

Geospatial Experience-Oriented Notation (GEON):

A Semantic Format for LLM-Native Spatial Intelligence

James Williams
University of Nottingham
Nottingham, United Kingdom
james.williams4@nottingham.ac.uk

February 2026

Abstract

Existing geospatial data formats such as GeoJSON, Well-Known Text (WKT), and CityGML, are optimised for geometric computation and rendering. While effective for Geographic Information Systems (GIS), these approaches present limitations when used with Large Language Models (LLMs). For example, coordinate arrays carry no inherent semantic meaning, spatial relationships require computational geometry to extract, and the human experience of place is usually absent. This manuscript introduces foundational work on Geospatial Experience-Oriented Notation (GEON), a text-based format that bridges machine-optimised geospatial data and human-readable spatial descriptors. GEON encodes identity, geometry, purpose, experiential qualities, spatial relationships, temporal patterns, and data provenance in a readable and structured syntax designed for human comprehension and LLM reasoning. This manuscript presents the initial specification, reference implementations in Python, Rust, and JavaScript, and an empirical evaluation demonstrating how GEON achieves 20% fewer tokens than equivalent GeoJSON files, while encoding 31% more semantic facts per token. This manuscript explores the implementation and how LLMs reason about place-making, urban design interventions, and spatial intelligence tasks that existing formats struggle to support.

Keywords: geospatial data, large language models, spatial intelligence, urban computing, semantic representation, place-based computing

1 Introduction

Recent articles have explored the convergence of Large Language Models (LLMs) and geospatial data science, seeking to understand how natural language can support complex spatial data science tasks. This presents a novel opportunity to examine the ability of AI systems to reason about places, their qualities, and their relationships, using geospatial science [1, 2]. However, this convergence is hampered by a mismatch in data formatting, where GIS such as QGIS and ArcGIS rely on spatial data formats such as GeoJSON, WKT, Shapefiles, and CityGML encode spatial data, these encodings rarely match the design of LLMs for natural language, processing, and performing semantic reasoning. While LLMs can process and extract geospatial data from these formats (for example, the widespread use of GeoJSON), this work explores if a more semantic approach would benefit the accuracy of systems and the effectiveness of tokens in structuring place-based experiences.

For example, consider a traditional GeoJSON representation of a public square in Birmingham, United Kingdom:

Listing 1: An example traditional GeoJSON approach to defining the location of Victoria Square in Birmingham.

```
{
  "type": "Feature",
  "geometry": {
    "type": "Point",
    "coordinates": [-1.8984, 52.4774]
  },
  "properties": {
    "name": "Victoria Square"
  }
}
```

This representation tells an LLM *where* Victoria Square is, but nothing about what it is, how it feels, what it connects to, or how people experience it. This work is proposed to bridge this gap, supporting natural language and semantically rich descriptions of places. The semantic gap between coordinates and meaningful spatial understanding represents a potential bottleneck in human-AI spatial applications. With the following four challenges being the purpose of this work: (1) Semantic Opacity: coordinate arrays provide no meaning about what a place is or how it functions, (2) Relational Isolation: direct relations between features is implicit (3) Absence of Experience: the representation of the human experience is left to the LLM to determine, and (4) the Fragmentation of Context: the data provided to the LLM is presented in different schemas and systems.

This manuscript presents GEON (Geospatial Experience-Oriented Notation), which attempts to address these limitations by providing a format that embeds semantic meaning alongside geometric data, making spatial relationships more explicit, encoding experiential qualities using standardised scales, and integrates temporal and provenance information into a single document. Instead of relying on the knowledge base of data provided with LLMs, the proposed approach provides context (e.g., purpose, experience) that can support language-based reasoning and natural language presentation with the intention of limiting hallucinations. In comparison to the GeoJSON above, the proposed GEON structure is:

Listing 2: An example GEON representation of defining Victoria Square in Birmingham, United Kingdom.

```
PLACE: Victoria Square, Birmingham
TYPE: public_space
LOCATION: 52.4791, -1.9024
PURPOSE:
  - civic gathering
  - events and festivals
EXPERIENCE:
  openness: high
  sense_of_safety: high (daytime), moderate (nighttime)
  activity_density: moderate
ADJACENCIES:
  - Council House (west)
  - Birmingham Museum & Art Gallery (north)
```

This paper makes the following contributions: (1) a format specification for GEON including a text-based structured notation for encoding places; (2) controlled vocabularies for place types (10 categories), qualities (13 dimensions), and purposes (8 categories, 29 activities); (3) reference implementations in Python, Rust and JavaScript; and (4) an initial evaluation of GEON in comparison to existing data formats.

2 Background and Related Work

2.1 Geospatial Data Formats

GeoJSON [3] has become one of the dominant interchange formats for web-based geospatial applications. It encodes geometric primitives (Point, LineString, Polygon, MultiPolygon) as nested JSON arrays, with an unstructured `properties` object used to store associated metadata. While lightweight and widely supported, GeoJSON’s flat properties collection provides no semantic schema, resulting in the meaning of properties being entirely application-dependent. Well-Known Text (WKT) [4] geometry text markup language provides a compact text encoding for geometries but carries no metadata at all. Semantic information must be maintained in a separate system (typically through an ID in a relational database column alongside the WKT geometry column). Meanwhile, CityGML [5] offers a detailed XML-based schema for 3D urban models with multiple levels of detail (LOD). While semantically comprehensive, its verbose XML structure and complex schema make it less practical for LLM processing, as a single building can require hundreds of XML elements.

Geospatial data formats are also influenced through data sources, including OpenStreetMap (OSM) [6], which uses a flexible key-value tagging system that can encode associated metadata. However, OSM tags are inherently flat (no hierarchy), inconsistently applied across contributors, and do not support experiential or temporal data as a standard. Overture Maps [7] also provides a curated, schema-normalised layer across multiple sources, but its schema remains geometry-focused and does not address experiential encoding or more experiential values.

