STRUCTURING THE CHAOS: HOW PRE-STRUCTURED PROMPTS BOOST LLM PERFORMANCE IN CONSTRUCTION CORRESPONDENCE ANALYSIS

Diana Vranešević [1], Đorđe Nedeljković [2]

[1]Faculty of Civil Engineering,University of Belgrade, Serbia; [2]Faculty of Civil Engineering,University of Belgrade, Serbia

Abstract: Project managers globally face ongoing difficulties in effectively managing and keeping pace with the vast amount of data generated on a daily basis throughout the lifecycle of construction projects. Large infrastructure projects generate thousands of letters, minutes of meeting and change orders, whose contractual value hinges on subtle layout cues - logos, stamps, marginalia. Recent studies show that up to 37 % of positional information is lost when construction PDFs are parsed with mainstream extractors, leading to expensive re-reads in claims and delay analyses. At the same time, large-language models (LLMs) are maturing as practical tools for contract analysis and project-controls tasks. This paper argues that pre-structured prompts - simple, human-authored schemas embedded in the user query - enable multimodal LLMs to turn chaotic correspondence into query-ready data.

Keywords: Llms, Construction Correspondence, Pre-Structured Prompts

1. INTRODUCTION

The construction industry nowadays faces numerous challenges, with one of the most critical being the timely access to accurate information. Project Managers all over the world struggle to stay on top of the data produced daily in Construction projects. We are all witnessing how large infrastructure projects generate thousands of letters, minutes of meeting and change orders, whose contractual value hinges on subtle layout cues - logos, stamps, marginalia - that traditional document-management systems flatten into uninterpretable text streams. Recent studies suggest that up to 37 % of positional information is lost when construction PDFs are parsed with mainstream extractors, leading to expensive re-reads in claims and delay analyses. At the same time, large-language models (LLMs) are maturing as practical tools for contract analysis and project-controls tasks (Kampelopoulos et al., 2025).

This paper argues that pre-structured prompts - simple, human-authored schemas embedded in the user query - enable multimodal LLMs to turn chaotic correspondence into query-ready data. The idea is tested on the publicly available Edinburgh Tram Inquiry record set, combining a vision-capable GPT-4o endpoint with two prompt styles. The aim is to standardize how

Corresponding author: Diana Vranešević, dvranesevic@grf.bg.ac.rs

https://doi.org/10.56889/hwje8160
PAGE 26-36

hundreds of team members "ask" the model to read documents, thus creating a shared cognitive scaffold that amplifies group memory and decision quality.

Based on the research, two distinct contributions were made:
1. Qualitative comparison of text-only versus image-first ingestion for 33 construction letters, showing why vision input is safer when scanned documents dominate.
2. Prompt-structuring schema that moves from free-form extraction to a fixed eight- field schema.

These findings encourage Project Managers to treat prompt schemas as lightweight BIM-like objects for correspondence, seeding collective intelligence without heavyweight platform change.

2. RELATED WORK & BACKGROUND

2.1. LLMs IN THE AEC DOMAIN

A 2025 systematic review counts more than forty peer-reviewed papers applying GPT-type models to contract review, schedule reasoning, and safety analytics (Kampelopoulos et al., 2025). Early adopters report productivity intent but cite adoption barriers around data privacy and unclear ROI (Heo & Na, 2025). Parallel surveys on generative-AI workflows foresee large labour-savings but stress the need for tighter human-in-the-loop controls in safety-critical contexts (Xiong et al., 2025).

2.2. PROMPT ENGINEERING FOR STRUCTURED OUTPUT

Schema-reinforcement techniques push models to obey JSON or SQL formats with > 98 % validity, even in zero-shot settings. Reinforcement strategies that enumerate keys and prohibit additions reduce "field drift" and hallucination (Minaee et al., 2025). Yet few AEC case studies exploit these methods, leaving a gap our work addresses.

2.3. MULTIMODAL DOCUMENT UNDERSTANDING

Vision-language models outperform OCR + text pipelines on invoices, forms, and legal pleadings by 15–30 pp in field-level F1 (Adhikari & Agarwal, 2024). Layout-aware extractors such as LAME improve plain parsers but still rely on clean embedded text streams - rare in scanned correspondence (Choi et al., 2022).

