour, while an operator needs dependable records of availability and transactions. These requirements share spatial and temporal entities, yet differ in their evidential standards. Treating them as interchangeable creates risks: a prediction may be mistaken for observation, a payment may be mistaken for physical occupancy, or a digital rule may be mistaken for the authoritative legal order.

Citation: Rahman SD. ParkIQ for predictive parking guidance and accountable kerbside management. J Artif Intell Robot. 2026; 3(2): 1044.

Parking research provides substantial foundations. Vlahogianni et al. [15] distinguish the survival of a currently free space from regional occupancy forecasting. Yang et al. [16] model block-level occupancy using spatial, temporal and contextual information. More recent work investigates heterogeneous information and meta-learning for regional prediction [13]. These contributions support further investigation of forecasting, but their results cannot be transferred directly to individual UK bays, different horizons or operational decisions. Economic modelling also connects saturated kerbside parking with search traffic, while illustrating that pricing and allocation influence outcomes alongside information provision [2].

UK policy provides a practical interoperability context. The Department for Transport's digital traffic regulation order service exposes machine-readable extracts of legal orders; these extracts are not themselves the legal documents [7]. The National Parking Platform connects participating parking operators and payment services [8]. ParkIQ is therefore positioned as a proposed complementary decision layer, rather than as the first attempt to digitise parking or an assumed replacement for existing services. No exhaustive market review or priority claim is made.

The research problem is whether a common, versioned representation of the kerbside can improve guidance and evidence traceability while preserving a strict boundary between estimated availability and legally actionable observations. This paper makes three specification-level contributions. First, it defines a six-layer architecture with explicit state, rule and provenance contracts. Second, it specifies a reproducible candidate forecasting pipeline and a decision interface that distinguishes probability, data quality and legal eligibility. Third, it links those design choices to falsifiable hypotheses and operational safeguards, including a comparison between conventional signed logs and optional ledger anchoring. Commercial sustainability is assessed as a cost question, not as demonstrated adoption or revenue impact.

All ParkIQ capabilities described below are proposed. Present-tense statements about schemas and algorithms define the design; they do not report an implemented system. The paper contains no empirical results and has not been preregistered. England is the initial reference jurisdiction for public civil parking enforcement; implementation in Scotland, Wales, Northern Ireland or private parking settings requires separate legal and operational mappings.

Related work and contribution boundary

Prediction and decision support

Vlahogianni et al. [15] use on-street sensor data from Santander to study free-space duration and regional occupancy forecasts. Yang et al. [16] combine graph convolution and recurrent modelling with traffic and weather information for block-level prediction in Pittsburgh. Kuang et al. [13] consider heterogeneous weather information and meta-learning across parking lots. Together, these studies show why spatial scale, input availability and transfer conditions must be specified. They do not establish that the ParkIQ ensemble will outperform either a simpler model or contemporary spatial models.

The Temporal Convolutional Network (TCN) follows the causal sequence-modelling family evaluated by Bai et al. [3]; XGBoost provides a contextual tree-boosting comparator and ensemble component [4]. Neither component is novel. Their combination is an engineering hypothesis whose merit depends on performance, calibration and resource cost. Calibration is evaluated separately from discrimination because high classification accuracy need not imply reliable probabilities [11]. The proposed contribution lies in the connection between forecast outputs, time-indexed eligibility and auditable decision records.

Trust mechanisms and their limits

Al Amiri et al. [1] propose a parking system using consortium blockchain and private information retrieval to protect discovery and reservation privacy. ParkIQ addresses a different question: whether independently verifiable record commitments add value to an enforcement-support workflow. Hashes can reveal later alteration of a committed record; they cannot verify sensor correctness, establish lawful processing or prove a contravention. A conventional signed append-only log is therefore the reference implementation to be evaluated. A distributed ledger is warranted only if independent organisations need shared verification and its benefits exceed governance and operating costs.

Table 1 relates the selected literature and policy sources to the proposed contribution. This focused comparison clarifies how prior work informs ParkIQ; it is not a systematic review or a claim that other systems lack equivalent features.

Table 1: Selected foundations and the questions retained for ParkIQ.

Foundation	Established focus	ParkIQ question requiring evaluation
Vlahogianni et al. (2016)	Free-space survival and regional forecasting	Can bay-level targets remain reliable at longer horizons?
Yang et al. (2019)	Multi-source block-level prediction	Do contextual inputs improve calibrated bay forecasts?
Kuang et al. (2025)	Heterogeneous information and meta-learning	Does transfer help a new location after accounting for shift?
Al Amiri et al. (2019)	Private discovery and reservation using blockchain	Does ledger anchoring improve evidence verification over signed logs?
DfT (2025a, 2025b)	Digital rules and payment interoperability	Can shared identifiers support consistent guidance and traceability?

Design methodology

We use a requirements-led architectural method. The unit of design is a bay-associated decision record containing the information available at a specified time. Driver guidance, operator coordination and enforcement support define separate use cases. For each use case, the specification identifies inputs, admissible decisions, unresolved states and the evidence needed to evaluate correctness. These requirements are analytical choices for this framework; they are not findings from stakeholder interviews or a completed field study.

