DESIGN DOCUMENT: Secure Future INSURANCE

Machine Learning System Design

1. Problem Definition

1.1 Origin

Secure Future is an insurance provider operating in multiple countries, serving hundreds of thousands of auto insurance customers. The company processes thousands of claims monthly, most related to vehicle collisions.

When a customer reports an accident, the insurer must first assess the damage before authorising repairs and/or payouts. At present, this process is slow and looks something like this:

- A claims adjuster must travel to inspect the vehicle.
- Photos and forms are manually reviewed.
- Paperwork often delays even the first decision on the claim by several days.

For example:

- **Best case** – A customer submits photos, an adjuster visits the same day, and Secure Future can start the claim within 24 hours. This is optimal but quite rare.
- **Bad case** – Adjuster visit delayed by several days. Repairs and payouts stall, frustrating customers.
- **Worst case** – The process is so delayed that the customer moves to a competitor which offers quicker handling.

In the age of instant services, even waiting for a few days for a decision feels unacceptable to customers. The goal is to reduce the time from claim submission to initial damage report from **days to under 5 minutes**.

1.2 Relevance and Reasons

1.2.1 Current Workflow Analysis

Current claims handling:

- Customer submits a claim online or by phone.
- A claims handler reviews it and schedules a physical inspection.
- An adjuster visits the site, photographs damage, and estimates cost.

- Report is entered into the system, starting repairs or payout.
  Issues:
- Delay before first contact with customer.
- Long "dead time" before payout approval.
- The model does not cover all the parts classes yet (e.g. fender, pillar, chassis).
- Low recall/precision for some types of damage → defects may be missed, especially small cracks or damage hidden in shadows.
- No complete calibration for rare cases (e.g. rare cars with no records, records for motorcycle vehicles, non-standard body shapes or tuning).

1.2.2 How much does Secure Future lose because of delays?

- **Customer churn:** Some policyholders switch insurers after a bad claim experience.
- **Operational cost:** Adjuster site visits cost time and travel expenses.
- **Inefficient resource allocation:** Highly skilled adjusters spend time on simple, obvious cases that could be automated.

Estimates from internal audits suggest potential savings in the **millions annually** if even 30% of simple claims were automated.

1.2.3 Other reasons to act now

- Competitive edge: Faster claim handling is a key differentiator in the insurance market.
- Regulatory goodwill: Faster settlements improve compliance with service-level agreements.
- Cross-industry potential: The same damage-assessment engine could be licensed to other insurers or rental car companies.

2. Previous Work

Questions before starting:

- Has Secure Future already tried automated photo-based damage estimation?
- Is there an internal pilot tool (e.g. basic photo upload in claims portal) that we can expand?
- Do we have limits we must respect (e.g. certain claims always require in-person checks for fraud prevention)?
- Can we integrate with existing CRM/claims platforms without major rework?

From early discussions, no full automation exists yet — only a manual "upload photo" step that still requires fully human validation.

3. Metrics

The metrics are aligned with the key business objective: minimise friction in claims processing, keep clients from leaving, and maximise retention-driven revenue.

3.1 North Star Metric

Retention rate — percentage of customers who stay with us over time. Faster, more transparent claims $\rightarrow$ higher retention $\rightarrow$ higher LTV.

3.2 Operational Efficiency Metrics

- **Time-to-First-Decision (TTFD):** Avg. time from claim submission (photo, description) to the first automated assessment. Target: $<$ 5 minutes.
- **First-Pass Automation Rate (FPAR):** Share of claims fully processed without human intervention. Higher FPAR = adjusters focus only on complex cases.
- **Adjuster Productivity:** Number of cases handled per adjuster per month. Expected to grow as routine cases shift to automation.
- **Operational Cost per Claim:** Total cost of processing a claim (labor + trips + overhead). Target: continuous reduction.

3.3 Accuracy & Quality Metrics

- **Error/Discrepancy Rate:** Difference between automated vs. manual assessment (in monetary terms). Goal: minimise underestimation risk.
- **Customer Satisfaction (CSAT/NPS):** Direct feedback loop validating automation impact.

3.4 Revenue Protection Metrics

- **Churn Cost Avoidance:** Estimated money saved by preventing customer churn thanks to faster and more accurate claims. Bridges retention with hard $$ impact.

4. Validation Schema

4.1 Requirements and Assumptions

To ensure reliable evaluation of the ML system, we explicitly list the key assumptions:

- New claims data is collected **daily**.
- Data can arrive with a **delay of up to several days**.
- Labels (segmentation masks, damage detection annotations, repair cost data) are provided asynchronously with the data.

- Recent data is more relevant for the prediction task (concept drift is expected).
- Seasonality strongly affects distribution (peaks in summer and winter).

4.2 Validation Strategy

We employ a rolling validation approach designed to capture both **short-term changes** and **long-term stability**:

- **Monthly rolling split:** At the beginning of each month, the previous month's data is moved to a validation set, which ensures evaluation against the most up-to-date distributions and allows us to monitor the effects of **concept drift** and seasonality.
- **Golden holdout set:** A static benchmark dataset ("golden set") is maintained and updated quarterly. This dataset includes diverse vehicle types, rare cases (e.g., unusual tuning, rare car models), and challenging conditions (shadows, occlusion). It provides a consistent baseline for long-term performance tracking.

4.3 Evaluation Protocol

- **Primary evaluation:** Each model iteration is evaluated monthly against an ongoing validation set.
- **Drift detection:** Compare your monthly results to the golden set to see if the performance gains were real or just adjusted to fit the short-term data distribution.
- **Business-aligned metrics:** Model evaluation includes ML metrics (IoU, precision, recall, RMSE for cost estimation) that are translated into business KPIs (error rate, time to first decision, etc.).

4.4 Update Cycle

- Models are retrained or fine-tuned on a monthly cadence.
- Golden set is refreshed quarterly but always keeps a portion of **frozen historical cases** to ensure comparability.
- Monitoring dashboards track both ML metrics (e.g., segmentation IoU, detection recall) and system metrics (e.g., TTFD, CSAT).

4.5 Benefits

- It allows you to distribute short terminations while maintaining a stable benchmark.
- Mitigates the risk of silent model degradation due to seasonality and label delays.
- Keeps evaluation grounded in both **ML performance** and **business outcomes**.