Evaluating Generative Engine Optimisation (GEO) for O-1A Immigration-Law Firms: A Semi-Synthetic, Reproducible and Ethically Grounded Comparative Study

Clarity Infra Research Group | London, UK | 8 October 2025 |

Abstract

This paper extends the Clarity Infra (2025) study on generative engine optimisation (GEO) by applying the framework to U.S. O-1A immigration-law firms. We contextualise the O-1A visa as a non-immigrant category reserved for individuals with *extraordinary ability* in science, education, business or athletics; applicants must demonstrate sustained national or international acclaim and provide extensive documentation [1]. Using a semi-synthetic dataset of 80 simulated law-firm webpages and 20 anonymised public excerpts from the USCIS policy manual and American Bar Association (ABA) ethics rules, we compare three embedding models (MiniLM-L6-v2, OpenAI's text-embedding-3-small and E5-base-v2). Each model's Top-3 retrieval accuracy is assessed under baseline SEO and GEO-optimised conditions. Consistent with Paper 1, GEO improves retrieval rates by approximately 47±5 percent, with bootstrapped 95 % confidence intervals and significant t-test results ($p < 0.05$). We construct a five-point Transparency & Accountability Index derived from NIST AI 600-1, OECD AI principles and ABA Rules 7.1–7.3, 1.6 and 5.3 [2] [3]. We find a positive correlation between ethical compliance and visibility gains: pages that explicitly state data provenance, attorney responsibility and AI-usage disclosures achieve higher retrieval scores. The paper concludes with regulatory implications, emphasising transparency, bias mitigation and lawyer supervision of AI. All materials—including synthetic datasets, analysis scripts and JSON-LD metadata—are available via GitHub and Zenodo to support reproducibility.

Keywords

Generative engine optimisation; O-1A visa; legal technology; schema.org; AI search; ethical AI; structured data.

1 Introduction

Retrieval-augmented generation (RAG) has emerged as a powerful paradigm for large language models (LLMs), reducing hallucinations by grounding generation in relevant documents [4]. In the legal domain, however, RAG systems still hallucinate or misground information. A 2024 study by Stanford HAI found that purpose-built legal research systems still produced incorrect information more than 17 % of the time, while some systems hallucinated in over one-third of responses [5]. Nevertheless, generative AI is rapidly entering legal practice; nearly three-quarters of U.S. lawyers plan to use generative AI tools [6]. For immigration-law firms, discoverability within AI-powered search engines is essential, particularly for O-1A petitions for individuals of extraordinary ability [1]. Traditional search-engine optimisation (SEO) strategies are insufficient in this new landscape. BigDog ICT's guide warns that SEO "tricks" are losing relevance because generative models compile answers from multiple sources [7].

Instead, firms must structure content so that AI models can parse, cite and include it in their responses [8]. This study therefore evaluates how GEO—an integrated strategy of content structuring, schema optimisation and multi-source linking—improves the visibility of O-1A law-firm content in AI-powered search systems.

The O-1A visa is a highly specialised non-immigrant category for individuals with extraordinary ability in science, education, business or athletics. The USCIS policy manual defines "extraordinary ability" as being among "a small percentage who have risen to the very top of the field" and requires sustained national or international acclaim [1]. Petitioners must show that the beneficiary will continue work in the area of extraordinary ability and provide extensive evidence, such as awards, membership in organisations requiring outstanding achievements and published material about the beneficiary [9]. Since 2024, the U.S. government has sought to modernise O-1A policies to attract talent in critical technologies. A 2025 policy alert clarifies that corporations can petition for O-1 beneficiaries and emphasises examples of evidence for individuals working in emerging technologies [10]. An FOIA release notes that Executive Order 14110 (Safe, Secure and Trustworthy AI) directs USCIS to update guidance for O-1A categories in science, technology, engineering and mathematics [11]. These developments heighten competition among law firms and underscore the need for AI-centric marketing.

This paper answers the following research questions: (1) How does GEO improve AI search retrieval of O-1A law-firm content compared with baseline SEO? (2) Do embedding models differ in their responsiveness to GEO? (3) What is the relationship between ethical compliance—particularly transparency, confidentiality and supervision—and AI visibility? (4) How can semi-synthetic datasets support reproducible legal-tech research? Through a controlled simulation of 100 webpages and cross-model retrieval evaluation, we demonstrate that GEO yields statistically significant improvements in Top-3 retrieval accuracy across MiniLM, E5-base and OpenAI embeddings. We further demonstrate that pages scoring higher on our ethical index correlate with better retrieval outcomes.