We develop the design in four linked steps: defining physical and legal state variables; allocating sensing, learning and rule

evaluation to separate services; specifying provenance and failure behaviour at their interfaces; and connecting each proposed benefit to an observable endpoint. The literature establishes the candidate methods and comparison points, while the UK application context constrains the proposed use of their outputs. The result is an architecture and research protocol, not empirical validation.

Research questions and problem definition

Targets and information boundaries

Let b denote a mapped bay, t a forecast origin and h one of {15, 30, 45, 60, 75, 90} minutes. Define y(b,t) as 1 when an independently labelled bay is physically occupied and 0 when free. Unknown observations are missing, not free. The primary forecasting target is $p(b,t,h) = P[y(b,t+h) = 1 \mid I(t)]$, where I(t) contains only information available to the platform by t. This is binary probabilistic prediction, distinct from predicting an aggregate occupancy fraction or the time until a vehicle departs.

For an initial five-minute grid, the endpoint label is the audited state at the target timestamp, accepted only when an observation lies within a declared tolerance of 30 seconds. Otherwise, it is missing. This tolerance is a proposed protocol choice and must be frozen before evaluation. A separate duration target, if later studied, requires its own censoring rules. Payment transactions and app interactions must not substitute for physical ground truth.

Define eligibility e(b,a,u,d) as permitted, prohibited or unknown for user requirements u, arrival a and planned duration d. It depends on the applicable rule version, vehicle and permit conditions, and any relevant restriction changes during the stay. Forecast availability is 1 − p; it is not eligibility. Guidance excludes prohibited bays and labels unresolved legal status explicitly. An eligible recommendation remains a forecast, not a reservation or a guarantee that a space will remain free.

Decision objective

A candidate ranking score is $J(b) = w_1T(b) + w_2C(b) + w_3p(b,t,h) + w_4W(b)$, where T is estimated travel time, C is the disclosed parking charge, W is walking distance and the non-negative weights reflect stated user preferences. Terms must be normalised to documented scales. Accessibility, vehicle compatibility and legal eligibility are hard constraints rather than penalties that can be traded away. The forecast horizon corresponds to estimated arrival; uncertain arrival times require sensitivity analysis across adjacent horizons. For arrivals outside the supported horizon, the service should avoid presenting extrapolated probabilities as validated predictions.

This score defines a reproducible initial guidance policy, not an optimal routing theorem. Competition among drivers, failed parking attempts and spillover traffic can invalidate independent-bay assumptions. A field evaluation must therefore compare actual search outcomes rather than infer them from forecast scores alone. Forecast outputs are prohibited as evidence of an individual parking contravention.

Research questions

RQ1 asks whether the candidate ensemble improves probabilistic forecasts over strong temporal, contextual and spatial baselines. RQ2 asks whether incorporating explicit freshness and rule status reduces unsuitable recommendations. RQ3 asks whether shared provenance improves case reconstruction without increasing unsupported referrals. RQ4 asks whether guidance reduces observed search time under controlled operating conditions. RQ5 asks whether transfer and adaptation remain useful under location shift, missing data and demand change. Section 7 translates these questions into hypotheses with defined endpoints and comparators.

Proposed system architecture

Six layers with explicit contracts

The six layers in Figure 1 share identifiers and timestamps but maintain separate responsibilities. Layer 1 ingests occupancy sensors and, where justified, camera-derived events, payment feeds and opt-in app observations. Layer 2 maintains a bay registry with geometry, ownership, use type and source provenance. Layer 3 stores versioned restrictions and eligibility logic. Layer 4 estimates current state, records freshness and resolves uncertainty. Layer 5 produces probabilistic forecasts. Layer 6 exposes guidance, operator functions and officer decision support through scoped interfaces.

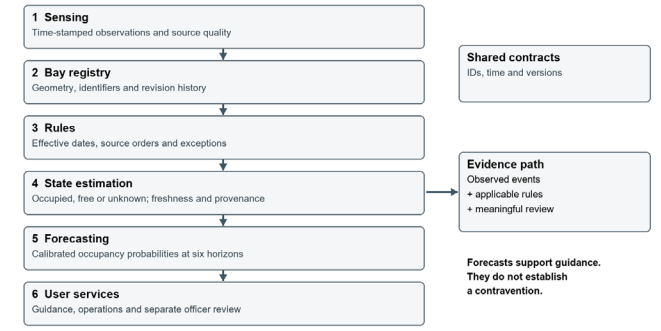


Figure 1: Proposed six-layer architecture. The evidence path uses observed events and versioned rules; forecasting supports guidance only. The diagram represents a specification, not a deployed system.

Mapping data can provide a spatial starting point, but does not establish a complete bay inventory or the legal status of individual spaces. Each bay requires authority or operator validation, a stable identifier and a history of geometry changes. Split, merged and retired bays retain lineage so that model training does not silently join incompatible entities. Electric vehicle charger operational status, connector compatibility and physical bay occupancy are separate fields; an available charger need not imply an accessible empty bay.

