urban science

MDPI

Article

Reading Significance: Using AI to Study Historic Recognition

Melissa Rovner * and Emily Talen

Social Sciences Division, The University of Chicago, Chicago, IL 60637, USA
* Correspondence: melrovner@uchicago.edu; Tel.: +1-805-857-0084

Abstract

The National Register of Historic Places (NR) is a structured artifact of meaning-making that encodes disciplinary values linking architectural and cultural significance to wealth and stylistic distinction. In doing so, it systematically underrepresents vernacular, working-class, and the built environments of racially and ethnically marginalized communities. This paper uses artificial intelligence (AI) to examine how that meaning is constructed. We analyze the preservation record across three scales: a national dataset of 100,117 NR listings (1966–2025), a state-level profile of Illinois's 1997 NR listings, and a close analysis of Lake Forest, Illinois, a community whose exceptional concentration of NR-listed estate architecture makes it an ideal site for examining how preservation significance has been defined and what it excludes. Two parallel AI methods are applied to eighteen Lake Forest nomination documents and their associated photographs. Natural Language Processing (NLP) analyzes nomination text to trace how preservation professionals connect buildings to cultural value; blind AI image analysis examines the same properties to assess how a model trained on cultural imagery constructs visual meaning independently. NLP analysis reveals a corpus dominated by architectural description, with social history, landscape, and labor systematically underrepresented. The visual analysis confirms and amplifies the nomination record's class-based assumptions while reproducing the same omissions regarding labor, diversity, and community context. These findings inform debates about AI's potential to audit existing listings and support nominations for underrepresented property types, while showing that without deliberate corrective design and policy reform, such tools are as likely to replicate the preservation system's inequities as to repair them.

Keywords: historic preservation; national register of historic places; natural language processing; AI visual analysis

Check for updates

Academic Editors: António Ribeiro Amado and Francesca Poggi

Received: 19 March 2026
Revised: 28 April 2026
Accepted: 11 May 2026
Published: 15 May 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Artificial intelligence (AI) has rapidly entered urban studies, offering powerful new methods for analyzing text, images, spatial data, and social processes at unprecedented scale [1]. Recent applications have focused largely on three domains: computational discourse analysis of planning documents and records of public engagement [2]; ethical and governance issues surrounding AI in planning practice [3]; and machine-learning analysis of the built environment using imagery and geospatial data [4]. This paper contributes to these emerging applications by examining an as-yet-understudied dimension of urbanism: the computational analysis of how architectural significance is constructed, whose values it encodes, and what it renders invisible.

Our AI-based evaluation of architectural significance focuses on a selected set of buildings listed on the National Register of Historic Places (NR), the U.S.'s official list of

preservation-worthy historic places. We first apply Natural Language Processing (NLP) to eighteen NR nomination documents from Lake Forest, Illinois, tracing how preservation professionals assign cultural value to buildings while foregrounding and excluding particular themes. We then apply AI-based image analysis to photographs of the same properties, without accompanying nomination text, to assess how a model trained on cultural imagery constructs visual meaning. Comparing these two records reveals where textual and visual interpretations converge, where they diverge, and what both overlook, particularly in relation to labor, landscape, and community context.

Illinois provides an instructive state-level context for this analysis (Figure 1). As home to Chicago, whose architectural heritage has shaped both the national preservation movement and the discipline of architecture history, the state's NR record reflects the outsized cultural authority of high-style buildings designed by named architects. Chicago's Loop and North Shore produced many of the canonical buildings that shaped what historic significance looked like in the formative decades of the NR. At the same time, Illinois's demographic history as the primary northern destination of the Great Migration, and the presence of one of the nation's most significant African American communities in Chicago's South Side, makes the state's underrepresentation of ethnic heritage designations especially legible as an institutional choice.



Figure 1. Map situating Lake Forest, north of the city of Chicago along the shoreline of Lake Michigan, in the northeast corner of the state of Illinois, United States.

A critical dimension of this inquiry concerns what AI systems can see in the built environment and whose values that vision reproduces. If AI image analysis can accurately read architectural character from photographs, and if NLP tools can analyze the language of existing NR nominations, both could support a more accessible nomination process, helping to identify and document overlooked property types and community histories. That potential is limited, however, by training data. Models trained on canonical architectural photography and existing landmark documentation will likely reproduce disciplinary biases, favoring legible stylistic typologies and formal compositions associated with high-

style architecture over the heterogeneity and informality of vernacular and working-class building stock. Even in a wealthy community like Lake Forest, the built environment includes worker housing and other vernacular forms tied to the labor that sustained it. The question animating this study is not only whether AI can read a building, but what it has been trained to see, and whose built environments fall outside the frame of recognition.

1.1. Interpreting Heritage Value

Our study ties into critiques of traditional preservation practice and the recognition that heritage can be interpreted as a socially constructed process [5,6]. Understanding the relationship between built forms and the written descriptions that underpin official recognition is a window into the interpretive process itself: How do planners, preservationists, and historians "read" historic buildings? What values are encoded in that reading, and what is elided?

The NR record is itself a structured artifact, shaped by forces well beyond the historical significance of the properties it contains. The pace of NR listings has been more a function of cultural geography and legislative history than of historical significance: the 1970s and 1980s account for more than half of all listings, driven by the bicentennial of 1976 and the Economic Recovery Tax Act of 1981, which introduced a federal rehabilitation tax credit and generated a wave of nomination activity; the Tax Reform Act of 1986 produced an equally sharp contraction. These patterns are visible in the national data and bear directly on how the Lake Forest corpus should be interpreted.

The nomination of a building or district to the National Register of Historic Places is an act of interpretation, a practice that has been widely critiqued for excluding certain types of heritage [5]. Rooted in Beaux-Arts values that privileged formal composition, stylistic pedigree, and monumental scale, nomination reports determine what is emphasized and what is omitted, constructing narratives of architectural significance that shape how sites are perceived, protected, and reproduced in public memory. Nationally, NR listings have historically clustered in wealthy, predominantly White communities, reflecting both the resource-intensive nature of the nomination process and the discipline's longstanding bias toward monumental and high-style architecture [7–9]. Vernacular buildings—structures whose forms derive from local building traditions, functional requirements, and available materials rather than from a named architect working within a recognized stylistic idiom—working-class neighborhoods, and landscapes associated with racially and ethnically marginalized communities have been systematically underrepresented [10–12].

Roland Barthes argued that the city is a discourse, that urban space functions as a system of signs readable as text, and that its streets, monuments, and zones constitute a language whose grammar can be analyzed [13]. This semiotic approach has been developed within critical heritage studies into a broader critique of how heritage institutions construct value [14]. Smith argues that the dominant Western preservation framework systematically privileges expert knowledge, formal aesthetics, and monumental scale over vernacular, working-class, and non-Western traditions of meaning-making [15]. Zukin similarly demonstrates how heritage value is deployed as an instrument of urban power, determining whose histories are made permanent and whose are rendered invisible [16]. These frameworks are directly relevant to preservation practice, where the selection of what to landmark and how to describe it reflects deeply ideological assumptions about which architectures and histories deserve permanence.

These dynamics are particularly legible in our case study site. Lake Forest, one of the nation's earliest planned residential communities (est. 1861), was designed as a refuge for Chicago's mercantile elite, with a spatial argument that reflected social order and exclusion. With large lots, deed restrictions, and land prices that effectively confined residence to the

city's wealthiest families, Lake Forest was designed to enforce spatial segregation from the outset. The labor force that built and maintained these estates lived in adjacent communities such as North Chicago and Waukegan, whose built environments remain relatively absent in the preservation record. The gap between Lake Forest's listing density and its neighbors is among the most visible expressions of the NR system's uneven patterns of recognition.

1.2. Computational Analysis in Urban Planning

The application of NLP to planning documents has grown substantially over the past decade. Brinkley and Stahmer [17] produced the first large-scale application, using topic modeling on 461 California general plans to surface structural patterns—housing content siloed from environmental justice, environmental justice language adopted before state mandate—that manual review could not have detected at that scale. Mleczko and Desmond [18] extended this approach nationally, constructing a zoning and land use database from municipal codes for approximately 2600 U.S. municipalities. Salazar-Miranda and Talen [19] similarly used NLP to assess form-based coding reforms across more than 2000 U.S. cities. These studies have established NLP as a cost-efficient, longitudinally replicable method for analyzing large administrative document collections.

More recent work has examined how large language models expand these possibilities. Fu [2], reviewing 55 planning NLP studies, identified fragmentation and inadequate validation as the field's primary weaknesses. A subsequent benchmarking study by Fu and colleagues [20] found that GPT-4 substantially outperformed conventional NLP on public comment analysis, achieving 94.1% accuracy against 55.8% for lexicon-based methods, while cautioning that LLMs reproduce training-data biases that can marginalize underrepresented voices. Nie and colleagues [21] demonstrated NLP's capacity to surface structural inequities, finding that greater real estate involvement in local planning meetings correlated negatively with citizen empowerment outcomes across more than 4000 transcripts.

A parallel body of scholarship has examined the ethical dimensions of AI in planning practice. Sanchez, Brenman, and Ye [3] reviewed AI ethics literature as applied to urban planning, finding a striking near-absence of dedicated guidance for practicing planners and arguing that AI systems trained on historical data inevitably inherit past discriminatory patterns. Peng and colleagues [22] propose a four-phase typology of AI in urban planning, from AI-assisted analysis through AI-autonomized decision-making, arguing that human oversight remains essential at every level. The present study positions itself within the AI-assisted phase, using computational tools to augment rather than replace human interpretive analysis of the preservation record.

