

M4 U2 Assignment: NLP for Construction Document Analysis

Master's in AI for Architecture & Construction (Zigurat)

Original Unstructured Document

Construction Contract – Excerpt (Unstructured) This agreement is entered into on the 12th of March 2024 between GreenBuild Developments Ltd. (“the Client”) and Apex Civil Works (“the Contractor”) for the construction of a mixed-use commercial building located at Riverside Business Park, London. The Contractor shall commence work on or before 1 April 2024 and achieve practical completion no later than 31 December 2024. Any delay caused by weather conditions, supply chain disruptions, or regulatory approvals must be communicated to the Client in writing within five working days. Failure to do so may result in penalties of up to £5,000 per week. The total contract value is £4.2 million. Payments will be made in monthly instalments subject to the submission and approval of progress claims. The Client reserves the right to withhold up to 10% of each payment as retention until final completion. The Contractor is responsible for maintaining site safety and complying with all applicable health and safety regulations. Any incident resulting in injury must be reported within 24 hours. The Contractor shall indemnify the Client against any claims arising from negligence, accidents, or failure to comply with statutory obligations. In the event of a dispute, both parties agree to attempt resolution through mediation before pursuing legal action. This agreement shall be governed by the laws of England and Wales.

Part 1: Information Extraction (Structured Output)

Extraction Methodology

A rule-based NLP pipeline was implemented using Python’s `re` module with the following extractors: - **Date extraction:** Multiple regex patterns for various date formats - **Monetary value extraction:** GBP currency pattern matching - **Obligation extraction:** Modal verb pattern detection (shall, must, is responsible for) - **Risk/penalty extraction:** Keyword-based sentence filtering - **Party extraction:** Named entity patterns with role labels - **Location/project type:** Contextual phrase matching

Structured JSON Output

```
{  
  "document_type": "Construction Contract Excerpt",  
  "parties": {  
    "Client": "GreenBuild\nDevelopments Ltd.",
```

```

    "Contractor": "Apex Civil Works"
  },
  "project": {
    "type": "mixed-use commercial building",
    "location": "Riverside Business\nPark, London"
  },
  "key_dates": {
    "contract_date": "12 March 2024",
    "commencement_date": "1 April 2024 (on or before)",
    "practical_completion": "31 December 2024 (no later than)"
  },
  "financial_terms": {
    "contract_value": "\u00a34.2 million",
    "payment_method": "Monthly instalments subject to approved progress claims",
    "retention": "Up to 10% of each payment until final completion",
    "delay_penalty": "Up to \u00a35,000 per week"
  },
  "key_activities_and_clauses": [
    "Construction of a mixed-use commercial building",
    "Monthly progress claim submissions required for payment",
    "Written notification of delays within 5 working days",
    "Site safety maintenance and H&S regulation compliance",
    "Incident/injury reporting within 24 hours",
    "Mediation before legal action for disputes"
  ],
  "risks_and_issues": [
    "Delay penalties of \u00a35,000/week for unreported delays",
    "Weather conditions may cause schedule disruption",
    "Supply chain disruptions as recognized delay cause",
    "Regulatory approval delays",
    "10% payment retention risk to contractor cash flow",
    "Liability for negligence, accidents, or statutory non-compliance"
  ],
  "obligations_and_actions": {
    "contractor": [
      "Commence work on or before 1 April 2024",
      "Achieve practical completion by 31 December 2024",
      "Notify Client of delays in writing within 5 working days",
      "Maintain site safety and comply with H&S regulations",
      "Report injuries within 24 hours",
      "Indemnify Client against negligence/accident claims"
    ],
    "client": [
      "Make monthly payments upon approved progress claims",
      "May withhold up to 10% retention per payment",
      "Attempt mediation before legal proceedings"
    ]
  }
}

```

```

    ]
  },
  "dispute_resolution": "Mediation before legal action",
  "governing_law": "Not found"
}

```

Part 2: Document Summary (164 words)

This construction contract, dated 12 March 2024, is between GreenBuild Developments Ltd. (Client) and Apex Civil Works (Contractor) for a mixed-use commercial building at Riverside Business Park, London. The contract value is £4.2 million.

Work must commence by 1 April 2024, with practical completion due by 31 December 2024. Delays from weather, supply chain issues, or regulatory approvals must be reported in writing within five working days, or penalties of up to £5,000 per week may apply.