Every event carries an event identifier, source identifier, bay or zone identifier, event time, ingestion time, schema version and quality metadata. Store timestamps in UTC; derive local calendar and restriction features using Europe/London time with explicit daylight-saving handling. Idempotent ingestion removes duplicate event identifiers. Late events may correct historical state but cannot retrospectively enter the feature set of an already issued forecast. The original decision record remains reproducible.

State estimation and rule provenance

The state estimator returns occupied, free or unknown, with observation age and source-specific reliability. Reliability is estimated against independent audits rather than inferred from vendor confidence alone. Conflicting signals remain visible in provenance. A stale heartbeat, ambiguous bay assignment or unresolved contradiction produces unknown status. The initial freshness limit is a proposed 60 seconds for live guidance, to be tested against measured sensor cadence and connectivity. The limit is not a guarantee that sensors achieve that rate.

Zone-entry ANPR does not establish which bay is occupied, and phone GPS generally cannot resolve adjacent spaces reliably. Such observations can inform zone-level demand only unless bay association is independently validated. Occupancy sensors cannot establish permit entitlement, loading activity or disability-related exemptions. These limits are represented in the state schema rather than repaired through optimistic confidence thresholds.

Rule records contain both the period when a restriction applies and the time when the platform learned it. They link to the originating order or authorised operator policy, geometry, revision, exceptions and validation status. Updates should be event-driven where possible, with reconciliation checks; a nightly import alone is insufficient for temporary changes. If a digital extract and source order conflict, the disputed rule is withheld from automated eligibility decisions pending authoritative review. Historical case reconstruction uses the rule and information available at the relevant event time.

Table 2: Minimum information contracts proposed for reproducibility.

Record	Required content	Audit or failure behaviour
Bay	Stable ID, geometry, type, provenance, validity dates	Retain revision and split or merge lineage
Observation	Source, event and ingestion times, state, quality	Preserve conflicts; expire stale observations
Rule	Source order, effective dates, recorded date, exceptions	Unknown if unresolved; retain prior versions
Forecast	Origin, horizon, probability, model and feature versions	Record freshness and served output unchanged
Case	Observations, rule version, reviewer, reasons, evidence digest	Trace correction, disclosure and disposition

Candidate data inputs

We begin with a minimum dataset comprising validated bay geometry, physical occupancy observations, independently audited labels, rule versions and calendar features. Additional inputs require availability and licence checks before inclusion. Weather must be the forecast issued by t, not subsequently observed weather. Events and closures require publication timestamps; local-authority sources may be necessary where national feeds do not cover local roads. School dates require location-specific calendars. Traffic, charging and payment inputs need explicit coverage and latency descriptions.

Candidate inputs are separated into predictors, labels and evaluation records to prevent circular learning. Historical utilisation may be derived from valid past occupancy. Forecast errors become monitoring signals only after their outcomes mature. Interaction logs are affected by the guidance policy and are not neutral measures of demand. Emission factors belong in a separate impact calculation, not the occupancy target. Cross-city aggregates are optional transfer inputs subject to disclosure-risk assessment; aggregation is not proof of anonymisation. ANPR, identity and precise journeys are excluded from the default forecast feature set.

Forecasting and model governance

Candidate model specification

Our initial TCN takes 288 five-minute steps, covering the preceding 24 hours, with occupancy, missingness and observation-age channels. Six residual blocks each contain two causal convolutions with kernel size 3 and dilations 1, 2, 4, 8, 16 and 32. The resulting receptive field is 253 steps, within the supplied history. Calendar features and past same-time daily and weekly occupancy summaries provide longer-period context without claiming that one day of input captures annual seasonality. Padding and normalisation must not expose future samples.

Each of the six output heads returns an occupancy probability trained with binary cross-entropy on valid labels. The contextual XGBoost component uses lag summaries, calendar variables, bay type, published weather and event information, rule features and missingness indicators. Separate horizon models are proposed. Bay identifiers are excluded in the primary transfer experiment to avoid treating unseen identifiers as learned location characteristics. A known-bay identifier ablation may be reported separately.

Combine the two components as $q(h) = \alpha(h)pTCN(h) + [1 - \alpha(h)]pXGB(h)$, with $\alpha(h)$ constrained to [0,1]. Fit ensemble weights on a chronological validation block after component training. Fit a per-horizon logistic calibration map on a later, separate calibration block. The map is $p(h) = \text{sigmoid}[a(h) + b(h) \text{ logit } q(h)]$, with clipped q used for numerical stability. Choosing an endpoint weight of zero or one is permitted; the ensemble should not be retained merely because it was proposed. Report raw and calibrated scores to show whether calibration improves reliability or conceals component weaknesses. Apply the same calibration opportunity and chronological partition to probabilistic baselines so that the comparison does not favour the ensemble by construction.

Table 3: Proposed tuning space to be fixed before test access. Values are design choices, not fitted parameters.