2 Literature Review

2.1 Generative engine optimisation and schema markup

Generative engine optimisation is a novel approach to search that aligns content with the requirements of generative AI. Traditional SEO focuses on keyword density and backlink strategies. However, BigDog ICT emphasises that generative AI models compile answers from "across the web," rendering legacy SEO tactics obsolete; law firms must restructure websites so that AI models can cite them [7]. GEO involves structuring web pages with clear headings, authoritative content and accurate metadata, enabling generative models to parse and integrate the material [8]. The guide lists major AI answer engines—Google's AI Overview, Bing Copilot, ChatGPT and Anthropic Claude—and notes that visibility within these systems requires content that is structured, sourced and verifiable [12].

Schema markup underpins GEO by embedding structured data in web pages using vocabulary such as Schema.org. LaFleur Marketing explains that schema markup clarifies key details—e.g., services, locations and attorney credentials—so search engines can interpret them [13]. For law firms, schema markup improves both general search and AI search by enabling pages to be categorised correctly and to appear in relevant results [14]. Benefits include rich snippets (e.g., FAQs and star ratings), improved local SEO and enhanced user experience [15]. PaperStreet adds that to optimise for AI search, content should use hyper-focused queries matching user intent, establish topical authority through comprehensive coverage, and implement structured data for services and attorney profiles [16]. Technical SEO remains important; websites must be crawlable, with clear architecture, properly formatted metadata and alt text for images to ensure AI readability [17]. Collectively, these works suggest that GEO comprises three pillars: (1) content structuring and authority building, (2) schema and metadata optimisation, and (3) corpus linking and multi-source consensus.

2.2 Retrieval-augmented generation and embedding models

RAG combines retrieval and generation to ground LLM outputs. The RAG evaluation survey by Yu et al. (2024) emphasises that evaluating RAG is challenging because of its hybrid structure and reliance on dynamic external knowledge sources [4]. The authors propose an analysis framework (RGAR) that examines retrieval accuracy, generation quality and additional requirements such as faithfulness [18]. They highlight metrics like relevance, accuracy and faithfulness, and emphasise the need for more comprehensive evaluation benchmarks [18]. In law, RAG is touted as a solution to hallucination, but evidence is mixed. Stanford HAI's study on legal research tools shows that retrieval-augmented systems reduce errors compared with general LLMs but still hallucinate frequently [5]. The article notes that providers often claim "hallucination-free" outputs without transparent evidence [19].

Embedding models translate text into vectors for dense retrieval. Mono Software's evaluation on the SQuAD dataset shows that the e5-large-v2 model achieved 89.5 % Top-3 accuracy, e5-base-v2 achieved 88.2 %, and MiniLM-L12-v2 achieved 73.7 % [20]. This suggests that E5-family models outperform MiniLM but require more computational resources. TigerData compared OpenAI's embeddings with open-source models; OpenAI's large model achieved 80.5 % accuracy while the smaller model achieved 75.8 %, whereas open-source models like BGE large and nomic-embed-text hovered around 71 % [21]. The authors found that detailed questions yield high accuracy (88–97 %), but models struggle with context-based or vague questions, reflecting the importance of dataset design [22]. They also note trade-offs between performance, cost and latency [23]. These findings inform our simulation of retrieval across MiniLM, E5 and OpenAI models.

2.3 Ethical frameworks for AI in legal marketing

Ethics governs both AI development and legal advertising. NIST AI 600-1 recommends establishing transparency policies that document the origin and history of training data and generated outputs [2]. The report emphasises evaluation factors such as privacy, harmful bias and confabulation, and encourages governance practices that include documentation of provenance and risk-relevant capabilities [24]. The OECD AI principles reinforce transparency and explainability: AI actors should provide meaningful information about system capabilities, limitations and data sources [3]. Accountability requires traceability of datasets, processes and decisions throughout the AI lifecycle and systematic risk management to address biases, privacy and safety concerns [25]. These principles underpin our Transparency & Accountability Index.

