APM Research Fund Series

Association for Project Management

AI-Driven Governance: Intelligent Stakeholder Strategy for Mega-Infrastructure Projects

Authored by:
Jin Xue, Wei-Ting Hong, Qian-Cheng Wang,
Yunping Liang, Jennifer Whyte and Kerui Lyu

May 2026

Because when projects succeed, society benefits

Contents

Acknowledgements

This research was financially supported by the Association for Project Management (APM). The authors would also like to acknowledge the European Academy of Management (EURAM) for its support, as well as the Engineering Vacation Research Internship Program at the University of Sydney.

About the authors

Jin Xue, Lecturer, University of Sydney, Australia

Dr Xue is a Lecturer in the School of Project Management at the University of Sydney. He specialises in pioneering research on stakeholder management in the dynamic and complex environment of megaprojects with the data-driven methods of natural language processing, network analysis and organisational simulation. He has a knowledge background in project management, information technology and mathematical modelling.

Wei-Ting Hong, Postdoctoral Fellow, University of Sydney, Australia

Dr Hong is currently a Postdoctoral Research Associate at the John Grill Institute for Project Leadership (JGIPL) at the University of Sydney. His research focuses on expert–AI collaboration, digital transformation, data-driven decision-making and policy analysis across various domains, including project analytics, construction and infrastructure management, public transportation and railway safety.

Qian-Cheng Wang, Assistant Professor, City University of Hong Kong, Hong Kong SAR

Dr Wang is a Presidential Assistant Professor at the City University of Hong Kong (CityU). Prior to joining CityU, he earned his PhD in Land Economy from the University of Cambridge. His research centres on the application of digital twin technology to enhance decision-making in smart cities and the built environment. Dr Wang also serves as an Associate Editor for the *Environment and Behavior* and *Computational Urban Science* journals.

Yunping Liang, Assistant Professor, University of Nebraska–Lincoln, US

Dr Liang's research focuses on enhancing decision-making to address economic and financial issues in the development of a sustainable built environment. His research has been sponsored by federal agencies in the US and Canada, and the outcomes have been frequently seen in niche journals. He has been the recipient of national and international awards, and he has leadership roles in international professional organisations related to construction and transportation.

Jennifer Whyte, Professor, University of Sydney, Australia

Jennifer Whyte is a Professor in the School of Project Management, in the Faculty of Engineering at the University of Sydney. She is an expert on complex projects, infrastructure delivery, systems integration and digital innovation. Her current research is on future-making and infrastructure megaproject ecologies: how projects can be set up and delivered to make better societal futures.

Kerui Lyu, PhD Candidate, University of Sydney, Australia

Kerui Lyu is currently a PhD Candidate in the School of Project Management at the University of Sydney. He received a bachelor's degree (honours) in Management and Organizational Study from the University of Western Ontario and a master's degree in Project Management from the University of Sydney. His research focuses on stakeholder engagement and social impact in construction projects.

Glossary

Knowledge graph (KG): A structured data model that represents knowledge as a network of entities (nodes, such as stakeholders and events) and relationships (edges), enabling machines to reason about the connections in data.

Large language model (LLM): An artificial intelligence system trained on vast datasets encompassing books, websites, articles and code, to understand and generate human language.

LightRAG: The graph-enhanced evolution of the retrieval-augmented generation (RAG) architecture used in this research. While standard RAG frameworks enhance LLMs by connecting them to an external, authoritative knowledge base (such as project documents) to reduce hallucinations, LightRAG further integrates knowledge graphs into the retrieval process to effectively handle both specific entity-focused queries and abstract conceptual inquiries.

Social network analysis (SNA): A methodological approach utilising networks and graph theory to investigate social structures and relationships (ties) between stakeholders (nodes), allowing for the dynamic analysis of stakeholder interaction patterns.

Strategic stakeholder engagement map: A diagnostic output of this AI framework, visually represented as a scatter plot comparing a required stakeholder engagement score (awareness, x-axis) against the realised stakeholder engagement score (outcome, y-axis) evidenced by project documents.

Executive summary

1.1 Project overview

Mega-infrastructure projects operate in volatile environments where stakeholder interests change rapidly. Management of these relationships, particularly regarding environmental, social and governance (ESG) issues, is often fragmented and difficult to track.

