Valuing the Unrepeatable

The deployment of multimodal valuation for unique ultra-luxury residences at the parcel level

Nikky Lewis
Curator Index
July 2026

Introduction

Our earlier study established that multimodal machine learning can improve ultra-luxury (defined as properties that sold at $10M+) residential valuation relative to structured-only models, by incorporating the listing photography and marketing language that carry the experiential and aesthetic factors driving price in this segment. This claim held as evidence conveyed that adding visual signals to a structured baseline reduced valuation error, and it continues to hold in our initial deployed system, where the full multimodal configuration remains the strongest performer.

That earlier work demonstrated the claim under controlled conditions, on a dataset with repeated exposure to comparable properties. Carrying the framework into deployment surfaced a requirement those conditions had not tested. This being that a deployed system must value unique properties, identified at the parcel level, that it has never seen, under the uneven multimodal coverage the market actually presents. Rebuilding the framework for these conditions purely on unique properties established the system's current operating performance and is the subject of this paper.

The deployed system currently values individual ultra-luxury residential properties in Los Angeles's Platinum Triangle out-of-sample at a **median error of 17.3 percent,** operating as a single calibrated model tested across properties in this market. In a segment where properties routinely list and then undergo substantial price reductions over many months before they clear, a first-pass valuation at this level of accuracy is an overall strong result. The system reaches it through three capabilities that work in sync. It integrates the visual record the market already produces with structured assessor and parcel attributes, so the estimate reflects the factors that actually differentiate homes at this tier. It stands on a proprietary foundation of over 10,000 transaction prices reconstructed from public records and validated against known sales at essentially zero median error, extending its view into the off-market transactions that public sources suppress. In addition, it reports every valuation as a calibrated range, so each estimate carries an explicit, empirically grounded account of its own precision. Together these produce decision infrastructure for the most complex segment of the residential market.

Understanding of results

The contribution of the multimodal layer is measured by holding everything else constant. Both the multimodal and structured only configurations use the same model class, the same training procedure, the same held-out properties, and the same evaluation only differing in the information available to them. It is worth noting what the structured baseline contains. Rather than the four to six core property attributes that a conventional automated valuation model or hedonic specification typically relies on, it includes continuous spatial coordinates, county parcel geometry, floor-area ratio, assessor quality class, and effective year, sixteen structured features in total. This is a deliberately demanding comparison as testing a new signal against a thin baseline inflates the apparent gain and tells you little, so the structured layer was developed to the strongest form available before the visual layer was evaluated against it.

Configuration	Information Available	Median Error
Structured Only	Assessor, engineered attributes, spatial coordinates, parcel geometry	22.1%
Full Multimodal	Structured layer plus visual signal	17.3%

The improvement is 4.8 percentage points, a reduction of roughly one fifth of the error of the structured baseline. This is the same comparison the earlier study made under controlled conditions, now repeated on unique properties the system has never seen, and it points the same direction.

The consequence is that the visual layer's contribution is measured net of what location alone can explain. Because the baseline can already place a property precisely on the market's price surface, any improvement the imagery provides must come from information that spatial position does not carry, which is what the aesthetic and architectural qualities of a specific home represent.