A third body of literature examines computer vision and deep learning applied to the built environment. Zhang and colleagues [1] term this domain "urban visual intelligence," situating current computational methods within a century-long tradition of visual urban analysis and proposing a hierarchy from street-level to city-scale analysis that encompasses scene classification, style identification, and place perception modeling. Ibrahim, Haworth, and Cheng [23] identified a critical gap in cross-layer research, calling for a shift from prediction toward AI-assisted decision-making and policy guidance.

More relevant to this study is work using AI to interpret the built environment for social meaning rather than categorical classification. Romero-Vega and colleagues [24] used CNN-based object detection to model human perception of inequality through architectural cues in Raleigh neighborhoods, providing a methodological precedent for AI readings of built environment as social text. Giannoulaki and Gruen [25], in a systematic review of 88 studies at the intersection of computer vision and architectural analysis, found that computational approaches are most valuable when grounded in well-defined spatial and visual criteria from architectural practice. Moradi and Biloria [26] raised critical questions

about whose perceptual norms are encoded in AI systems analyzing street-view imagery and what biases are reproduced at scale.

NLP and computer vision have developed largely in parallel within urban studies, each generating substantial methodological literature but only rarely applied in combination to the same corpus. The equity critique of the National Register likewise has an established scholarly foundation, where scholars have documented the structural underrepresentation of vernacular, working-class, and non-White built environments in the official preservation record [8–12]. What this literature has not yet attempted is to bring computational text analysis and AI image analysis to bear on the same set of preservation documents simultaneously, examining how meaning is constructed around architectural form across both linguistic and visual registers. Addressing that gap is the contribution of the present study.

2. Materials and Methods

This study employs a three-tier analytical framework. We first analyze the national NR listing database to establish the structural patterns against which the Lake Forest corpus can be interpreted. Second, we apply natural language processing to the full set of Lake Forest nomination documents. The third tier uses AI-based image analysis to examine photographs of the same properties under a blind protocol. The second and third tiers together form the core comparative analysis.

2.1. National and State Database Analysis

The national dataset was obtained from the National Park Service's National Register database export (as of June 2025), comprising 100,117 property records with fields including property name, state, county, city, category of property, area(s) of significance, listed date, and other names. The analysis was conducted in Python 3.13 using the pandas 3.0.1 and openpyxl 3.1.5 libraries for data manipulation and spreadsheet processing. Areas of significance were parsed from semicolon-delimited strings and normalized for consistent counting; because individual listings may include multiple areas, percentages reflect the share of listings in which a given area appears rather than the share of a fixed total. Decade-by-decade distributions were computed from the listing date field. It is important to note that this dataset counts each individually listed property and historic district as a single record; contributing properties within historic districts—buildings that are part of a designated district but not individually listed—are not included as separate records in the NPS export and are therefore excluded from all counts reported here. State-level analysis for Illinois was conducted by filtering the state field, yielding 1997 records, and replicating the decadal and area-of-significance analyses within that subset. This count is consistent with the Illinois Historic Preservation Agency's own database (HARGIS), which records 1973 individually listed properties as of its most recent update—a difference of 24 records attributable to the HARGIS dataset predating the June 2025 NPS export by approximately one update cycle. Use-type detection was performed via keyword matching on the concatenated property name and other names fields, with residential listings identified through terms including house, residence, home, dwelling, farmhouse, cottage, manor, villa, mansion, and estate.

2.2. Lake Forest NLP Corpus: Document Preparation and Preprocessing

The local corpus comprises 18 of the 20 National Register nomination PDFs for properties listed in Lake Forest, Illinois. Two listings are excluded from the NLP and visual corpora: the Lake Forest Cemetery, excluded due to its specific use-type as a non-architectural landscape feature that falls outside the evidentiary register this study is designed to examine;

and the Mildred and Abel Fagen House (listed 2021), excluded because its nomination report has not yet been digitized and was not available through the NPS online system at the time of data collection. Lake Forest's 20 NR entries comprise five historic district listings—Lake Forest Historic District, Green Bay Road Historic District, Vine–Oakwood–Greenbay Historic District, West Park Neighborhood Historic District, and Deerpath Hill Estates Historic District—and fifteen individually listed properties. All five district nominations are included in the corpus; district-level nominations contain photographs and narrative descriptions of multiple contributing properties and therefore serve as a proxy for analysis of those buildings. An additional 275 contributing properties recorded in state survey data do not appear as discrete entries in the national dataset and are excluded from the analysis.

The 18 listings in the corpus were produced between the early 1970s and the mid-1990s. PDFs were processed using pdfplumber 0.11.9 for text extraction, followed by a two-pass named entity normalization process to consolidate variant forms of architect names, firm names, and place references into standardized tokens. The consolidated text was lowercased, stripped of non-alphabetic characters, and tokenized by whitespace. A two-tier stop-word list, comprising standard English stop words supplemented with NR boilerplate terms such as nomination, section, and continuation, was applied, yielding a working corpus of 122,129 total tokens and 15,104 unique types, with a type–token ratio of 0.124. Named entity recognition was performed using a structured registry of 24 architect and landscape designers associated with Lake Forest estates, including Howard Van Doren Shaw, Kersey Coates Reed, David Adler, Jens Jensen, O.C. Simonds, and Frederick Law Olmsted, along with a set of architectural firms and local institutions including Holabird and Roche, Pond and Pond, Lake Forest College, and Onwentsia Club. Entities were resolved to underscore-joined canonical tokens prior to frequency analysis, preventing fragmentation across abbreviated and full-name variants.

2.3. Thematic Classification

Terms were assigned to one of seven inductively derived thematic categories designed to capture the principal discursive registers of NR nominations (Table 1): (1) Built Fabric, Materials and Elements; (2) Architectural Style and Form; (3) Ownership and Social Class; (4) Landscape and Site; (5) Architectural Practice and Design Authorship; (6) Social and Community History; and (7) Adaptive Reuse and Integrity. Definitions and example terms are included in Table 1. Category membership was established through close reading of the nomination corpus and cross-referenced with the NPS documentation standards. Thematic distribution was computed as the percentage of all classified substantive tokens falling into each category, with an unclassified residual retained for high-frequency terms not clearly assignable to a single domain.

Table 1. Thematic Classifications by category, according to their respective definitions and example terminology.

#	Category	Definition	Example Terms
1	Built Fabric, Materials & Elements	Physical description of building components and materials. Distinguished from Category 2 by its focus on tangible physical substance rather than stylistic classification.	windows, doors, walls, roofs, brick, stone
2	Architectural Style & Form	Stylistic attribution and formal composition; vocabulary concerned with the building's place within architectural history rather than its physical fabric.	revival, colonial, tudor, massing, symmetry

Table 1. *Cont.*

#	Category	Definition	Example Terms
3	Ownership & Social Class	Terms denoting property ownership, social identity, and class position. Distinguished from Category 6 by its focus on individual owners and patrons rather than collective or institutional life.	estate, gentleman, club, prominence, family
4	Landscape & Site	Vocabulary of the designed and natural environment surrounding the structure.	garden, shore, gate, lane, tree, landscape
5	Architectural Practice & Design Authorship	References to the professional work of architects, landscape designers, and builders.	architect, firm, plan, studio, builder
6	Social & Community History	Vocabulary of public and collective life; the property's relationship to civic or religious institutions. Distinguished from Category 3 by its focus on institutional rather than individual identity.	school, church, library, community, association
7	Adaptive Reuse & Integrity	Preservation-specific evaluation language relating to alterations, physical condition, and material authenticity.	alteration, integrity, restoration, character

2.4. Collocation, Style Detection, N-Gram, and Evaluative Qualifier Analysis

Collocations were computed using a symmetric plus-or-minus five-word window centered on each of the 25 highest-frequency content words across the concatenated corpus, with the top six collocates per target term reported. A term-theme co-occurrence heat map was constructed by computing, for each of the top 15 corpus terms, the raw co-occurrence count against each thematic category's keyword set within the same window, with cell intensities normalized to a 0–5 scale.

Architectural style detection employed a phrase-matching routine scanning the raw, pre-tokenized text of each nomination for 34 canonical style labels drawn from architectural history nomenclature, including Tudor Revival, Colonial Revival, Georgian Revival, French Renaissance Revival, Prairie Style, Arts and Crafts, Beaux-Arts, and Shingle Style, as well as landscape traditions such as the Jensen School and Olmstedian design. Results were aggregated both as corpus hits and as document frequency across the 18 nominations.

Bigrams and trigrams were extracted from the lemmatized, stop-word-filtered corpus to surface recurring multi-word expressions not captured by single-term frequency analysis, particularly compound style labels, property-name references, and formulaic description patterns. A set of 41 evaluative and significance-signaling terms, including original, historic, fine, distinguished, exceptional, intact, craftsmanship, integrity, and unaltered, was used to track the qualitative register of nomination language within each thematic category, with top collocating evaluative terms extracted using the same five-word window.

2.5. AI Visual Analysis

The visual analysis was conducted using Claude Sonnet (claude-sonnet-4-20250514, Anthropic), a Large Language Model (LLM) with multimodal image analysis capabilities. In accordance with the journal's requirement for disclosure of generative AI use in research, the model served as the primary analytical instrument for the visual component of the study, rather than as an aid to writing or text editing. All visual analysis outputs were produced by the model and treated as analytical data. It bears emphasis that AI is used

here to augment critical human interpretation rather than to automate analytical judgment: the structured prompt protocol, thematic categories, and interpretive framework were designed by the research team, and all model outputs were evaluated and analyzed by the authors against the NLP findings and the broader preservation literature.