The objective of this project is to develop an AI-driven systematic framework to enhance stakeholder management by understanding the complex and dynamic nature of stakeholder engagement and deriving actionable insights for project managers.

In achieving this, the project delivered an innovative solution that leverages artificial intelligence (AI) and large language models (LLMs) to analyse official engagement records from case projects across three countries, which enabled stakeholder entity and stakeholder-event extraction. The core outputs in this framework include:

- a strategic stakeholder engagement map to illustrate the alignment between intended and actual engagement, revealing scenarios of proportionate engagement, over-engagement or under-engagement
- data-driven knowledge graphs (KGs) as visual networks that link stakeholders directly to specific events
- an LLM-powered chatbot that answers issue-level questions using KGs and source texts
- a web-based platform designed to enhance stakeholder management practices for industrial practitioners, which integrates:
    - Knowledge Base, where users upload and manage project document sets and trigger KG construction
    - an interactive KG visualiser that renders stakeholders (organisations) and events (documented actions or decisions) as a network
    - an LLM-powered chatbot that answers issue-level questions using KGs and source texts.

1.2 Key findings

To evaluate the models developed, three megaproject case studies were selected, and their publicly available reports were used as data input. Within these cases, the strategic stakeholder engagement map has shown different insightful results where alignment, over-engagement and under-engagement scenarios can all be observed. In parallel, KGs are constructed with two node types – stakeholders and events – and event-mediated edges enabling centrality-based filtering and phase snapshots.

1.3 Recommendations

This research bridges the gap between academic theory and industrial practice by offering a tangible web-based stakeholder knowledge platform. For corporate partners and project practitioners, this platform could be used for the following:

- **Monitor the stakeholder engagement strategy execution gap**: Use the strategic engagement map to monitor the awareness–outcome gap in an existing strategy. Shift resources away from 'over-engaged' areas to address 'under-engaged' risks.
- **Monitoring dynamic networks**: Move beyond static stakeholder registers. Use KGs to visualise how stakeholder relationships evolve over time.
- **Manage the project dataset:** Use the platform as a knowledge management tool to upload and break down historical project reports into structured, reusable evidence for future analysis.
- **Utilising the web platform**: Upload project records and instantly generate visual insights. Ask the AI chatbot specific questions (e.g. 'Who raised environmental concerns in Phase 1?'), and review answers linked directly to the original documents for verification.

Introduction

2.1 Background and significance

Mega-infrastructure projects span long time horizons, mobilise substantial public and private resources, and require complex societal interaction between heterogeneous communities, regulators, contractors and supply chains (Ford et al., 2024; Omotayo et al., 2024; Xue et al., 2025). In such settings, stakeholder dynamics – the way actors' interests, salience and influence evolve across project phases (Shen & Xue, 2021) – become central to performance and legitimacy.

Moreover, project managers are facing the need to shift stakeholder engagement strategy to better respond to stakeholder complexity (K. Mok et al., 2017) and evolving stakeholder influence in projects (Gonzalez-Porras et al., 2021). Traditional approaches, such as salience models or static stakeholder networks, offer useful snapshots, yet they struggle to represent how stakeholder issues intensify, diffuse or reconfigure over time as project environment changes (Luo et al., 2018; Mok et al., 2015). Conventional instruments, such as surveys and interviews, are costly to repeat and difficult to scale, leaving many organisations without a longitudinal, auditable view of how engagement requirements translate into realised outcomes. Lastly, although both academic and industry advances have contributed to the development of platforms for project knowledge management (Elwakil & Zayed, 2018; Iheukwumere-Esotu & Yunusa-Kaltungo, 2022; Kivrak et al., 2008), limited effort has been directed towards creating similar tools tailored to stakeholder management in projects.

Nevertheless, the availability of publicly disclosed project documents and advances in LLMs create a methodological opening, where professional judgement can be fused with machine-assisted evidence retrieval so that engagement strategies can be calibrated using traceable, project-specific data. Building on this opportunity, this study introduces a web-based stakeholder knowledge platform that ingests official project records, automatically structures and indexes them, constructs and visualises KG-based stakeholder networks (including phase-specific snapshots), and supports provenance-linked question-and-answer functionality through an LLM-enhanced chatbot so practitioners can monitor engagement strategy and the evolving network in one place.