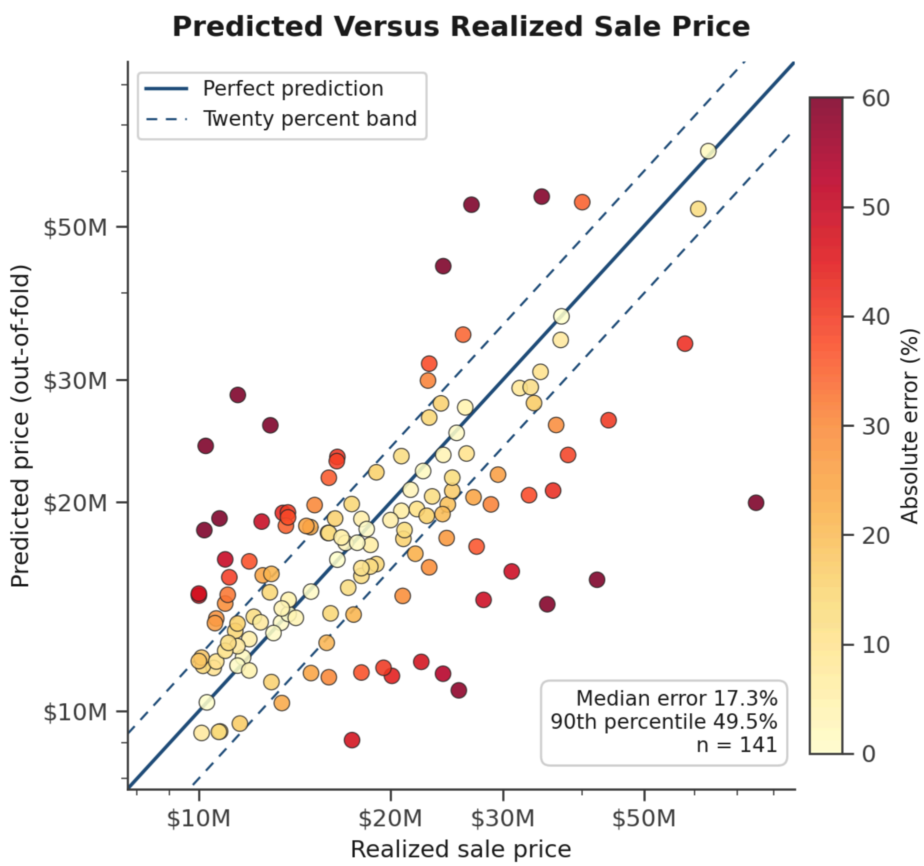

Figure 1: Predicted versus realized sale price, out-of-fold, for the 141 held-out properties. Each point is a property the model priced without having trained on it. The solid line marks perfect prediction and the dashed lines a twenty percent band on either side.

Figure 1 places every held-out valuation against its realized sale price. Because these are out-of-fold predictions, each point represents a property the model priced without having seen it, so the plot is a direct picture of deployable accuracy rather than in-sample fit. The majority of properties fall inside the twenty percent band, and the cloud is centered on the line of perfect prediction rather than tilted above or below it indicating that the system is close to unbiased across the price range and does not systematically over-value or under-value at any tier.

The dispersion is not uniform and the pattern is informative. Predictions are tightest in the lower portion of the range where comparable transactions are most plentiful, and they widen as price rises and comparable evidence thins. This reflects the broader relationship in which accuracy strengthens with the depth of local comparable evidence which is now visible at the level of individual properties. It is also the reason behind the system reporting wider intervals on the most expensive and most singular homes where the widening is a property the model recognizes and prices honestly rather than one it conceals.