The protocol was designed as a blind reading: the model received only the photographic images contained in each property's nomination PDF and was provided no property name, reference number, nomination text, architect attribution, or any other contextual information. The full analytical workflow proceeded as follows. (1) PDF extraction: nomination PDFs were obtained from the NPS online database for each of the 18 Lake Forest properties in the corpus. (2) Image extraction: photographs were extracted from each PDF using PyMuPDF 1.27.2.3, with each image-bearing page rendered as a JPEG at 150 DPI. A minimum file size threshold of 8000 bytes was applied to exclude decorative elements and page artifacts; a maximum of 20 images per property was set to manage token limits while ensuring representative coverage. (3) Blind prompt delivery: extracted images were passed to the model with no accompanying text, property name, or nomination context. The same six-part structured prompt was applied identically to all 18 properties. (4) Output structuring: model responses for each property were saved as structured JSON files capturing responses across all six analytical lenses. (5) Corpus aggregation: individual JSON outputs were combined into a corpus-level dataset. (6) Comparative analysis: corpus-level AI visual findings were systematically compared against the NLP thematic and frequency results by the research team, with divergences and convergences interpreted in relation to the preservation literature.

3. Results

3.1. National and State Context

We begin by situating the Lake Forest case within broader national and state patterns of National Register designation. Architecture is the dominant "Area of Significance" noted across the full dataset, appearing in approximately 74% of all listings and peaking at 82.7% in the 1980s (Figure 2). Its dominance reflects structural features of the NR form itself, which dedicates more space, detailed prompting, and technical vocabulary to physical description and architectural analysis than to any other dimension of significance. Social History, the next most significant category, accounts for roughly 12% of listings nationally but grew from 6 to 7% in the 1970s and 1980s to over 15% in recent years, reflecting the gradual absorption of social and cultural history frameworks into nomination practice. Equity-oriented categories evolved more slowly. Ethnic Heritage-Black designations are effectively absent from the record through the 1980s, appear at under 1% through the 2000s, and reach approximately 3% in the partial 2020s dataset. Across six decades, these categories account for roughly 1% of all listings, a figure bearing no relationship to the demographic weight of Black Americans in U.S. history or to the scale of the associated architectural and cultural heritage.

Turning to the state level, Illinois contributes 1997 properties to the national dataset, putting it in the mid-tier of states in terms of listing count. Its temporal distribution closely tracks the national pattern, with the same legislative inflection points producing the same surges and contractions. While the national share of listings designated under Architecture declines sharply after the 1980s peak, the Illinois share remains elevated, averaging above 75% through the 2010s and 2020s. This reflects the outsized influence of Chicago on the state's preservation culture, the strength of the city's high-style architectural heritage, and the institutional priorities of state preservation programs trained in architectural history. The equity category trajectories in Illinois lag behind national patterns. Ethnic Heritage-Black designations are absent from the Illinois record through the entire 1990s and

remain below 1% through the 2000s, despite the state's demographic history as a primary destination of the Great Migration and home to one of the nation's most historically significant African American communities.

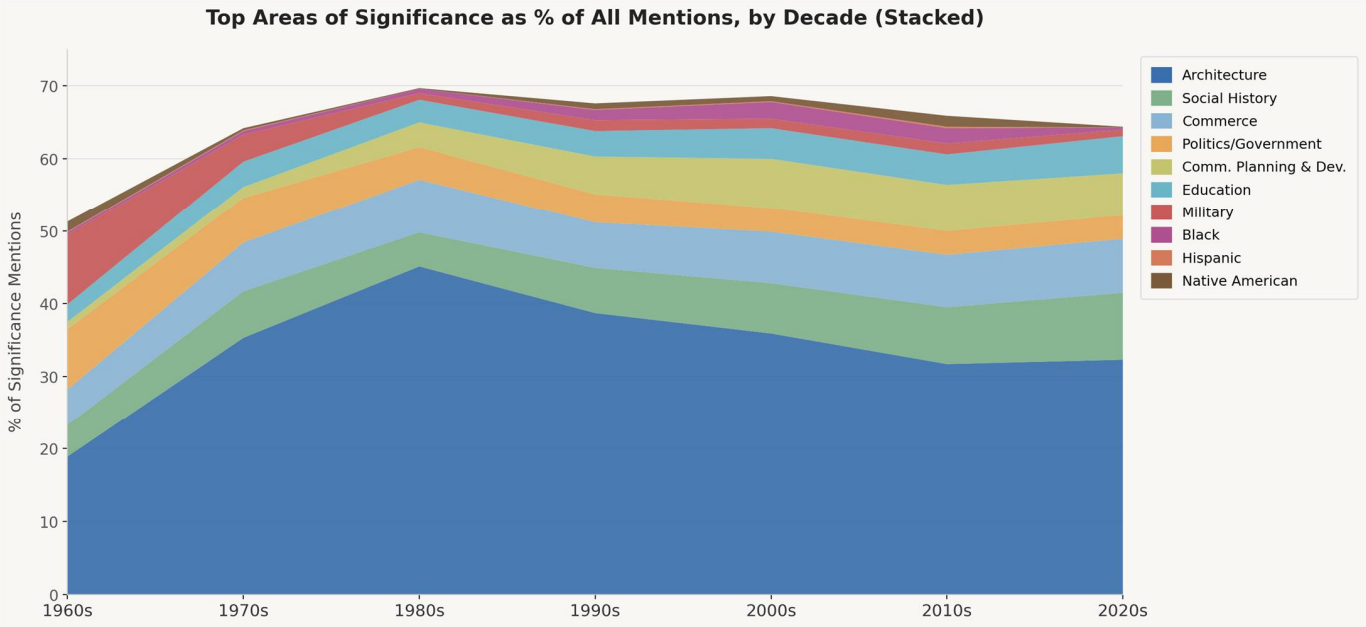


Figure 2. Areas of significance by decade across all U.S. National Register of Historic Places listings since its founding. Refer to legend at top right associating areas of significance with color. The area of each color field is proportional to the percent of mentions of that significance area from 1960–2020.

Lake Forest's 20 NR listings place it twelfth among Illinois municipalities by absolute count, but its listing density of approximately 10.3 per 10,000 residents substantially exceeds Chicago's 1.5 and is comparable only to smaller historic cities like Galena and Jacksonville, where local economies are organized around heritage tourism. All 20 listings are either residential properties, historic districts, or estate complexes, and the corpus is temporally clustered in two waves corresponding to the national surges. The Lake Forest Historic District contains over 800 contributing structures, including 559 single-family homes reflecting the region's planning to suit an elite Chicago audience (Figure 3). This profile, a highly coherent set of architect-designed residential properties listed during the peak decades of architecture-dominant nomination practice, makes Lake Forest an unusually controlled case for examining how the NR system constructs architectural significance.

3.2. Lake Forest NLP Findings

The Lake Forest NLP corpus comprises 18 nomination documents totaling 122,129 tokens and 15,104 unique types, with a type–token ratio of 0.124. The five highest-frequency substantive terms are house (n = 3583, 2.93%), design (n = 1123, 0.92%), style (n = 837, 0.69%), window (n = 826, 0.68%), and estate (n = 809, 0.66%). Twenty terms achieve universal document frequency across all 18 nominations, including landscape, brick, structure, building, estate, architect, architecture, site, residence, garden, style, design, house, preservation, and plan. These universally shared terms constitute the stable lexical infrastructure of the nomination genre across the corpus. Named entity analysis identifies John Anderson (n = 411, DF 15/18) and Howard Van Doren Shaw (n = 387, DF 14/18) as the most frequently appearing design professionals, followed by Kersey Coates Reed (DF 8/18), David Adler (DF 7/18), and Jens Jensen (DF 4/18).