Payments are made monthly based on approved progress claims, with the Client retaining up to 10% until final completion. The Contractor bears full responsibility for site safety, health and safety compliance, and must report any injury incidents within 24 hours. The Contractor also indemnifies the Client against claims arising from negligence or statutory non-compliance.

Disputes are to be resolved through mediation before legal proceedings, governed by the laws of England and Wales. This contract establishes clear obligations for both parties regarding schedule, payment, safety, and dispute resolution.

Part 3: Prompt Engineering Documentation

Prompt 1 — Initial Prompt (Naive)

Prompt: “Extract information from this construction contract.” **Issue:** Too vague — no structure, no field specification, no output format. **Output Quality:** Generic paragraph with missed details.

Prompt 2 — Improved Prompt (Structured Request)

Prompt: “Analyze the following construction contract excerpt. Extract and return a JSON object with these fields: parties (client and contractor names), project type, location, key dates (contract date, start date, completion deadline), financial terms (value, payment method, retention, penalties), contractor obligations, client obligations, risks, dispute resolution mechanism, and governing law.” **What Changed:** Added explicit output format (JSON), specified exact fields to extract, and separated obligations by party. **Why It Improved:** Structured

prompts reduce hallucination and omission. Specifying JSON ensures machine-readable output. Field-level detail prevents the model from summarizing instead of extracting.

Prompt 3 — Refined Prompt (Domain-Aware + Role Assignment)

Prompt: “You are a senior construction contract analyst. Given the following unstructured contract excerpt, perform a detailed information extraction. Return a structured JSON with: (1) Parties and roles, (2) Project scope and location, (3) Key dates and deadlines, (4) Financial terms including penalties, (5) Risks and liabilities identified, (6) Obligations per party, (7) Dispute resolution and governing law. Flag any ambiguous or missing clauses that a construction manager should verify manually.” **What Changed:** Added role assignment (‘senior construction contract analyst’), numbered extraction categories for clarity, and a meta-instruction to flag ambiguities. **Why It Improved:** Role assignment activates domain-specific reasoning in LLMs. Numbered categories improve completeness. The ‘flag ambiguities’ instruction adds a critical-thinking layer, making the output more useful for professional review.

Prompt 4 — Summary-Specific Prompt

Prompt: “Summarize the following construction contract in under 200 words. The summary should be accurate, professionally written, and suitable for a construction manager or project stakeholder. Focus on: parties, scope, key dates, financial terms, safety obligations, and dispute resolution. Avoid legal jargon where possible.” **What Changed:** Specified word limit, target audience, key focus areas, and tone guidance (avoid jargon). **Why It Improved:** Audience-aware prompts produce more relevant summaries. Word limits prevent verbose output. Focus areas ensure no critical information is omitted.

Part 4: Critical Commentary (290 words)

Critical Commentary on AI-Assisted Construction Document Analysis

The AI-generated outputs demonstrated strong capability in extracting structured data from unstructured contract text. Key dates, financial terms, and party identification were accurately captured, and the JSON structure provided a clear, machine-readable format suitable for integration into project management systems or contract databases.

However, several limitations were observed. The extraction pipeline relies on pattern matching and keyword detection, which means it can miss implicit obligations or contextual nuances. For example, the phrase “subject to the submission and approval of progress claims” implies a procedural dependency that a rule-based system may not fully interpret without domain-specific training. Similarly, the indemnification clause contains legal subtleties around “failure

to comply with statutory obligations” that require expert legal interpretation beyond what NLP can reliably provide.

The document summary was concise and professionally appropriate, but it necessarily omits granular detail. A stakeholder relying solely on the summary might overlook the specific conditions under which delay penalties apply or the exact scope of the indemnification clause.

From a professional standpoint, AI tools are highly useful for accelerating initial document review, flagging key clauses, and producing first-draft summaries. They reduce the time required for contract intake and can standardize how information is captured across multiple documents on a project.

Nevertheless, significant risks exist in relying on AI for construction contract analysis. Misinterpreted penalty clauses could lead to financial exposure. Incomplete extraction of safety obligations could create compliance gaps. AI cannot assess the commercial reasonableness of terms or detect missing clauses that industry practice would expect. Therefore, AI outputs should always be reviewed by qualified professionals — quantity surveyors, contract administrators, or construction lawyers — before being used for decision-making. AI is a powerful assistant, not a replacement for professional judgment.

Generated using Python NLP pipeline — M4 U2 Assignment