2.2 Place Theory

The distinction between *space* (geometric extent) and *place* (meaningful location with identity, history, and human significance) has deep roots in geography. Tuan [8] explores both space and place and expresses how space becomes place through experiences. Relph [9] defines place as the centre of deep human, emotional, and existential meaning, rooted in experiences rather than just physical location. Norberg-Schulz [10] introduced the concept of *Genius loci* as the ‘spirit of the place,’ referring to the unique, distinctive atmosphere, character, and essence of a specific location.

In urban design, Lynch [11] demonstrates that people navigate cities through the use of cognitive maps built from paths, edges, districts, nodes, and landmarks; categories that are semantic, not geometric. Hillier and Hanson [12] propose Space Syntax theory to analyse how spatial configuration influences movement and social encounter, rather than being a neutral backdrop. Gehl [13] explored how the physical qualities of spaces, such as: enclosure, scale, permeability, shape human activity patterns.

These theoretical frameworks, in addition to literature on spatial information [14], the place-based alignment to spatial data and the intersection between digital understandings of space and place [15, 16], inform GEON’s design. If places are constituted by experience, relationships, and meaning rather than spatial geometries, then a format aspiring to represent places must encode these qualities directly.

2.3 LLMs and Geospatial Reasoning

Recent work has begun exploring the intersection of LLMs and geospatial intelligence. Mai et al. [17] demonstrated that embedding geographic knowledge improves spatial question-answering. Manvi et al. [1] showed that LLMs can extract geospatial knowledge when prompted with appropriate context such as additional data from OSM. Hu [18] found that knowledge-guided prompting improves location extraction from disaster reports posted on social media. Roberts

et al. [2] evaluated GPT-4’s geographic reasoning capabilities, finding general positives in the models understanding of the world, while suffering from hallucinations.

A common thread in this literature is that LLM performance on spatial tasks depends critically on how geographic information is represented in the prompt. Additionally, whether contextual data is passed to the model, or if the contextual data presented in the prompt can be effectively parsed by an LLM. Dense coordinate arrays may waste context-window tokens on syntactically necessary but semantically empty structural characters (braces, brackets, commas), while failing to provide the relational and experiential context that supports reasoning.

2.4 TOON and Semantic Data Formats

GEON is inspired by TOON (Token-Oriented Object Notation) [19], a format designed to make structured data more accessible to LLMs by prioritising human readability and semantic clarity over machine-parsing efficiency. TOON demonstrates how indentation-based formats can be simultaneously human-readable and LLM-processable while consuming fewer tokens than equivalents. GEON attempts to extend this principle into the geospatial domain, adding domain-specific controlled vocabularies, spatial relationship primitives, and experiential encoding.

3 The GEON Format

3.1 Design Principles

GEON is guided by seven design principles:

1. **Semantic richness over geometric precision.** GEON treats coordinate accuracy as adequate rather than primary. A few metres of positional uncertainty is acceptable when accompanied by rich semantic descriptions.
2. **Human readability.** The format should be comprehensible without GIS training. An urban planner, architect, or community organiser should be able to read and write GEON documents.
3. **LLM accessibility.** The structure should support natural language reasoning: field names should be self-explanatory, values should use natural language where appropriate, and syntactic overhead should be minimised.
4. **Interoperability.** Bidirectional conversion with GeoJSON and other standard formats must be straightforward, ensuring GEON complements as opposed to replaces existing infrastructure.
5. **Hierarchical flexibility.** Places contain sub-places. For example: A building contains floors; a market contains stalls.
6. **Temporal awareness.** Places change over time. Footfall varies by hour, character shifts across seasons, and identity evolves across decades.
7. **Experiential grounding.** How a place is perceived, its openness, noise, safety, pace; is as important as its coordinates. These qualities must be encoded using standardised, comparable scales.

3.2 Syntax

GEON documents are UTF-8 encoded text files (extension `.geon`). The syntax uses indentation-based hierarchy (2 spaces per level), key-value pairs separated by colons, and dash-prefixed list items:

Listing 3: GEON syntax elements and structure.

```

KEY: value (* key-value pair *)
SECTION: (* section header *)
  sub_key: sub_value (* nested key-value *)
LIST_FIELD: (* list field *)
  - item one (* list item *)
  - item two (* list item *)
CONTAINS: (* nested places *)
  - PLACE: Child Place (* embedded GEON block *)
    TYPE: building
    LOCATION: 52.48, -1.90

```

3.3 Field Categories

GEON defines over 30 fields organised into six categories. Three fields are required (PLACE, TYPE, LOCATION); five are recommended (PURPOSE, EXPERIENCE, ADJACENCIES, CONNECTIVITY, SOURCE); the remainder are optional. Table 1 summarises the core fields.

Table 1: GEON core field categories and their fields.

Category	Field	Status	Type
Identity	PLACE	Required	String
	TYPE	Required	Controlled vocab
	ID	Optional	URI / UUID
Geometry	LOCATION	Required	Coordinate (lat, lon)
	BOUNDARY	Optional	Coordinate list
	EXTENT	Optional	Bounding box
	ELEVATION	Optional	Value + unit
	AREA	Optional	Value + unit
Semantic	PURPOSE	Recommended	List of activities
	EXPERIENCE	Recommended	Key-value scales
	CHARACTER	Optional	Descriptive list
Relational	ADJACENCIES	Recommended	Named neighbours
	CONNECTIVITY	Recommended	Network access
	CONTAINS	Optional	Nested GEON blocks
	PART_OF	Optional	Parent reference
	VIEWSHEDS	Optional	Visible landmarks
Temporal	TEMPORAL	Optional	Time-varying patterns
	LIFESPAN	Optional	Temporal extent
Provenance	SOURCE	Recommended	Data origins
	CONFIDENCE	Optional	Quality indicators
	UPDATED	Optional	ISO 8601 timestamp

Additionally, seven *extended domain fields* support specialised applications: BUILT_FORM (e.g., building heights, materials), ECOLOGY (e.g., habitat type, canopy cover), INFRASTRUCTURE (e.g., utilities, digital connectivity), DEMOGRAPHICS (e.g., population, diversity), ECONOMY (e.g., rents, business types), VISUAL (e.g., image references, landmark visibility), and VERTICAL_PROFILE (e.g., floor-by-floor occupancy).

