



From Field Geometry to Jewelry Interfaces: Multimodal AI-Assisted Shape Translation and Evidence-Grounded Agricultural Storytelling

Evelyn Liu¹, Lumeng Han^{*2}

¹Design and Technology, Parsons School of Design, NY, USA

²International Agricultural Development, University of California, Davis, CA, USA

Email Address: lmhan@ucdavis.edu (^{*2})

Abstract. *This study develops a field-to-ornament framework connecting agricultural field geometry, computational jewelry design, and evidence-grounded interface storytelling. The main design pipeline uses 350 Fields2Benchmark polygons from the Netherlands, Lithuania, and Estonia as authoritative sources for contour translation. Five abstraction levels are evaluated through geometric similarities and a Geometry Preservation Score (GPS). The Balanced setting retains mean GPS 93.46, compared with 95.06 for Subtle and 88.06 for Abstract. The non-Trace ranking persists across 15,000 sampled weight-scale combinations, although orientation-only weighting reverses Subtle and Balanced. Balanced remains a heuristic demonstration choice without evidence of aesthetic preference. Optional image-recovery experiments assess the limitations of deriving contours when vector geometry is unavailable. On 53 overlay-suppressed annotated map plates, Random Forest achieves mean IoU 0.5417, boundary F1 0.2119, and absolute area error 53.56%. A separate 350-tile AI4Boundaries subset yields modest labelled-extent recovery on 70 aerial test tiles; contour smoothing does not correct prediction errors. Four provenance-card designs are instantiated for 30 fields, producing 120 deterministic audits. Balanced disclosure scores 20.00 and the unsupported minimal claim 100.00 under the proposed overclaim rubric. Sensitivity and controlled content-layout comparisons characterize the rubric's behavior, without measuring comprehension or trust. The framework documents how field geometry can inform two-dimensional ornament concepts while distinguishing design provenance from claims about crop identity, ownership, farming practice, material origin, or environmental benefit.*

Keywords. *agricultural field geometry; computational jewelry design; UI/UX design; visual communication; design provenance; sustainability communication.*

INTRODUCTION

Agricultural fields are both productive territories and visible cultural forms. Their boundaries record topography, infrastructure, ownership history, cultivation practice, and repeated human intervention, but a field seen from above can also be read as a line, surface, rhythm, and silhouette. Satellite and vector datasets make these forms available to computational design. The design challenge is not simply to place an agricultural image beside a jewelry product. It is to translate the geometry of a real field into an ornament while preserving enough of the source form for the relationship to remain inspectable.

Automatic field-boundary extraction has become an important remote-sensing problem because parcel geometry supports agricultural inventories, field-level statistics, and digital services. Earlier approaches detected cropland parcels from spectral and spatial cues, while later convolutional methods learned field interiors and boundaries at several spatial resolutions (Graesser & Ramankutty, 2017; Persello et al., 2019; Waldner & Diakogiannis, 2020). These methods establish a technical foundation for treating a field contour as structured data. They also show why image provenance matters: apparent model performance depends on sensor characteristics, spatial resolution, labeling, and whether reference information is already visible in the input.

The Fields2Benchmark collection offers an unusually direct bridge between agricultural geometry and design. It pairs 350 satellite-basemap renderings with WKT field polygons from the Netherlands, Estonia, and Lithuania. The polygons derive from the wider EuroCrops data

infrastructure, while the collection was assembled to support field-coverage planning for agricultural vehicles (Mier et al., 2023, 2024; Schneider et al., 2023). For the present study, the same pairings support a different but traceable question: how can a field boundary become a jewelry contour and a consumer-facing design-provenance record?

The dataset also imposes a necessary evidentiary limit. It establishes a relationship among a field identifier, a country prefix, a rendered image, and a vector polygon. It does not establish crop type for the selected record, the identity of a farm or owner, a particular agricultural practice, the source of jewelry material, or an environmental outcome. The paper therefore uses the term design provenance: the geometry of a documented agricultural field acts as a visual source for an ornament. A claim about material origin would instead require evidence that the ornament's substance came from the field or its crops.

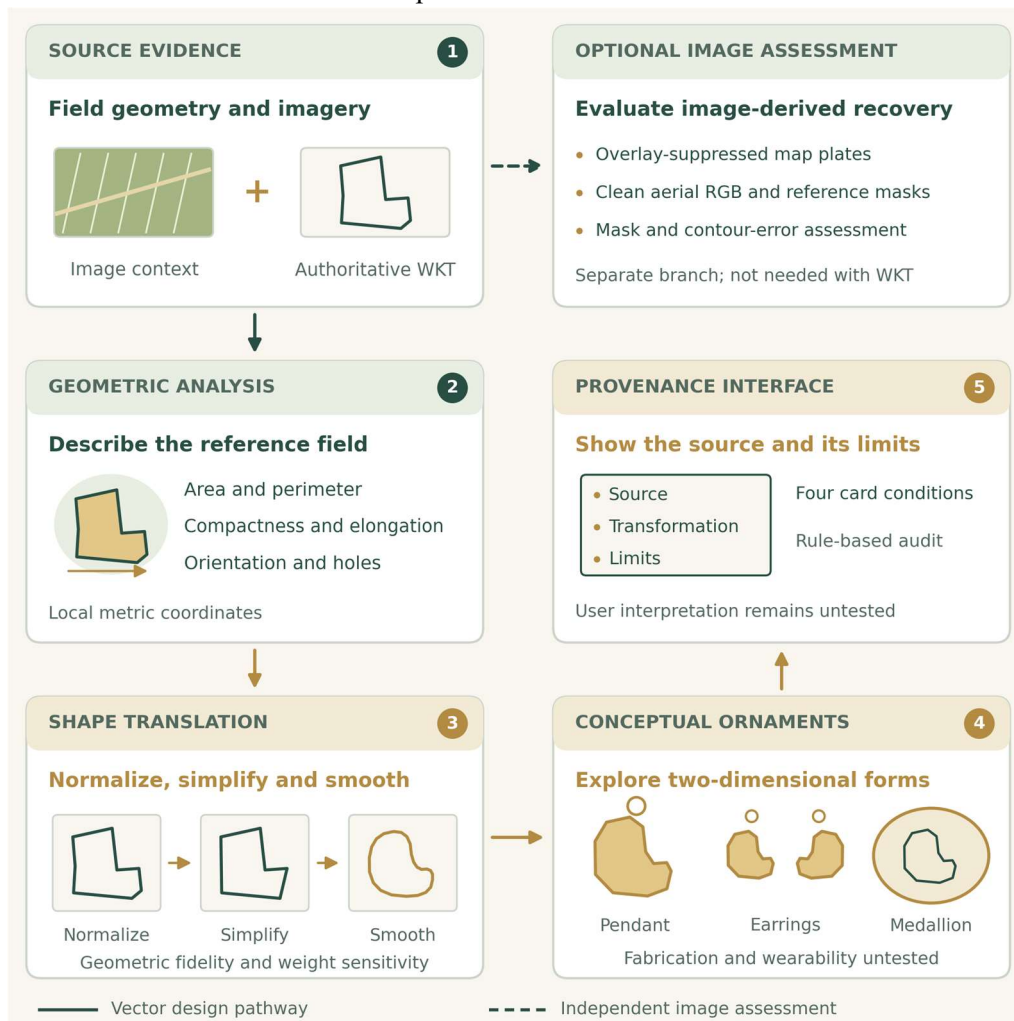


Figure 1. Field-to-ornament framework. Authoritative WKT geometry supports contour translation and provenance interfaces; optional map and aerial-image recovery assess the limitations of image-derived contours.

This distinction matters because environmental and agricultural imagery can lend authority to claims that the underlying data do not support. Vague phrases about regenerative farming, a greener future, or responsible land can make a product appear environmentally verified even when the evidence shows only a shape relationship. Research on greenwashing and environmental advertising has repeatedly shown that vague claims, emotional imagery, and weakly specified evidence can affect how consumers interpret sustainability communication (Delmas & Burbano, 2011; Lyon & Montgomery, 2015; Schmuck et al., 2018). An agricultural

jewelry interface must therefore explain what was transformed, how strongly the derived form resembles its source, and which claims remain outside the evidence.

This study proposes a multimodal field-to-ornament framework in which raster imagery, vector geometry, model output, and language occupy distinct roles. The main pipeline translates WKT polygons into simplified and smoothed ornament silhouettes and records their source and transformation in provenance cards. Optional image experiments examine the reliability of a fallback route for cases without usable vectors: recovery from overlay-suppressed map plates and from unannotated aerial orthophotos. The interface comparison audits four designs for readability, specificity, disclosure, hierarchy, and overclaim risk. Figure 1 distinguishes the authoritative vector route from these image-recovery assessments.

The study makes three contributions. It provides a repeatable transformation from field polygons to families of ornament contours, with explicit geometric measurements and weight sensitivity. It examines image recovery as an optional source of geometry and measures the effect of contour translation on predicted aerial extents. It also compares four provenance-card designs through an explicit rule-based audit and analysis of their visual organization. Together, these contributions position field geometry as a source of form whose evidentiary limits remain visible.

LITERATURE REVIEW

Agricultural parcel delineation has moved from rule-based edge and region methods toward learned semantic and instance representations. Graesser and Ramankutty (2017) combined remote-sensing features to detect field parcels from Landsat imagery. Persello et al. (2019) used fully convolutional networks and combinatorial grouping to delineate smallholder fields whose boundaries were visually indistinct. Waldner and Diakogiannis (2020) framed field extraction as a multitask segmentation problem involving field extent, boundary probability, and distance to the nearest boundary. Related work combined learned boundary evidence with graph-based contour growth and examined fully convolutional architectures for agricultural boundary detection (Taravat et al., 2021; Wagner & Oppelt, 2020). Scalable systems have further emphasized consolidation of local boundary predictions and generalization across images, places, and seasons (Waldner et al., 2021).

Open, harmonized data are central to that progress. AI4Boundaries paired imagery and parcel labels across multiple European settings to support comparable model development (d'Andrimont et al., 2023). EuroCrops addressed a complementary problem by harmonizing crop and parcel information across the European Union (Schneider et al., 2023). Fields2Cover showed how vector field geometry can drive coverage-path planning for unmanned agricultural vehicles (Mier et al., 2023). These projects demonstrate that a field polygon is not only a segmentation target; it can also serve as a computational input for downstream operations. The present study extends that downstream role into computational form-making while preserving the difference between a reference polygon and an image-derived prediction.