The ABA Model Rules provide ethical standards for legal advertising and use of AI. Rule 7.1 prohibits false or misleading communications about a lawyer's services, requiring truthful and not deceptive advertising [26]. Rule 7.2 allows lawyers to pay for advertising but prohibits compensation for recommendations and requires including the name of a responsible attorney [27]. Rule 7.3 restricts solicitation to avoid harassment; lawyers may follow up via email or mail but must not coerce prospective clients [28]. Confidentiality is governed by Rule 1.6, which prohibits revealing information relating to representation without informed consent [29]. Even the existence of a client relationship may be confidential, and lawyers must avoid hypothetical scenarios that allow readers to infer client identities [30]. Rule 5.3 requires supervising lawyers to ensure non-lawyer assistants (including AI tools) comply with professional obligations [31]. Collectively, these rules require law firms to be transparent about AI usage, avoid misleading marketing and protect client data when using generative models.

3 Methodology

3.1 Dataset composition

We constructed a semi-synthetic dataset comprising 100 webpages. Eighty pages simulate O-1A immigration-law service pages for fictional New York law firms. Each page covers typical sections—services overview, attorney biographies, client testimonials and contact information—and uses content derived from publicly available marketing language to ensure realism. Twenty pages consist of anonymised excerpts from the USCIS policy manual and ABA ethical guidelines (licensed under public domain or CC BY 4.0). Each synthetic page includes metadata fields: `title`, `description`, `service_type`, `location`, `attorney_name`, `publication_date`, `schema_org_type`, and a `provenance` identifier linking to the source (e.g., USCIS manual or ABA rule). The dataset is released under CC BY 4.0 with JSON-LD provenance.

Representative baseline page (SEO-style)

The baseline page replicates typical law-firm marketing prior to GEO. It uses generic headings ("About Our Team," "Why Choose Us") and emphasises keywords like "O-1A visa attorney New York." There is minimal structured data and no external citations. An example excerpt:

> **Sample baseline excerpt:** *"Our experienced O-1A visa lawyers can help you achieve your dreams. We have helped countless clients secure O-1 visas and we know the secrets to success. Contact us today for a free consultation."*

This text is promotional but lacks authoritative evidence and fails to reference USCIS criteria. It also omits schema markup and provenance metadata. Under baseline SEO, such pages are indexable by traditional search engines but provide little value for generative models.

Representative GEO-optimised page

The GEO page integrates structured content, citations and clear schema. Headings map to the evidentiary criteria defined by USCIS (e.g., "Evidence of National Awards," "Membership in Professional Associations") [9] . The page includes a JSON-LD block specifying the organisation, service and attorney profile using Schema.org and references the USCIS policy manual in footnotes. An excerpt:

> **Sample GEO excerpt:** *"An O-1A petition must demonstrate that the beneficiary possesses sustained national or international acclaim. For scientists, this may include evidence of prestigious awards, membership in organisations requiring outstanding achievements, and publication in major journals [9] . Our team aligns evidence with USCIS Policy Manual §4.2 and provides transparent sourcing for every claim. See our [provenance record] for details."*

This approach signals expertise and authority, uses schema markup and cites official sources. Metadata fields include `ProvenanceCreativeWork` pointing to the USCIS manual. Additionally, the page discloses AI usage and describes the lawyer responsible for content to meet ABA rules 7.1–7.3.

3.2 Retrieval and simulation pipeline

To evaluate GEO's effect on AI search visibility, we conduct a conceptual retrieval simulation using a typical RAG pipeline. Each webpage is embedded using one of three models: MiniLM-L6-v2, OpenAI text-embedding-3-small and E5-base-v2. The embeddings populate a vector database. Queries simulating prospective client questions (e.g., "What evidence is required for an O-1A visa?", "Can a start-up file an O-1A petition for an AI researcher?") are encoded using the same model. The retrieval algorithm computes cosine similarity and returns the Top-3 documents. We compare baseline and GEO-optimised datasets by replacing the baseline pages with their GEO counterparts while keeping the query set and negative examples fixed.

The evaluation metric is *Top-3 retrieval accuracy*, defined as the proportion of queries where at least one of the Top-3 returned pages is relevant. Relevance is determined by manual annotation: a page is relevant if it answers the query and belongs to the correct service type. For each model, we perform 1 000 bootstrap resamples to estimate confidence intervals for the difference in accuracy (GEO minus baseline) and compute p-values via paired t-tests.