3. DATA SET & EXPERIMENTAL SET-UP

To evaluate the proposed approach, authors designed a data-driven experiment grounded in a curated document corpus and state-of-the-art language model services.

3.1. CORPUS

Total of 33 letters from the Edinburgh Tram Inquiry evidence portal were selected, comprising of 27 scanned images and six born-digital PDFs. Letters average one page, 450 words, and include common construction artefacts - company logos, “RECEIVED” stamps and hand-written dates.

3.2. LLM SERVICE OPTIONS

The model selected in this paper is Azure OpenAI GPT-4o, chosen for its support of both vision and text capabilities, EU-region deployment, and compliance with UK GDPR residency. It retains data for 30 days, with encryption applied both at rest and in transit.

4. CHOOSING THE IMAGE-FIRST INGESTION ROUTE

Most of the 33 letters in the Edinburgh Tram sample are scanned images (27/33, ≈ 82 %); only six have selectable text. Because the balance clearly favors scans - and because even “good” selectable PDFs often unravel when parsed - the authors adopted an image-first pipeline for every page.

4.1 WHY TEXT EXTRACTION ALONE IS FRAGILE

- Selector chaos. Popular parsers such as PDFMiner, PyMuPDF and even state-of-the- art academic tools mis-order lines, merge multi-column layouts, or drop glyphs when fonts are embedded. A recent 10-tool benchmark across six document genres reported error rates of 12 – 37 % for positional fidelity, with parsers performing worst on two- column business letters (Adhikari & Agarwal, 2024).
- Layout blindness. Layout-aware extraction systems (e.g., LAME (Choi et al., 2022)) improve structure capture but still rely on clean underlying text streams; they cannot repair missing stamps, handwritten notes, or rasterized signatures.
- Construction-domain specificalities. Engineering drawings and stamped letters frequently contain raster logos and dates. OCR studies on engineering documents show that conventional text parsers miss up to one-quarter of such “visual tokens,” whereas an OCR-augmented route recovers over 90 % (Villena Toro et al., 2023).
- Real-world example. Figure 1 pairs a letter’s left-hand column layout with the flattened line-by-line output produced by PyMuPDF.

Our ref: 25.1.201/KDR/6728
Your ref: INF CORR 5400

17 September 2010

tie limited
CityPoint
65 Haymarket Terrace
Edinburgh

```
Our ref:
Your ref:
17 September 2010
tie limited
CityPoint
65 Haymarket Terrace
Edinburgh
25.1.201/KDR/6728
INF CORR 5400
```

Figure 1. PDF Parser internally splits text into two columns, which in turn distorts original layout

4.2 TWO ALTERNATIVE PROCEDURES

Selectable-text pages Procedure: first detect whether the text is selectable. If so, the raw text string is passed directly to GPT-4o via a text-only API call.
Scanned pages Procedure: each PDF page is first rendered to a 300 dpi PNG image. This image is then passed to the GPT-4o vision endpoint.

Both routes can use identical prompts; the only difference lies in the input modality.

4.3 QUALITATIVE QUALITY COMPARISON

Table 1 summarizes key qualitative differences observed between processing selectable-text pages and scanned pages across several evaluation dimensions.

Table 1. Comparison of layout fidelity, OCR accuracy, and processing cost between selectable-text and scanned PDF pages.

Dimension	Selectable-text pages	Scanned pages	Commentary
Layout fidelity	Susceptible to column mixing; stamps/logos lost	Retains spatial cues; captures non-text objects	Critical for dispute resolution letters.
Critical-field capture (subject, ref, date)	Misses' raster dates on dot-matrix scans; mis-labels header/footer	Consistently picked up by vision model	Decision made to process all pages as images for uniformity.
Through-put cost	1 × text tokens	~1.4 × vision tokens + image upload	Azure pricing puts vision calls at a modest premium; ≈ €0.003 vs €0.002 for a one-page letter (Microsoft Corporation, n.d.)