2.2 Objective of the research

The primary aim of this study is to develop and demonstrate an AI-driven systematic framework to enhance stakeholder management by understanding the complex and dynamic nature of stakeholder engagement and deriving actionable insights for project managers. Under this aim, this research will be trying to achieve the following objectives:

First, it aims to reveal stakeholder complexity at the level of issues by comparing the project professional's judgement in required stakeholder engagement effort to actual engagement efforts from historical project records. In achieving this, this study leverages customised LLM-enhanced retrieval and analytics on stakeholder-related issues across projects in three countries, based on official project reports containing stakeholder engagement records. By comparing required engagement efforts and actual engagement evidence, a strategic stakeholder engagement map is developed, which provides project teams with practical insights to adjust engagement strategies and allocate resources more effectively. To facilitate this process, a project-level ESG framework developed by Zhang et al. (2024) was adopted as a structured basis for defining stakeholder issues in three case studies.

The focus on ESG-related issues is particularly important because construction projects are consistently subject to strong institutional pressures concerning environmental, social and governance performance (Zhang et al., 2015). However, project teams frequently struggle to translate these broad external expectations into concrete stakeholder engagement practices (Derakhshan et al., 2019; Montalbán-Domingo et al., 2019; Zhang et al., 2024). The Project Management Institute highlights that clear reporting and communication of ESG outcomes not only enhances the visibility of project impacts, but also fosters buy-in for broader sustainability goals and helps maximise the long-term value of ESG initiatives (PMI, 2023). In addition, ESG reporting provides stakeholders with tangible evidence of an organisation's commitment to sustainable development and green practices (Li et al., 2018; Yang & Han, 2023). In this regard, examining ESG issues, and recognising that different issues demand varying levels of stakeholder engagement effort, becomes a critical step in understanding and improving engagement strategies in mega-infrastructure projects.

After constructing the strategic stakeholder engagement map, this study further seeks to develop a retrieval-augmented generation pipeline (LightRAG) and translate it into a practitioner-facing web-based platform that together support both project management researchers and industrial practitioners with visual explorations of stakeholder networks and interactive query through an LLM-powered chatbot.

2.3 Scope and limitation

The scope of analysis is defined, first, by an intentional reliance on publicly available, official project documents – environmental impact assessments, business plans, board minutes and agendas, contractor/progress reports, consultation summaries and annual reports. This choice prioritises traceability and auditability across large, complex corpora and aligns with prior studies that have successfully used similar sources to analyse stakeholder engagement in infrastructure projects (K. Y. Mok et al., 2017a; Xue et al., 2020b). As with any disclosure-based approach, the outcome measures may reflect documented visibility rather than the full universe of informal interactions (Branco & Rodrigues, 2008; Hassan, 2018).

Second, representation in official records can vary across stakeholder groups. Some minority or marginalised communities may be under-represented relative to institutional actors. This possibility is mitigated, though not eliminated, by the fact that the selected cases operate in jurisdictions where social procurement is widely embedded in the construction sector (Barraket, 2020; Cravero, 2017; Loosemore, 2016; Loosemore & Reid, 2019). Social procurement refers to organisations using their purchasing power to generate social value, such as creating employment opportunities for minority groups (Loosemore & Reid, 2019). The selected case studies were obliged to document opportunities for minority-owned businesses and targeted employment programmes. In such settings, many inclusion initiatives are captured in formal documentation. Nevertheless, complementary sources (e.g. statutory submissions, community reports or targeted interviews) could further enrich the perspective on groups whose participation is less visible in official records. Taken together, these boundaries do not detract from the study's contribution. They clarify the lens: the analysis provides a rigorous, provenance-linked view of how engagement appears in the institutional record, which is precisely the evidence base most relevant to accountability, audit and strategy calibration. Future work can extend this foundation by integrating carefully vetted supplementary sources to broaden coverage while preserving the same standards of transparency and reproducibility.