3.4 Controlled Vocabularies

To ensure consistency and comparability across GEON documents, the specification defines controlled vocabularies for three key dimensions:

Table 2: GEON experiential quality dimensions and their ordinal scales.

Domain	Dimension	Scale
Spatial	openness	very_low → very_high
	enclosure	very_low → very_high
	permeability	very_low → very_high
	legibility	very_low → very_high
Sensory	noise_level	very_quiet → very_loud
	visual_complexity	very_simple → very_complex
	air_quality	very_poor → very_good
Social	activity_density	deserted → crowded
	social_diversity	very_low → very_high
	sense_of_safety	very_unsafe → very_safe
	territoriality	very_private → very_public
Temporal	pace	very_slow → very_fast
	temporal_stability	very_transient → very_permanent

3.4.1 Place Types

Ten primary categories classify the TYPE field: **public_space** (squares, plazas, parks), **street** (roads, lanes, boulevards), **building** (structures with roof and walls), **transport_hub** (stations, terminals), **infrastructure** (utilities, facilities), **natural_feature** (rivers, hills, woodlands), **district** (neighbourhoods, quarters), **landmark** (monuments, notable structures), **threshold** (gateways, bridges, boundaries), and **hybrid** (mixed-use entities).

3.4.2 Experiential Qualities

The EXPERIENCE field encodes 13 phenomenological dimensions, each with a standardised ordinal scale (Table 2). These dimensions draw on environmental psychology research into how physical qualities of space influence human perception and behaviour [11, 13, 20].

These scales are deliberately qualitative. While quantitative measurements (e.g., decibel levels for noise) are available for some dimensions, GEON targets the level of granularity at which humans and LLMs reason about place quality: “loud” is more actionable for an urban designer than “72 dB”.

3.4.3 Purpose Categories

Eight functional categories organise the PURPOSE field, containing 29 specific activity types: Economic (commerce, retail, services, production, agriculture), Civic (governance, community, education, health, emergency), Social (gathering, celebration, protest, exchange, encounter), Cultural (arts, heritage, performance, exhibition, worship), Recreational (play, sport, leisure, contemplation, exercise), Residential (dwelling, sleeping, domesticity), Circulation (movement, waiting, transition, parking), and Ecological (habitat, biodiversity, environmental services).

3.5 Hierarchical Structure

GEON supports hierarchical place descriptions through the CONTAINS field, which embeds complete GEON blocks as sub-places:

Listing 4: Hierarchical nesting via CONTAINS.

```
PLACE: Grand Central Birmingham
TYPE: building
```

```

LOCATION: 52.4774, -1.8984
CONTAINS:
- PLACE: Specific Store
  TYPE: building
  LOCATION: 52.4776, -1.8983
  PURPOSE:
    - retail (department store)
  AREA: 25000 sqm
- PLACE: Shopping Centre
  TYPE: building
  LOCATION: 52.4773, -1.8985
  PURPOSE:
    - retail (mixed)
    - food and beverage

```

This enables multi-scale modelling: a district contains streets; streets contain buildings; buildings contain floors or functional zones. The reciprocal PART_OF field provides upward references without requiring nesting.

3.6 Relational Encoding

Unlike GeoJSON, where spatial relationships must be computed from geometry, GEON makes relationships explicit through specified fields:

ADJACENCIES: Named neighbours with optional distance and direction (“St Martin’s Church (100m west)”). CONNECTIVITY: Structured network access such as pedestrian entries, vehicular access, cycling routes, public transport connections. VIEWSHEDS: Visible landmarks with direction and distance, encoding perceptual (not just geometric) relationships [11], and CONTAINS / PART_OF: hierarchical containment as discussed above.

This explicit encoding enables LLMs to reason about spatial relationships without performing geometric computations, supporting a more focused natural-language-based spatial reasoning.

4 Reference Implementation

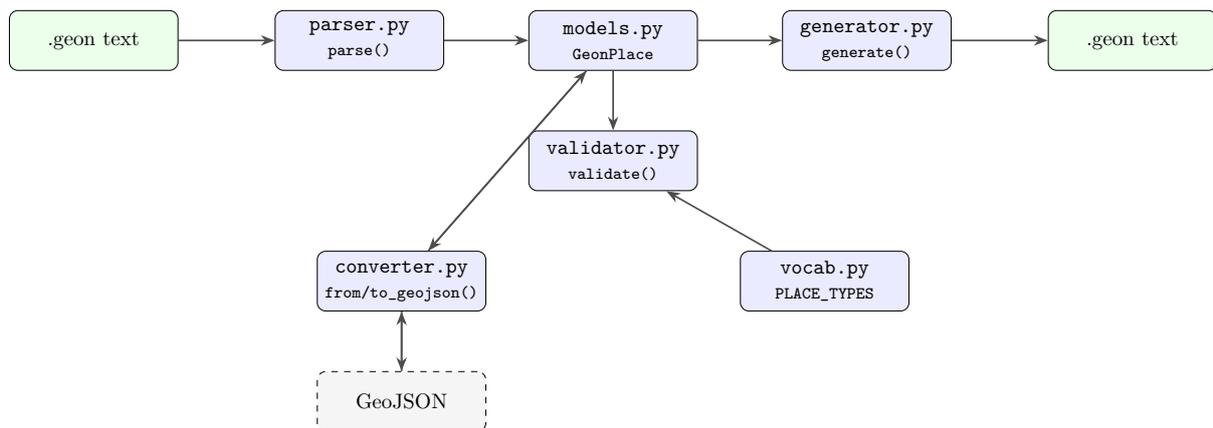


Figure 1: Architecture of the GEON Python reference implementation. The central `GeonPlace` data model connects parsing, generation, validation, and conversion modules.

A reference implementation is provided in three languages: Python ¹ (for data science and scripting), Rust ² (for high-performance systems), and JavaScript/TypeScript ³ (for web applications). All implementations share the same data model and support the full GEON specification. For the purposes of this manuscript, the Python implementation is described in detail as presented in Figure 1.