Computational design provides the second foundation. Parametric, generative, and algorithmic design all convert explicit relationships into families of forms rather than treating a single drawing as the final object (Caetano et al., 2020). Jewelry has long been a useful domain for this approach because its contours, repetitions, symmetries, openings, and settings can be encoded as rules. Kielarova et al. (2013) applied a generative system to jewelry design, and later combined shape grammar with evolutionary design in an interactive workflow (Kielarova et al., 2015). Di Nicolantonio et al. (2020) connected generative design with printable mass-customized jewelry, while Jeong et al. (2021) demonstrated how parametric form and additive manufacturing

can support fashion products. This literature supports the use of field geometry as a controllable source shape rather than as a motif copied by hand.

Contour abstraction introduces a design trade-off. A field polygon may contain narrow notches, holes, or irregular edges that become difficult to reproduce at ornament scale. Line simplification reduces contour detail, while subdivision replaces sharp transitions with smoother curves (Chaikin, 1974; Douglas & Peucker, 1973). These operations change the relationship between the source and its derived silhouette. Geometric measurements describe those changes, but attractiveness, recognizability, and suitability for fabrication require separate evaluation.

Agricultural landscapes also carry social and cultural meanings that cannot be reduced to geometry. Photo-elicitation research shows how images of agricultural landscapes prompt accounts of farming, community, and place (Beilin, 2005). Landscape-character frameworks identify coherence, complexity, openness, naturalness, and other visual concepts that shape interpretation (Ode et al., 2008; Tveit et al., 2006). Ecological appearance and ecological function may align, but they can also diverge, making visual appeal an unreliable proxy for environmental performance (Gobster et al., 2007). Stephenson's (2008) cultural-values model similarly treats landscape meaning as an interaction among forms, practices, and relationships. Visual storytelling can bring such relationships into public discourse, but it also participates in constructing what a place is understood to mean (Spiegel, 2020). These insights support landscape storytelling while warning against turning shape alone into an environmental judgment.

Research on greenwashing sharpens that warning. Delmas and Burbano (2011) identify organizational and market drivers that encourage misleading environmental communication. Parguel et al. (2011) show how external ratings can influence the credibility of sustainability claims, and Atkinson and Rosenthal (2014) demonstrate the importance of label source and claim specificity for consumer trust. Lyon and Montgomery (2015) emphasize that greenwash includes both what organizations say and the strategic ends such communication serves. Later typologies distinguish unsupported claims, selective disclosure, vague language, and misleading imagery (de Freitas Netto et al., 2020; Seele & Gatti, 2017). In advertising experiments, affective environmental language can generate favorable responses even when the reasoning basis is weak (Schmuck et al., 2018). For agricultural jewelry interfaces, the practical implication is direct: a field image should not imply regenerative practice, local ownership, low impact, or agricultural material content unless those propositions are independently documented.

Human-AI interaction research offers tools for presenting these limits. Guidelines for human-AI interaction recommend making system capabilities and uncertainty visible and supporting appropriate reliance (Amershi et al., 2019; Yang et al., 2020). Model cards, datasheets, and data cards formalize related practices at the model and dataset levels (Gebru et al., 2021; Mitchell et al., 2019; Pushkarna et al., 2022). Interactive model cards and social-transparency research further show that documentation becomes more useful when people can connect an output to its source, process, and context (Crisan et al., 2022; Ehsan et al., 2021). Transparency is not automatically beneficial in unlimited amounts; its organization affects comprehension and trust (Kizilcec, 2016).

These concerns are particularly relevant to generative and language-assisted design. AI chains can separate a complex generation task into inspectable steps and preserve user control over intermediate outputs (Wu et al., 2022). Co-creation research likewise treats AI experience design as an iterative process of specifying, inspecting, and refining system behavior (Subramonyam et al., 2021). Designers need information not only about model accuracy but also

about provenance, failure modes, and the relationship between model outputs and intended experiences (Liao et al., 2023). The field-to-ornament framework follows this staged logic: images and polygons create evidence; geometry operations create a silhouette; structured evidence constrains copy; and the finished card is audited for unsupported claims.

The resulting research gap is specific. Field-boundary studies evaluate geometric recovery; computational jewelry studies generate form; landscape research explains place-based meaning; and responsible-AI and greenwashing research examine disclosure. Few studies connect all four into a single, measurable interface pipeline. The present study addresses that gap without assuming that agricultural appearance proves agricultural sustainability.

METHODS

Data sources and evidence scope

The empirical study used Fields2Benchmark version 1.0, released by Wageningen University & Research on December 19, 2024 (Mier et al., 2024). Files with the same stem pair a PNG map plate with a longitude–latitude WKT polygon. Inspection found 350 complete pairs: 140 from the Netherlands, 113 from Lithuania, and 97 from Estonia. Two polygons, `nl_field_22` and `nl_field_32`, contained self-intersecting zero-area linework. Validity repair retained the single positive-area polygon in each case, with no area loss; all 350 records were included. Table 1 summarizes the source inventory and evidence scope.

Table 1. Dataset inventory and evidence scope.

Item	Observed value
Paired PNG/WKT records	350
Country distribution	Netherlands 140; Lithuania 113; Estonia 97
Raster files	Annotated satellite-basemap PNG map plates
Raster dimensions	786–860 px wide; 592–606 px high
Map-panel dimensions	607–641 px wide; 524–533 px high
Vector files	Longitude–latitude WKT POLYGON geometries
Interior-hole distribution	0 holes: 301; 1: 40; 2: 7; 3: 1; 4: 1
Directly supported evidence	Field ID, country prefix, paired PNG, paired WKT
Derived geometric evidence	Area, perimeter, compactness, elongation, orientation, convexity, boundary complexity, holes, vertices
Unsupported by the files	Crop type, farm ownership, farming practice, material origin, environmental impact, capture date, satellite platform
<i>Note. PNG map plates include axes, scale information, attribution, and a colored field boundary. Interior-hole counts refer to the valid polygonal geometries; all source holes were retained.</i>	

The WKT polygon is the authoritative source for geometry descriptors and the main design examples. Its paired PNG supplies landscape context and an input to a map-recovery experiment. Recovery is unnecessary when a usable vector already exists; its practical role is to explore a fallback for imagery-only sources. Because the reference boundary is rendered over each map plate, the visible overlay was suppressed before prediction. A separate aerial assessment evaluates image inputs without this annotation and follows their predicted contours through abstraction.

The aerial assessment used 350 AI4Boundaries RGB orthophotos from data released in 2022 (d'Andrimont et al., 2023). Each 512×512 -pixel tile has 1 m resolution. The subset includes 50 tiles from each of Austria, Catalonia, France, Luxembourg, the Netherlands, Slovenia, and Sweden, retaining the original split membership: 30 training, 10 validation, and 10 test tiles per region, or 210/70/70 overall. Within each region and split, tile identifiers were ranked by SHA-256 of the fixed string '20240919:' followed by the identifier, and the quota was retained. The target is the union of labelled parcel extents in the first of four label bands. GSAA labels may omit agricultural parcels, so this target is not an exhaustive agricultural-land inventory. All seven regions occur in training and testing.

Map-plate recovery

Preprocessing detected the dark map frame and cropped its interior. WKT coordinates were transformed using the Web-Mercator latitude relation, then registered to the visible outer-overlay centerline by least-squares estimation of horizontal and vertical scale and translation. Inner rings used the same transform. Across 350 records, 99.969% of colored overlay pixels lay within five pixels of the rasterized WKT boundary on average; the median was 100.000%. Very thin branches can remain visible as strokes while disappearing at mask resolution. Detected plot colors were dilated by four pixels and filled from nearest valid neighboring pixels. Images and reference masks were resized to 256×208 pixels. Country-stratified random splitting used seed 20240919. The training set contained 245 fields, the validation set 52, and the held-out test set 53. Table 2 shows that all three countries were represented in every subset. The validation set selected thresholds for the learned pixel classifiers; the test set was reserved for final comparisons.

Table 2. Country-stratified train, validation, and test split

Country	Train	Validation	Test	Total
Netherlands	98	21	21	140
Lithuania	79	17	17	113
Estonia	68	14	15	97
All	245	52	53	350

Note. The split was generated with seed 20240919 and stratified by country

Source geometry

Each WKT polygon was transformed to a locally centered metric plane using the mean longitude and latitude of all ring vertices, excluding repeated closures, and an Earth radius of 6,371,008.8 m. Area excluded holes, whereas perimeter included all rings. Compactness was $4\pi A/P^2$. Elongation was the long-to-short edge ratio of the minimum-area rotated rectangle. Orientation was the dominant axis of the uniformly filled polygon's exact second moments, subtracting holes, expressed on a 0–180° east-referenced scale. This avoids weighting densely digitized boundary sections more strongly. Convexity was area divided by convex-hull area; boundary complexity was perimeter divided by convex-hull perimeter. Vertex count included all rings without repeated closures.

Table 3. WKT-derived field geometry by country.

Country	n	Median area (ha)	Median perimeter (m)	Median compactness	Median elongation	Median convexity	Median vertices
Netherlands	140	2.371	766.484	0.567	2.023	0.959	26.0
Lithuania	113	2.118	701.200	0.585	1.762	0.965	26.0
Estonia	97	1.925	696.028	0.518	1.683	0.907	74.0
All fields	350	2.134	756.738	0.553	1.847	0.954	33.5

Note. Boundary-complexity medians were 1.014 for the Netherlands, 1.029 for Lithuania, 1.083 for Estonia, and 1.031 overall. Measurements use a local planar approximation.

Table 3 summarizes the country distributions. The full collection ranged from 0.498 to 25.745 ha and from 282 to 2,996 m in perimeter. Compactness ranged from 0.079 to 0.834, elongation from 1.002 to 17.275, convexity from 0.544 to 1.000, and boundary complexity from 1.000 to 2.403. These ranges provide a substantive test of whether the translation pipeline can retain both compact and highly elongated source forms. Figure 2 shows the country composition, and Figure 3 shows the positively skewed area distribution on a logarithmic axis. The figures establish the sample balance and geometric scale before any predictive or design result is considered.

Image models and evaluation

Five methods were evaluated on overlay-suppressed map crops. The median rectangular prior used the median training-mask bounding box. Otsu thresholding retained the high excess-green class (Otsu, 1979). Per-image MiniBatch K-means used four clusters over RGB, excess green, and seven-pixel local grayscale standard deviation; the modal cluster in an 11×11 central

seed was retained. Pixel logistic regression and Random Forest used eleven features: RGB, grayscale, excess green, seven-pixel local mean and standard deviation, Sobel gradient magnitude, normalized x and y, and radius from the image center. Each training field contributed 600 positive and 600 negative pixels, totaling 294,000 samples. Logistic regression used standardized features and averaged stochastic gradient descent with log loss, alpha 0.0001, maximum 2,000 iterations, and tolerance 0.0001. Random Forest used 120 trees, depth 18, minimum leaf size 4, square-root feature sampling, and balanced subsamples (Breiman, 2001).