Component	Candidate settings	Selection rule
TCN	Channels 32 or 64; dropout 0.1 or 0.3; learning rate 0.001 or 0.0003	Validation log loss; maximum 100 epochs; patience 10
XGBoost	Depth 3 or 6; rate 0.03 or 0.1; maximum 1000 trees; subsample 0.8; column sample 0.8	Validation log loss; early stopping after 50 rounds
Ensemble	One convex weight per horizon	Minimise validation log loss after fitting components
Calibration	Logistic intercept and slope per horizon	Fit on later calibration data only
Repeatability	Seeds 11, 23, 37, 53 and 71	Report all seed scores and training cost

All omitted library settings must be serialised in the experiment configuration with dependency versions. The tuning budget and early-stopping data are recorded for every comparator. Class weighting is not applied by default because it changes the probability estimation objective; any weighted variant requires separate calibration and reporting. Probabilities are clipped to $[10^{-6}, 1 - 10^{-6}]$ for numerical log-loss calculation only.

Time separation and leakage controls

The first proposed offline study uses at least 180 consecutive days, subject to access and sufficient label quality. The final 28 days form an untouched test block. The preceding 14 days are reserved for calibration and the preceding 28 for model selection and ensemble fitting; earlier observations provide training data. These durations define an initial design, not an acquired dataset or proof of annual representativeness. If data are insufficient, revise and register the protocol before test access rather than recycling the test block as validation.

At each boundary, exclude forecast origins whose 90-minute target window crosses into the next block. Fit imputers, scalers, encoding, feature selection and model parameters within the training partition only. Historical context at a later origin is permitted when it would have been available in real time; future labels are not. Rolling-origin development folds provide additional stability checks entirely before the final test. A separate spatial holdout evaluates unseen streets, and leave-one-city-out testing evaluates transfer only when genuinely independent city data exist.

The primary offline test freezes the model throughout the final block. An adaptive experiment is reported separately as a forward replay in which every retraining event uses only matured past outcomes. It must not select a new champion based on later test performance. Forecast identifiers, cut-off times and versioned feature snapshots make these distinctions auditable.

Monitoring and abstention

Forecast probability, observation reliability and predictive uncertainty are different quantities. A probability of 0.65 is not a universal confidence threshold. The interface displays occupancy probability, observation age and any unavailable inputs. Operational abstention uses validated data-quality rules and calibration diagnostics; report the fraction of bays withheld alongside error on served bays. Otherwise an apparently accurate service could achieve its score simply by omitting difficult locations.

Nightly retraining is an experimental candidate, compared with weekly and frozen models. A 90-day window and heavier weighting of recent observations are ablations, not default evidence of improvement. Promotion requires rolling validation, calibration and subgroup checks, resource-budget compliance and a rollback path. Retraining failure retains the last accepted model while exposing stale service status. Drift monitoring considers changes in occupancy prevalence, missingness and residuals; it cannot establish the cause of a change by itself.

Inference and decision trace

At each forecast origin, the proposed inference procedure first retrieves the bay revision and constructs an as-of snapshot of observations and contextual inputs. It rejects duplicate or future-dated records, records missing channels and applies the declared freshness policy. The TCN and contextual model then generate six probabilities, which are combined and calibrated using the stored horizon-specific parameters. Each output retains its forecast origin, model version and feature-snapshot identifier. A failure of inference or an unresolved input condition returns a documented fallback or unavailable status rather than a silently fabricated probability.

The guidance service next evaluates eligibility for the intended arrival and stay. It ranks only candidates that meet the hard constraints and returns the forecast, timestamp and restriction explanation together. It records the candidates considered, exclusion reasons and final recommendation so that an unsuccessful recommendation can be reconstructed. Enforcement support reads the observation and rule services through a separate interface; it cannot consume a forecast as proof of an event. This information-flow restriction is a proposed property to test with interface checks and negative test cases.

For example, a physically free bay may become restricted during a requested stay. The guidance service should exclude that bay even if the predicted occupancy probability is low. Conversely, an occupied bay with uncertain permit entitlement should not generate a confirmed contravention. These are illustrative logical cases, not observations or experimental results. They show why forecasting quality and decision correctness require separate evaluation.

Accountable enforcement support and UK relevance

Decision boundaries

ParkIQ provides a proposed case referral and evidence assembly, with meaningful officer review before any enforcement action in the initial pilot. Observations first establish whether a potentially relevant event exists. The rules service then checks the applicable restriction and known exceptions. Missing evidence, uncertain vehicle-to-bay association or unresolved entitlement requires further observation or closure. A sensor score alone cannot establish liability. The reviewer records reasons and can reject the referral; review is not a mandatory approval of the system's suggestion.

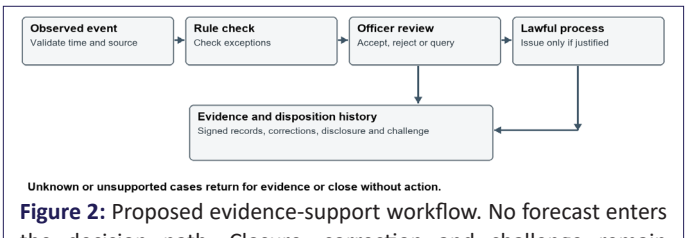


Figure 2: Proposed evidence-support workflow. No forecast enters the decision path. Closure, correction and challenge remain possible; referral does not imply that a notice will be issued.