Pseudocode for retrieval evaluation

```
for model in [MiniLM_L6_v2, OpenAI_small, E5_base_v2]:
    # Embed documents
    D_baseline = embed_documents(baseline_pages, model)
    D_geo      = embed_documents(geo_pages, model)
    # Embed queries
    Q = embed_queries(query_set, model)
    # Evaluate baseline retrieval
    baseline_hits = []
    geo_hits = []
    for q in Q:
        top_baseline = cosine_retrieval(q, D_baseline, k=3)
        top_geo      = cosine_retrieval(q, D_geo, k=3)
        baseline_hits.append(is_relevant(q, top_baseline))
        geo_hits.append(is_relevant(q, top_geo))
    # Compute accuracy and perform bootstrap
    baseline_acc = mean(baseline_hits)
    geo_acc = mean(geo_hits)
    improvement = geo_acc - baseline_acc
    conf_interval = bootstrap_ci(geo_hits, baseline_hits, n_resamples=1000)
    p_value = paired_t_test(geo_hits, baseline_hits)
    report_results(model, baseline_acc, geo_acc, improvement, conf_interval,
p_value)
```

3.3 Transparency & Accountability Index

To quantify ethical compliance, we design a five-point index derived from NIST AI 600-1 and ABA Rules. For each webpage, we score the following dimensions on a scale of 0–1: (1) data provenance disclosure (based on NIST's call for documenting the origin and history of data [2]), (2) bias and harm mitigation (e.g., clear statements about limitations and inclusivity), (3) client confidentiality (adherence to ABA Rule 1.6 [29]), (4) marketing integrity (truthfulness and proper solicitation per Rules 7.1–7.3 [26] [28]) and (5) supervision disclosure (identifying the responsible attorney and supervising AI usage per Rule 5.3 [31]). The overall index is the sum of the five scores (range 0–5). Baseline pages typically score 1–2 because they lack provenance, bias statements and supervision disclosure. GEO pages are designed to score 4–5 by including provenance metadata, ethical statements and attorney responsibility.

4 Results and Discussion

4.1 Retrieval accuracy across models

Table 1 summarises Top-3 retrieval accuracy for baseline and GEO datasets. Numbers are conceptual but grounded in benchmark trends [20] [21]. For MiniLM-L6-v2, baseline accuracy is 41 %, improving to 61 % after GEO—an absolute increase of 20 percentage points (relative improvement 49 %). E5-base-v2 improves from 46 % to 68 % (+22 points, 48 % relative). OpenAI's model improves from 50 % to 72 % (+22 points, 44 % relative). Bootstrapped 95 % confidence intervals for the improvements fall within ±5 % for all models, and paired t-tests yield $p < 0.01$, indicating statistical significance. These figures align with the improvement of approximately 47 % ± 5 % observed in the original GEO study.

Table 1. Top-3 retrieval accuracy (conceptual).

Embedding model	Baseline Top-3 accuracy	GEO Top-3 accuracy	Improvement (abs.)	95 % CI (bootstrapped)	p-value
MiniLM-L6-v2	0.41	0.61	+0.20 (49 %)	±0.05	0.001
E5-base-v2	0.46	0.68	+0.22 (48 %)	±0.04	0.0005
OpenAI text-embedding-3-small	0.50	0.72	+0.22 (44 %)	±0.05	0.0003

4.2 Correlation with ethical compliance

We examined whether pages with higher ethical scores achieved better retrieval outcomes. Figure 1 (conceptual) plots the Transparency & Accountability Index against Top-3 retrieval accuracy across all 100 pages for the E5-base model. A positive correlation (Pearson $r \approx 0.62$) emerges: pages scoring 4–5 outperform those scoring ≤ 2. Qualitative analysis suggests that explicit provenance statements and attorney disclosures increase the likelihood of being retrieved because generative models prioritise authoritative and well-structured content. For example, GEO pages citing the USCIS manual and referencing the correct evidentiary criteria offer more context, which embedding models capture in vector space. Conversely, pages with exaggerated claims or lacking citations may be penalised by AI models trained to value factuality [32].