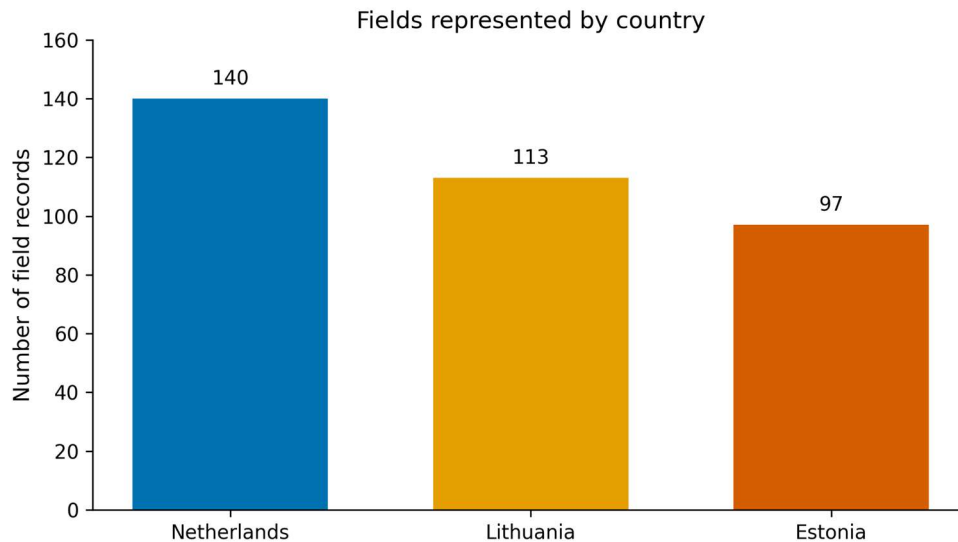


Figure 2. Distribution of the 350 paired field records by country.

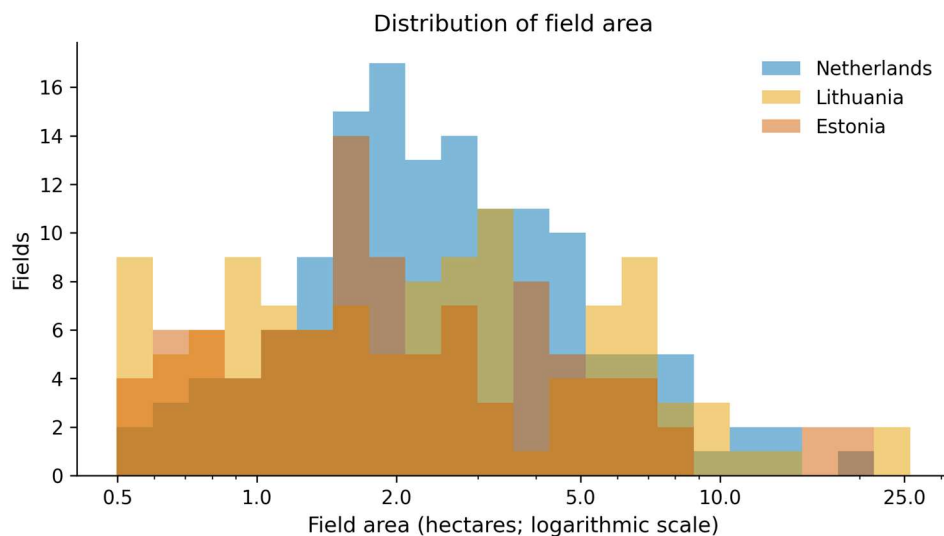


Figure 3. Field-area distribution for the Netherlands, Lithuania, and Estonia on a logarithmic hectare scale.

Probability thresholds from 0.30 to 0.70 in 0.05 increments were selected by mean validation IoU. Both learned classifiers selected 0.45, yielding validation IoU 0.5616 for logistic regression and 0.5731 for Random Forest. A 3×3 opening and 5×5 closing preceded central-component selection; the largest component was retained if none intersected the seed. The rectangular prior remained unchanged. Spatial features and component selection exploit the centered presentation, limiting transfer to arbitrary scenes. Table 4 records the configurations.

Predictions were evaluated using IoU, Dice, boundary F1 at three-pixel tolerance, symmetric Hausdorff distance, and absolute percentage area error. These measures capture region overlap, boundary alignment, worst-case boundary separation, and scale error (Taha & Hanbury, 2015). Means and percentile 95% intervals used 3,000 bootstrap resamples of the 53 test fields with seed 20240919.

Table 4. Field-mask recovery configurations.

Method	Input evidence	Learned component	Selection rule
Median rectangular prior	Training-mask extents	Median normalized rectangle	Fixed
Otsu excess-green	Excess-green image	None	Per-image Otsu threshold
K-means color-texture	RGB, excess green, local texture	Four color-texture clusters	Central-seed cluster
Pixel logistic regression	Eleven pixel features; 294,000 training samples	Averaged SGD logistic model	Validation threshold 0.45
Pixel Random Forest	Eleven pixel features; 294,000 training samples	120-tree balanced forest	Validation threshold 0.45
<i>Note. Every image method operated on the overlay-suppressed 256×208 map crop. The task is map-plate field-mask recovery, not raw-satellite parcel delineation</i>			

Aerial extent recovery

The aerial experiment compared a constant extent prior, Otsu excess-green, six-cluster K-means with training-label mapping, logistic regression, and Random Forest. Learned models used 252,000 pixels sampled uniformly without replacement from the 210 training tiles. Fourteen RGB, color-ratio, grayscale, excess-green, local mean, local standard-deviation, and gradient features excluded image coordinates. Logistic regression used standardized inputs and L2 regularization ($C = 1$). The forest used 120 trees, depth 18, minimum leaf size 8, and square-root feature sampling. Mean validation IoU selected the constant prior, Otsu polarity, and probability thresholds from 0.10 to 0.90 in steps of 0.05. Selected thresholds were 0.35 for K-means mapping, 0.30 for logistic regression, and 0.75 for Random Forest. A 3×3 opening followed by closing retained all components.

All 70 aerial test tiles were retained, including 19 without labelled extent. IoU and Dice were one for two empty masks and zero when only one was empty. Extent-boundary F1 used a Euclidean tolerance of three native pixels (3 m), excluding image-frame edges. It measures union borders, not internal boundaries between adjacent labelled parcels. Relative area error used the 51 positive-label tiles; false-positive coverage was summarized separately for empty-label tiles. Mean intervals used 3,000 bootstrap resamples within the seven regions. The test therefore assesses variation within this subset, rather than unseen-region transfer.

Logistic regression had the highest validation IoU and was fixed for downstream contour analysis. Its exterior and hole contours were simplified at 1% of each contour's bounding-box diagonal and smoothed once with Chaikin subdivision. Reference extents underwent the same operations separately. All polygonal components remained in the original coordinate frame. The resulting overlap measures spatial distortion of contour unions, without matching parcel identities or evaluating jewelry form.

Contour translation and geometric fidelity

The vector experiment used all 350 polygons in a deterministic transformation. Each polygon was centered at its area centroid, rotated to place its dominant axis horizontally, and scaled to a unit bounding-box diagonal. Outputs remained in that source frame. Topology-preserving Douglas–Peucker simplification operated on the complete polygon, including holes (Douglas & Peucker, 1973). Closed-ring Chaikin subdivision replaced each edge with points at one-quarter and three-quarters of its length (Chaikin, 1974). Five settings were defined: Trace (tolerance 0, no smoothing), Subtle (0.005, one iteration), Balanced (0.010, one iteration),

Expressive (0.020, two iterations), and Abstract (0.040, three iterations). An invalid smoothing iteration was stopped, retaining the preceding valid polygon and recording the applied count.

Fidelity was evaluated before adding jewelry fittings or other hardware. Source and translated silhouettes used a shared 192×192 pixel-center grid spanning 110% of the source bounding box's longer side. Measurements included IoU, Dice, area, elongation, compactness, orientation, and boundary distance. Approximately 2,048 arclength samples were distributed over all rings, with at least 32 per ring and ring-length weights. Normalized symmetric Chamfer distance C averaged directed sample-to-boundary distances and divided by the source diagonal; sampled Hausdorff distance retained the maximum. Output vertex counts include Chaikin sampling points, while simplified counts describe structural reduction before smoothing.

GPS combines five bounded similarities. For area, elongation, and compactness, similarity is $\min(\text{source}, \text{output})/\max(\text{source}, \text{output})$. Boundary similarity is $\exp(-kC)$, with $k = 10$ by default. Orientation similarity is $1 - \Delta\theta/90$, where $\Delta\theta$ is the smaller undirected axis difference in degrees. $\text{GPS} = 100\sum w_j s_j$, with nonnegative weights summing to one. Equal weights of 0.20 provide a transparent reference convention in the absence of empirical preference weights. The components describe complementary geometric properties, although they are correlated and are not assumed to be perceptually equivalent. GPS is reported alongside overlap and distance measures.

Sensitivity analysis used equal weights, five focused settings (0.60 for one component and 0.10 for each other), five leave-one-out settings, and five pure-component extremes. These 16 settings and 5,000 Dirichlet(1,1,1,1,1) weight draws were evaluated at $k = 5, 10$, and 20, giving 48 named and 15,000 sampled combinations. Rankings compared mean scores of the four non-Trace levels; the identity Trace score is necessarily maximal. Additional checks omitted orientation or excluded fields with an axis eigenvalue ratio below 1.10 at any level or interrupted smoothing. Paired field-level bootstrap resampling (3,000 draws; seed 20240919) estimated intervals for mean GPS and level differences. These analyses assess numerical sensitivity, not aesthetic preference.

Provenance-card design and audit

Thirty fields were selected by ranking polygon area within each country, breaking ties by field identifier, and retaining ten ranks from $\text{round}(\text{linspace}(0, n - 1, 10))$. Four templates per field produced 120 cards. A presented a minimal environmental claim without traceable evidence. B provided a compact data label with identifiers, paired filenames, geometry, and two scope limits. C used a narrative-led agricultural landscape story with limited evidence. D separated source evidence, shape translation, and five scope exclusions. The unit of analysis is a field-populated template, not an independent human judgment.