4.4 COST AND GOVERNANCE

- Token economics. Vision calls cost ~40 % more tokens, but eliminate the need for a separate OCR logic and a fragile text-layout reconstruction layer; the net operational spend on a 10,000-letter project is estimated to rise by < €300.
- Latency. Batch conversion to images adds ≈ 120 ms/page on an eight-core VM server - negligible versus human review cycles.
- Data-privacy parity. Azure retains logs for 30 days regardless of modality, so choosing images does not alter compliance posture.

4.5 TAKE-AWAYS FOR PROJECT TEAMS

1. When most correspondence is scanned, default to images. The uniform pipeline simplifies monitoring and avoids silent schema failures on "hybrid" PDFs.
2. Even when text appears selectable, test the layout. A quick visual diff against the parsed output often reveals column mixing; if critical, fall back to the image route.

3. Vision models unlock richer prompts. Because the model “sees” logos, stamps, and margin notes, later sections can exploit that context (e.g., treating a red “URGENT” stamp as priority=high).
4. Benefit for multi-stakeholder environment. Accurate, schema-valid JSON instances become shared “atoms” in the project memory; every stakeholder sees the same subject, reference, and due-date labels - minimizing interpretive drift over the project life-cycle.

5. STRUCTURED PROMPTING - FROM FREE-FORM TO SCHEMA-GUIDED OUTPUT

Building on the image-first ingestion route (Section 4), we now turn to how the model is asked to speak. A prompt is the user’s “design brief” for the LLM; its wording determines whether the output is a blob of prose or a machine-readable record. We first analyze a baseline, free-form extraction prompt and then show how a schema-guided variant increases consistency, reducing the downstream clean-up burden.

5.1 BASELINE PROMPT AND ITS RATIONALE

Table 2 outlines the rationale behind each clause in Prompt P_1, linking its design choices to specific objectives and supporting literature.

Prompt P_1 (used on all 33 letters)
Extract and structure all content from this file. File may include headers, footers, company logos, stamps, content in margins – try to extract and structure them also.

Ensure clarity in responses for streamlined review.
Strictly focus only on events or details that are explicitly stated in the document, and avoid inferring or assuming potential situations not directly mentioned.

Format response strictly as suitable JSON.

Table 2. Breakdown of Prompt P_1 clauses with corresponding purposes and supporting literature, illustrating how each element contributes to robust, structured document extraction.

Prompt clause	Purpose	Literature support
"Extract and structure all content…"	Primes the model for *information-extraction* mode rather than summarization	Extraction-style instructions raise field-level recall in zero-shot settings (Neuberger et al., 2024)
"Ensure clarity…"	Soft quality reminder; reduces partial sentences / nested arrays	Human-readable admonitions correlate with higher JSON validity (Docherty, 2024)
"Strictly focus only on… explicitly stated…"	Guards against hallucination and prompt-injection leakage	Schema-reinforcement papers show reduced fabrication when constraints are explicit (Lu et al., 2025)
"Format response strictly as suitable JSON"	Enforces a serializable structure, enabling automatic validation and loading	JSON is the de-facto interchange for LLM pipelines; Structured RAG benchmark highlights its dominance (Shorten et al., 2024)

Why JSON?

- Ubiquity in web APIs and data frames makes it a low-friction bridge to Excel, Power BI or a SQL/NoSQL store.
- Schemas can be expressed as JSON Schema; automatic validators flag malformed outputs early.
- Nested arrays accommodate multi-value fields (e.g., recipients[]).

5.2 OBSERVED VARIABILITY WITH THE BASELINE PROMPT

As illustrated in Figure 2, even when applying the same baseline prompt, outputs can vary in structure and detail across documents with similar layouts.