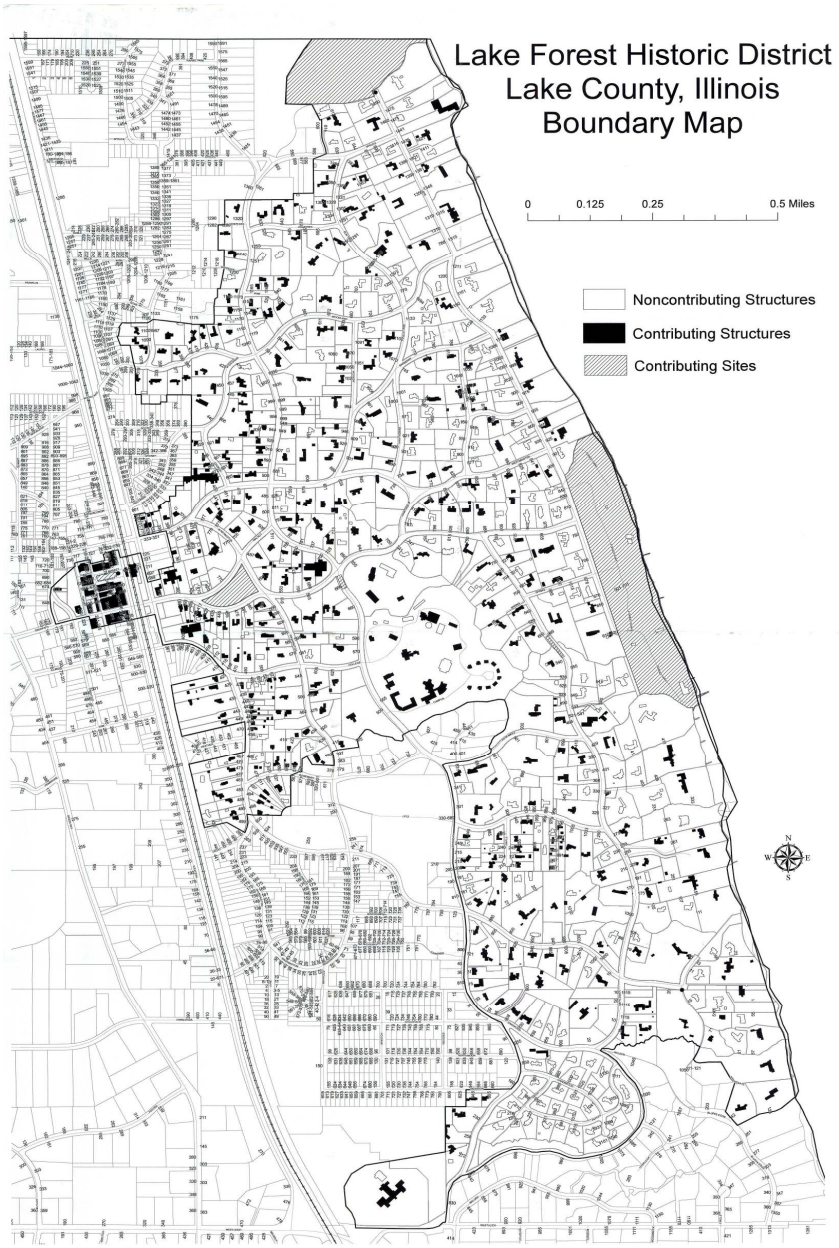


Figure 3. Boundary map of the Lake Forest Historic District, Lake Forest, Illinois, showing the concentration and extent of contributing structures along the Lake Michigan shoreline. Source: National Register of Historic Places Nomination Form, Lake Forest Historic District, National Park Service, U.S. Department of the Interior. Public domain.

The thematic distribution of the corpus across seven categories (Table 2), shows aggregate token weight, percentage of the substantive corpus, top terms by frequency, and dominant evaluative qualifiers within each category. Built Fabric, Materials and Elements dominate the corpus at 8.3% of substantive tokens, with building materials vocabulary specifically accounting for 2.3% of the total corpus, or 27% of the Built Fabric category. Architectural Style and Form follows at 5.3%. Together these two categories establish physical description and style classification as the primary evidentiary register of the nomination record. Ownership and Social Class registers at 2.1%, present across all 18 nominations and reflecting the degree to which elite patronage and estate identity frame significance arguments throughout the corpus. Landscape and Site accounts for 1.9%, a modest figure given Lake Forest's historically significant designed landscapes associated with Jens Jensen, O.C. Simonds, and Olmsted-influenced estate commissions. Social and

Community History accounts for 1.0% of tokens and appears in only 14 of 18 nominations. Adaptive Reuse and Integrity is the most sparsely represented category at 0.4%.

Table 2. Thematic distribution within Lake Forest nomination records, showing aggregate token weight of each thematic category as a percentage of the total corpus.

Thematic Category	Tokens	% Corpus	Top Terms (Freq)	Top Evaluative Qualifiers
Built Fabric, Materials & Elements	10,187	8.3%	window, building, room, wall, roof, door, garage	original (84), prominent (14), historic (13), elaborate (11), typical (8)
Architectural Style & Form	6451	5.3%	design, style, revival, architecture, colonial, french, tudor	historic (85), original (54), fine (29), significant (27), period (25)
Ownership & Social Class	2571	2.1%	estate, country, residence, family, club, prominence, society	historic (29), original (20), prominent (12), period (8), elegant (7)
Landscape & Site	2320	1.9%	landscape, garden, shore, gate, site, lane, tree	original (26), historic (22), important (9), prominent (9), significant (5)
Architectural Practice & Authorship	1863	1.5%	architect, plan, firm, office, builder, studio, practice	historic (55), original (23), prominent (12), distinguished (5), significant (5)
Social & Community History	1179	1.0%	school, library, business, church, community, association, commercial	historic (20), important (9), significant (9), major (6), period (6)
Adaptive Reuse & Integrity	501	0.4%	preservation, alteration, character, integrity, demolition, restoration, renovation	historic (122), original (4), integrity (4), major (3), significant (2)

The style detection engine identified 28 distinct architectural style labels across the 18 nominations (Table 3). Tudor Revival dominates with 415 corpus hits across 15 nominations (83%), followed by Colonial Revival (294 hits, 56% of nominations), Georgian Revival (213 hits, 78%), French Renaissance Revival (107 hits, 61%), and Norman Revival (104 hits, 67%). Prairie Style achieves document frequency in 10 of 18 nominations but accumulates only 31 corpus hits, less than one tenth of Tudor Revival's total.

Table 3. Architectural style weighting in the Lake Forest NR nominations by total phrase occurrences and document frequency, with top 3 matched phrases indicated.

Rank	Canonical Style	Corpus Hits	DF (/18)	% Nominations	Matched Phrases (Top 3)
1	Tudor Revival	415	15	83%	"tudor" (239), "tudor revival" (148), "english tudor" (19)
2	Colonial Revival	294	10	56%	"colonial revival" (279), "colonial revival style" (12), "dutch colonial revival" (3)
3	Georgian Revival	213	14	78%	"georgian" (139), "georgian revival" (63), "georgian style" (10)
4	French Renaissance Revival	107	11	61%	"french renaissance" (59), "french renaissance revival" (33), "french provincial" (12)

Table 3. *Cont.*

Rank	Canonical Style	Corpus Hits	DF (/18)	% Nominations	Matched Phrases (Top 3)
5	Norman Revival	104	12	67%	"norman" (79), "french norman" (21), "norman style" (4)
6	Gothic Revival	95	5	28%	"gothic" (55), "gothic revival" (25), "collegiate gothic" (14)
7	Beaux-Arts	79	9	50%	"beaux-arts" (32), "ecole des beaux-arts" (23), "beaux arts" (15)
8	Arts and Crafts	75	14	78%	"arts and crafts" (40), "craftsman" (28), "craftsman bungalow" (5)
9	Queen Anne	70	4	22%	"queen anne" (61), "queen anne style" (9)
10	Romanesque Revival	69	1	6%	"romanesque" (27), "richardsonian" (21), "richardsonian romanesque" (17)
11	Art Deco	57	5	28%	"art deco" (56), "art moderne" (1)
12	Italian Renaissance Revival	56	4	22%	"italian renaissance" (36), "italian renaissance revival" (20)
13	Italianate	48	3	17%	"italianate" (44), "italianate style" (4)
14	Neoclassical	45	4	22%	"classical revival" (40), "neoclassical" (4), "neo-classical" (1)
15	Vernacular	44	3	17%	"vernacular" (43), "vernacular architecture" (1)
16	Picturesque	42	10	56%	"picturesque" (40), "romantic picturesque" (1), "picturesque style" (1)
17	Shingle Style	37	9	50%	"shingle style" (28), "shingle-style" (9)
18	Prairie Landscape/Jensen School	33	4	22%	"jens jensen" (32), "prairie landscape" (1)
19	Prairie Style	31	10	56%	"prairie school" (18), "prairie style" (11), "prairie house" (2)
20	Second Empire	21	3	17%	"second empire" (21)
21	Bungalow	21	3	17%	"bungalow" (20), "bungalow style" (1)
22	Olmstedian Landscape	21	2	11%	"olmsted" (21)
23	Richardsonian Romanesque	17	1	6%	"richardsonian romanesque" (17)
24	Spanish Colonial Revival	6	3	17%	"spanish revival" (3), "spanish colonial revival" (1), "spanish mission" (1)
25	Federal Style	5	2	11%	"federal style" (3), "federal period" (1), "adam style" (1)
26	Greek Revival	4	3	17%	"greek revival" (3), "temple front" (1)
27	Modernist/Modern Movement	2	1	6%	"modernist" (1), "modern movement" (1)
28	Stick Style	1	1	6%	"stick style" (1)

A note on the terminology in Table 3: the style labels used here follow the analytical vocabulary of NR nominations themselves rather than a strict architectural history typology, which produces a list that mixes formal movements, regional traditions, vernacular building types, and aesthetic sensibilities. Several entries merit clarification. "Vernacular" in this context refers to buildings whose forms derive from local building traditions, available materials, and functional requirements rather than from a trained architect working within

a named stylistic idiom—it describes a relationship to production and knowledge rather than a specific formal vocabulary. "Picturesque" is not a building style but a design philosophy associated with the English landscape tradition that values variety, irregularity, and the integration of buildings into naturalistic settings; it appears in nominations to describe site and landscape treatment rather than architectural form. "Bungalow" denotes a building type—small, single-story or story-and-a-half domestic structure with wide eaves and horizontal massing—that cuts across multiple stylistic movements; its presence alongside Neoclassical and Beaux-Arts reflects the nomination corpus's use of both type and style labels rather than a conflation of those categories.

The dominant evaluative terms across the corpus (Figure 4) are historic (n = 85), original (n = 84), fine (n = 29), distinguished, and exceptional. These terms operate in distinct evaluative registers: "original" functions primarily as a condition assessment, affirming that a physical element survives in unaltered form and thereby satisfies the NR integrity standard; "historic" by contrast functions as a period attribution, characterizing materials, features, or patterns as belonging to the property's period of significance rather than making a claim about their survival. "Prominent" operates as a visual–spatial descriptor applied to features that are architecturally dominant or publicly legible; "elaborate" denotes formal complexity and decorative richness, functioning as a qualitative marker of craft investment rather than position or scale. That both pairs appear frequently in the same thematic categories reflects the NR form's tendency to layer physical description, temporal attribution, and aesthetic evaluation within single passages. Collocation analysis confirms the concentration of evaluative language within the Built Fabric and Architectural Style categories: the strongest collocates of house, design, and style are material and formal descriptors, while social, community, and labor terms generate weak collocate relationships with the corpus's highest-frequency terms. The top bigrams (Figure 4) are colonial revival (288 occurrences), revival house (199), hill estate (183), and country house (164). The top trigrams are colonial revival house (86) and double hung window (85).