4.1 Data Model

The core data structure is `GeonPlace`, a Python dataclass with typed fields mirroring the specification (Section 3.3). Supporting types include `Coordinate` (WGS84 latitude/longitude pair) and `Extent` (bounding box).

4.2 Parser

The parser transforms GEON text into `GeonPlace` objects via a multi-stage pipeline. The process begins with tokenisation, where lines are split into `(indent_level, content)` tuples while removing empty lines. A recursive descent block parser then processes key-value pairs and nested structures, using indentation to determine hierarchical boundaries. For each section encountered, children are classified as either list items (identified by dash prefixes) or nested key-value blocks, enabling the distinguishing between different structural elements within the document.

When `contains` blocks are detected during parsing, the system extracts embedded GEON blocks, re-bases their indentation to zero, and recursively parses them as independent documents. This indentation re-basing demonstrates that nested places within `contains` may appear at arbitrary indentation depths, but must be parsed identically to top-level places. The parser detects the actual field indentation of nested content, subtracts this base offset, and feeds the normalised text back through the recursive parser. Finally, raw string values undergo type coercion, converting coordinates to `Coordinate` objects.

4.3 Generator

The generator produces deterministic GEON text from `GeonPlace` objects. Fields are emitted in a fixed order matching the specification sections (identity, geometry, semantic, relational, temporal, provenance, extended), ensuring consistent output regardless of internal object ordering. Indentation depth is tracked as a parameter, incremented for nested structures. The generator uses round-trip idempotence: `parse(generate(place)) = place` for all valid inputs.

4.4 Validator

The validator checks GEON documents against the specification with three severity levels: (1) **ERROR**: Blocks valid interpretation based on missing required fields, out-of-range coordinates; (2) **WARNING**: Recommends compliance based on unrecognised type values, unclosed boundary polygons, experiential values outside controlled vocabulary; and (3) **INFO**: Suggests completeness based on missing recommended fields (purpose, experience, source). Validation is recursive: nested places within `CONTAINS` are validated with hierarchical error path reporting.

¹<https://github.com/jwilliamsresearch/GEON/tree/main/geon-py>

²<https://github.com/jwilliamsresearch/GEON/tree/main/geon-rs>

³<https://github.com/jwilliamsresearch/GEON/tree/main/geon-js>

4.5 GeoJSON Converter

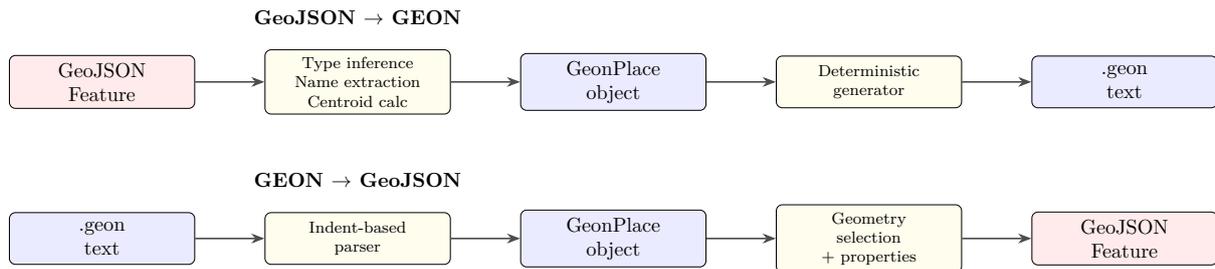


Figure 2: Bidirectional conversion pipeline between GeoJSON and GEON..

Bidirectional conversion between GEON and GeoJSON is supported through two complementary processes as demonstrated in Figure 2. When converting from GeoJSON to GEON, the converter extracts geometry (converting GeoJSON `[lon, lat]` to GEON `lat, lon` order), computes centroids for Polygon/MultiPolygon features, infers `TYPE` from OSM-style property tags using 88 mapping rules (e.g., `amenity=park` \rightarrow `public_space`), extracts `PURPOSE` from functional tags, and annotates `SOURCE` with conversion provenance.

In the reverse direction, converting from GEON to GeoJSON, the converter uses `BOUNDARY` (as Polygon) if available, falling back to `LOCATION` (as Point). Semantic fields are serialised into the GeoJSON `properties` object, while nested places from `CONTAINS` are flattened into a `FeatureCollection` to maintain compatibility with the GeoJSON specification.

5 Evaluation

GEON is evaluated along three dimensions: token efficiency, semantic density, and context-window scaling. All measurements use the `tiktoken` library [21] with the `cl100k_base` encoding (used by GPT-4 and compatible with Claude-family tokenisers) [22].

5.1 Token Efficiency

Table 3: Token comparison across formats for Birmingham Bullring Markets (16 semantic fields, 38 extractable facts). Tokeniser: `cl100k_base`.

Format	Tokens	Chars	Lines	Structured?
GEON	536	1,731	53	Yes
GeoJSON (pretty)	667	2,386	85	Yes
GeoJSON (compact)	475	1,781	1	Yes
WKT + metadata	436	1,546	28	Partial
Natural language	436	1,929	1	No

Using a representative and example place description, the Birmingham Bullring Markets, a richly-annotated public space with 16 distinct semantic fields. This place description is tested in five formats to measure the token consumption. The GEON representation includes identity, geometry, purpose, experience (7 dimensions), character, adjacencies, connectivity, temporal patterns, source, and confidence fields.