The templates were populated directly from the field records. Their wording, amount of evidence, disclosure, and layout varied together, so the primary comparison concerns complete card designs. It does not isolate a causal effect of information architecture. The cards used fixed copy; generative language-model outputs were not evaluated. Full text and feature flags were retained for the audit. D identified the relationship as design provenance and explicitly excluded crop type, farm ownership, farming practice, material origin, and environmental impact.

The audit measured word count, Flesch Reading Ease (FRE), Flesch–Kincaid grade, hierarchy, specificity, disclosure, and overclaim risk. Specificity gave equal weight to six evidence slots: field ID; country; paired PNG; paired WKT; area and perimeter; and a shape descriptor. Disclosure awarded 20 points per explicit scope exclusion. Hierarchy awarded 10 for a headline, 20 for evidence labels, 20 for a key-value block, 10 for a design-use label, 20 for a

scope label, and 20 for a visually segmented evidence–design–limits sequence. These are author-defined structural criteria.

Readability included the title and body rows. Alphabetic words and numbers were tokens; underscores and slashes split filenames. Each key-value row counted as at least one sentence or fragment. Syllables were estimated from vowel groups, with a terminal silent-e adjustment and a minimum of one per token. With W words, N sentences/fragments, and Y estimated syllables, $FRE = 206.835 - 1.015(W/N) - 84.6(Y/W)$, and $grade = 0.39(W/N) + 11.8(Y/W) - 15.59$. These text indices are especially sensitive to short headings and filenames. Overclaim risk was calculated as follows:

$$Risk = clip(60 - 0.20S - 0.20D + 5V + 10U + 20A, 0, 100).$$

S denotes specificity, D disclosure, V vague environmental-language categories, U unsupported claim categories, and A absolute overclaims. The baseline of 60 is a conventional starting score for copy without evidence or exclusions. Complete specificity and disclosure each offset 20 points. Unsupported categories receive twice the penalty for vague language, and absolute overclaims four times that penalty. This ordering encodes intended severity rather than measured severity ratios. Scores are neither probabilities of deception nor participant measures of comprehension or trust. The predefined template flags make this an audit of the evidence schema rather than a classifier for arbitrary prose.

The sensitivity grid varied the baseline over $\{40, 60, 80\}$, specificity and disclosure weights independently over $\{0.10, 0.20, 0.30\}$, vague-language penalties over $\{2.5, 5, 7.5\}$, and unsupported-claim penalties over $\{5, 10, 15\}$, totaling 243 combinations. Clipped and unclipped scores, ranges, ties, and ordering were recorded. No card contained an absolute overclaim, so its penalty could not be evaluated. Two controlled diagnostics retained D 's full text in segmented and linear layouts, then held the segmented layout and evidence fixed while including five, two, or zero exclusions. The accompanying research archive contains the contour operations, scoring rules, model configurations, split manifests, per-field and per-tile results, full card texts, and figure-generation code. Feature definitions, validity checks, random seeds, and software versions are recorded with each analysis.

RESULTS

The file audit confirmed one-to-one correspondence for all 350 field identifiers. Figure 4 compares representative overlay-suppressed map crops with their WKT-derived masks. The examples make the distinction between modalities visible: the top row supplies agricultural image context after the rendered contour has been removed, while the bottom row supplies the reference geometry used for evaluation and design translation.

Table 5. Held-out field-mask recovery after boundary-overlay suppression.

Method	Mean IoU	Mean Dice	Boundary F1 (3 px)	Hausdorff (px)	Area error (%)
Median rectangular prior	0.4256	0.5834	0.0794	88.63	152.90
Otsu excess-green	0.2987	0.4055	0.2302	110.20	102.17
K-means color-texture	0.4208	0.5494	0.2410	86.57	72.90
Pixel logistic regression	0.5228	0.6724	0.1407	58.02	60.19
Pixel Random Forest	0.5417	0.6876	0.2119	59.19	53.56
<i>Note. Values are means across 53 country-stratified test fields. Figure 6 displays 95% bootstrap confidence intervals for IoU, Dice, and boundary F1. Higher is better for the first three metrics; lower is better for Hausdorff distance and area error.</i>					

The geometry distribution was broad enough to support visibly different ornament silhouettes. Across all fields, the median area was 2.134 ha and the mean was 3.272 ha, indicating a right-skewed distribution driven by a smaller number of large fields. Median perimeter was

756.738 m. Median compactness was 0.553, median elongation was 1.847, median convexity was 0.954, and median boundary complexity was 1.031.

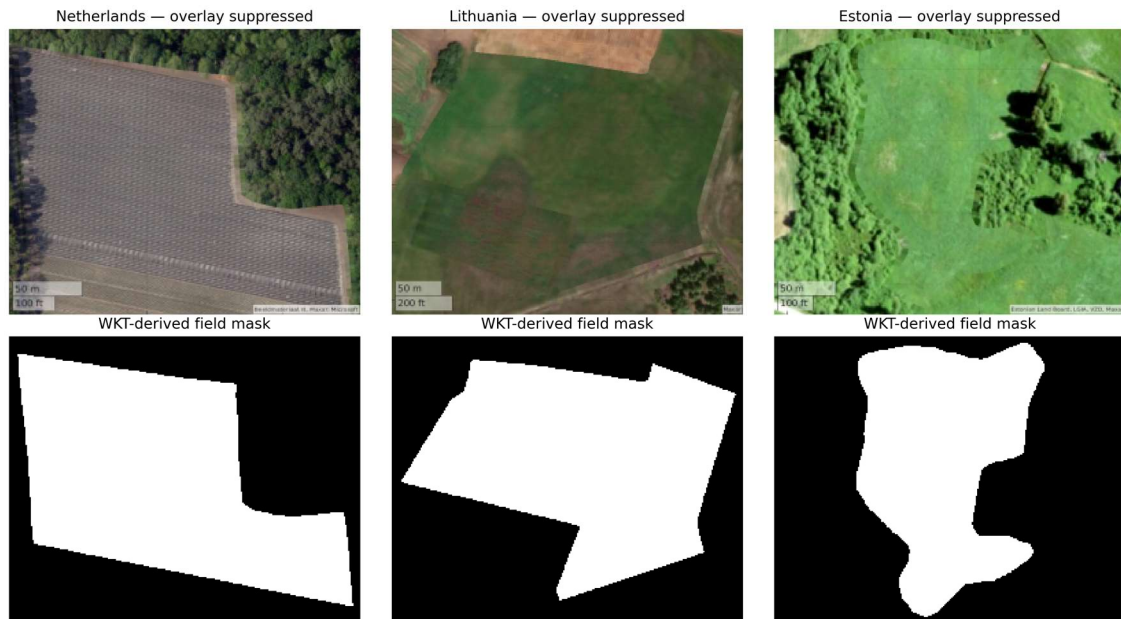


Figure 4. Representative overlay-suppressed agricultural map crops and aligned WKT-derived field masks from the Netherlands, Lithuania, and Estonia.

The Netherlands had the highest median elongation at 2.023, whereas Lithuania had the highest median compactness at 0.585. Estonia had a substantially higher median vertex count (74) than the Netherlands and Lithuania (26 each), and its median boundary complexity was also highest. Figure 5 plots compactness against elongation and shows that the design space includes compact, near-equiaxed forms as well as long and irregular parcels.

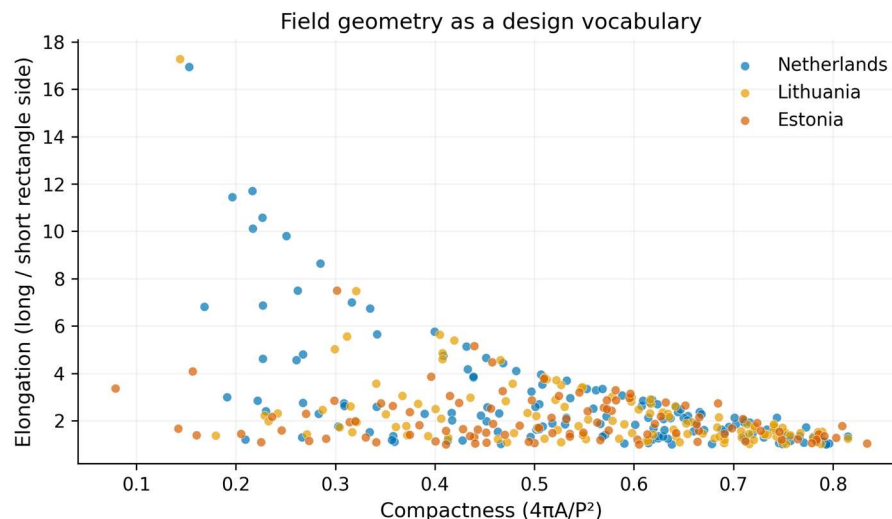


Figure 5. Compactness and elongation of all 350 WKT field polygons, colored by country.

Table 5 reports the held-out map-recovery results. Random Forest achieved the greatest mean overlap (IoU 0.5417; Dice 0.6876) and the smallest mean absolute area error (53.56%). K-means achieved the highest boundary F1 (0.2410), while logistic regression had the smallest Hausdorff distance (58.02 pixels). Random Forest's mean IoU had a 95% field-bootstrap interval

of 0.5000–0.5841. Modest boundary agreement and large area errors limit these contours to exploratory candidates requiring review. The results concern overlay-suppressed annotated map plates; filling artifacts and centered framing constrain their interpretation. Figure 6 shows that region overlap and boundary agreement rank methods differently. A method can recover much of an interior while shifting its boundary or losing narrow notches. Reporting both therefore makes the limitations of an image-derived contour more explicit than a single overlap score.