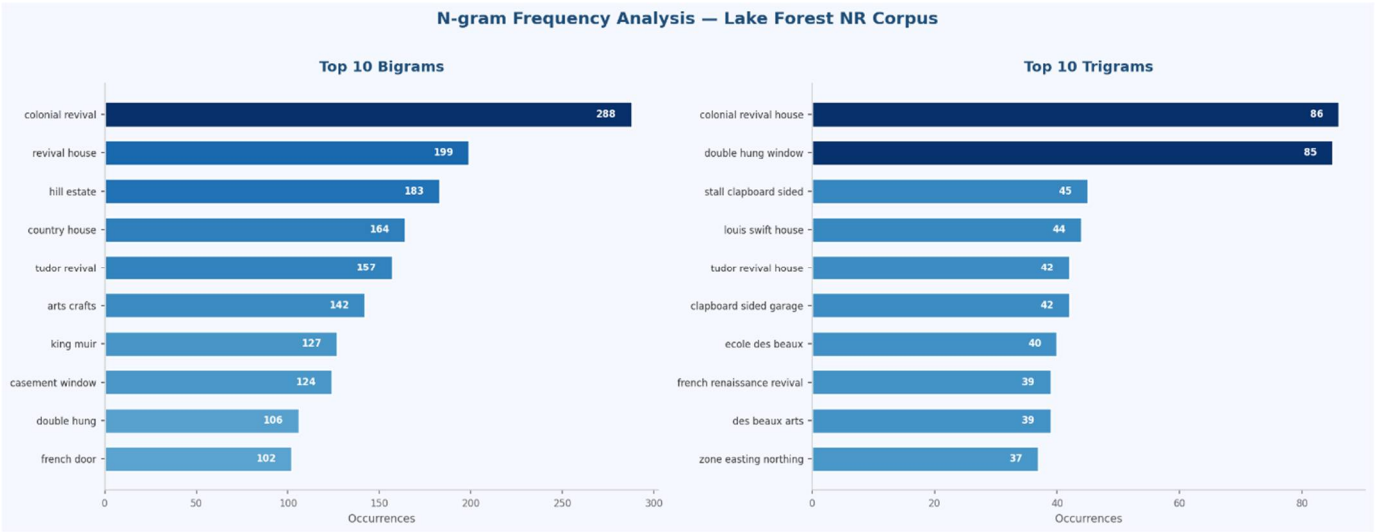


Figure 4. Chart demonstrating frequency of bigrams and trigrams within the Lake Forest NR nomination document corpus's thematic language.

3.3. AI Visual Analysis Findings

3.3.1. Protocol and Prompt Structure

Each property was analyzed through a structured six-part prompt delivered to Claude Sonnet (claude-sonnet-4-20250514) alongside the extracted property photographs, with no accompanying text, property name, or nomination context provided. The prompt asked

the model to analyze: (1) architectural style: identify primary and secondary styles with a confidence rating and any relevant historiographic framing; (2) materials and physical fabric: identification of primary and secondary exterior and interior materials; (3) landscape and site character: read spatial organization, planting character, boundary treatment, and the relationship between structure and site; (4) social and ownership cues: what the photographs suggest about the socioeconomic status, use patterns, and class position of the property's occupants, including reading of scale, service infrastructure, ancillary structures, and exclusivity markers; (5) condition and integrity: overall assessment of physical condition and degree of alteration from historic character; and (6) semiotic reading: "What cultural narratives, social values, and ideological associations does the architectural and landscape language of this property construct? What does the photographic record emphasize, and what does it render invisible or absent? What kind of social world is this property designed to produce, sustain, and project?" A cross-theme synthesis was generated for each property following the six-lens analysis. Outputs were structured as JSON files per property (see Appendix A: Clifford Milton Farm example) and aggregated into a corpus-level dataset for comparison with the NLP findings.

3.3.2. Style Identification

Across all 18 properties the model identified primary architectural styles with high confidence, converging with the nomination's primary style characterization in 17 of 18 cases. The one partial divergence was Deerpath Hill Estates, where the model identified English Garden City Movement and Arts and Crafts Revival as the primary reading, while the nomination argued planned residential district significance. The model's reading is not incorrect but operates at a different analytical register, oriented toward architectural typology rather than planning history. In several cases the model's style vocabulary exceeded the nominations'. For the Armour Ogden House, the model introduced the historiographic category Country Place Era, situating the property within the national movement of designed private estates from the 1880s through 1930s, a framing absent from the nomination text. For the Lake Forest Historic District, the model declined to assign a single primary style, instead identifying a Historic District Ensemble with seven secondary styles operating in aggregate.

The model similarly produced consistent readings with the nomination NLP analysis across the other four evaluative dimensions: site character was rated formal for all 18 properties; scale was rated large-estate for 14 properties and substantial for 4; condition was rated excellent for 13 properties, good for 4, and mixed for 1; and service infrastructure was rated visible for all 18 properties.

3.3.3. Social and Ownership Cues

The model's readings of social and ownership cues drew on a consistent vocabulary of visual elements across the corpus. Boundary features were read as deliberate spatial instruments in every property where they appeared. Gates, perimeter walls, stone pillars, screened approach drives, and elaborately sequenced entries were identified not merely as physical enclosures but as social infrastructure: elements that establish "clear boundaries between the private enclave and the broader urban context" and produce what the model described at one property as "semi-public displays of prosperity while maintaining appropriate privacy for elite domestic life." Gatehouses, where present, were read as "both practical barrier and symbolic threshold separating elite domestic space from public realm".

Scale was read not merely as a quantitative measure but as a social argument. Properties where the primary structure occupied only a portion of a large lot, with the remainder devoted to formal gardens, stable complexes, service drives, and recreational amenities,

were consistently read as performing a spatial claim about the relationship between wealth and land. For one estate, the model identified an ornamental fountain as creating "a ceremonial approach axis," reading the property as demonstrating "not merely wealth but social legitimacy rooted in historical precedent." The model identified "servants' quarters" visible in floor plan photographs at another property as direct evidence of "a complex domestic hierarchy".

Material quality constituted a third register of class reading. The model identified high-quality craftsmanship as a finding across all 18 properties, with specific evidence varying by property: Flemish bond brickwork, ashlar limestone, slate roofing, leaded glass, wrought iron detailing, and custom millwork were each cited in multiple analyses as accumulating evidence of commissioned rather than speculative construction. At one working farm property, the model read the "sophisticated structural engineering" of the circular barn alongside "high-quality craftsmanship with coursed limestone masonry" as evidence of a gentleman's farm where "functions are elevated through sophisticated architectural treatment," presenting farming as "both productive enterprise and cultural statement".

3.3.4. Semiotic Readings

The semiotic readings produced overlapping cultural associations organized around the dominant style registers of the corpus. Tudor, Norman, and French Chateau Revival properties were consistently read as invoking English and continental aristocratic heritage. The medieval revival vocabulary was identified as signaling "permanence, tradition, and inherited status," with the model reading these styles as strategies for legitimizing new wealth through association with pre-industrial social hierarchies. One Norman and French Chateau Revival estate was characterized as "an elaborate cultural performance of aristocratic taste and social distinction, employing medieval European architectural languages to construct an aura of ancient nobility".

Colonial and Georgian Revival properties, which constitute the largest single style cluster in the corpus, generated readings of a specifically American variant of the same logic. The model consistently interpreted symmetrical facades, classical porticos, and formal landscape axes as invoking "founding-era legitimacy" and "colonial American gentility," positioning owners within narratives of historical stewardship rather than commercial achievement. One Georgian estate was read as functioning as "both residence and cultural institution," its photographic record presenting "a preserved historical artifact rather than a lived-in domestic space".

The corpus's single modernist property generated a reading that maintained the class identification while inverting the style logic. International Style architecture was read as projecting "values of rational planning, technological progress, and rejection of historical revival styles," with the model observing that "the modernist vocabulary itself functions as a marker of cultural capital, distinguishing the owners as forward-thinking and architecturally informed." The Arts and Crafts and Prairie School property similarly produced a reading of tension between democratic design ideals and elite patronage, identifying a moment when "American domestic architecture sought to balance democratic ideals with clear markers of social distinction." Both outlier readings confirm the corpus-wide pattern, identifying class position and social distinction as legible across all style registers.

3.3.5. Absent Content

The model identified absent content with equal consistency across all 18 analyses without specific prompting within the semiotic lens. The photographic record contained no signs of daily domestic labor, no service vehicles or utility infrastructure, no depiction of the working landscapes required to sustain the properties shown, no representation of

community context beyond property boundaries, and no social diversity of any kind. At one agricultural estate the model noted the absence of "working farm equipment, signs of active agricultural use, and evidence of the labor force required to maintain such an extensive operation," producing "an idealized view of agricultural life divorced from its productive realities." At multiple residential properties the model observed that images present "effortless perfection" by excluding "the maintenance required to achieve this level of preservation." The model consistently characterized this pattern as a property of the nomination photography constructing "a vision of what those buildings are for, one from which the conditions of their production have been systematically excised".