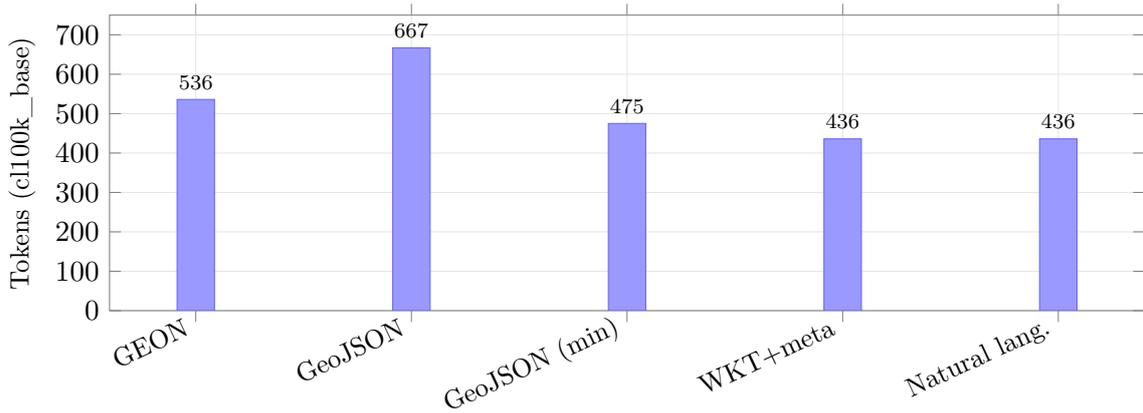


Figure 3: Token counts for the same spatial data (Birmingham Bullring Markets, 16 semantic fields) encoded in different formats. GEON uses 20% fewer tokens than pretty-printed GeoJSON. While having higher token usage than WKT and Natural Language.

Table 3 presents the results, with Figure 3 visualising these. GEON uses 536 tokens (20% fewer) compared to pretty-printed GeoJSON (667 tokens) for identical semantic content. Compact (minified) GeoJSON achieves 475 tokens but produces a single line document, defeating the purpose of human interpretation. Natural language prose (436 tokens) is more compact but sacrifices parse-able structure entirely.

While GEON is the most token-efficient structured format. Its advantage over GeoJSON comes from eliminating JSON’s syntactic overhead: braces, brackets, quotes around keys, trailing commas. GEON’s indentation-based structure carries equivalent hierarchical information at lower token usage. Additionally, GEON enables natural language variations of the data to be quickly generated, while also including relational data directly, which is not standard practice for GeoJSON and WKT.

5.2 Semantic Density

Table 4: Semantic density analysis for Victoria Square (based on 38 example source facts).

Format	Tokens	Facts	Facts/100 tok	Lossless?
GEON	431	38	8.8	Yes
GeoJSON	565	38	6.7	Yes
OSM tags	72	12	16.7	No
CSV row	68	9	13.2	No

Measuring token count alone would be insufficient, what also matters is how much meaning each token carries. Where semantic density is defined as the number of independently extractable facts per 100 tokens. Using a richer example (Victoria Square, Birmingham) with 38 annotated facts across 8 categories (identity, geometry, purpose, experience, character, relations, temporal, provenance), semantic density can be measured as presented in Table 4 and Figure 4.

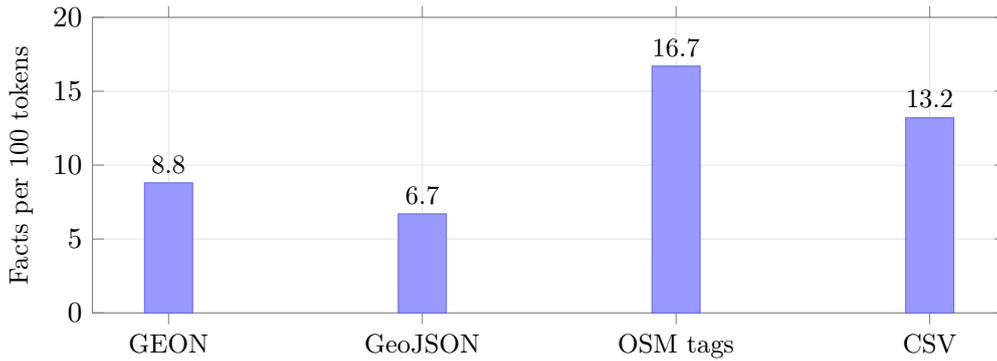


Figure 4: Semantic density (extractable facts per 100 tokens) across formats. GEON achieves the highest density from the fully-structured formats. OSM tags and CSV are denser but lossy as they cannot encode experiential qualities.

GEON achieves 8.8 facts per 100 tokens compared to GeoJSON’s 6.7, a 31% improvement in semantic density. OSM tags (16.7) and CSV (13.2) achieve higher raw density but are lossy, as they do not include the experiential semantic tags (e.g., experiential qualities, temporal patterns, provenance metadata, hierarchical relationships). Based on this demonstration, GEON is the densest format that preserves the full semantic richness of the defined place descriptions. However, alternative formats such as OSM tags and CSV retain significant higher facts per 100 tokens.

5.3 Projecting Context Window Scaling

The practical implication of token efficiency is that more places fit within an LLM’s context window, enabling richer spatial reasoning, place-based description formation, or comparative writing. Figure 5 projects how many complete place descriptions fit in various context windows.

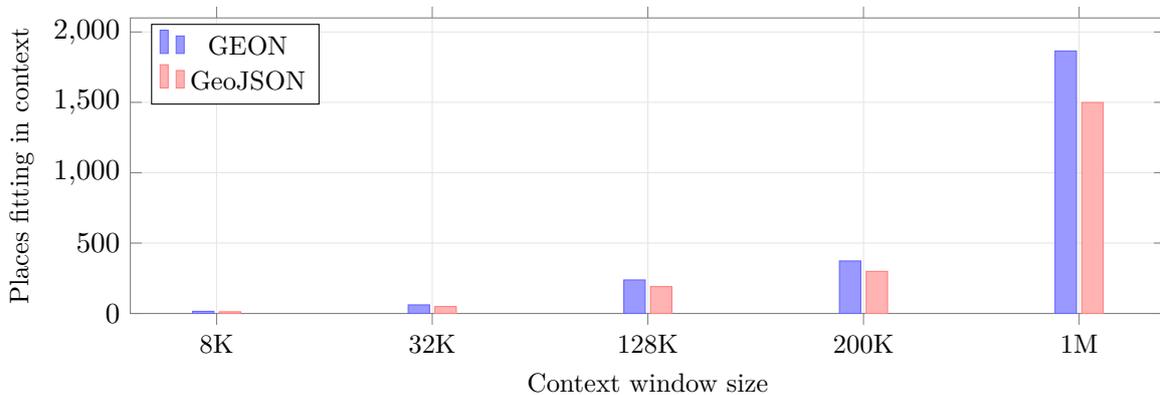


Figure 5: Projection of the number of complete place descriptions (536 tokens GEON, 667 tokens GeoJSON) fitting in different LLM context windows.