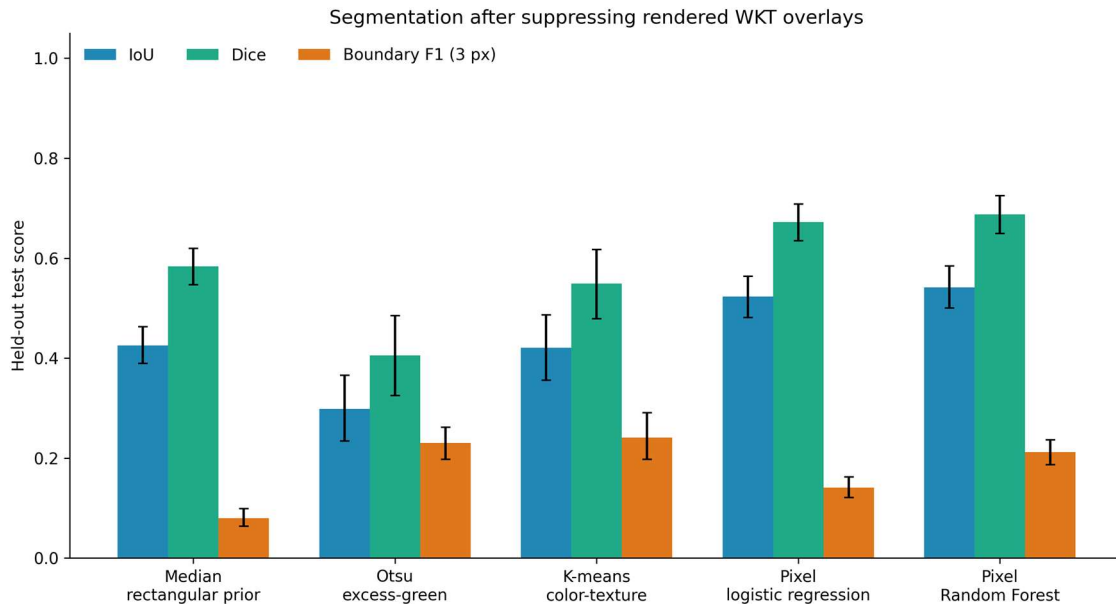


Figure 6. Test-set IoU, Dice, and three-pixel boundary F1 for five field-mask recovery methods after rendered-boundary suppression; error bars show 95% bootstrap confidence intervals.

The contour experiment exposed a controlled trade-off between abstraction and source fidelity. Trace preserved the source and scored 100.00. Subtle and Balanced reduced contour complexity while retaining mean GPS 95.06 and 93.46, respectively. Expressive and Abstract scored 90.35 and 88.06. Table 6 reports the comparison. Output vertex count rises at the most abstract levels because Chaikin subdivision adds curve samples; mean simplified counts continued to decline from 15.54 for Subtle to 6.95 for Abstract. All 49 fields with holes retained them. The validity safeguard stopped smoothing in 20 field-level cases across seven fields, preserving the last valid contour.

Table 6. Geometry preservation across five contour-abstraction levels.

Level	Tolerance / diagonal	Requested iterations	Mean output vertices	Silhouette IoU	Normalized Chamfer	Axis change (°)	Mean GPS
Trace	0.000	0	52.61	1.0000	0.0000	0.00	100.00
Subtle	0.005	1	30.79	0.9422	0.0066	1.51	95.06
Balanced	0.010	1	22.07	0.9171	0.0095	1.47	93.46
Expressive	0.020	2	34.11	0.8740	0.0146	1.70	90.35
Abstract	0.040	3	54.53	0.8344	0.0193	2.15	88.06

Note. Each level includes all 350 fields. GPS uses equal weights and $k = 10$. Counts are requested smoothing iterations; the preceding valid shape was retained when the safeguard intervened. Balanced mean GPS has a 95% paired field-bootstrap interval of 93.17–93.74.

Table 7 summarizes GPS sensitivity. All 15,000 sampled combinations and all 33 equal, focused, and leave-one-out combinations retained Subtle > Balanced > Expressive > Abstract. Three orientation-only extremes instead placed Balanced slightly above Subtle. The paired mean difference between Subtle and Balanced was 1.60 points (95% interval 1.41–1.78). Omitting

orientation or excluding the seven fields with a near-isotropic axis at any level, the seven with interrupted smoothing, or both groups retained the default order. This indicates stability over the tested conventions without establishing universal weight invariance or perceptual validity.

Table 7. Sensitivity of the non-Trace GPS ranking.

Weight family	Combinations across k = 5, 10, 20	Default order retained	Observed departure
Equal, focused, leave-one-out	33	33/33	None
Pure-component extremes	15	12/15	Orientation alone favors Balanced over Subtle
Dirichlet weight draws	15,000	15,000/15,000	None
<i>Note. The default order is Subtle > Balanced > Expressive > Abstract. The same 5,000 Dirichlet draws were reused at each boundary-distance scale; these are parameter scenarios, not independent empirical observations.</i>			

Balanced was chosen as the middle parameter setting for the principal demonstration, not as the geometric or aesthetic optimum. Figure 7 follows nl_field_702198 from its WKT source through normalization and abstraction to pendant, mirrored-earring, and medallion concepts. The sample retained GPS 91.98 and silhouette IoU 0.8884. These measurements concern the shared contour before decorative fittings; the applications remain two-dimensional concepts without fabrication or wearer evaluation.

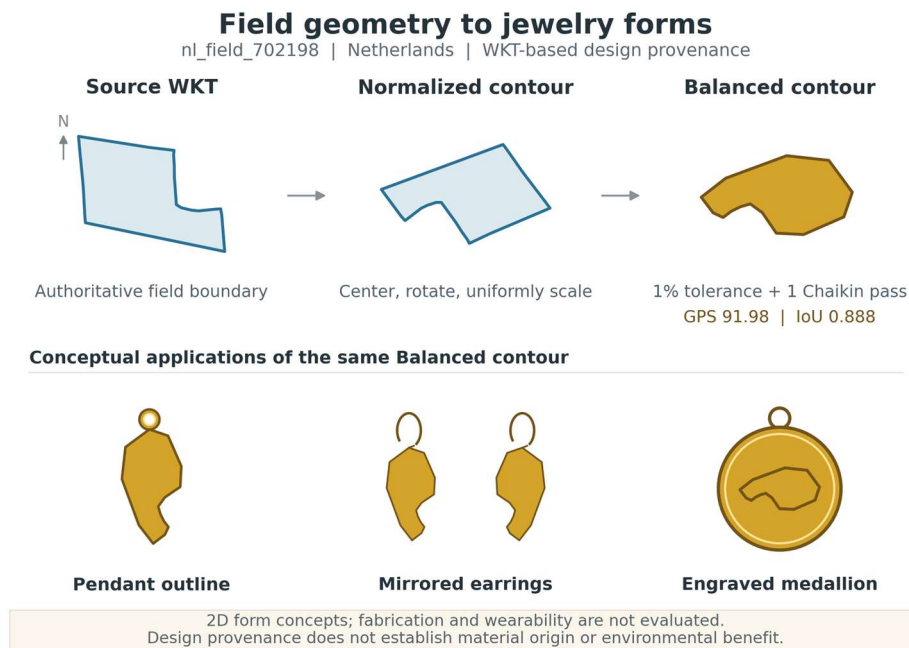


Figure 7. WKT-derived translation of nl_field_702198 through normalization and Balanced abstraction to pendant, mirrored-earring, and engraved-medallion concepts. Blue denotes the source contour and gold the derived form; hardware and material appearance are schematic.

Table 8 summarizes the four provenance-card conditions. The source record is constant within each four-card set, while evidence coverage, wording, disclosure, and visual organization vary. The audit compares these combined design choices. The audit separated the templates according to their specified content (Table 9). A combined four vague environmental-language categories and two unsupported associations, scoring 100.00 risk with no evidence or disclosure. B contained all six evidence slots and two exclusions, yielding hierarchy 80.00, specificity 100.00, disclosure 40.00, and risk 32.00. C retained country and qualitative shape evidence but no exclusions, giving specificity 33.33 and risk 58.33. D scored 100.00 for hierarchy, specificity,

and disclosure and 20.00 risk. These structural scores were constant across each template's 30 field substitutions.

Table 8. Agricultural jewelry provenance-card conditions.

Condition	Design intent	Evidence shown	Disclosure strategy	Intended audit profile
A Minimal claim	Emotional environmental claim	None	None	Highest risk
B Data label	Compact traceable record	Field ID, country, image/WKT files, geometry	Excludes crop type and material origin	High specificity; partial disclosure
C Story led	Narrative-first landscape story	Country and qualitative shape	None	Moderate risk
D Balanced disclosure	Evidence, transformation, and limits	All six evidence categories	States all five scope exclusions	Lowest risk

Note. Each condition was generated for 30 fields, producing 120 cards.

B had the highest FRE (60.89), followed by C (45.31), D (34.99), and A (18.94). D contained 91 words, compared with 57 for B. Its additional scope information therefore increases textual demand under this implementation. The indices do not indicate whether a reader notices the limits or understands the difference between design and material provenance. Across 243 coefficient settings, $D \leq B < C < A$ was retained throughout. Strict $D < B < C < A$ held in 234 settings; nine produced $D = B = 0$ because of clipping. Any clipping occurred in 126 settings. Risk ranges were 60–100 for A, 0–66 for B, 32.50–84.17 for C, and 0–60 for D. The stable ordering partly follows from the rubric's construction: D adds favorable disclosure slots without additional penalties. The baseline shifts unclipped scores equally, while clipping can create ties. Absolute-overclaim severity remains untested because its count was zero.

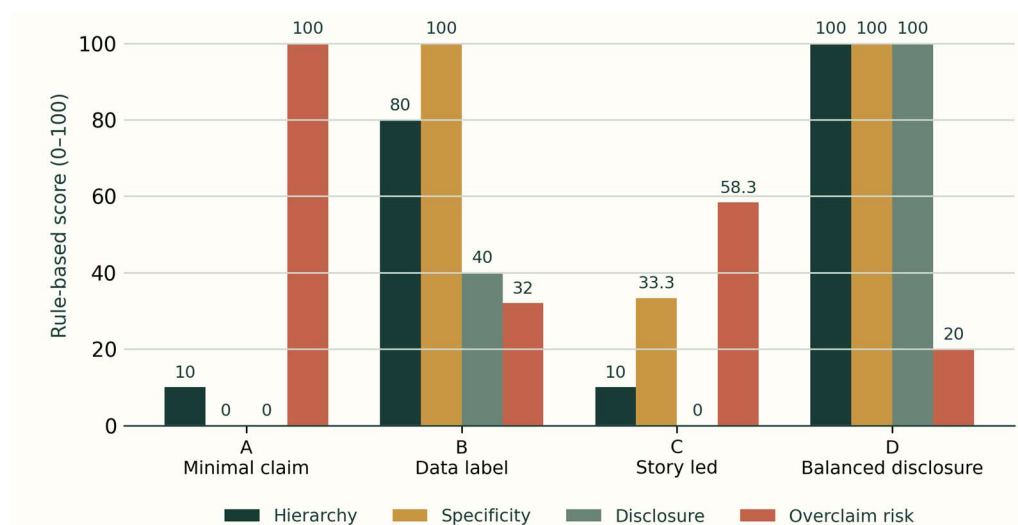


Figure 8. Deterministic interface-audit scores for four agricultural jewelry provenance-card conditions.