4. Discussion

The parallel analytical tracks of this study converge on a shared finding: that the Lake Forest preservation record is organized around the estate as the fundamental unit of historic significance, that architectural style functions as the primary evidentiary register within which significance is argued, and that class legibility is produced continuously across both linguistic and visual modes. Where the two tracks diverge is equally instructive, and the divergences reveal something important about what each method is designed to see, what the nomination system is designed to produce, and what both systematically fail to acknowledge. The study therefore suggests that preservation documentation is not simply a record of historic significance but a system that actively produces it. This leads to a pointed question about AI's role in the future of preservation practice: will AI reinforce or reshape those patterns?

The dominance of the Built Fabric, Materials & Elements thematic category reflects the disciplinary conventions of the National Register—specifically, Criterion C architectural significance, which directs evaluators toward formal qualities, materials, and stylistic attribution. The collocate patterns make this legible: evaluative modifiers like original, intact, and ornate simultaneously describe physical condition and assert preservation value, functioning not as neutral description but as a combined act of documentation and significance-making. AI visual analysis amplifies the same elements, and not only because of what the nomination photographs contain. The model has been trained on the very documentary record that has historically emphasized wealthy, high-style architecture, encoding its biases before the analytical protocol begins. While NR policy over the past two decades has incentivized nominations for properties associated with women, persons of color, and working-class histories, this study shows those identities absent from both the textual and photographic record.

Both analytical tracks identify the bounded private domain of the architect-designed residence as the primary spatial unit of significance—over the neighborhood, the streetscape, or the community. The underrepresentation of landscape and site vocabulary is particularly striking given Lake Forest's planning identity: the curvilinear streets, ravine system, and designed relationship between built form and natural landscape are central to its historical character as an early planned suburb, yet landscape language registers at a fraction of the weight of architectural element language, and Jens Jensen's landscape work does not generate the descriptive density that building elements do. This gap illustrates how computational analysis of nomination language can serve as an audit tool for thematic coverage, surfacing corpus-level patterns that individual close reading would not detect.

The Ownership & Social Class thematic category registers a share of corpus language that, while secondary to built fabric description, is analytically significant for what it reveals about how significance is framed. Terms like estate, gentleman, and prominent consistently anchor historical significance to the identity of wealthy patrons rather than to the labor that built the estates, the domestic workers who maintained them, or the broader community

whose exclusion was constitutive of Lake Forest's planned social order. This pattern is not unique to Lake Forest. The NR Criteria for Evaluation link significance to "association with the lives of persons significant in our past," a formulation that has historically operated to amplify the biographical record of elite owners at the expense of other social histories. The Lake Forest corpus makes the pattern measurable: social history language is thin, and the social history that is present is predominantly the history of privilege.

The AI visual analysis amplifies and extends this reading. Where nomination language describes architectural elements, the visual analysis interprets what those elements are doing socially. The model identifies boundary features as spatial instruments of exclusion, reading scale as a claim about the relationship between wealth and land, and framing material quality as evidence of commissioned rather than speculative construction. It also generates interpretive readings of style that nomination language does not attempt: Tudor and Norman Revival are identified as strategies for legitimizing new wealth through association with pre-industrial social hierarchies; Colonial Revival symmetry is read as an invocation of founding-era legitimacy; modernist formal language is characterized as cultural capital that inverts the style logic while maintaining the class logic. These readings reflect the model's training on architectural history, art history, and critical discourse about the built environment. That training is itself a form of bias, but it also produces analytical value that pure frequency analysis cannot.

The most revealing finding of the study is what neither analytical track is able to capture. The NLP corpus contains no vocabulary of labor, construction workers, craft process, or the domestic service workforce that maintained the Lake Forest estates. The AI visual analysis identifies servant's quarters and service drives as class-legibility markers but cannot recover the social histories of the people who occupied them. As the model itself observed, the photographic record presents "effortless perfection" by excluding "the maintenance required to achieve this level of preservation." This is a structural limitation: neither track can see what the nomination photography was constructed not to show.

Community context is similarly absent from both records, with the Social and Community History thematic category accounting for only 1% of NLP tokens across the corpus. Racial and ethnic diversity is invisible in both records, not because it was absent from the history of Lake Forest or its adjacent communities but because it was systematically excluded from the nomination documentation. This is precisely the condition that Roland Barthes's semiotic framework helps to name; the nomination form is not a transparent window onto historic significance but a text organized by a grammar of exclusion. What the nomination form selects, describes, and evaluates reflects a disciplinary consensus about architecture and history and what deserves permanence. The national data confirms that grammar. Architecture appears as an area of significance in 74% of all NR listings, while Ethnic Heritage-Black designations account for roughly 1% across six decades. The Lake Forest corpus intensified these same patterns, concentrating architectural language while leaving community and racial history largely absent.

This is where AI's analytical contribution becomes most legible. The value of applying computational methods to the NR corpus is not simply efficiency—it is the capacity to make the political starting points of architectural evaluation visible at a scale impossible through manual review. The frequency distributions in Table 1, the evaluative qualifier patterns, and the AI visual analysis outputs do not merely describe what nominations contain; they reveal the choices embedded in what nominations foreground and omit. The thematic dominance of built fabric and style over social history, labor, and community is not a neutral reflection of what buildings are—it is a record of what the preservation system has been organized to see and validate. The AI visual analysis adds a second register of evidence: when a model trained on cultural imagery independently reads the same properties through class,

scale, and boundary, and arrives at conclusions that mirror and amplify the nomination language's own social assumptions, it confirms that those assumptions are not incidental but structural. Meaning, as both the NLP and visual findings demonstrate, is not given in the building itself but constructed through the frameworks—disciplinary, institutional, and now computational—brought to bear on it.

Implications for Preservation Practice

AI visual analysis demonstrates reliable identification of architectural style, material character, landscape organization, and physical condition across all 18 properties analyzed. These outputs correspond directly to the descriptive requirements of the standard NR nomination form Sections 7 and 8 [27]. If AI tools can generate the physical description of a building from photographs with sufficient accuracy, the labor-intensive survey work that currently represents one of the primary resource barriers to nomination preparation could be substantially expedited, partially reducing the comparative advantage currently enjoyed by well-resourced communities.

One especially hopeful advantage of this automation is the potential to redress omissions in historic recognition. The time preservation professionals save on description could be redirected toward the historically underrecognized sites, persons, and narratives the NR record has structurally bypassed: vernacular buildings, working-class neighborhoods, landscapes of Black and Indigenous history, immigrant community environments, and women's histories encoded in domestic and institutional space.

But there is a limit to this potential, which our study has made clear. AI visual analysis can identify absences in the photographic record but it cannot recover what was never photographed. It cannot supply the historical context, community memory, or significance arguments that constitute the interpretive core of a nomination. A model trained primarily on canonical architectural photography and existing landmark documentation will reliably identify Tudor Revival from a photograph. It will not know what to do with a two-flat, a shotgun house, a storefront church, or a community garden whose significance is inseparable from the social history of the people who used it. The existing NR corpus, which provides the richest available dataset of documented historic properties, encodes the system's biases directly into any model trained on it. Without deliberate corrective design, AI tools introduced into the nomination pipeline are as likely to replicate the system's inequities as to repair them.

A corrective design would involve training models on vernacular architectural typologies, non-monumental building scales, and the material heterogeneity characteristic of working-class building stock. It would require incorporating oral histories, community documentation, and first-voice accounts of historic meaning as training inputs alongside physical surveys. It would require developing semiotic frameworks tuned to the cultural associations and historical significance of underrepresented communities rather than to the formal legibility of high-style architecture. And it would require treating the nomination form itself as a design problem, asking whether an instrument optimized for describing high-style architecture is the right one for documenting a civil rights movement site, a labor hall, or a vernacular landscape shaped by generations of collective practice. The good news is that the AI tools this study employs are powerful precisely because they are flexible enough to be retrained and redirected. Whether that flexibility is activated in the service of equity or in the amplification of existing biases is a policy choice, not a technical inevitability.

It is worth being precise, however, about what AI can and cannot do in relation to the political dimensions of registration bias. The underrepresentation of racially and ethnically marginalized communities in the NR record is not primarily a documentation problem—it

is a resource and power problem. The nomination process is expensive, technically demanding, and dependent on institutional infrastructure that has historically been concentrated in communities with capital and professional networks. AI tools that reduce the documentation burden can lower one barrier, but they cannot change who controls the nomination pipeline, who funds it, or whose significance frameworks govern it. An AI model that can describe a vernacular building from photographs still requires that someone with authority and resources initiate and submit a nomination; it cannot substitute for the community relationships, political will, and sustained investment that equitable preservation practice requires. The contribution of AI to reversing registration bias is therefore partial and conditional: it can expand the pool of documentable properties and surface patterns of exclusion at the corpus scale, as this study demonstrates, but it operates within a political and institutional structure that remains unchanged unless directly addressed through policy reform, funding reallocation, and the intentional centering of underrepresented communities in preservation decision-making.

5. Limitations

Several methodological and conceptual limitations bear acknowledgment. First, the Lake Forest corpus is intentionally narrow: 18 nominations drawn from a single wealthy suburb, representing a highly coherent and elite slice of the NR record. The patterns identified here—architectural dominance, class legibility, the systematic absence of labor and social diversity—are argued to be structurally characteristic of the NR system as a whole, and the national and state-level data support that claim, but the intensity and consistency of those patterns in Lake Forest is partly a function of the corpus's homogeneity. How these methods perform on a more heterogeneous corpus—one that includes working-class neighborhoods, rural vernacular properties, or sites of underrepresented heritage—remains to be tested.