Based on this rudimentary demonstration, an LLM processing a neighbourhood’s worth of spatial data (e.g., 50 places) uses approximately 26,800 tokens with GEON versus 33,350 with GeoJSON; a saving of 6,550 tokens that can be allocated to system prompts, chain-of-thought reasoning, or additional spatial context. In the future, demonstration will be based on combinations of many place descriptions, to further examine context window scaling.

6 Use Cases

6.1 LLM-Assisted Place Analysis

An urban designer provides a GEON document describing Victoria Square and asks an LLM to assess it:

Listing 5: GEON input for place analysis.

```
PLACE: Victoria Square, Birmingham
TYPE: public_space
LOCATION: 52.4791, -1.9024
AREA: 7500 sqm
EXPERIENCE:
  openness: high
  enclosure: medium
  activity_density: moderate
  sense_of_safety: high (daytime), moderate (nighttime)
ADJACENCIES:
  - Council House (west)
  - Birmingham Museum & Art Gallery (north)
TEMPORAL:
  weekday_footfall: 1000-2000 people/hour
  events: Christmas market (Nov-Dec), summer festivals
```

To validate GEON’s effectiveness for spatial reasoning, we prompted two LLMs (NVIDIA Nemotron Nano 9B V2 and Grok 4.1 Fast with reasoning) through OpenRouter ⁴ with: “*Based on the following GEON place notation, please reason about the place and write a short 1 paragraph description.*”

Nemotron Nano 9B V2 ⁵ responded: “Victoria Square in Birmingham is a vibrant public space spanning 7,500 square meters, characterized by its high openness and medium enclosure, offering a balanced blend of accessible urban flow and defined boundaries. With moderate activity density and high daytime safety, it serves as a lively hub for locals and visitors, particularly during weekdays when footfall ranges between 1,000–2,000 people per hour. The square’s adjacency to the Council House to the west and the Birmingham Museum & Art Gallery to the north enhances its cultural and civic significance. Seasonal events like the Christmas market (November–December) and summer festivals further amplify its dynamic character, transitioning to moderate nighttime safety. This mix of historical landmarks, pedestrian activity, and event-driven energy positions Victoria Square as a key social and cultural anchor in the city.” [23, 24]

Grok 4.1 Fast ⁶ responded: “Victoria Square in Birmingham is a lively 7,500 sqm public space at the heart of the city center, defined by its high openness and medium enclosure, fostering an inviting atmosphere amid moderate activity density. Flanked by the imposing Council House to the west and the cultural Birmingham Museum & Art Gallery to the north, it draws steady weekday footfall of 1,000–2,000 people per hour, exuding a strong sense of safety during daylight hours that dips to moderate after dark. The square truly thrives during seasonal highlights like the bustling Christmas market in November–December and vibrant summer festivals, transforming it into a communal heartbeat for locals and visitors alike.” [23, 25]

Both models synthesised experiential qualities, spatial relationships, and temporal patterns into coherent place narratives. Notably, both LLMs accurately extracted and integrated the temporal safety variation (high daytime, moderate nighttime), connected adjacencies to cultural

⁴<https://openrouter.ai/>

⁵<https://openrouter.ai/nvidia/nemotron-nano-9b-v2>

⁶<https://openrouter.ai/x-ai/grok-4.1-fast>

significance (Council House and Museum & Art Gallery), and interpreted the experiential scales (openness, enclosure, activity density) as meaningful place characteristics rather than abstract data fields. The Nemotron model foregrounded the square as a “social and cultural anchor”, while Grok highlighted its role as a “communal heartbeat”, both demonstrated consistent spatial reasoning grounded in semantic structure.

6.2 Additional Applications

GEON’s semantic richness supports diverse spatial intelligence tasks beyond urban design. In emergency response, suitably labelled ACCESSIBILITY and CONNECTIVITY fields could enable identification of evacuation routes and assembly points without geometric analysis. Tourism or active travel applications could use VIEWSHEDS and EXPERIENCE qualities to generate personalised itineraries matching visitor preferences (e.g., “quiet safe places” or “vibrant social areas”) [16].

7 Discussion

7.1 Design Trade-offs

GEON prioritises semantic richness and human interpretability over geometric precision and quantitative rigour. The format accepts general coordinate accuracy or bounding boxes rather than in-depth geometries, targeting use cases where understanding a place matters more than precisely targeting it. Similarly, experiential qualities use ordinal scales (“loud”, “moderate”) matching human perception granularity rather than continuous measurements (72 dB). Additionally, compound temporal values (high (daytime), moderate (nighttime)) attempt to preserve nuance while remaining interpretable to both LLM systems and end users. However, for GIS users specifically interested in a complex polygons such as OpenStreetMap features [6] or 3D building footprints [5], GEON would be less suitable.

GEON’s flexibility creates inherent variability, for example, some documents may only contain minimal features (3 required fields), while others may be fully implemented and complex (30+ fields). Future work in this regard may look to explore specific build toolkits, taking and combing data from multiple sources (e.g., [6, 7]) to form in-depth GEON documents. Built-in validation distinguishes required, recommended, and optional content with appropriate severity levels, but this will not automatically build the types of in-depth data needed for large scale LLM tasks. Section 5 does demonstrate the effectiveness of GEON in supporting natural language spatial reasoning over serving as a universal geospatial interchange format.

7.2 Relationship to Existing Standards

GEON does not seek to replace established formats but occupies a complementary niche in the spatial human-AI interactions and prompting. GeoJSON [3] excels at geometric interchange for rendering and spatial databases, while GEON targets semantic description for reasoning and design. CityGML [5] provides detailed 3D models for simulation and visualisation, while GEON provides experiential descriptions for qualitative and structured text analysis. Finally, OSM [6] tags offer crowdsourced functional metadata while GEON attempts to curate experiential information in a human readable format.