2.4 Research methodology

This study undertakes a mixed-methods methodology, including data-driven approaches (quantitative research) and case studies (qualitative research). With the open-sourced official project documents from parliament and council libraries in Australia, the UK and the US, this study developed two data-driven models that utilised LLMs and KGs for better understanding and responding to stakeholder complexity and dynamics in stakeholder engagement. To quantify the stakeholder engagement level, this study adopts a dual approach. First, project-management researchers assign 1–10 stakeholder awareness scores (required engagement) to ESG-defined issues from official project records. Second, realised engagement (outcome) is extracted from the same corpus using an LLM-enhanced retrieval pipeline (LightRAG) that links distinct stakeholders to documented events with full provenance. In this way, a scatter plot is developed to serve as a strategic stakeholder engagement map where stakeholder awareness (x) and outcome (y) are then compared in scatter plots with a ±0.10 equivalence band around y = x. This map enables an explicit test of alignment or misalignment in stakeholder engagement strategies. The gap could further be used as guidance for adjusting engagement strategy and resource allocation.

Beyond retrieval, LightRAG also functions as the engine of constructing KGs for stakeholder mapping to better understand stakeholder dynamics. It extracts stakeholder entities and event relationships from the retrieved passages, linking them into a structured network of stakeholders and the events that connect them. The full workflow is deployed in a web-based platform for industrial practitioners that unifies document management, KG construction and visualisation, and an LLM-guided chatbot that answers users' queries based on the KGs constructed within the case study projects.

For the evaluation of the model developed in this study, three megaproject case studies were selected that differ in sectoral, regulatory and ownership contexts: the Scarborough Energy Project (Australia), High Speed Two (HS2, United Kingdom) and California High-Speed Rail (United States).

Literature review

This literature review on stakeholder engagement in mega-infrastructure projects highlights three central themes.

First, section 3.1 highlights current research on stakeholder dynamics and complexity, which has shown that traditional models – such as salience frameworks, static stakeholder maps and survey-based approaches – struggle to capture how stakeholder issues evolve across project phases, largely due to bottlenecks in data availability and the static nature of existing tools.

Second, section 3.2 highlights work on stakeholder engagement themes and models, which points to the need for more dynamic and systematic methods that move beyond single-stakeholder perspectives and one-off surveys, with recent efforts turning to natural language processing (NLP) and AI-based techniques. However, challenges of interpretability and depth remain.

Third, section 3.3 highlights studies of industry practices and trends. It reveals that while digitisation and social network analysis have advanced the field, existing approaches remain labour-intensive, fragmented and insufficient for capturing longitudinal stakeholder interactions.

Together, this body of literature underscores an urgent call for methods that can systematically trace stakeholder issues over time, integrate heterogeneous data sources, and produce outputs that are both auditable and useful for practice.

In response, this study introduces an LLM-enhanced retrieval framework (LightRAG) and KGs to address these gaps. LightRAG provides a scalable and provenance-preserving way to extract stakeholder events and outcomes from large corpora of official project documents, overcoming the limits of static models and subjective surveys. KGs, in turn, organise these extracted entities and relationships into structured, event-mediated networks that reveal both the breadth of participation and the structural positions of stakeholders. By combining LightRAG's analytic capacity with the interpretability of KGs, the study creates a data-driven and auditable framework for analysing stakeholder complexity, enabling both researchers and practitioners to better understand alignment, misfit and the evolving dynamics of stakeholder engagement in megaprojects.

3.1 Definition and importance of addressed topics

Stakeholder dynamics and complexity are central to managing stakeholder engagement in large infrastructure projects. 'Stakeholder dynamics' is the term used to describe how stakeholders change over the course of a lengthy project; stakeholder dynamics may cause profound shifts in the behaviours and mindsets of project participants at different phases (Aaltonen et al., 2015; Rowley, 1997). Existing literatures have provided the classical 'power, legitimacy, urgency' model and stakeholder influence mapping technique to evaluate stakeholder dynamics in each project phase (Mitchell et al., 1997; Olander, 2007). However, few studies have examined stakeholder dynamics at the level of stakeholder issues, limiting dynamic stakeholder engagement in practice (Shen & Xue, 2021). The bottleneck comes from data availability (K. Mok et al., 2017). The conventional model of stakeholder attributes is static and depends on the cognitive data that participating stakeholders supply (Xue et al., 2020b). Given the length and complexity of mega-infrastructure projects (Mok et al., 2015), it is challenging to use questionnaire surveys to track changes in stakeholder issues in a dynamic environment, because statistical evaluation necessitates a large number of samples. In this regard, an advanced method for analysing stakeholder dynamics in megaprojects is desperately needed.