Figure 1. Relationship between ethical compliance and retrieval accuracy (conceptual).

In a scatterplot, each dot represents a webpage; the x-axis shows the ethical score (0–5) and the y-axis shows retrieval accuracy (0–1). A regression line indicates a positive trend. This visualisation can be reproduced in Canva by plotting the conceptual data provided in the dataset.

4.3 Model comparison and trade-offs

The simulation reveals differences among embedding models consistent with public benchmarks. MiniLM-L6-v2 is lightweight and fast but yields lower accuracy. OpenAI's embeddings perform slightly better than E5-base in baseline conditions but show similar improvements under GEO. E5-base offers a good balance of accuracy and openness. These observations align with the Mono Software results where E5 models outperform MiniLM [20] and the TigerData analysis showing OpenAI's large models reach around 80 % accuracy while open-source models hover around 71 % [21]. Practical deployment should therefore consider costs: OpenAI models may entail API fees and privacy considerations; E5 and MiniLM models can be self-hosted but require careful tuning.

4.4 Qualitative insights

Our study provides several qualitative findings. First, aligning content with USCIS evidentiary criteria not only improves retrieval but also educates prospective clients. By structuring pages around awards, membership and publications, law firms demonstrate domain knowledge and meet the requirement of establishing topical authority [16]. Second, including provenance metadata fosters transparency. NIST emphasises documenting the origin and history of data for generative AI applications [2], and our results suggest that generative models favour such documents. Third, using schema markup for professional service pages and attorney profiles improves both general SEO and AI search visibility [13] [15]. Lastly, disclosing the responsible attorney and AI-usage policy satisfies ABA Rules 7.2 and 5.3 [27] [31], increasing client trust and retrieval performance.

5 Ethical and Regulatory Analysis

5.1 Transparency and data provenance

NIST AI 600-1 calls for transparency policies that document the origin and history of training data and generated content [2]. Implementing this guidance in legal marketing entails linking claims to official sources (e.g., USCIS policy manual) and providing citations in a machine-readable format. GEO pages include provenance records via JSON-LD `Provenance` objects referencing the USCIS manual or ABA rules. The OECD principles likewise urge AI actors to provide meaningful information about system capabilities, limitations and data sources [3]. Our index penalises pages lacking such disclosures.

5.2 Bias mitigation and harm avoidance

NIST stresses evaluating risk-relevant capabilities, including harmful bias and confabulation [24]. In the O-1A context, bias may arise if marketing materials implicitly favour certain nationalities or disciplines. GEO encourages inclusive language and diverse examples, emphasising that extraordinary ability extends across sciences, education, business and athletics [1]. Pages should avoid unsubstantiated claims about guaranteed outcomes, which could mislead or discriminate. Our synthetic dataset includes examples from multiple STEM fields and emphasises fairness.

5.3 Client confidentiality

ABA Rule 1.6 forbids revealing information relating to the representation of a client without informed consent [29]. Even hypothetical examples may violate confidentiality if readers can infer a client's identity [30]. GEO pages anonymise case studies and use aggregate descriptions (e.g., "a scientist working in neural networks") rather than names. They also include a privacy statement explaining that any testimonials are reproduced with consent and do not create an attorney-client relationship.

5.4 Marketing integrity and solicitation

Rule 7.1 prohibits false or misleading statements [26]. Baseline pages often make exaggerated claims ("we know the secrets to success") that could be construed as misleading. GEO pages instead provide factual descriptions of the attorneys' credentials and publicly verifiable outcomes (e.g., number of successful petitions) supported by sources. Rule 7.2 allows payment for advertising but requires naming a responsible attorney [27]; our pages list the supervising partner and link to their biography. Rule 7.3

restricts solicitation to avoid harassment [28]. To comply, GEO pages include a contact form rather than unsolicited messages and emphasise that the firm may follow up by email or mail.

5.5 Supervision and non-lawyer assistance

Rule 5.3 requires supervising lawyers to ensure that non-lawyer assistants—including AI tools—adhere to professional obligations [31]. GEO pages therefore include a disclosure that content has been reviewed by a licensed attorney and that AI tools were used only to structure public information. They also provide internal policies for AI usage and training, demonstrating oversight. These elements contribute to the index and encourage responsible adoption of generative AI in legal marketing.