7.3 Limitations

GEON’s current state presents several important constraints. The experiential scales encode inherently subjective perceptions without established inter-rater reliability protocols, meaning observers may assign different values to the same place. This should be considered on a project-by-project base, using statistical methods to determine high or low rankings. The fixed vocabulary

lacks mechanisms for domain-specific customisation beyond predefined extended fields, limiting specialised applications. While the evaluation demonstrates how GEON enables coherent place-based reasoning in contemporary LLMs, comparative studies quantifying task performance improvements over GeoJSON remain future work.

8 Conclusions and Future Work

This manuscript presented Geospatial Experience-Oriented Notation, a demonstration format exploring how text-based semantic encoding might better support LLM-based spatial reasoning compared to traditional geometry-focused formats. The initial demonstration introduces four key design elements: (1) experiential encoding through 13 standardised phenomenological dimensions grounded in environmental psychology; (2) explicit relational structure via named spatial relationships (e.g., adjacencies, connectivity, viewsheds, containment) that reduce reliance on computational geometry; (3) token-efficient semantic density achieving 8.8 facts per 100 tokens (31% denser than GeoJSON); and (4) integrated provenance with source attribution and confidence levels embedded alongside spatial data.

8.1 Future Work

The evaluation demonstrated GEON’s potential advantages in token efficiency and semantic density, while the use case validation with contemporary LLMs (Nemotron Nano 9B V2 and Grok 4.1 Fast) showed that models can effectively synthesise GEON’s structured experiential and relational fields into coherent place narratives. However, limitations constrain GEON’s current applicability. Future work will now focus on implementing variations of GEON in applied projects and building a database of pre-constructed places. Additionally, validation is planned alongside traditional formats with LLMs to determine the changes in output quality, semantic accuracy, and task completion (for example, aligning with work benchmarking TOON [26]).

8.2 Availability

The GEON specification, reference implementations (Python, Rust, JavaScript), interactive web toolkit, and all evaluation data are available under open licences:

- Specification: Creative Commons Attribution 4.0 (CC BY 4.0)
- Code: MIT Licence
- Repository: <https://github.com/jwilliamsresearch/geon>
- Demonstrator: <https://jwilliams.science/GEON>

Acknowledgements

GEON is inspired by TOON (Token-Oriented Object Notation) and builds on decades of research in environmental psychology, space syntax, geographic information science, and human-computer interaction.

The author thanks colleagues at the University of Nottingham and Birmingham Newman University for discussions on spatial cognition and urban computing. This work was completed at Birmingham Newman University, and submitted at the University of Nottingham.

This article and the GEON library used Claude Sonnet 4.5 through <https://claude.ai/> to refine language use and support software library development. All final decisions and intellectual contributions are the author’s own.

References

- [1] R. Manvi, S. Khanna, G. Mai, M. Burke, D. Lobell, and S. Ermon, “Geollm: Extracting geospatial knowledge from large language models,” 2024. [Online]. Available: <https://arxiv.org/abs/2310.06213>
- [2] J. Roberts, T. Lüddecke, S. Das, K. Han, and S. Albanie, “Gpt4geo: How a language model sees the world’s geography,” 2023. [Online]. Available: <https://arxiv.org/abs/2306.00020>
- [3] H. Butler, M. Daly, A. Doyle, S. Gillies, T. Schaub, and S. Hagen, “The GeoJSON Format,” RFC 7946, Aug. 2016. [Online]. Available: <https://www.rfc-editor.org/info/rfc7946>
- [4] J. R. Herring, “OpenGIS Implementation Standard for Geographic Information – Simple Feature Access – Part 1: Common Architecture,” *Open Geospatial Consortium*, 2011, oGC.
- [5] G. Gröger, T. H. Kolbe, C. Nagel, and K.-H. Häfele, *OGC City Geography Markup Language (CityGML) Encoding Standard*. Open Geospatial Consortium, 2012. [Online]. Available: https://portal.opengeospatial.org/files/?artifact_id=47842
- [6] M. Haklay and P. Weber, “Openstreetmap: User-generated street maps,” *IEEE Pervasive Computing*, vol. 7, no. 4, pp. 12–18, 2008.
- [7] Overture Maps Foundation, “Overture Maps Introduction,” <https://docs.overturemaps.org/>, 2026, accessed: 2026-02-14.
- [8] Y.-f. Tuan, *Space and place : the perspective of experience*. Minneapolis :: University of Minnesota Press, 1977.
- [9] E. Relph, *Place and placelessness*, ser. Research in planning and design ; 1. London: Pion, 1976.
- [10] C. Norberg-Schulz, *Genius loci : towards a phenomenology of architecture*. New York: Rizzoli, 1980.
- [11] K. Lynch, *The image of the city*. London: MIT P, 1960.
- [12] B. Hillier and J. Hanson, *The social logic of space*. Cambridge: Cambridge University Press, 1984. [Online]. Available: <https://doi.org/10.1017/CBO9780511597237>
- [13] J. Gehl, *Life between buildings : using public space*. Washington, DC: Island Press, 2011.
- [14] F.-B. Mocnik, “Putting geographical information science in place – towards theories of platial information and platial information systems,” *Progress in Human Geography*, vol. 46, no. 3, pp. 798–828, 2022. [Online]. Available: <https://doi.org/10.1177/03091325221074023>
- [15] J. Williams and G. Priestnall, “Placecrafter: Curating urban functional regions through platial clustering of openstreetmap points of interest,” in *Proceedings of OSMScience 2025*, Manila, Philippines, 2025.
- [16] J. Williams, S. Cavazzi, J. Pinchin, A. Hazzard, G. Priestnall, and A. Ballatore, “Context for leisure walking routes: A vision for a spatial-platial approach,” in *Spatial Data Science Symposium 2022 Short Paper Proceedings*, Online, 2022. [Online]. Available: <https://doi.org/10.25436/E20W2J>
- [17] G. Mai, K. Janowicz, L. Cai, R. Zhu, B. Regalia, B. Yan, M. Shi, and N. Lao, “Se-kge: A location-aware knowledge graph embedding model for geographic question answering and spatial semantic lifting,” *Transactions in GIS*, vol. 24, no. 3, pp. 623–655, 2020. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1111/tgis.12629>

