

Week 2: Counter-Metrics in the Wild

Bálint Décsi

January 30, 2026

1. The Case: Zillow's Algorithmic House Flipping Disaster

Article: *Zillow's home-buying debacle shows how hard it is to use AI to value real estate* (CNN coverage)

Link: <https://edition.cnn.com/2021/11/09/tech/zillow-iBuying-home-zestimate>

In 2021, Zillow shut down its iBuying program after losing over \$500 million in mere months. The company attempted to use its "Zestimate" algorithm to make instant cash offers on homes, aiming to flip them for a profit. However, the algorithm failed to account for market volatility and "unobservable" defects, leading Zillow to purchase thousands of homes at inflated prices that they could ultimately only sell at a loss.

2. The "What Breaks" Framework

- **Primary Metric (The Target): Purchase Volume / Growth.** Zillow incentivized teams to acquire homes aggressively to gain market share.
- **Directly Worsened: Unit Economics (Margin).** To hit volume targets, the algorithm (and human overrides) began making higher offers. CAC went up, but only because they were overpaying as the model was trained on listing prices, not real sale prices.
- **Quality Degraded: Asset Quality (Adverse Selection).** The model neither could see "soft" defects (e.g., bad smells, weird layouts) that human buyers notice. Sellers with "lemon" houses flocked to Zillow's algorithm because it offered more than the open market, while owners of pristine homes sold elsewhere.
- **Segments Harmed: Internal Employees.** The collapse led to a layoff of 25% of the workforce.
- **Long-term Metrics: Company Valuation & Trust.** Zillow's stock price plummeted, and trust in the "Zestimate"—their flagship consumer feature—was severely damaged.

3. Counter-Metrics: What Should They Have Tracked?

To catch this "Adverse Selection" loop earlier, they needed metrics that proxied for the *quality* of the purchase, not just the quantity.

- **Tradeoff Metric: Inventory Aging (Days on Market).**
Why: A "Bad Buy" is revealed when it doesn't resell. If Zillow holds homes longer than the local average, they overpaid.
Threshold: Hard stop on buying in any zip code where Zillow's average "Days on Market" exceeds the local average by 15 %.
- **Guardrail Metric: Short-Term Cash Flow.**
Why: Positive cash flow is one of the goals of flipping houses. If purchase volume increases but cash flow turns negative, it's a red flag.
Threshold: Weekly cash flow must remain positive; if negative for two quarters, pause buying.
- **Guardrail Metric: Model Error.**
Why: If actual sale prices consistently are lower than predictions, the model training lacked

external validity.

Threshold: Weekly monitoring of prediction errors with a threshold for acceptable deviation.

4. Personal Connection & Capstone Application

I have seen a similar pattern in the telecom sector where teams optimize for customer acquisition without measuring long-term value. Agents would offer aggressive discounts to possible pre-paid customers in hope they turn post-paid subscribers, inflating CAC while destroying the Average Revenue Per User (ARPU).

For my capstone with possibly the public sector climate project, this is a crucial lesson. If we optimize strictly for **Model Accuracy** (e.g., predicting flood levels), we might build a model that is great at predicting "no flood" (since floods are rare) but misses the catastrophic events.

Capstone Counter-Metric: I must track **Recall on Extreme Events** as a guardrail, ensuring we don't optimize for global accuracy at the expense of safety in critical edge cases.