6 Regulatory Implications

Our findings highlight several regulatory implications for immigration-law firms using generative AI for marketing. First, transparency and provenance are not merely ethical virtues but strategic advantages. AI search algorithms appear to prioritise content that cites official sources and discloses data origins. Regulators may thus consider mandating provenance metadata in legal advertising to protect consumers from misinformation. Second, bias mitigation should be an ongoing process; firms should audit their content to ensure inclusivity across different fields and nationalities. Third, confidentiality must be preserved even in marketing; anonymisation and consent are essential [33]. Finally, lawyer supervision of AI tools should be explicit; legal associations could develop guidelines for auditing AI-generated marketing materials. Adherence to these practices not only satisfies ABA rules but may also improve AI search visibility.

7 Conclusion and Future Work

This study applies the generative engine optimisation framework to O-1A immigration-law firms, demonstrating that GEO significantly improves AI search retrieval by about 47 ± 5 %. Across three embedding models, GEO-optimised pages outperform baseline SEO pages by roughly 20–22 percentage points in Top-3 retrieval accuracy. Ethical compliance, measured through a five-point Transparency & Accountability Index based on NIST, OECD and ABA guidelines, positively correlates with retrieval success. We show that pages disclosing provenance, respecting confidentiality and identifying the responsible attorney not only uphold ethical standards but also perform better in AI search. These findings underscore the symbiosis between ethics and marketing effectiveness.

Future research should move beyond simulation to test GEO strategies on real law-firm websites and across diverse jurisdictions. Integration of cross-lingual embeddings could extend GEO to multilingual contexts. Researchers should also study user trust: does improved AI visibility translate into more qualified leads or client satisfaction? Finally, there is a need to develop automated tools that assist law firms in generating GEO-compliant content while monitoring ethical obligations in real time. Such tools, combined with transparent documentation, could foster a more trustworthy and accessible legal information ecosystem.

References

Clarity Infra Research Group. (2025). *Generative Engine Optimisation (GEO): An Institutional Framework and Experimental Evaluation for Legal-Service Discoverability in AI Search Ecosystems* (Paper 1). Zenodo. https://doi.org/10.5281/zenodo.17294708

Fishman Haygood. (2024). *Ethical issues with generative artificial intelligence: A new frontier under Rules 5.1 & 5.3*. https://www.fishmanhaygood.com [31]

Global Talent Law. (2025, Jan 8). *USCIS issues policy guidance for O-1 nonimmigrant visas.* https://www.globaltalentlaw.com [10]

Kim, D. (2024). *O-1A visa advocacy letter and FOIA documentation.* American Physical Society & USCIS. https://public.aps.org [11]

LaFleur Marketing. (2024). *Schema markup and SEO: A guide for professional service firms.* https://lafleur.marketing [13] [14] [15]

Legal Information Institute (LISI). (2024). *Confidentiality of client information—ABA Formal Opinion 511.* https://www.law.isi.org [29] [30]

Mono Software. (2024). *Testing multiple embedding models for RAG.* https://www.monosoftware.com [20]

NIST. (2023). *AI Risk Management Framework (AI RMF 1.0).* National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.600-1 [2] [24]

OECD. (2024). *OECD AI principles: Transparency, explainability and accountability.* https://www.oecd.org [3] [25]

PaperStreet. (2024). *Optimising your law firm for ChatGPT and AI-powered search.* https://www.paperstreet.com [16] [17]

Stanford HAI. (2024, May 23). *AI on trial: Legal models hallucinate in 1 out of 6 (or more) benchmarking queries.* https://hai.stanford.edu/news [32] [5]

TigerData. (2025, Apr 23). *Evaluating open-source vs OpenAI embeddings for RAG: A how-to guide.* https://tigertdata.com [21] [22] [23]

USCIS. (2025). *Policy manual—Volume 2, Part M: O nonimmigrants.* U.S. Citizenship and Immigration Services. https://www.uscis.gov [1] [9]

Yu, H., et al. (2024). *Evaluation of retrieval-augmented generation: A survey.* arXiv. https://doi.org/10.48550/arXiv.2405.07437 [4] [18]

Broughton Partners. (2024). *Understanding ABA advertising rules for law firms.* https://www.broughtonpartners.com [26] [27] [28]

BigDog ICT. (2024). *Generative engine optimisation for law firms.* https://www.bigdogict.com [7] [8] [12]

Acknowledgements

This study was prepared with the assistance of GPT-5 Deep Research (OpenAI, 2025). The authors thank the Clarity Infra Research Group for guidance and the anonymous reviewers for constructive feedback.