'Stakeholder complexity' refers to the complex interdependencies between stakeholders and their relevant project issues at different project phases (K. Mok et al., 2017). Social network analysis (SNA) has been employed to visualise stakeholder complexity and manage stakeholder relationships in megaprojects (Xue et al., 2020b). The existing SNA enables the representation of the interdependencies of complicated stakeholder relationships and prioritises stakeholder issues when analysing stakeholder complexities. However, the current network-based stakeholder analysis does not provide longitudinal evidence of how stakeholder complexities evolve in various project phases (Luo et al., 2018). The bottleneck comes from the lack of a reliable database to reveal the changing project environment (Mok et al., 2015). Meanwhile, there is not yet an efficient approach to enhance the network-based method to evaluate the evolution of stakeholder complexities in mega-infrastructure projects.

Big data has become a way for researchers to address stakeholder complexities in mega-infrastructure projects. Government data provides stakeholders with access to critical project information, and this publicly available data allows stakeholders to align their expectations and goals, thereby reducing uncertainty and fostering trust between interested parties (Xue et al., 2020b). The use of government data allows for the early identification and quantification of interdependencies between stakeholders, which is critical for managing systemic risk (Moussa & El-Dakhakhni, 2022). Government data can inform the evolution of stakeholder relationships, helping to identify moments of underperformance and adjust strategies accordingly (Xue et al., 2023). This dynamic approach ensures that stakeholder concerns are consistently addressed, leading to increased satisfaction and collaboration (Bahadorestani et al., 2020). While government data can simplify stakeholder interactions, data alone cannot solve all complex problems. Stakeholder engagement and negotiation remain critical to navigating the diverse interests inherent in large projects.

Therefore, researchers started to make breakthroughs with social media data (Zhou et al., 2021). Jiang et al. (2015) and Tang et al. (2017) introduced classical NLP techniques (e.g. topic modelling and sentiment analysis) into online public participation studies in mega-construction projects, which shows the potential of big data-driven engagement studies. Then Chung et al. (2023) leveraged social media data to study online stakeholder engagement strategies, shedding light on big data analysis for stakeholder engagement. Nevertheless, current social media data has shortcomings. First, it cannot show the full picture of the project, as it is unofficial and scattered. Second, it is not objective; it shows the subjective views of active users. To overcome these issues researchers tried to undertake big data-driven stakeholder studies with official project documents from the legislative council (Xue et al., 2020a, 2020c). In a dynamic and complex project environment, official project documents offer a trustworthy and impartial dataset for analysing stakeholder engagement. The research is still in its infancy, though, and needs a thorough data-driven framework before it can suggest strategies for engaging stakeholders in the complex and dynamic world of mega-infrastructure projects.

The existing framework used to analyse stakeholder participation in infrastructure projects comprises three dimensions: situations, implementation and participation. This provides a comprehensive structure for understanding and evaluating stakeholder engagement projects (Xiao & Hao, 2023). However, while the public's participation in infrastructure projects is gaining recognition, there is still a need to develop a deeper and more systematic understanding of the field. Existing guidelines are partially applicable to social media, but researchers and practitioners would benefit from more comprehensive data-driven frameworks for proposing stakeholder engagement strategies in the dynamic and complex environment of large infrastructure projects (Chung et al., 2023).

3.2 Stakeholder engagement themes and models

As stakeholder dynamics are a primary feature of project management, there is a strong need to track the dynamics of stakeholder interactions (De Schepper et al., 2014). At present, stakeholders face different challenges and needs throughout the project life cycle. Therefore, an analytical framework for stakeholder dynamics is needed. This view is echoed by Castelblanco and Guevara (2024). They point out that traditional literature mostly focuses on the interests of a single stakeholder. Most researchers fail to adequately consider the interaction between stakeholders. Traditional methods require empirical data to conduct stakeholder analysis, often obtained at a given time through questionnaire surveys and interviews (K. Mok et al., 2017), which have innate deficiencies in that they do not reflect the dynamic patterns of interactions between stakeholders and the relevant issues that occur over time in mega-infrastructure projects.