For England, camera-based postal civil parking enforcement is restricted to specified circumstances and approved devices. A generic ANPR camera and confidence threshold do not create authority to issue a PCN. The statutory ten-minute grace period applies in defined designated parking-place circumstances after permitted parking ends; it is not a configurable universal 5–30-minute allowance. Revenue raising is not the purpose of civil enforcement [6]. ParkIQ therefore needs a jurisdiction and contravention-specific configuration that preserves applicable service, observation, exemption and appeal requirements. Public penalty charge notices and private contractual parking charges are separate workflows.

Warnings may assist consenting users, but delivery or app registration cannot be a prerequisite for fair treatment. The system should neither invent a legal entitlement through an app countdown nor remove an existing one when a message fails. Keeper information is requested only through an authorised organisation and permitted route, with the required justification; platform credentials are not a substitute for access authority (DVLA, n.d.). A registered keeper must not automatically be labelled the observed driver.

Evidence integrity and threat model

The proposed evidence store contains original observations, relevant imagery where lawful, clock-quality information, device identity, the rule snapshot, reviewer reasons and disclosure history. A canonical manifest lists file digests and versioned metadata. The service signs the manifest and chains successive log entries through their predecessor digests. Verification recomputes digests and checks signatures, authorised keys and chronology. Corrections create linked records rather than overwriting history. Retention expiry can delete underlying personal evidence while retaining only metadata that remains lawful and necessary.

The threat model includes post-capture alteration, record deletion, compromised device keys, clock manipulation, unauthorised access and a dishonest participant. Signed logs address some alteration and attribution risks; independent replication or witnessing can improve detection of omissions. Neither mechanism proves that an event was captured truthfully. Missing capture, colluding witnesses and compromised signing keys remain residual risks requiring operational controls, revocation and investigation.

The optional permissioned ledger stores only carefully assessed batch commitments, not raw plates, imagery or journeys. Even hashes and linkable identifiers may remain personal data. Membership, quorum, key rotation, outage behaviour, controller roles and deletion implications must be specified before this variant is implemented. The audit experiment compares it with independently witnessed signed logs under identical injected faults, reporting detection rate, verification time, storage, recovery and administrative cost. Blockchain is not assumed to make a record admissible or a system compliant.

Data protection and inclusive services

ANPR processing requires a documented necessity and proportionality assessment, a data protection impact assessment, suitable transparency and purpose-based retention (ICO, n.d.). The proposal excludes a blanket seven-year retention rule. Each data class needs a justified schedule distinguishing routine occupancy, unsuccessful referrals, live cases and material retained for a defined dispute. Raw evidence remains access-controlled and encrypted; forecast data are separated from identities. A data-sharing agreement and technical restriction should govern each recipient's purpose and access.

The Data (Use and Access) Act 2025 changes the framework for significant solely automated decisions, while retaining safeguards [5]. The implementation must map the applicable current provisions and processing regime before deployment. An architecture alone cannot establish legal compliance. Meaningful review, understandable reasons, correction and accessible challenge are design requirements irrespective of any later proposal for greater automation.

Public-facing services should target WCAG 2.2 AA and publish an accessibility statement in line with public-sector guidance (GDS, n.d.). Non-colour map cues, keyboard and screen-reader support, readable restrictions and non-smartphone routes are part of the design. Driver guidance must not encourage interaction while driving. Accessibility testing should include disabled users, while collecting sensitive attributes only with an appropriate research basis and safeguards.

Policy interoperability and charging

The D-TRO service is a candidate source interface for rule extracts, subject to checking spatial coverage, update delay and source-order reconciliation [7]. Payment integration should investigate compatibility with participating National Parking Platform services rather than assume a new proprietary payment account is necessary [8]. Neither integration has been implemented or agreed.

The LEVI programme supports charging provision for residents without off-street parking in England [14]. ParkIQ could contribute observations of physical bay occupation and turnover to charging planning, but connector telemetry alone cannot establish access. Evaluation should measure blocked charger time, accessible-bay availability and displacement to nearby streets. Policy value must be established through these outcomes, not by quoting national EV adoption figures as evidence of ParkIQ impact.

Evaluation protocol

Study status and staged gates

No city partner, instrumented site, participant cohort or experimental dataset is established by this manuscript. A possible feasibility stage covers 200–500 bays, but this is a planning envelope rather than a powered sample. Hardware procurement, permissions, ethics review, data agreements, mapping validation and a statistical analysis plan are prerequisites. The protocol should be deposited in a timestamped registry before confirmatory data analysis, with amendments and deviations reported.