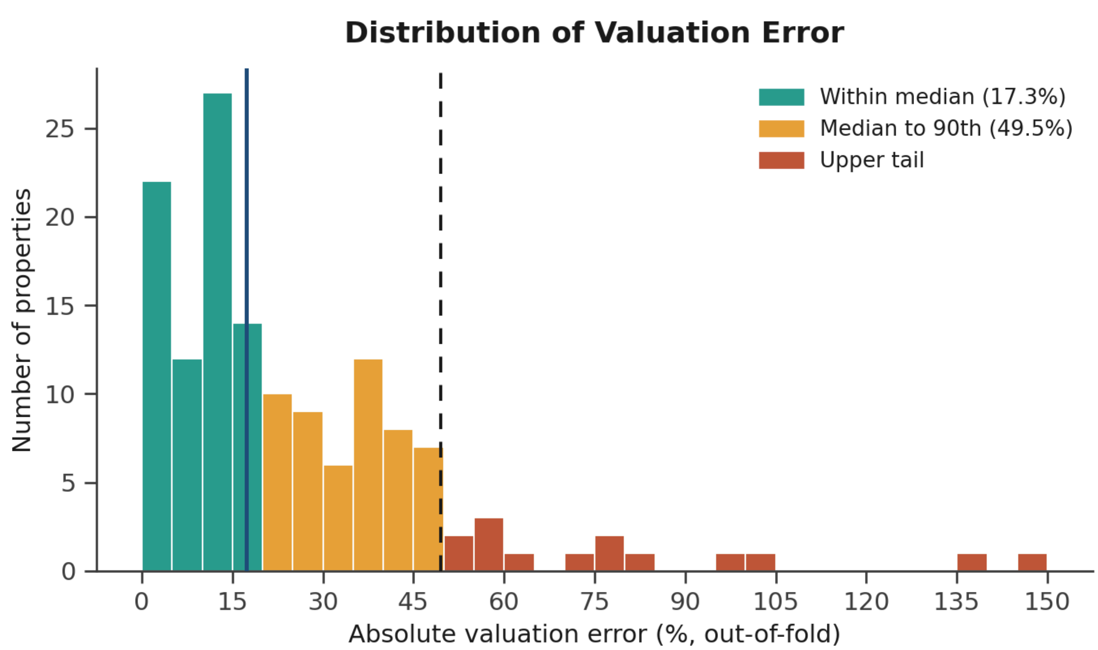

Figure 2: Distribution of absolute valuation error across the same held-out properties, with the median and ninetieth percentile marked.

Figure 2 summarizes the same errors as a distribution. The mass concentrates below twenty percent, with a median of 17.3 percent and a ninetieth percentile of 49.5 percent. The distribution is right skewed, which means the typical valuation is tight while a minority are looser, and a small number of outlier properties exceed one hundred percent error. Those extreme cases are not distributed at random as they correspond to properties that sit far outside the body of the market in size, configuration, or transactional state, or to records with underlying data anomalies, and they are precisely the cases the system's wide intervals and diagnostic checks are built to flag. The concentration of the distribution matters for interpreting the headline figure as the 17.3 percent median is representative of the bulk of the portfolio rather than an average rescued by a handful of exceptional predictions.

Model features in detail

Multimodal signal integration: The model combines multiple sources of data into a single estimate. The structured layer draws on assessor records, engineered attributes, and county parcel geometry, including lot size and quality class extracted from official GIS shapefiles. The visual layer converts listing photography into numerical representations, using a single well-chosen primary image together with a set of interpretable aesthetic scores that measure the presence of specific value-relevant features such as views and grounds. The visual features earn a substantial place in the model.

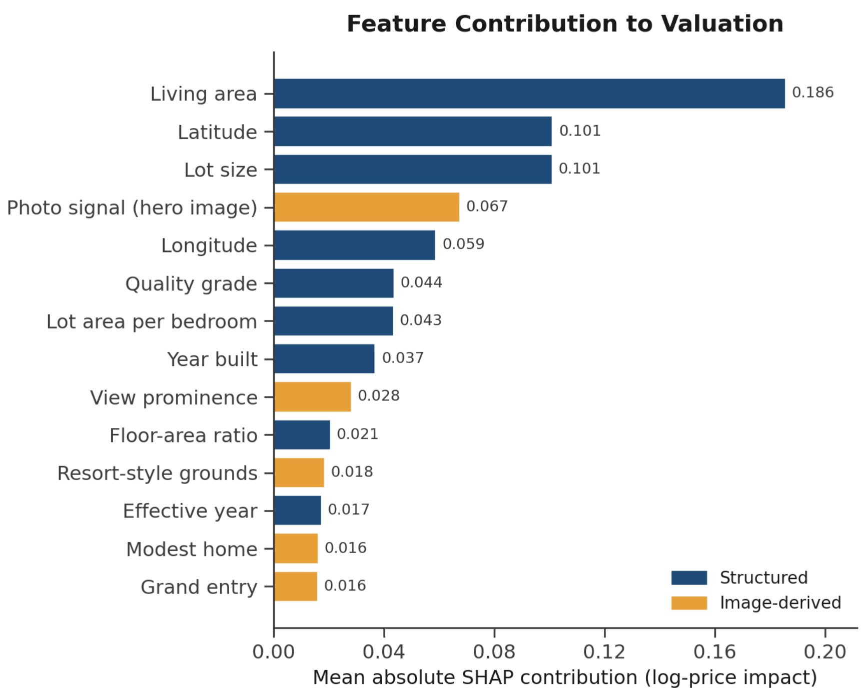

Figure 3: Mean absolute SHAP contribution by feature, in log-price terms. The photo signal is the net contribution of the hero-image representation, computed by summing each property's signed component contributions before averaging.