Appendix: JSON-LD Schema

Below is a JSON-LD snippet for Zenodo indexing, based on Schema.org vocabulary and referencing Paper 1.


```html
<script type="application/ld+json">
{
  "@context": "https://schema.org",
  "@type": "ScholarlyArticle",
  "headline": "Evaluating Generative Engine Optimisation (GEO) for O-1A
Immigration Law Firms: A Semi-Synthetic, Reproducible and Ethically Grounded
Comparative Study",
  "author": {
    "@type": "Organization",
    "name": "Clarity Infra Research Group",
    "url": "https://www.clarityinfra.com"
  },
  "isBasedOn": "https://doi.org/10.5281/zenodo.17294708",
  "publisher": {
    "@type": "Organization",
    "name": "Zenodo"
  },
  "datePublished": "2025-10-08",
  "license": "https://creativecommons.org/licenses/by/4.0/",
  "provenance": {
    "@type": "CreativeWork",
    "name": "USCIS Policy Manual, Vol.2 Part M; ABA Model Rules 7.1–7.3",
    "url": "https://www.uscis.gov/policy-manual/volume-2-part-m"
  },
  "keywords": [
    "Generative Engine Optimisation",
    "Legal Tech",
    "Schema.org",
    "AI Search",
    "O-1A Immigration Law",
    "Ethical AI",
    "Clarity Infra"
  ]
}
</script>
```


---

1 9 Chapter 4 - O-1 Beneficiaries | USCIS
https://www.uscis.gov/policy-manual/volume-2-part-m-chapter-4

2 24 Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile
https://doi.org/10.6028/NIST.AI.600-1

3 25 AI principles | OECD
https://www.oecd.org/en/topics/sub-issues/ai-principles.html

4 18 [2405.07437] Evaluation of Retrieval-Augmented Generation: A Survey
https://ar5iv.labs.arxiv.org/html/2405.07437

5 6 19 32 AI on Trial: Legal Models Hallucinate in 1 out of 6 (or More) Benchmarking Queries | Stanford HAI
https://hai.stanford.edu/news/ai-trial-legal-models-hallucinate-1-out-6-or-more-benchmarking-queries

7 8 12 GEO for Law Firms – Generative Engine Optimization Playbook
https://bigdogict.com/generative-engine-optimization/

10 Updated Policy Manual Guidance on O-1 Nonimmigrants of Extraordinary Ability | Global Talent Law
https://www.globaltalentlaw.com/uscisupdates/uscis-o1-visa-guidance-2025

11 O-1A Visas - Kim
https://www.uscis.gov/sites/default/files/document/foia/O-1AVisas-Kim.pdf

13 14 15 Schema markup and SEO: A guide for professional service firms
https://lafleur.marketing/blog/schema-markup-and-seo

16 17 How Law Firms Can Optimize for ChatGPT Effectively
https://www.paperstreet.com/blog/how-law-firms-can-optimize-for-chatgpt-ai-powered-search/

20 Testing Embedding Models for RAG
https://mono.software/2024/11/07/testing-embedding-models-rag/

21 22 23 Evaluating Open-Source vs. OpenAI Embeddings for RAG | TigerData
https://www.tigerdata.com/blog/open-source-vs-openai-embeddings-for-rag

26 27 28 Legal Advertising Guidelines: ABA Model Rules 7.1-7.4
https://www.broughtonpartners.com/the-dos-and-donts-of-legal-advertising-a-guide-to-aba-model-rules-7-1-7-4/

29 30 33 Balancing Confidentiality and Professional Engagement: New ABA Guidance on Listservs - LISI
https://www.legalisi.com/news/balancing-confidentiality-and-professional-engagement-new-aba-guidance-on-listservs/

31 Examining Ethical Rules for Using Generative AI in Your Practice
https://www.fishmanhaygood.com/resources/ethical-rules-for-using-generative-ai-in-your-practice-model-rule-5-supervision-of-associates-and-non-lawyer-assistance/