Stage A validates schemas and rule logic using explicitly synthetic fixtures, including daylight-saving changes, overlapping restrictions, revoked permits, delayed events and sensor failure. These tests can establish software behaviour only. Stage B collects observational data and independently audits labels before offline model comparison. Stage C runs silent referrals alongside existing operations, without changing enforcement decisions. Stage D evaluates guidance in a controlled field study. Multi-city transfer and dynamic pricing are later studies rather than scheduled commitments. Each stage requires acceptable data quality and governance before progression.

Ground truth and sampling

An independent occupancy audit should sample bay-time observations across streets, time of day, weather, bay type and sensor-reported state, including unknown periods. Stratified sampling oversamples rare transitions and sensor disagreements, with known inclusion probabilities retained for weighted population estimates. Annotators should be blinded to forecast outputs. Double annotation on a prespecified subset, adjudication of disagreements and inter-annotator agreement quantify label reliability. Record occlusion and time uncertainty rather than forcing binary labels.

Sensor quality is assessed against this audit; the sensor being tested cannot be its own unquestioned ground truth. For forecasting, report valid labelled counts at each horizon, missingness by reason and sensitivity to observation tolerance. Long occupancy runs generate correlated examples: millions of timestamps are not millions of independent observations. Bay, street and day counts must accompany sample totals. Access to any public dataset would require a separate provenance, licence and target-compatibility assessment.

Hypotheses and endpoints

Table 4 specifies directional hypotheses. The primary confirmatory comparison is H1 at 30 minutes. Other horizons and hypotheses are secondary unless separately powered and registered. All effect estimates require uncertainty intervals; neither favourable point estimates nor non-significant differences alone establish success.

Table 4: Proposed hypotheses and evaluation contracts. None has been tested or preregistered.

Hypothesis	Comparator and endpoint	Interpretation
H1 Ensemble improves forecasting	Best development-selected baseline versus ensemble; mean Brier score of ensemble minus baseline at 30 minutes	Support requires an interval below zero; report all horizons
H2 Rule and freshness checks improve guidance	Same forecasts with versus without the checks, replayed safely; unsuitable recommendation rate and coverage	Compare at matched coverage; also report the full coverage–error trade-off
H3 Provenance improves review	Usual evidence pack versus versioned pack; correct reconstruction rate and review time in blinded tasks	Report unsupported-referral rate and reviewer disagreement
H4 Guidance reduces search time	Cluster-assigned guidance versus usual information; time from declared search start to confirmed parking	Estimate intention-to-treat difference including failed searches
H5 Transfer improves data efficiency	Target-only versus transferred model at equal local-data budgets; Brier score and calibration	Report negative transfer and subgroup changes, not only average gain

H2 is initially a replay or simulation comparison; unsafe guidance is not deliberately served to drivers. Suitability labels must be independently adjudicated against complete rules and user constraints. H3 uses authorised or synthetic cases with known review answers, randomised presentation order and reviewers blinded to experimental labels where feasible. The ledger-versus-signed-log study is a separate engineering comparison, not proof of H3 or legal validity.

Forecast metrics and comparators

For N valid forecast–label pairs, Brier score is $BS = (1/N)\Sigma(p_i - y_i)^2$. Log loss is $LL = -(1/N)\Sigma[y_i \log p_i + (1 - y_i)\log(1 - p_i)]$. Both are reported by horizon, bay type, location and occupancy prevalence. Also report reliability diagrams, calibration intercept and slope, precision–recall performance and, at a threshold chosen on development data, sensitivity, specificity and balanced accuracy. Plain accuracy alone can reward an always-occupied classifier. Report micro averages and equal-bay macro averages so high-volume locations do not dominate every conclusion.

Required baselines are last-observation persistence, a smoothed bay-by-time-of-week frequency model, logistic regression with the same available features, XGBoost alone and TCN alone. Persistence is included as a classification benchmark; probabilistic scores use prespecified development-fitted smoothing to avoid meaningless infinite losses. Where enough connected locations exist, include a graph-based temporal baseline motivated by Yang et al. [16]. Compare like targets: a block-level published percentage error is not a bay-level accuracy benchmark.

The ablation plan removes contextual feeds, calibration, quality indicators and each ensemble component in turn. Evaluate sensor dropout at proposed rates of 10%, 30% and 50%, delayed messages and contiguous outages, with multiple documented random seeds. Artificial faults supplement, rather than replace, naturally occurring failures. Cross-city experiments use local-data budgets of 7, 14, 28 and 56 days when available, with an identical target test period. Hyperparameters are selected without target-test access.

Use paired cluster bootstrap intervals for forecast differences, resampling street-level series and contiguous day blocks to reflect spatial and temporal dependence. The final resampling design and block length must be fixed using development residuals. For operational effects use inference at the randomisation unit; individual forecast rows are not valid independent replicates. H1 has a two-sided 95% interval, with improvement supported only when its upper bound is below zero. Apply a prespecified multiplicity correction, such as Holm adjustment, to any family of additional confirmatory tests; label remaining analyses exploratory.