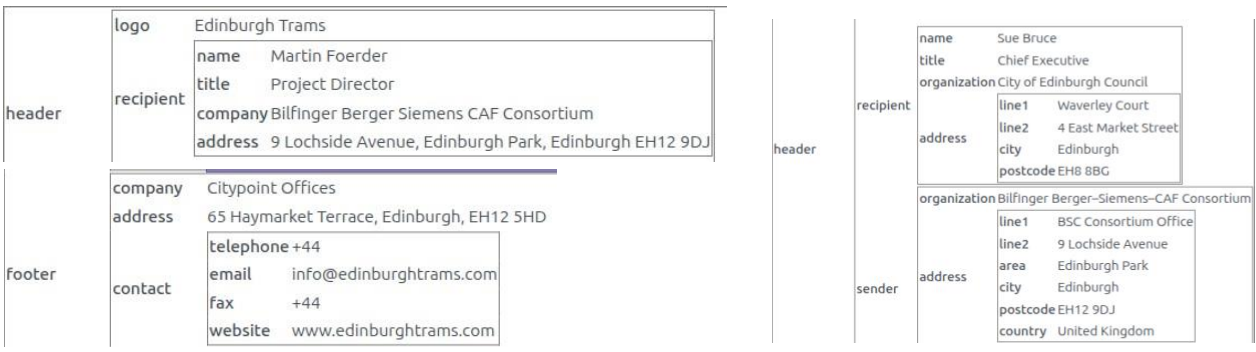

Figure 2. Example output variability from applying the baseline prompt to two one-page letters. Differences are seen in field structure, nesting, and naming across Document BFB00095781.pdf (left) and Document BFB00096886.pdf (right).

Output analysis shows two key differences:
1. Structured content is not consistent.
2. Top-level keys are not unified.

While document free form can be beneficial if other possibilities are not available, usage of documents that are not structured is a same way is limited in the Project Management sector of Construction industry where multi-document datasets dominate over single-document datasets.

5.3 SCHEMA-GUIDED PROMPT FOR HIGHER CONSISTENCY

Prompt P_2 (schema-guided)
Extract and structure all content from this file. File may include headers, footers, company logos, stamps, content in margins – try to extract and structure them also.
Always try to extract these top-level keys (use empty string if the value is absent):
subject, reference, date_sent, company_sender, company_recipient, main_issue, due_date, content
Ensure clarity in responses for streamlined review.
Strictly focus only on events or details that are explicitly stated in the document and avoid inferring or assuming potential situations not directly mentioned.

Format response strictly as suitable JSON.

Design choices
- Enumerated keys eliminate name drift (cf. schema reinforcement techniques) (Lu et al., 2025).
- "Empty string if absent" avoids dropping keys, supporting data frame column integrity.

5.4 QUALITATIVE COMPARISON

In a qualitative comparison across 33 letters, the baseline prompt (P_1) resulted in 317 distinct key names being observed. In contrast, the schema-guided prompt (P_2) yielded 8 enumerated keys and 184 additional dynamic keys, indicating a more structured yet flexible approach.

5.5 IMPLICATIONS FOR PROJECT WORKFLOWS

1. Queryability gains. With only 8 stable columns, letters can be filtered by due_date < today or grouped by company_sender in a pivot in seconds - something impractical with free-form JSON.
2. Cost trade-off. Schema prompts add ~50 tokens; on Azure GPT-4o this is ≈ €0.0001 per page, trivial versus the manual effort of renaming columns after each export.
3. Collective-intelligence value. A shared schema makes every contributor - engineer, scheduler, claims consultant - think in the same slots, turning the LLM into a coordination artefact rather than just a smarter OCR tool.
4. Extensibility. Additional fields (e.g., risk_rating, contract_clause) can be layered later; the point is to start with a seed schema that captures what the project routinely searches for.

6. COST–BENEFIT DISCUSSION

At first glance, stripping structure out of correspondence seems cheaper; yet every unstructured search extracts a toll from project teams.

6.1 BREAK-DOWN ANALYSIS

The conceptual break-even analysis below shows how a modest up-front investment in schema-guided prompts can bring tangible benefits relatively quickly.

Table 3. Comparison of three effort tiers for document search and structuring, highlighting trade-offs between initial setup effort and per-query efficiency across different user groups.