The controlled diagnostics further separated content from arrangement. With identical full text, segmented and linear layouts both scored 20 risk, although hierarchy changed from 100 to 80. With the segmented layout fixed, five, two, and zero exclusions produced risks of 20, 32, and 40. Thus, text-based risk responds to disclosure by construction and does not capture whether visual arrangement changes interpretation. Table 10 states the audit rules. They assess correspondence between the authored card schema and available evidence. Neither the penalty scale nor its sensitivity analysis measures actual environmental benefit, consumer trust, or the probability of an unsupported inference.

Table 9. Deterministic interface-audit results for 30 fields per condition.

Condition	n	FRE	FK grade	Words	Hierarchy	Specificity	Disclosure	Risk
A	30	18.94	12.61	16.00	10.00	0.00	0.00	100.00
B	30	60.89	6.34	57.00	80.00	100.00	40.00	32.00
C	30	45.31	9.59	32.00	10.00	33.33	0.00	58.33
D	30	34.99	9.71	91.00	100.00	100.00	100.00	20.00

Note. Values are template means. FRE and FK grade are text diagnostics; higher FRE indicates easier wording under the specified formula. Hierarchy, specificity, disclosure, and risk are author-defined 0–100 scores. Lower risk means fewer rubric penalties, not a measured reduction in misunderstanding.

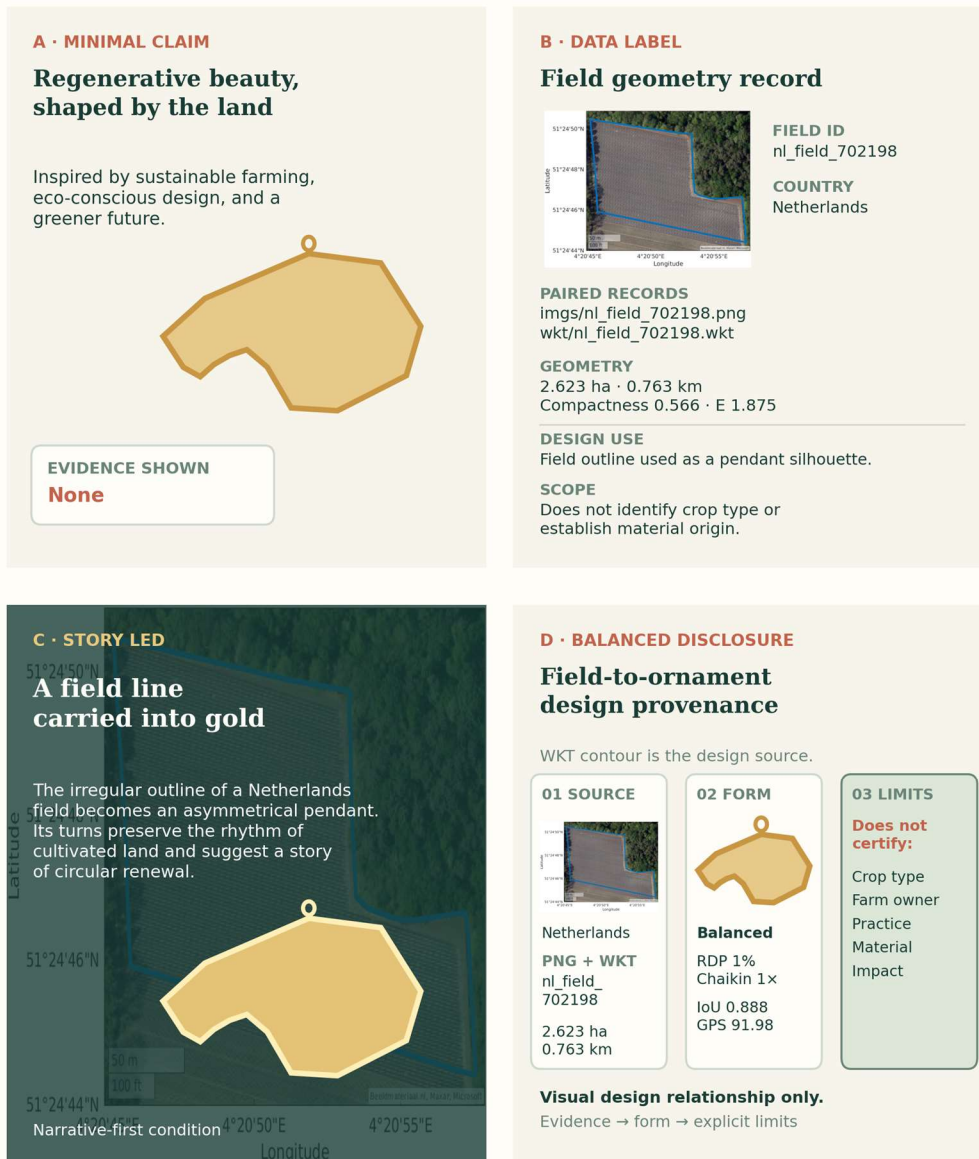


Figure 9. Abridged provenance-card designs for nl_field_702198: A, unsupported minimal claim; B, compact data label; C, narrative-led story; D, source, translation, and disclosure. Table 11 gives the detailed record. Environmental wording in A is an experimental condition, not a verified claim about the field.

Figure 8 compares the deterministic audit scores. B combines complete evidence slots with a compact text, whereas D includes all five scope exclusions. Figure 9 shows abridged views of the four designs for the same field; the full texts form the basis of the readability audit. Their layout differences provide hypotheses about locating evidence and reading disclosures that require human evaluation. Table 11 gives the detailed provenance record for nl_field_702198. Geometry and fidelity values derive from its WKT polygon. The map supplies visual context,

while the source vector establishes the relationship to the pendant, earring, and medallion contours.

Table 10. Evidence and overclaim audit rubric.

Criterion	Operational rule	Risk direction
Evidence specificity	Six equally weighted slots: field ID, country, PNG, WKT, area/perimeter, and shape	More evidence lowers risk
Disclosure completeness	Twenty points for each explicit limit: crop, owner, practice, material, and impact	More disclosure lowers risk
Vague environmental language	Counts categories such as green future, regenerative, responsible land, and sustainable farming without evidence	More vague categories raise risk
Unsupported capability claim	Counts claims about verified crops, farms, practices, material sourcing, or environmental effects	More unsupported categories raise risk
Absolute overclaim	Counts categorical certification or impact statements not established by the files	Each absolute claim raises risk substantially
Information hierarchy	Headline 10; evidence labels 20; key-value block 20; design-use label 10; scope label 20; visually segmented evidence–design–limits sequence 20	Separate structural score; absent from risk formula
<i>Note. Risk = clip(60 - 0.20 × specificity - 0.20 × disclosure + 5 × vague categories + 10 × unsupported categories + 20 × absolute overclaims, 0, 100)</i>		

Aerial recovery and contour translation

Table 12 reports the complementary aerial experiment. Random Forest achieved mean extent IoU 0.4427 (95% interval 0.3688–0.5186), Dice 0.5282, and extent-boundary F1 0.2014. Logistic regression achieved IoU 0.4215 (0.3400–0.5012). Their paired IoU difference was 0.0212 (–0.0574 to 0.1043), providing no clear separation. On the 51 tiles with positive labels, the ranking reversed: logistic regression achieved IoU 0.5589 and Random Forest 0.4507. Random Forest correctly predicted no extent in eight of the 19 empty-label tiles, compared with one for logistic regression. Empty-label coverage therefore materially affects the aggregate overlap comparison.

Table 11. Design-provenance record for nl_field_702198.

Field	Value
Field ID / country	nl_field_702198 / Netherlands
Source records	imgs/nl_field_702198.png; wkt/nl_field_702198.wkt
Area / perimeter	2.623007 ha / 0.762960 km
Shape descriptors	Compactness 0.566247; elongation 1.875186; orientation 148.120034°; 22 vertices
Source basis	Authoritative WKT contour; annotated map supplies landscape context
Shape translation	Balanced: tolerance 1% of the normalized diagonal; one Chaikin iteration
Sample shape fidelity	Silhouette IoU 0.8884; GPS 91.98
Jewelry design use	Two-dimensional pendant silhouette, mirrored earring pair, or engraved medallion motif
Disclosure	This record documents a visual design relationship. It does not certify crop type, farm ownership, farming practice, material origin, or environmental impact.
<i>Note. Measurements refer to the named field and its shared translated contour. Schematic hardware and material appearance are excluded from the fidelity assessment</i>	

For the validation-selected logistic model, contour translation changed mean IoU from 0.4215 to 0.4086, a paired decrease of 0.0129 (95% interval –0.0172 to –0.0091). Translating reference extents retained IoU 0.9486 across all tiles and 0.9294 across positive-label tiles. Figure 10 shows the metric comparison and the lowest-, median-, and highest-IoU positive-label examples for this model. Smoothing preserved much of an available reference shape but did not correct image-recovery errors. Union overlap also leaves individual parcel identity unresolved.

DISCUSSION

The principal design contribution is an inspectable relationship between a field boundary and an ornament contour. The WKT polygon determines the form, transformation parameters describe its abstraction, and the card identifies the supporting record. Image recovery, geometric

measurement, and interface auditing serve distinct purposes within that relationship. Together they support a research prototype for computational form-making and provenance presentation.

Table 12. Labelled-extent recovery on AI4Boundaries aerial test tiles.

Method	IoU (all 70)	IoU (positive 51)	Dice	Extent- boundary F1	Area error (%)	Empty-tile FP (%)
Constant extent prior	0.3501	0.4805	0.4265	0.0143	5670.90	100.00
Otsu excess-green	0.2116	0.2904	0.3043	0.1410	2283.14	45.06
K-means label mapping	0.3605	0.4947	0.4348	0.1487	761.69	25.96
Pixel logistic regression	0.4215	0.5589	0.4949	0.1521	118.53	8.66
Pixel Random Forest	0.4427	0.4507	0.5282	0.2014	61.96	2.98

Note. IoU, Dice, and extent-boundary F1 use all 70 test tiles. Area error uses 51 positive-label tiles; false-positive coverage uses 19 empty-label tiles. The smallest positive reference occupies 103 pixels, making percentage area error sensitive to small denominators. These are extent-union labels, which may omit agricultural parcels.