Controlled operational evaluation

Our preferred guidance trial assigns comparable street clusters or zones to guidance access or usual information, after a common baseline period. Stratify assignment by baseline occupancy and land use; record allocation, eligibility and intervention dates. Keep tariffs and enforcement policy stable during the primary comparison. Buffer areas and adjacent-street monitoring help measure spillover, but cannot guarantee independence when drivers move between zones. Record app exposure and cross-zone trips; retain assigned groups in the primary intention-to-treat analysis.

Search starts when a participant declares the beginning of a parking search and ends at independently corroborated parking. Use consented traces or diary confirmation with an audit subset and report disagreement between methods. Abandoned searches are retained as failures or censored observations under a prespecified time-to-event analysis, not dropped. Report restricted mean time over a common window, completion probability and walking distance. This avoids presenting faster successful searches as an overall benefit when more drivers fail to park.

A matched before–after difference-in-differences study is a fallback if randomisation is infeasible. It requires pre-trend checks, concurrent-change records, sensitivity analyses and weaker causal language. Trial duration and cluster count require simulation-based power analysis using baseline variation, intra-cluster correlation, expected spillover, attrition and a

stakeholder-agreed minimum worthwhile change. A count of 200–500 bays alone cannot establish power. A single-city pilot cannot support a national effect estimate.

Compliance is measured as independently observed contraventions per 100 occupied bay-hours, with stable observation effort. Issued notices measure both behaviour and enforcement intensity and are unsuitable as a sole compliance endpoint. Report unsupported referrals, cancellations, challenges and outcomes with explicit denominators and follow-up windows. A low appeal rate may reflect difficulty challenging a notice; include accessibility and user-comprehension measures. Compare similar public-enforcement settings rather than using private-ticket appeal figures as a universal baseline.

Environmental analysis uses observed changes in search distance or time and declared vehicle and emission-factor assumptions. Distinguish modelled tailpipe estimates from measured pollutants, and report rebound, displaced trips and infrastructure energy as limitations. No carbon-credit entitlement follows from an estimated avoided emission. Economic analysis reports tariffs, usage, collection costs, refunds and service costs separately; penalty income is neither the success target nor evidence of improved welfare.

Reliability and reproducibility outputs

Candidate engineering targets are a 95th-percentile map response within 200 ms, sensor-to-map delay within 10 seconds, and monthly guidance availability of 99.9%. These targets require a declared number of concurrent users, bay count, payload size, hardware, measurement locations and outage definition. The 10-second target applies to successfully delivered events; missing or offline devices are reported separately. Report tail latency, error rate, stale-bay coverage and recovery under replay, network loss and backend outage. None of these service levels has been measured.

The proposed reproducibility release includes the bay and event schemas, rule fixtures, data dictionary, extraction and labelling code, split manifests, seeds, dependency lockfiles, full model configurations, baseline implementations, trained artefact hashes and commands for reproducing every table. Raw personal data are not promised for open release. Where lawful sharing is impossible, provide disclosure-reviewed aggregates, synthetic fixtures and a controlled-access procedure if approved. Synthetic data must be labelled and must not substantiate field-performance claims. At present no ParkIQ code, trained model or dataset is supplied with this paper.

Discussion of AI deployment and operational feasibility

The architecture locates learning where uncertainty about future occupancy is unavoidable, while using explicit rules and review where decisions require traceable reasons. This allocation creates a falsifiable trade-off: additional interfaces may improve consistency and accountability, but also increase integration cost, failure modes and latency. The evaluation should therefore compare the integrated system with simpler alternatives rather than equate architectural breadth with technical superiority.

An explanation to a driver should identify the forecast horizon, data age and relevant restriction; it should not present feature importance as a causal account of demand. An explanation to a reviewer should identify the underlying observations and rule version. The two explanations serve different decisions. Measuring whether users understand them is part of evaluating the system, rather than a property guaranteed by displaying more information.

The potential public value is improved access, reduced search, better charging-bay use and more accountable review. Officer deployment could become better informed, but the design does not demonstrate that staff can be removed or that detection replaces the other functions of civil enforcement officers. Any workload assessment must include training, exception handling, equipment maintenance and challenges. Procurement should separate performance assessment from incentives to generate notices.

Commercial options include authority service licences, operator subscriptions and disclosed booking fees. Prices, adoption and willingness to pay are untested. The preferred starting arrangement is a service fee independent of notice volume to reduce incentive conflict. Secondary data licensing requires purpose and disclosure assessment; carbon-credit income is excluded unless a separate recognised methodology and verification process establish eligibility. International licensing is outside the current evaluation.

For an evaluation period, define incremental operating cost as hardware and installation amortisation plus connectivity, maintenance, hosting, rule curation, support, audit and review costs, less displaced costs actually evidenced. Report these components and sensitivity to sensor lifespan, failure, utilisation and scale. Revenue and social benefit are separate accounts: higher collections may be transfers rather than net public benefit. Private capacity can be considered only where owners have authority, access arrangements and verified availability; nominal empty capacity is not demonstrated rentable supply.