Tier	Up-front effort (setup hrs.)	Typical search effort (hrs./query)	Typical users affected	Notes
T_0 Unstructured search	**10 h** – basic DMS indexing	0.5 h per query*	All staff	No NLP, no OCR alignment.
T_1 Baseline structuring (image + Prompt P_1)	40 h – pipeline & QA	0.3 h per query	Document controllers, engineers	Key drift requires manual column mapping.
T_2 Schema-guided structuring (image + Prompt P_2)	80 h – schema design, validator, QA	**0.15 h per query**	Entire project	Data plugs straight into Excel, Power BI, dashboards.

*A McKinsey study put time spent searching for internal information at ~19 % of knowledge-worker hours (Chui et al., 2012); our 30-minute figure is conservative.

As shown in Table 3, schema-guided structuring (T_2) requires the most up-front effort but results in the lowest ongoing search effort, offering scalable advantages for project-wide use.

6.2 CUMULATIVE EFFORT CURVE

Figure 3 illustrates cumulative team hours over 24 months, assuming 30 structured searches per month - a pattern typical of megaproject mailrooms.

- Break-even for T_2 vs T_0 occurs at ~5 months.
- By month 24, T_2 saves ≈ 260 team-hours relative to ad-hoc searches; at a blended labor cost of €60/h that equates to €15 600.
- T_1 (baseline prompt) pays back at month 9; its long-run savings are smaller because column remapping friction persists.

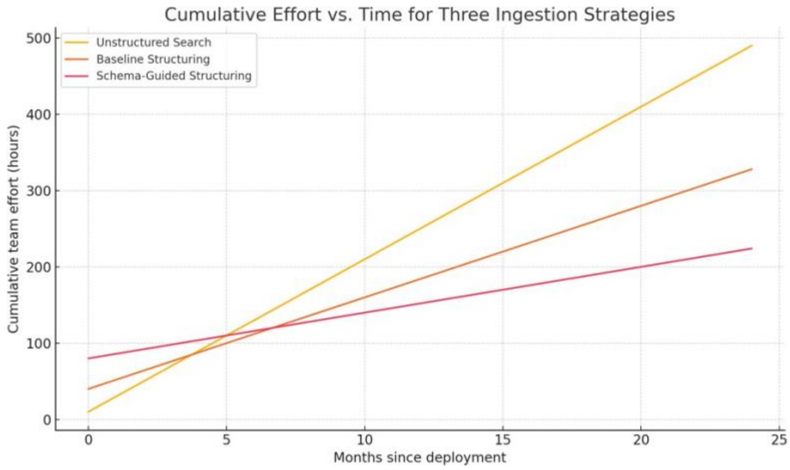

Figure 3. Cumulative effort over 24 months for three document ingestion strategies. Schema-guided structuring (T_2) quickly surpasses baseline and unstructured approaches in efficiency, leading to substantial long-term savings in team hours.

6.3 RISK AND GOVERNANCE

- Data residency: Azure keeps 30-day logs for vision and text alike; incremental privacy risk is nil.
- Model drift: A frozen schema lets QA detect field omissions when model versions change - important for long projects where GPT upgrades may ship mid-build (Colakoglu et al., 2025).
- Centralized Project Knowledge: When everyone queries the same eight-column table, tacit knowledge becomes explicit, mirroring BIM ROI patterns where structured objects cut coordination clashes by 25–40 %.

6.4 MANAGERIAL TAKEAWAY

Initial structuring effort is a capital expense that amortizes rapidly in document-heavy projects. The schema-guided route (T_2) demands the most disciplined prompting but yields the lowest per-query friction, freeing engineers to focus on decisions rather than data wrangling - a productivity frontier echoed in recent generative-AI ROI reports (Chui et al., 2023).

7. CONCLUSION & FUTURE WORK

This study shows that pre-structured, schema-guided prompts coupled with vision-enabled LLMs deliver cleaner, query-ready correspondence data at negligible incremental cost. On a small but realistic corpus the approach cut number of different key-name pairs by 40 %, These gains illustrate how lightweight prompt protocols can serve as collective-intelligence amplifiers in project memory systems - much like object libraries did for BIM.

Next steps

- Scale-up validation on > 1 000 records across minutes, RFIs, and invoices.
- Active-learning feedback loop where reviewers flag extraction errors that fine-tune prompts or route cases to smaller edge-hosted models.