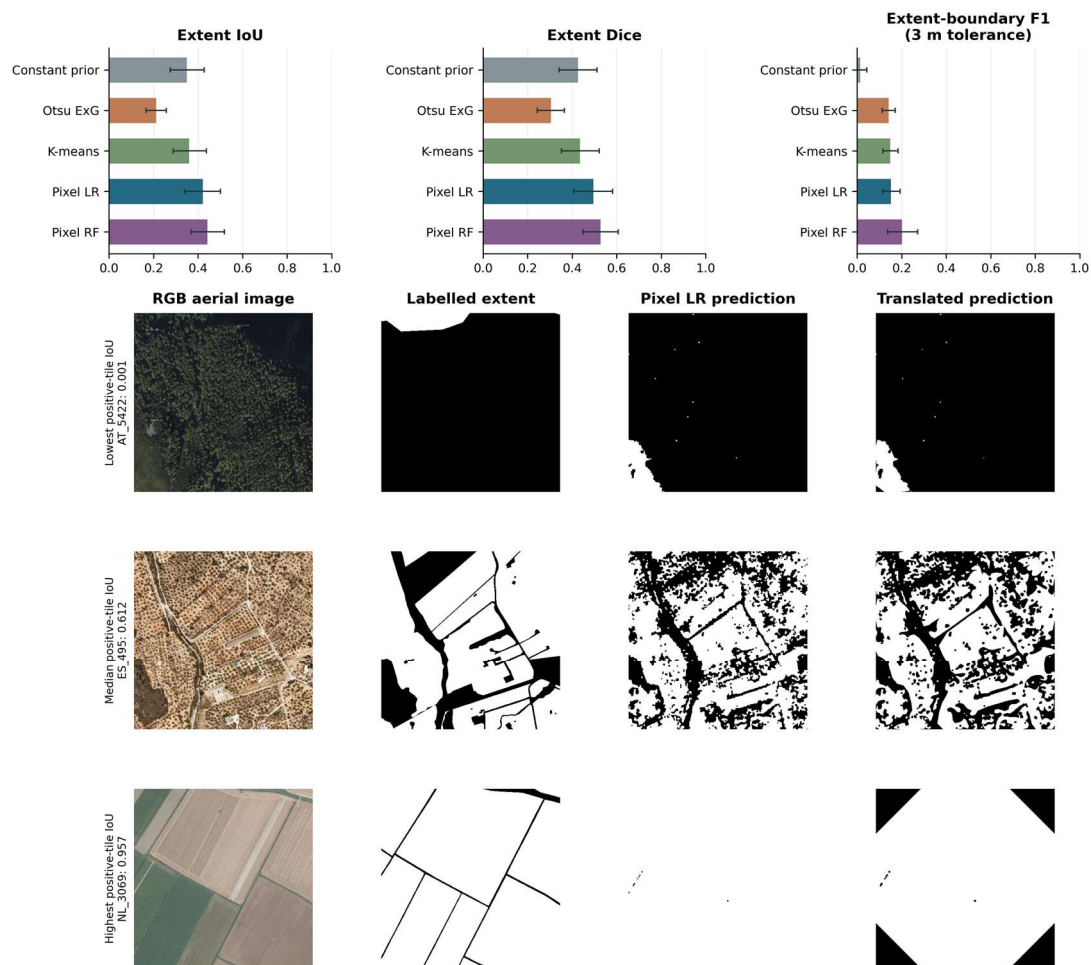


Figure 10. Aerial labelled-extent recovery and contour translation. Top: test-tile means with 95% region-stratified bootstrap intervals. Bottom: lowest-, median-, and highest-IoU positive-label examples for the validation-selected logistic model, with reference extent, prediction, and translated prediction. The masks represent spatial unions, not individual parcel identities.

The map experiment remains constrained by its annotated inputs. Suppression removes the visible reference stroke but can leave filling artifacts, and registration relies on the known source geometry. Its modest boundary agreement and large area errors limit practical contour use. The aerial assessment removes the rendered-overlay issue, but introduces different resolution, label coverage, and multi-parcel composition. Scores from these two tasks are therefore not directly

interchangeable. In particular, aerial extent-union overlap does not establish individual parcel delineation.

The vector reference remains the source of the main design examples. A predicted mask adds no provenance evidence when that polygon is already available. Image recovery has a practical role when a designer has imagery but no suitable vector boundary; the aerial contour experiment makes the limitation of that route concrete. Translation retained reference contours far more closely than predicted extents, and smoothing did not repair prediction error. An imagery-only design would therefore need to identify its source as a model-derived contour and undergo boundary and identity checking before making a field-specific provenance claim.

The abstraction levels offer a controlled way to explore form. Expressing tolerance relative to field size makes the operation comparable across sources. GPS describes retained geometric properties, with sensitivity analysis revealing both broad rank stability and the orientation-only exception. Balanced remains a heuristic demonstration setting; Subtle has higher mean geometric fidelity, and Trace is an identity. Selecting among them for visual character, recognizability, or wearability requires designer and user judgment beyond the metric. The workflow relates to parametric jewelry research, in which one rule system generates a family of forms (Di Nicolantonio et al., 2020; Kielarova et al., 2015). Its input is documented landscape geometry. The pendant, mirrored earrings, and engraved medallion illustrate possible two-dimensional uses of that source. Material thickness, added fittings, structural balance, fastening, and contact with the body remain to be evaluated before fabrication.

Figure 9 exposes four approaches to presenting that relationship. In A, a large environmental headline, negative space, and gold ornament silhouette dominate; no field record can be inspected. B places a map beside a compact key-value panel, giving identifiers and geometry a clear location, while its smaller scope statement follows the design-use row. C integrates the map as a narrative background and foregrounds the ornament, leaving source identifiers and evidentiary limits less prominent. D separates source, translation, and exclusions into three labelled regions; its distinct disclosure panel sits alongside the form. These choices concern the visibility and specificity of evidence, issues also examined in environmental-label and model-documentation research (Atkinson & Rosenthal, 2014; Crisan et al., 2022; Parguel et al., 2011).

The compositions suggest different scanning paths. A and C lead with association and narrative, B with a compact record, and D with a sequence from source to form to limits. D makes more information available but may increase reading effort and cognitive load through denser text and additional labels. These are design hypotheses, not observed viewing behavior. The fixed-content diagnostic confirms that the text-based risk score cannot detect an effect of arrangement. Human-AI guidance and transparency research provide a rationale for testing evidence location, comprehension, and appropriate reliance across these layouts (Amershi et al., 2019; Kizilcec, 2016; Liao et al., 2023).

The structured record could also support language-assisted copy development (Li, 2024; Nie & Zheng, 2023). A model might propose alternative wording while source identifiers, derived geometry, and disclosure scope remain explicit inputs. The present cards use fixed templates, and the audit covers only predefined categories. Evaluation of generated copy would require varied outputs and independent assessment of unsupported implications. Prompt chaining and co-creation research offer useful approaches to making such intermediate decisions inspectable (Subramonyam et al., 2021; Wu et al., 2022). For graphic design, the central distinction is between

documenting a source of form and documenting the origin or effects of a product's materials. The field record supports the former. A provenance interface can acknowledge the source, quantify its transformation, and explain that scope without implying sustainable production. Whether readers understand and retain this distinction remains an empirical question.

Limitations

Several limitations delimit the findings. Fields2Benchmark contains a selected European collection assembled for coverage-path research, and its annotated maps cannot be restored to exact original imagery. The aerial subset is small and region-balanced; original split membership was preserved, but spatial independence was not newly established, and unseen-region transfer was not assessed. Its labels may omit agricultural parcels. Local planar geometry is not a cadastral survey, and geometric fidelity does not measure aesthetic value. The ornament concepts have no physical validation of wall thickness, strength, balance, fastening, comfort, or casting tolerance. Finally, the card conditions vary wording, evidence, and layout together. Their deterministic scores and repeated field substitutions provide no participant evidence of comprehension, trust, authenticity, perceived sustainability, or purchase intent.

Future validation should involve farmers, jewelry designers, and consumers. Farmers could assess the meaning and appropriateness of source descriptions; designers could compare abstraction levels for recognizable form, aesthetic preference, and suitability for further development. Consumer tasks should measure comprehension, evidence-location accuracy, perceived authenticity, trust, perceived sustainability, and confusion between design, material, and environmental provenance. Randomized presentation of matched-content layouts would help separate wording and disclosure effects from visual organization. Fabricated prototypes require a separate structural and ergonomic assessment.

CONCLUSION

This study connects field geometry, contour abstraction, and provenance-card design. Its 350 WKT records support a direct, traceable route from agricultural boundary to ornament concept. Optional recovery from overlay-suppressed map plates achieved Random Forest IoU 0.5417, with boundary F1 0.2119 and area error 53.56%; these results remain limited to annotated-map recovery. The aerial experiment likewise showed modest labelled-extent recovery, and subsequent smoothing did not repair prediction errors.

The five-level vector pipeline made geometric changes measurable. Balanced retained mean GPS 93.46 and its position below Subtle persisted across the sampled weight scenarios, with an orientation-only exception among the extreme settings. It remains a heuristic demonstration choice, without evidence of aesthetic preference or manufacturing readiness. Across 120 cards, the proposed rubric distinguished evidence and disclosure, assigning risk 100.00 to the minimal claim and 20.00 to balanced disclosure. Sensitivity and controlled diagnostics explain that scoring behavior without establishing improved human understanding or trust. The contribution is an inspectable design-provenance record: a documented field can supply a source of jewelry form, while claims about crops, ownership, farming practice, material origin, or environmental benefit require additional evidence.

Data Availability Statement

Fields2Benchmark version 1.0 is available through Zenodo (Mier et al., 2024; <https://doi.org/10.5281/zenodo.14524735>). The December 19, 2024 record provides the paired imgs and wkt archives. Reuse follows the source-country terms listed with that record: public-

domain identification for the Netherlands, Creative Commons Attribution–ShareAlike 3.0 Estonia for Estonia, and a non-commercial limitation for Lithuania. AI4Boundaries imagery and labels are available from the European Commission Joint Research Centre (d'Andrimont et al., 2023; <https://jeodpp.jrc.ec.europa.eu/ftp/jrc-opendata/DRLL/AI4BOUNDARIES/orthophoto/>). The 350-tile subset retains original image and label files. Its selection manifest, original file URLs, checksums, analysis code, fitted models, per-record results, and card texts accompany the article in the research archive. Source attribution and reuse terms are retained with the dataset documentation.

AI Usage Declaration

OpenAI ChatGPT assisted with language editing, reference checking, and development of analysis and visualization code.