Limitations and threats to validity

The main limitation is maturity: the system has not yet been implemented or tested. A specification can expose design assumptions and make evaluation possible; it cannot demonstrate model accuracy, usability, lawful deployment, economic viability or external recognition. The proposed ensemble may be unnecessary if seasonal or tree-based models perform as well. The selected related work is not an exhaustive priority search, and no claim that ParkIQ is the first integrated platform is justified.

Construct validity depends on correctly distinguishing physical occupancy, legal eligibility and search completion. Sensor-derived labels can embed systematic errors; audit sampling and uncertainty reporting reduce but do not eliminate this risk. Internal validity in a field study is threatened by spillovers, weather, local events, tariff changes, enforcement intensity and selective app use. Interventions also change future training data, so a forecast evaluated in passive conditions may become miscalibrated when used for guidance.

External validity is limited by city layout, bay types, sensor coverage, driver behaviour and differing legal regimes. A single season cannot establish annual performance. Underrepresented streets and disabled users may experience worse service even when average metrics improve. Distributional analysis should report missingness, calibration, unsuitable recommendations and false referrals by operationally relevant groups, with uncertainty and disclosure controls; lack of statistical significance is not evidence of equal treatment.

Technical risks include mapping errors, delayed restrictions, network outages, compromised sensors, cyberattack and privacy leakage through trajectories or supposedly anonymous aggregates. A common data model can also propagate one incorrect rule across multiple services. Independent validation, version rollback and restricted evidence access are therefore central design requirements. Ledger anchoring introduces governance and availability dependencies without resolving errors at capture. Simulated fault tolerance is not evidence of resilience in a live authority environment.

Conclusion

ParkIQ is proposed as a testable architecture for connecting parking forecasts, time-indexed restrictions and accountable evidence handling. Its principal intellectual contribution is the specification of information boundaries: occupancy is not eligibility, probability is not observed evidence, and a cryptographic commitment is not proof of correctness. These boundaries support clearer implementation requirements and stronger evaluation than a capability checklist alone.

The next milestone is a versioned implementation and independently audited observational dataset, followed by the comparisons to be preregistered as described here. Demonstrated value would require improved calibration and decision outcomes against strong baselines, acceptable distributional effects, and operating costs justified by measured benefits. Until those studies are completed, performance, adoption and wider impact remain open questions.

Declarations

Funding: No funding was received for this work.

Data and code availability: This article presents a architecture and protocol. It reports no collected ParkIQ dataset, trained model or empirical study; no implementation repository is provided.

Ethics: No participant study is reported. Any subsequent collection of personal or participant data requires the relevant ethics and governance assessment before collection.

References

1. Al Amiri W, Baza M, Banawan K, Mahmoud M, Alasmary W, Akkaya K. Privacy-preserving smart parking system using blockchain and private information retrieval. arXiv preprint. 2019; arXiv:1904.09703.
2. Arnott R, Inci E. An integrated model of downtown parking and traffic congestion. J Urban Econ. 2006; 60: 418-442.
3. Bai S, Kolter JZ, Koltun V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv preprint. 2018; arXiv:1803.01271.
4. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016: 785-794.
5. Department for Science, Innovation and Technology (DSIT). Data (Use and Access) Act 2025: data protection and privacy changes. GOV.UK. 2025.
6. Department for Transport (DfT). Statutory guidance for local authorities in England on civil enforcement of parking contraventions. GOV.UK. 2022.
7. Department for Transport (DfT). Digital traffic regulation orders (D-TRO) service. GOV.UK. 2025.
8. Department for Transport (DfT). Government teams with parking giants to ensure drivers can use preferred apps in all car parks. GOV.UK. 2025.
9. Driver and Vehicle Licensing Agency (DVLA). Request information about another vehicle or its registered keeper. GOV.UK; [date unknown].
10. Government Digital Service (GDS). Understanding accessibility requirements for public sector bodies. GOV.UK; [date unknown].
11. Guo C, Pleiss G, Sun Y, Weinberger KQ. On calibration of modern neural networks. In: Proceedings of the 34th International Conference on Machine Learning. Proceedings of Machine Learning Research. 2017; 70: 1321-1330.
12. Information Commissioner's Office (ICO). Automatic number plate recognition (ANPR): guidance on video surveillance. [date unknown].
13. Kuang H, Deng K, Wang Q, Ye W, Qu H, Li J. Deep meta-learning approach for regional parking occupancy prediction considering heterogeneous and real-time information. Adv Eng Inform. 2025; 64: 102969.
14. Office for Zero Emission Vehicles (OZEV). Apply for Local Electric Vehicle Infrastructure (LEVI) funding. GOV.UK; 2023.
15. Vlahogianni EI, Kepaptsoglou K, Tsetsos V, Karlaftis MG. A real-time parking prediction system for smart cities. J Intell Transp Syst. 2016; 20: 192-204.
16. Yang S, Ma W, Pi X, Qian S. A deep learning approach to real-time parking occupancy prediction in transportation networks incorporating multiple spatio-temporal data sources. Transp Res C Emerg Technol. 2019; 107: 248-265.
17. Online policy and regulatory sources accessed 8 September 2026.