Second, the AI visual analysis is constrained by the photographic record it receives. Nomination photography was taken specifically to document architectural character for NPS review purposes; it is a curated selection that foregrounds formal qualities and physical condition. The model cannot see what was not photographed, and the systematic absences it identifies—labor, service infrastructure, community context—are absences in the documentation rather than analytical gaps in the model itself. Any AI visual analysis of NR nominations will inherit this constraint until the photographic conventions of the nomination process are themselves reformed.

Third, the LLM used for visual analysis (Claude Sonnet, claude-sonnet-4-20250514) is a general-purpose model trained on broad cultural corpora. Its architectural readings are shaped by training data that almost certainly overrepresents canonical, high-style, and Western architectural traditions—the same biases that characterize the NR corpus itself. The model's sophisticated semiotic readings and class-legibility analyses reflect genuine analytical capability, but that capability is not neutral: it is calibrated to the kinds of buildings and visual conventions that dominate the documentary record on which it was trained. The convergence between the model's readings and the nomination language's assumptions may partly reflect shared training-data bias rather than independent confirmation.

Fourth, the thematic classification of NLP tokens was developed inductively from the Lake Forest corpus and cross-referenced with NPS documentation standards; it has not been validated against an independent corpus or through inter-rater reliability testing. The category definitions reduce but do not eliminate interpretive ambiguity for terms that participate in multiple discursive registers, and replication using the same category set on a different corpus may require recalibration.

6. Conclusions

This study has demonstrated that AI methods—natural language processing and blind multimodal visual analysis—can be applied to the National Register corpus to surface structural patterns of meaning-making that are difficult to detect through manual review. Applied to eighteen nomination documents from Lake Forest, Illinois, and situated within a national dataset of 100,117 NR listings, the analysis reveals a preservation record organized around architectural form as the primary evidentiary register of significance, with social history, labor, landscape, and community context systematically underrepresented in both the textual and photographic record. The AI visual analysis amplifies and extends this finding by independently reading class position, social exclusion, and ideological framing across the same corpus, confirming that these are not accidental omissions but structural features of how the NR system constructs significance.

These findings carry practical implications for the use of AI in preservation practice. Computational tools can reduce the documentation burden of the nomination process and serve as audit instruments for identifying thematic gaps in existing listings. But their potential to support more equitable recognition is conditional on deliberate design choices: training data that includes vernacular, working-class, and non-Western building types; significance frameworks tuned to underrepresented heritage; and policy reforms that address the structural barriers—resource concentration, institutional gatekeeping—that AI cannot reach on its own. Without those choices, AI applied to preservation will most likely reproduce the inequities it encounters in the record it is trained on.

Several directions for future research follow from this work. The methods should be tested on corpora that include vernacular, working-class, and non-elite property types—industrial buildings, shotgun houses, storefront churches, immigrant neighborhood streetscapes—to assess whether the same tools can surface significance arguments for property types the NR system has historically undervalued. Alternative datasets, including oral histories, community archives, and first-voice documentation, should be incorporated as both training inputs and analytical corpora. Comparative application across different regions and national contexts would test generalizability and illuminate how preservation systems with different institutional histories construct significance differently. Finally, computational analysis used as an ongoing audit tool—tracking shifts in nomination language and thematic coverage as policy reforms take effect—represents a promising application of these methods to questions of accountability in heritage governance.

Author Contributions: Conceptualization, M.R. and E.T.; methodology, M.R.; software, M.R.; validation, M.R.; formal analysis, M.R.; investigation, M.R. and E.T.; resources, M.R. and E.T.; data curation, M.R.; writing—original draft preparation, M.R.; writing—review and editing, M.R. and E.T.; visualization, M.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: An example of data supporting reported results, including photographs and JSON output, is included in Appendix A. Python scripts and outputs across the Corpus are available on request of the Authors.

Acknowledgments: During the preparation of this study, the authors used Claude Sonnet (claude-sonnet-4-20250514, Anthropic) for the purposes of AI-based visual analysis of National Register of Historic Places nomination photographs, as described in Section 2.5 (AI Visual Analysis). All visual analysis outputs were reviewed and interpreted by the authors, and the authors take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

NR	National Register of Historic Places
NLP	Natural Language Processing
AI	Artificial Intelligence
NHPA	National Historic Preservation Act
LLM	Large Language Model

Appendix A

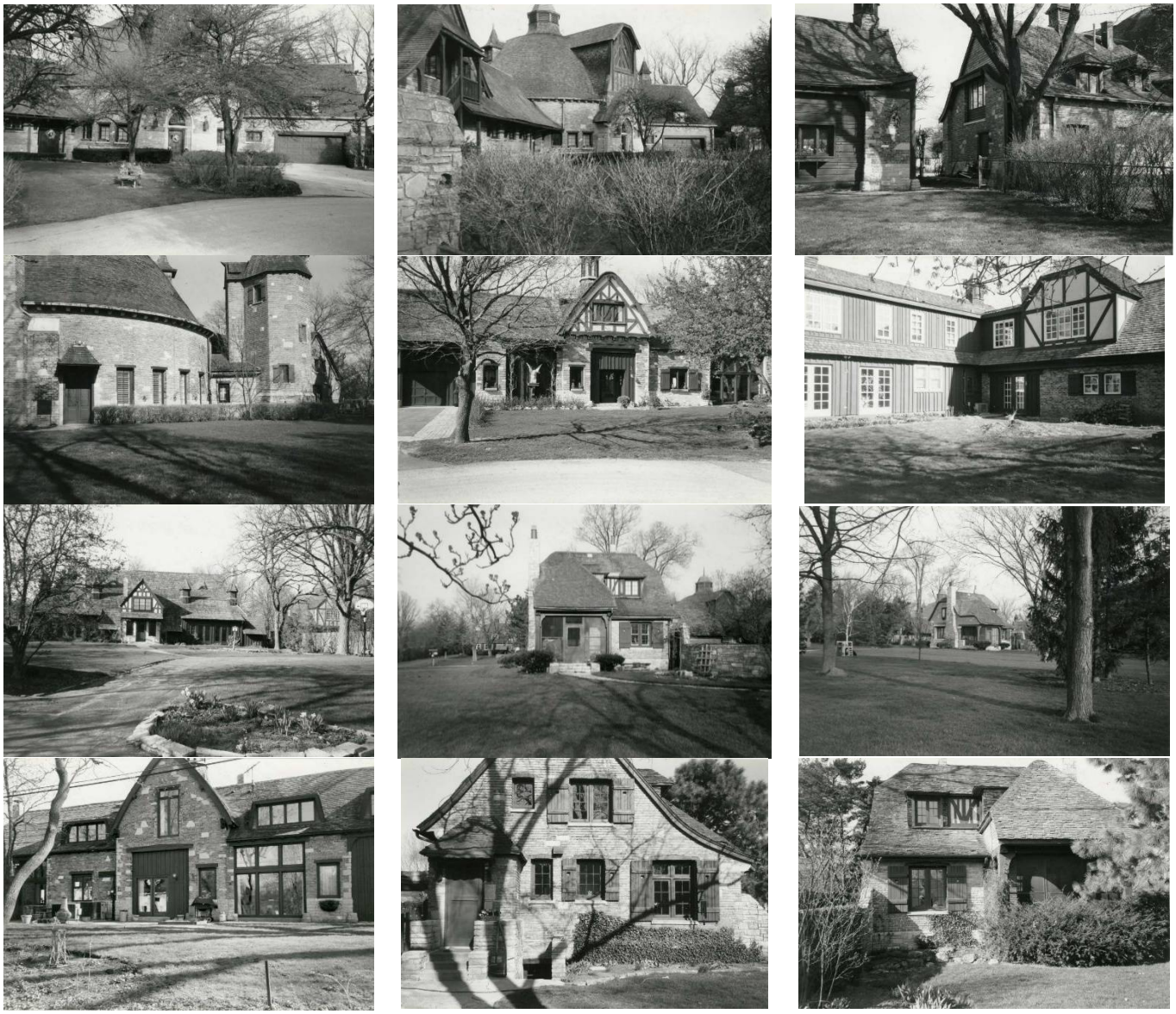


Figure A1. Selection of representative photos included in the National Register of Historic Places nomination report for Clifford Milton Farm in Lake Forest, IL, followed by the descriptive JSON produced through AI/LLM visual analysis using claude-sonnet-4-20250514, Anthropic [27].

{
“image_count_analyzed”: 24,
“architectural_style”: {

"primary_style": "Vernacular Farm Complex with Eclectic Revival Elements",
"secondary_styles": [
"Tudor Revival",
"Arts and Crafts",
"Colonial Revival",
"Prairie School influences"
],
"confidence": "high",
"style_evidence": "The complex displays a remarkable round barn with conical roof and cupola at its center, surrounded by residential and outbuilding structures that mix Tudor Revival half-timbering, Colonial Revival shuttered windows, and Arts and Crafts stonework. The main barn's circular plan with radiating wings represents innovative early 20th-century agricultural architecture, while the residential structures show deliberate stylistic eclecticism combining medieval revival elements with American vernacular traditions."
},
"materials_and_fabric": {
"primary_materials": [
"limestone",
"brick",
"wood shingles",
"timber framing"
],
"secondary_materials": [
"stucco",
"concrete",
"metal roofing",
"glass"
],
"notable_details": "The construction demonstrates high-quality craftsmanship with coursed limestone masonry, carefully detailed half-timber work with stucco infill, and cedar shingle roofing throughout. Windows feature multi-pane sash with functional shutters, and the barn complex shows sophisticated structural engineering in its circular form. The variety of materials suggests construction spanning multiple phases, with consistent attention to both functional and aesthetic considerations."
},
"landscape_and_site": {
"site_character": "formal",
"landscape_features": [
"manicured lawns",
"mature specimen trees",
"circular driveway",
"stone walls",
"formal gardens",
"connecting pathways"
],
"spatial_organization": "The buildings are arranged in a carefully planned composition around the central circular barn, with residential structures positioned to create intimate courtyards and vista points. The landscape treatment emphasizes the architectural ensemble through formal lawn areas and specimen tree placement, while stone walls

and pathways create clear circulation patterns that reinforce the hierarchical relationship between structures."