- [18] Y. Hu, G. Mai, C. Cundy, K. Choi, N. Lao, W. Liu, G. Lakhanpal, R. Z. Zhou, and K. Joseph, “Geo-knowledge-guided gpt models improve the extraction of location descriptions from disaster-related social media messages,” *International Journal of Geographical Information Science*, vol. 37, no. 11, pp. 2289–2318, 2023. [Online]. Available: <https://doi.org/10.1080/13658816.2023.2266495>
- [19] TOON Contributors, “TOON: Token-Oriented Object Notation,” <https://github.com/toon-format/toon>, 2026, accessed: 2026-02-14.
- [20] R. Ewing and S. Handy, “Measuring the unmeasurable: Urban design qualities related to walkability,” *Journal of Urban Design*, vol. 14, no. 1, pp. 65–84, 2009. [Online]. Available: <https://doi.org/10.1080/13574800802451155>
- [21] OpenAI, “tiktoken: A Fast BPE Tokeniser,” <https://github.com/openai/tiktoken>, 2026, accessed: 2026-02-14.
- [22] R. Sennrich, B. Haddow, and A. Birch, “Neural machine translation of rare words with subword units,” in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, vol. 1: Long Papers. Association for Computational Linguistics (ACL), Aug. 2016, pp. 1715–1725, 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016 ; Conference date: 07-08-2016 Through 12-08-2016. [Online]. Available: <https://mirror.aclweb.org/acl2016/>
- [23] OpenRouter, “Openrouter: A unified interface for llms,” 2026, accessed: 2026-02-14. [Online]. Available: <https://openrouter.ai/>
- [24] xAI, “Grok-4.1-fast,” 2026, accessed via OpenRouter: 2026-02-14. [Online]. Available: <https://openrouter.ai/x-ai/grok-4.1-fast>
- [25] NVIDIA, “Nemotron nano 9b v2,” 2026, accessed via OpenRouter: 2026-02-14. [Online]. Available: <https://openrouter.ai/nvidia/nemotron-nano-9b-v2>
- [26] E. Masciari, V. Moscato, E. V. Napolitano, G. M. Orlando, M. Perillo, and D. Russo, “Are llms ready for toon? benchmarking structural correctness-sustainability trade-offs in novel structured output formats,” 2026. [Online]. Available: <https://arxiv.org/abs/2601.12014>

A Complete GEON Example

The following listing shows a complete GEON document for Birmingham Bullring Markets, written as an example of the library by Claude Sonnet 4.5.

Listing 6: An example GEON document, generated by Claude Sonnet 4.5 to test token usage. Note: This is mock data

```
PLACE: Birmingham Bullring Markets
TYPE: public_space
ID: osgb:1000000347112034
LOCATION: 52.4777, -1.8933
BOUNDARY:
  - 52.4780, -1.8940
  - 52.4780, -1.8926
  - 52.4774, -1.8926
  - 52.4774, -1.8940
  - 52.4780, -1.8940
AREA: 4200 sqm
ELEVATION: 142m above sea level
PURPOSE:
```

- retail (fresh food, flowers, clothing)
- social gathering
- cultural heritage (market tradition since 1166)
- circulation (pedestrian link)

EXPERIENCE:

openness: medium
 enclosure: medium-high (surrounding buildings)
 activity_density: high (weekdays), very_high (weekends)
 noise_level: loud
 visual_complexity: very_high (stalls, signage, products)
 sense_of_safety: high (daytime), moderate (evening)
 social_diversity: very_high

CHARACTER:

- vibrant (energetic street market atmosphere)
- diverse (multicultural vendors and shoppers)
- authentic (working market, not tourist recreation)
- gritty (worn surfaces, informal trade)

ADJACENCIES:

- Bullring Shopping Centre (immediate south)
- St Martin's Church (100m west)
- Moor Street Station (200m east)
- Smithfield Market site (300m north)

CONNECTIVITY:

pedestrian_entries: 4 (north, south, east, west)
 vehicular_access: service only (05:00-11:00)
 cycling: Rea Valley Route (via Digbeth)

CONTAINS:

- **PLACE:** Outdoor Market
TYPE: public_space
LOCATION: 52.4777, -1.8935
PURPOSE: retail (fresh produce, flowers)
TEMPORAL:
 trading_days: Tue, Thu, Fri, Sat
 trading_hours: 09:00-17:00
- **PLACE:** Rag Market
TYPE: building
LOCATION: 52.4776, -1.8931
PURPOSE: retail (clothing, textiles, household)

PART_OF: Digbeth and Eastside

VIEWSHEDS:

- St Martin's Church spire (prominent, 100m west)
- Rotunda (visible, 200m southwest)
- Selfridges building (immediate south, iconic facade)

TEMPORAL:

weekday_footfall: 5000-8000 people/hour (Thursday peak)
 weekend_footfall: 8000-12000 people/hour (Saturday peak)
 seasonal_variation: +30% December, -20% January

BUILT_FORM:

market_hall_height: 2 stories

stall_configuration: modular (2m x 2m typical)
canopy: metal frame with fabric (outdoor market)
materials: brick (hall), steel and fabric (outdoor)

DEMOGRAPHICS:

vendor_count: ~120 (variable)
visitor_demographics: diverse (age, ethnicity, socioeconomic)
catchment: Birmingham city region + tourist visitors

ECONOMY:

stall_rents: 50-150 GBP per day (estimated)
employment: ~200 traders + support staff

SOURCE:

- Ordnance Survey MasterMap (2024-11)
- Birmingham City Council market data (2024)
- OpenStreetMap (2025-01)
- Field observation (2025-01-18, Saturday 11:00-13:00)
- Historical records (Birmingham Archives)

CONFIDENCE:

geometry: high (OS survey)
footfall: medium (council estimates)
experience_qualities: medium (single observation, winter)
economic_data: low (estimated from partial sources)

UPDATED: 2026-01-20T09:00:00Z