REFERENCES

- Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for human-AI interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Paper 3, pp. 1-13). Association for Computing Machinery. <https://doi.org/10.1145/3290605.3300233>
- Atkinson, L., & Rosenthal, S. (2014). Signaling the green sell: The influence of eco-label source, argument specificity, and product involvement on consumer trust. *Journal of Advertising*, 43(1), 33-45. <https://doi.org/10.1080/00913367.2013.834803>
- Beilin, R. (2005). Photo-elicitation and the agricultural landscape: “Seeing” and “telling” about farming, community and place. *Visual Studies*, 20(1), 56-68. <https://doi.org/10.1080/14725860500064904>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32. <https://doi.org/10.1023/A:1010933404324>
- Caetano, I., Santos, L., & Leitão, A. (2020). Computational design in architecture: Defining parametric, generative, and algorithmic design. *Frontiers of Architectural Research*, 9(2), 287-300. <https://doi.org/10.1016/j.foar.2019.12.008>
- Chaikin, G. M. (1974). An algorithm for high-speed curve generation. *Computer Graphics and Image Processing*, 3(4), 346-349. [https://doi.org/10.1016/0146-664X\(74\)90028-8](https://doi.org/10.1016/0146-664X(74)90028-8)
- Crisan, A., Drouhard, M., Vig, J., & Rajani, N. (2022). Interactive model cards: A human-centered approach to model documentation. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 427-439). Association for Computing Machinery. <https://doi.org/10.1145/3531146.3533108>
- d'Andrimont, R., Claverie, M., Kempeneers, P., Muraro, D., Yordanov, M., Peressutti, D., Batič, M., & Waldner, F. (2023). AI4Boundaries: An open AI-ready dataset to map field boundaries with Sentinel-2 and aerial photography. *Earth System Science Data*, 15(1), 317-329. <https://doi.org/10.5194/essd-15-317-2023>
- de Freitas Netto, S. V., Sobral, M. F. F., Ribeiro, A. R. B., & Soares, G. R. da L. (2020). Concepts and forms of greenwashing: A systematic review. *Environmental Sciences Europe*, 32(1), Article 19. <https://doi.org/10.1186/s12302-020-0300-3>
- Delmas, M. A., & Burbano, V. C. (2011). The drivers of greenwashing. *California Management Review*, 54(1), 64-87. <https://doi.org/10.1525/cm.2011.54.1.64>

- Di Nicolantonio, M., Rossi, E., & Stella, P. (2020). Generative design for printable mass customization jewelry products. In M. Di Nicolantonio, E. Rossi, & T. Alexander (Eds.), *Advances in additive manufacturing, modeling systems and 3D prototyping* (pp. 143-152). Springer. https://doi.org/10.1007/978-3-030-20216-3_14
- Douglas, D. H., & Peucker, T. K. (1973). Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 10(2), 112-122. <https://doi.org/10.3138/FM57-6770-U75U-7727>
- Ehsan, U., Liao, Q. V., Muller, M., Riedl, M. O., & Weisz, J. D. (2021). Expanding explainability: Towards social transparency in AI systems. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Article 82, pp. 1-19). Association for Computing Machinery. <https://doi.org/10.1145/3411764.3445188>
- Geburu, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., & Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM*, 64(12), 86-92. <https://doi.org/10.1145/3458723>
- Gobster, P. H., Nassauer, J. I., Daniel, T. C., & Fry, G. (2007). The shared landscape: What does aesthetics have to do with ecology? *Landscape Ecology*, 22(7), 959-972. <https://doi.org/10.1007/s10980-007-9110-x>
- Graesser, J., & Ramankutty, N. (2017). Detection of cropland field parcels from Landsat imagery. *Remote Sensing of Environment*, 201, 165-180. <https://doi.org/10.1016/j.rse.2017.08.027>
- Jeong, J., Park, H., Lee, Y., Kang, J., & Chun, J. (2021). Developing parametric design fashion products using 3D printing technology. *Fashion and Textiles*, 8(1), Article 22. <https://doi.org/10.1186/s40691-021-00247-8>
- Kielarova, S. W., Pradujphonphet, P., & Bohez, E. L. J. (2013). An approach of generative design system: Jewelry design application. In *2013 IEEE International Conference on Industrial Engineering and Engineering Management* (pp. 1329-1333). IEEE. <https://doi.org/10.1109/IEEM.2013.6962626>
- Kielarova, S. W., Pradujphonphet, P., & Bohez, E. L. J. (2015). New interactive-generative design system: Hybrid of shape grammar and evolutionary design-An application of jewelry design. In Y. Tan, Y. Shi, F. Buarque, A. Gelbukh, S. Das, & A. Engelbrecht (Eds.), *Advances in swarm and computational intelligence* (pp. 302-313). Springer. https://doi.org/10.1007/978-3-319-20466-6_33
- Kizilcec, R. F. (2016). How much information? Effects of transparency on trust in an algorithmic interface. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 2390-2395). Association for Computing Machinery. <https://doi.org/10.1145/2858036.2858402>
- Li, Y. (2024). Findable then explainable: Retrieval–summary integration for code intelligence on a lightweight CodeSearchNet subset. *Journal of Advanced Computing Systems*, 4(7), 65–82. <https://doi.org/10.69987/JACS.2024.40706>
- Liao, Q. V., Subramonyam, H., Wang, J., & Wortman Vaughan, J. (2023). Designerly understanding: Information needs for model transparency to support design ideation for AI-powered user experience. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Article 9, pp. 1-21). Association for Computing Machinery. <https://doi.org/10.1145/3544548.3580652>
- Lyon, T. P., & Montgomery, A. W. (2015). The means and end of greenwash. *Organization & Environment*, 28(2), 223-249. <https://doi.org/10.1177/1086026615575332>

- Mier, G., Casado Faulí, A. M., Valente, J., & de Bruin, S. (2024). Fields2Benchmark dataset (Version 1.0) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.14524735>
- Mier, G., Valente, J., & de Bruin, S. (2023). Fields2Cover: An open-source coverage path planning library for unmanned agricultural vehicles. *IEEE Robotics and Automation Letters*, 8(4), 2166-2172. <https://doi.org/10.1109/LRA.2023.3248439>
- Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 220-229). Association for Computing Machinery. <https://doi.org/10.1145/3287560.3287596>
- Nie, J., & Zheng, D. (2023). Ambiguity-aware HDFS log anomaly detection with retrieval-augmented failure narratives and selective refusal. *Journal of Advanced Computing Systems*, 3(1), 66–80. <https://doi.org/10.69987/JACS.2023.30105>
- Ode, Å., Tveit, M. S., & Fry, G. (2008). Capturing landscape visual character using indicators: Touching base with landscape aesthetic theory. *Landscape Research*, 33(1), 89-117. <https://doi.org/10.1080/01426390701773854>
- Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1), 62-66. <https://doi.org/10.1109/TSMC.1979.4310076>
- Parguel, B., Benoît-Moreau, F., & Larceneux, F. (2011). How sustainability ratings might deter “greenwashing”: A closer look at ethical corporate communication. *Journal of Business Ethics*, 102(1), 15-28. <https://doi.org/10.1007/s10551-011-0901-2>
- Persello, C., Tolpekin, V. A., Bergado, J. R., & de By, R. A. (2019). Delineation of agricultural fields in smallholder farms from satellite images using fully convolutional networks and combinatorial grouping. *Remote Sensing of Environment*, 231, Article 111253. <https://doi.org/10.1016/j.rse.2019.111253>
- Pushkarna, M., Zaldivar, A., & Kjartansson, O. (2022). Data cards: Purposeful and transparent dataset documentation for responsible AI. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 1776-1826). Association for Computing Machinery. <https://doi.org/10.1145/3531146.3533231>
- Schmuck, D., Matthes, J., & Naderer, B. (2018). Misleading consumers with green advertising? An affect-reason-involvement account of greenwashing effects in environmental advertising. *Journal of Advertising*, 47(2), 127-145. <https://doi.org/10.1080/00913367.2018.1452652>
- Schneider, M., Schelte, T., Schmitz, F., & Körner, M. (2023). EuroCrops: The largest harmonized open crop dataset across the European Union. *Scientific Data*, 10(1), Article 612. <https://doi.org/10.1038/s41597-023-02517-0>
- Seele, P., & Gatti, L. (2017). Greenwashing revisited: In search of a typology and accusation-based definition incorporating legitimacy strategies. *Business Strategy and the Environment*, 26(2), 239-252. <https://doi.org/10.1002/bse.1912>
- Spiegel, S. J. (2020). Visual storytelling and socioenvironmental change: Images, photographic encounters, and knowledge construction in resource frontiers. *Annals of the American Association of Geographers*, 110(1), 120-144. <https://doi.org/10.1080/24694452.2019.1613953>
- Stephenson, J. (2008). The cultural values model: An integrated approach to values in landscapes. *Landscape and Urban Planning*, 84(2), 127-139. <https://doi.org/10.1016/j.landurbplan.2007.07.003>

- Subramonyam, H., Seifert, C. M., & Adar, E. (2021). Towards a process model for co-creating AI experiences. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference* (pp. 1529-1543). Association for Computing Machinery. <https://doi.org/10.1145/3461778.3462012>
- Taha, A. A., & Hanbury, A. (2015). Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. *BMC Medical Imaging*, 15(1), Article 29. <https://doi.org/10.1186/s12880-015-0068-x>
- Taravat, A., Wagner, M. P., Bonifacio, R., & Petit, D. (2021). Advanced fully convolutional networks for agricultural field boundary detection. *Remote Sensing*, 13(4), Article 722. <https://doi.org/10.3390/rs13040722>
- Tveit, M., Ode, Å., & Fry, G. (2006). Key concepts in a framework for analysing visual landscape character. *Landscape Research*, 31(3), 229-255. <https://doi.org/10.1080/01426390600783269>
- Wagner, M. P., & Oppelt, N. (2020). Deep learning and adaptive graph-based growing contours for agricultural field extraction. *Remote Sensing*, 12(12), Article 1990. <https://doi.org/10.3390/rs12121990>
- Waldner, F., & Diakogiannis, F. I. (2020). Deep learning on edge: Extracting field boundaries from satellite images with a convolutional neural network. *Remote Sensing of Environment*, 245, Article 111741. <https://doi.org/10.1016/j.rse.2020.111741>
- Waldner, F., Diakogiannis, F. I., Batchelor, K., Ciccotosto-Camp, M., Cooper-Williams, E., Herrmann, C., Mata, G., & Toovey, A. (2021). Detect, consolidate, delineate: Scalable mapping of field boundaries using satellite images. *Remote Sensing*, 13(11), Article 2197. <https://doi.org/10.3390/rs13112197>
- Wu, T., Terry, M., & Cai, C. J. (2022). AI chains: Transparent and controllable human-AI interaction by chaining large language model prompts. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (Article 385, pp. 1-22). Association for Computing Machinery. <https://doi.org/10.1145/3491102.3517582>
- Yang, Q., Steinfeld, A., Rosé, C., & Zimmerman, J. (2020). Re-examining whether, why, and how human-AI interaction is uniquely difficult to design. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-13). Association for Computing Machinery. <https://doi.org/10.1145/3313831.3376301>