Figure 3 quantifies how much each feature moves the model's valuations, measured as mean absolute SHAP contribution in log-price terms. Of the sixteen most influential features, eight are image-derived, behaving exactly as domain knowledge predicts where strong view signals raise value, and dated-interior signals lower it. The ordering confirms the hierarchy the system is designed to preserve. Living area, geography, and lot size are overall the dominant drivers, exactly the classical hedonic quantities that should anchor any credible valuation. Image-derived features then occupy a substantial and legitimate share of the remaining influence, refining the estimate beneath the structural anchors rather than displacing them.

Two aspects deserve emphasis. First, the combined photo-signal contribution, aggregated from the latent components of the hero-image representation, ranks fourth among all features, behind only living area, latitude, and lot size. It is the strongest driver in the model that is not a measure of size or location, and it sits ahead of established structured attributes including longitude, quality grade, and year built. Examined component by component, each element of the image representation appears modest, which understates the role of the imagery. Reported as the single net quantity the system surfaces to a user, the photo signal is a leading contributor to valuation in this segment which is the result the earlier study predicted and the deployed system now confirms on unique properties it has never seen. Second, the named aesthetic scores, such as view prominence and resort-style grounds, appear as distinct and interpretable contributors alongside it. The visual signal is therefore not only predictive but legible, since it can be attributed to specific, nameable property qualities rather than to an opaque embedding.

A proprietary transaction-price foundation: A large share of ultra-luxury transactions close off-market, through entities, at prices that public sources suppress, delay, or obscure. The system addresses this directly by reconstructing transaction prices from the mechanics of the assessor's reassessment process. 10,000+ prices were reconstructed this way, and the method was validated against 153 known sales at a median error of essentially zero. The same process surfaced and corrected a data error in the curated records, a property whose price had been recorded off by an order of magnitude.

Calibrated valuation with confidence intervals: The system's primary output is a valuation together with empirically calibrated intervals so that every estimate is precise about both its central figure and the range around it. The intervals are constructed from the model's own out-of-sample residuals, which means their width is earned from demonstrated performance rather than assumed. This is a form of precision that a single-number estimate cannot offer, because a bare point silently conceals how much confidence stands behind it while a calibrated interval states it. A valuation is delivered as a central estimate with a range that reflects exactly how much the system knows about that specific property. On a home valued near eighteen million dollars, that range is measured in the low millions, and it is the quantity that actually governs a pricing or negotiation decision.

Full valuation decomposition: Every valuation the system produces can be broken into the contributions of its drivers, separating the structural, geographic, condition, and visual components of the estimate. This turns the model from a number generator into an explanation. Deliverables from the model's outputs show that a given home is valued where it is because of its scale and location, adjusted by a strong or weak visual profile, rather than being asked to trust an opaque figure. The visual contribution is reported as a single combined photo-signal component, a legible quantity that expresses how much the property's imagery moved its valuation.

Information-sensitive diagnostics: The system recognizes when its richer signals do not apply and relies on structured fundamentals rather than inventing value from sparse inputs. A parcel transacted as vacant land, for instance, carries little architectural or experiential information, and for such a property the visual signals contribute almost nothing while the structured attributes carry the estimate. This behavior bounds the approach and keeps the system disciplined where evidence is thin. The same diagnostic posture allows the system to flag records that appear stale or internally inconsistent, which is valuable in a market where public data is frequently out of date.

Taken together, these capabilities describe a system that is quantitative in the way institutional research is quantitative. Unlike a language model wrapped around a listing feed, the system is providing a calibrated statistical instrument with a proprietary data foundation, an interpretable structure, and a disciplined sense of where its knowledge is strong.

A note on method

The model predicts price on the natural logarithm scale. Working in logs is standard in valuation research and is necessary here because absolute dollar error scales mechanically with price level in a market spanning ten million to well over one hundred million dollars. Differences in log price correspond to proportional differences in price, so a fixed log-scale error means the same relative accuracy at every tier.

Accuracy is reported as the median absolute percentage error across held-out properties. For a property with realized price P and predicted price $\widehat{P}$,

$$APE = \frac{|\widehat{P} - P|}{P}$$

and the headline figure is the median of that quantity across all held-out properties. The median is used rather than the mean because a small number of extreme cases would otherwise dominate the summary and misrepresent typical performance.