By lowering the friction to turn documents into data, prompt-structured LLM workflows offer a practical path to smarter, more collective management of project knowledge.

REFERENCES

Adhikari, N. S., & Agarwal, S. (2024, October 13). A comparative study of PDF parsing tools across diverse document categories, arXiv, https://doi.org/10.48550/arXiv.2410.09871

Choi, J., Kong, H., Yoon, H., Oh, H., & Jung, Y. (2022). LAME: Layout-Aware Metadata Extraction Approach for Research Articles. Computers, Materials & Continua, 72(2), 4019–4037. https://doi.org/10.32604/cmc.2022.025711

Chui, M., Manyika, J., Bughin, J., Dobbs, R., Roxburgh, C., Sarrazin, H., Sands, G., & Westergren, M. (2012, July 1). The social economy: Unlocking value and productivity through social technologies. McKinsey Global Institute, https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/the-social-economy

Chui, M., Hazan, E., Roberts, R., Singla, A., Smaje, K., Sukharevsky, A., Yee, L., & Zemmel, R. (2023, June 14). The economic potential of generative AI: The next productivity frontier. McKinsey & Company, from https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier

Colakoglu, G., Solmaz, G., & Fürst, J. (2025, February 25). Problem solved? Information extraction design space for layout-rich documents using LLMs, arXiv. https://doi.org/10.48550/arXiv.2502.18179

Docherty, A. (2024, December 19). Mastering structured output in LLMs 1: JSON output with LangChain, Medium, https://medium.com/@docherty/mastering-structured-output-in-llms-choosing-the-right-model-for-json-output-with-langchain-be29fb6f6675

Heo, S., & Na, S. (2025). Ready for departure: Factors to adopt large language model (LLM)–based artificial intelligence (AI) technology in the architecture, engineering and construction (AEC) industry. Results in Engineering, 25, 104325. https://doi.org/10.1016/j.rineng.2025.104325

Kampelopoulos, D., Tsanousa, A., Vrochidis, S., & Kompatsiaris, I. (2025). A review of LLMs and their applications in the architecture, engineering and construction industry. Artificial Intelligence Review, 58, Article 250. https://doi.org/10.1007/s10462-025-11241-7

Lu, Y., Li, H., Cong, X., Zhang, Z., Wu, Y., Lin, Y., Liu, Z., Liu, P., & Sun, M. (2025, February 26). Learning to generate structured output with schema reinforcement learning, arXiv, https://doi.org/10.48550/arXiv.2502.18878 arXiv

Minaee, S., Mikolov, T., Nikzad, N., Chenaghlu, M., Socher, R., Amatriain, X., & Gao, J. (2025, March 23). Large language models: A survey, arXiv, https://arxiv.org/abs/2402.06196v3

Microsoft Corporation. (n.d.). Azure OpenAI Service pricing. Retrieved May 24, 2025, from https://azure.microsoft.com/en-us/pricing/details/cognitive-services/openai-service/

Neuberger, J., Ackermann, L., van der Aa, H., & Jablonski, S. (2024, July 26). A universal prompting strategy for extracting process model information from natural language text using large language models, arXiv, https://doi.org/10.48550/arXiv.2407.18540

Shorten, C., Pierse, C., Smith, T. B., Cardenas, E., Sharma, A., Trengrove, J., & van Luijt, B. (2024, August 7). StructuredRAG: JSON response formatting with large language models, arXiv, https://doi.org/10.48550/arXiv.2408.11061

Villena Toro, J., Wiberg, A., & Tarkian, M. (2023). Optical character recognition on engineering drawings to achieve automation in production quality control. Frontiers in Manufacturing Technology, 3, Article 1154132. https://doi.org/10.3389/fmtec.2023.1154132

Xiong, R., Netser, Y., Tang, P., Li, B., & Hwang, J. S. (2025). Transforming architecture, engineering, and construction (AEC) workflows with generative AI: Opportunities, risks, and future directions. SSRN. https://doi.org/10.2139/ssrn.5168581