},

"social_and_ownership_cues": {

"apparent_scale": "large-estate",

"exclusivity_markers": [

"architectural variety",

"high-quality materials",

"extensive grounds",

"multiple specialized buildings",

"formal landscaping"

],

"service_or_functional_infrastructure": "visible",

"social_reading": "This property represents a gentleman's farm or model agricultural estate, where working farm functions are elevated through sophisticated architectural treatment. The variety of building types suggests a complex domestic economy with specialized spaces for different activities, while the architectural eclecticism and high-quality materials signal significant financial resources. The careful integration of utilitarian and residential functions suggests an owner who valued both agricultural innovation and aesthetic refinement, presenting farming as both productive enterprise and cultural statement."

},

"condition_and_integrity": {

"apparent_condition": "good",

"integrity_observations": "The structures appear well-maintained with intact rooflines, original fenestration patterns, and preserved architectural details. Some evidence of adaptive modifications is visible, particularly in window installations and mechanical systems, but the overall architectural character remains coherent. The landscape shows signs of ongoing stewardship with maintained lawn areas and mature tree preservation."

},

"semiotic_reading": {

"dominant_cultural_associations": [

"progressive agriculture",

"Arts and Crafts movement",

"gentleman farming",

"architectural innovation",

"rural romanticism"

],

"symbolic_registers": "This complex embodies early 20th-century ideals of scientific agriculture wedded to aesthetic sophistication, representing a vision of farming as both modern enterprise and cultural pursuit. The architectural eclecticism signals cosmopolitan taste and historical consciousness, while the integration of innovative barn design with domestic comforts suggests an owner who saw agricultural work as compatible with refined living. The property communicates values of stewardship, innovation, and cultural refinement, positioning rural life not as provincial retreat but as site of progressive experimentation. The careful architectural treatment of utilitarian buildings elevates agricultural work to the level of cultural expression, embodying Arts and Crafts ideals of integrated work and life."

"what_is_foregrounded": "The photographs emphasize the architectural sophistication and material quality of the structures, focusing on stylistic details, craftsmanship, and the picturesque arrangement of buildings within the landscape setting.",

"what_is_backgrounded_or_absent": "Working farm equipment, signs of active agricultural use, and evidence of the labor force required to maintain such an extensive operation are largely absent from the visual record, creating an idealized view of agricultural life divorced from its productive realities."

},

"cross_theme_summary": "This property represents a sophisticated early 20th-century model farm complex that transforms agricultural architecture into a statement of cultural refinement and progressive ideals. The remarkable circular barn at its center demonstrates innovative agricultural engineering while the surrounding eclectic residential and outbuilding ensemble reflects Arts and Crafts values of integrated living and working environments. The high-quality materials, formal landscape treatment, and architectural variety signal substantial resources invested in creating a gentleman's farm where productive agriculture serves as the foundation for aesthetic and social distinction. The visual record presents this as an idealized rural estate where modern scientific farming methods are housed within architecturally sophisticated structures, embodying early 20th-century optimism about the compatibility of technological progress with traditional rural values. The careful photographic documentation emphasizes the property's architectural achievement while obscuring the labor relations and economic structures that made such an elaborate agricultural operation possible, creating a romanticized vision of farming as cultural practice rather than economic necessity".

References

1. Zhang, F.; Salazar-Miranda, A.; Duarte, F.; Vale, L.; Hack, G.; Chen, M.; Liu, Y.; Batty, M.; Ratti, C. Urban Visual Intelligence: Studying Cities with Artificial Intelligence and Street-Level Imagery. *Ann. Am. Assoc. Geogr.* **2024**, *114*, 876–897. [CrossRef]
2. Fu, X. Natural Language Processing in Urban Planning: A Research Agenda. *J. Plan. Lit.* **2024**, *39*. [CrossRef]
3. Sanchez, T.W.; Brenman, M.; Ye, X. The Ethical Concerns of Artificial Intelligence in Urban Planning. *J. Am. Plan. Assoc.* **2025**, *91*, 294–307. [CrossRef]
4. Chen, L.; Lu, Y.; Ye, Y.; Xiao, Y.; Yang, L. Examining the association between the built environment and pedestrian volume using street view images. *Cities* **2022**, *127*, 103734. [CrossRef]
5. Max, P.; Miller, M.R. Introduction. In *Bending the Future: 50 Ideas for the Next 50 Years of Historic Preservation in the United States*; University of Massachusetts Press: Amherst, MA, USA, 2016; pp. 1–43.
6. Avrami, E. *Preservation and Social Inclusion*; Columbia Books on Architecture and the City: New York, NY, USA, 2020.
7. Avrami, E. (Ed.) *Preservation and the New Data Landscape*; Columbia University Press: New York, NY, USA, 2019.
8. Bluestone, D. *Buildings, Landscapes, and Memory: Case Studies in Historic Preservation*; W.W. Norton: New York, NY, USA, 2011.
9. Page, M.; Mason, R. (Eds.) *Giving Preservation a History: Histories of Historic Preservation in the United States*; Routledge: New York, NY, USA, 2004.
10. Barile, K.S. Race, the National Register, and Cultural Resource Management: Creating an Historic Context for Postbellum Sites. *Hist. Archaeol.* **2004**, *38*, 90–100. [CrossRef]
11. Kaufman, N. Historic Places and the Diversity Deficit in Heritage Conservation. *CRM J. Herit. Steward.* **2004**, *1*, 68–85. Available online: https://npshistory.com/newsletters/crm/journal-v1n2.pdf (accessed on 16 April 2026).
12. Mekonnen, L. Integrity and Racial Inequity: Case Studies from the Field. National Trust for Historic Preservaion. Available online: https://savingplaces.org/stories/integrity-and-racial-inequity-case-studies (accessed on 7 September 2021).
13. Barthes, R. Semiology and the Urban. In *The City and the Sign: An Introduction to Urban Semiotics*; Gottdiener, M., Lagopoulos, A.P., Eds.; Columbia University Press: New York, NY, USA, 1988; pp. 87–98.
14. Hayden, D. *The Power of Place: Urban Landscapes as Public History*; MIT Press: Cambridge, MA, USA, 1995.
15. Smith, L. *Uses of Heritage*; Routledge: London, UK, 2006.
16. Zukin, S. *Naked City: The Death and Life of Authentic Urban Places*; Oxford University Press: New York, NY, USA, 2010.
17. Brinkley, C.; Stahmer, C. What Is in a Plan? Using Natural Language Processing to Read 461 California City General Plans. *J. Plan. Educ. Res.* **2021**, *44*, 632–648. [CrossRef] [PubMed]
18. Mleczko, M.; Desmond, M. Using Natural Language Processing to Construct a National Zoning and Land Use Database. *Urban Stud.* **2023**, *60*, 2564–2584. [CrossRef] [PubMed]
19. Salazar-Miranda, A.; Talen, E. An AI-Based Analysis of Zoning Reforms in US Cities. *Nat. Cities* **2025**, *2*, 304–315. [CrossRef]

20. Fu, X.; Sanchez, T.W.; Li, C.; Reu Junqueira, J. Deciphering Public Voices in the Digital Era: Benchmarking ChatGPT for Analyzing Citizen Feedback in Hamilton, New Zealand. *J. Am. Plan. Assoc.* **2024**, *90*, 728–741. [CrossRef]
21. Nie, X.; Liu, H.; Han, M.; Guan, C. Evaluating Citizen Participation in Local Public Meetings: Exploring a Large Language Model Approach Using Transcripts from the United States. *J. Plan. Educ. Res.* **2025**, *in press*. [CrossRef]
22. Peng, Z.-R.; Lu, K.-F.; Liu, Y.; Zhai, W. The Pathway of Urban Planning AI: From Planning Support to Plan-Making. *J. Plan. Educ. Res.* **2023**, *44*, 2263–2279. [CrossRef]
23. Ibrahim, M.R.; Haworth, J.; Cheng, T. Understanding Cities with Machine Eyes: A Review of Deep Computer Vision in Urban Analytics. *Cities* **2020**, *96*, 102481. [CrossRef]
24. Freitas, F.; Berreth, T.; Chen, Y.-C.; Jhala, A. Characterizing the Perception of Urban Spaces from Visual Analytics of Street-Level Imagery. *AI Soc.* **2022**, *38*, 1361–1371. [CrossRef] [PubMed]
25. Giannoulaki, M.; Gruen, A. Computer Vision-Based Analysis of Buildings and Built Environments: A Systematic Review of Current Approaches. *ACM Comput. Surv.* **2023**, *55*, 14. [CrossRef]
26. Moradi, L.; Biloria, N. Implications of Artificial Intelligence for Assessing the Built Environment. *J. Urban Technol.* **2025**, *32*, 163–191. [CrossRef]
27. National Park Service, U.S. Department of the Interior. National Register Nomination Form: Leonard, Clifford Milton, Farm, National Register of Historic Places Collection, Lake Forest, IL. Available online: https://npgallery.nps.gov/NRHP (accessed on 16 April 2026).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.