The intervals are constructed from the model's own out-of-sample residuals rather than from a distributional assumption. Let r denote the log-scale residual of a held-out property,

$$r = ln\, P - ln\, \widehat{P}$$

and let $q(\alpha)$ denote the α quantile of the absolute residuals $|r|$ across all held-out properties. The interval reported at coverage level α is then

$$\left[\widehat{P} \cdot e^{(-q(\alpha))},\ \widehat{P} \cdot e^{(+q(\alpha))}\right]$$

By construction, an interval built this way contains the realized price for approximately α of held-out properties, which is what makes the reported confidence empirical rather than assumed.

What this means for involved parties

A valuation that carries its supported range alongside its central estimate changes what each party to a transaction can do with it. The value of the system is realized in the decisions the estimate supports, and a calibrated figure supports better decisions than a bare one.

For Sellers: A central problem in this market is the cost of mispricing at listing. Ultra-luxury properties routinely debut at aspirational prices and then endure long periods of downward revision before they clear, and each visible price cut carries a reputational cost and lengthens time on market. A calibrated valuation addresses this at the point of decision. It provides a defensible central figure together with the realistic floor and the credible stretch around it, which is precisely the information required to choose a listing strategy. A seller who knows where the supported range sits can price to transact rather than price to discover, with an explicit understanding of the trade-off between a higher opening figure and a longer, costlier path to sale. The system also identifies which qualities of the property are carrying its value, including the contribution of its visual presentation which informs where preparation and marketing investment will actually be recovered.

For Buyers: The same calibrated range is a map of where value sits relative to the number on the listing. In a market defined by aspirational pricing, the gap between list and supported value is often wide and rarely quantified. The system quantifies it, identifying where a listed price falls inside, above, or below the model's supported range, with an explicit account of the confidence behind that judgment. The decomposition adds a second layer by showing which attributes drive the figure, which supports a more precise negotiating position than a bare estimate can. A buyer can distinguish a property that is expensive because it genuinely carries premium characteristics from one that is expensive because it is mispriced and that distinction is worth millions at this tier.

For the advising principal: The party who holds the relationship and prepares the deliverables, the system is decision infrastructure, not in the sense of replacing judgment, but equipping it. Every figure carried into a client conversation is backed by a calibrated interval and a legible decomposition, which converts an opinion into a defensible position. This compounds over time. A single point estimate can be shown to be wrong when a property trades away from it while a well-calibrated interval that repeatedly contains the outcome builds a record and in a small, relationship-driven segment that record is the durable edge. The calibration is not a hedge. It is what makes the figure defensible, deal after deal.

For the market more broadly: A valuation framework that narrows the range within which participants must search for a clearing price has structural implications. Much of the prolonged repricing that characterizes ultra-luxury listings reflects an information problem rather than irreducible randomness, a gap between the perceptual and experiential value that buyers respond to and the structured attributes that conventional valuation records. A system that internalizes the visual and contextual signal narrows that gap and anchors expectations closer to where the market actually clears. The effect is more credible pricing in a segment where informational opacity is unusually costly, and shorter, less punishing paths from listing to sale.

Conclusion

The model of deployment carries forward the claim from our earlier study which is that multimodal machine learning improves ultra-luxury residential valuation relative to structured-only models by reading the visual and textual record that conventional approaches ignore. The deployed system confirms it under materially harder conditions by valuing unique properties it has never seen at a median out-of-sample error of 17.3 percent against 22.1 percent for a structured-only baseline built the same way. The imagery is not a marginal refinement in this result as the photo signal ranked fourth among all features and stands as the strongest driver that is not a measure of size or location, while the named aesthetic scores move valuations in the directions domain knowledge requires.

In essence, the position this work takes is that valuation in this segment should be treated as a problem of representational completeness and calibrated uncertainty rather than purely of a search for a single confident number. Properties at this tier are differentiated by qualities that structured records do not capture but additional, alternative modalities do. A system that reads those qualities, prices them, and states honestly how much confidence stands behind each estimate produces something more useful than just a point prediction. At this level, it produces a defensible position, which is what a decision at this scale actually requires.