To overcome the challenges mentioned above, the classical NLP method, like the topic model, provides an effective method to track stakeholder dynamics by automatically mining the stakeholder behaviours from timestamped text documents relevant to the projects (Hong et al., 2024; Xue et al., 2020a). However, the topic model still faces the problem of unstable interpretability. This research will innovatively introduce the cutting-edge AI-based NLP technique of using LLMs to develop an LLM-enabled stakeholder-topic model, which will significantly improve the accuracy and comprehensiveness of learning stakeholder dynamics from large volumes of project data (Thirunavukarasu et al., 2023). Based on the emerging LLM, it provides deep insight into stakeholder behaviour.

This method addresses the inherent limitations of traditional questionnaire-based surveys by capturing stakeholder interactions in real-time. Recent literature suggests that LLM-based method could empower researchers to capture the emotions and decision-making dynamics of stakeholders. For instance, Lai et al., (2024) proposes a behaviour enhanced LLM to handle online comments from stakeholders. Stakeholders' emotional needs are further analysed in conjunction with their behavioural data. Such frameworks also enable project managers to navigate complex, discrete, and continuous decision-making processes (Lai et al., 2024). Furthermore, LLM-based optimisation systems have also been shown to significantly accelerate design speed by approximately 30 times when compared with traditional methods through the analysis and verification of numerous cases (Qin et al., 2024). Their LLM-based optimisation system can also guarantees the design scheme's affordability and safety (Qin et al., 2024).

KG techniques can be used to describe and visualise complex relationships among stakeholders and their dynamics over time (Chen et al., 2020). Bao et al. (2024) explored the use of KGs in project processes, showing that KGs can effectively manage the flow of information relevant to stakeholders. Researchers use KGs involving geometric and non-geometric information to describe nodes, relationships and properties between different stages of project development. The KGs can record multi-source heterogeneous data generated in the process of project development to provide data support for subsequent analysis and optimisation. Moreover, the combination of KGs and an LLM can provide a more comprehensive perspective for understanding the interaction between stakeholders. It enables project managers to make more forward-looking decisions in a dynamic environment (Zhou et al., 2024).

3.3 Stakeholder engagement trends and industry practices

Although previous studies have shown that it is feasible to tackle the challenges of stakeholder complexity through social network analysis (K. Mok et al., 2017), the traditional static network-based stakeholder analysis fails to reveal the change of complex stakeholder relationships in mega-infrastructure projects. To address this, recent research has emphasised dynamic stakeholder practices, particularly the collaboration between policy makers, companies and supply chain partners. These dynamic stakeholder interactions can contribute to the positive impact of digitisation on supply chain resilience and sustainable performance by driving digitisation standards and providing incentives (El Baz & Ruel, 2024). Similarly, the perspective of stakeholder dynamics extends the multi-stakeholder focus on public–private partnerships (PPP) through a relational governance model. For instance, in mature toll road markets, active stakeholder participation plays a key role in project outcomes for user-paid transport projects (Castelblanco & Guevara, 2024). At the same time, project process knowledge often presents challenges of complex structure, wide distribution and weak correlation. This makes sharing and reuse more difficult. In addition, existing methods for organising process knowledge are labour-intensive and inefficient. Therefore, stakeholders' inputs inevitably require a large amount of manual annotation (Wen et al., 2023).

Increasingly, project management practices are adopting a data-driven approach in the industry. This research will innovatively bring KG technology from the data science domain to analyse stakeholder complexity in a fast-changing project environment. A KG-enabled stakeholder network model will be developed, as the KG has the advantages of representing complex stakeholder relationships with structured knowledge and updating the knowledge of stakeholder complexity over time (Chen et al., 2020).

In addition, more studies are beginning to combine LLMs with KGs. This combination can improve the accuracy of stakeholder analysis. LLMs can support engineering analysis of materials as a tool. It is applied to retrieve key information related to disciplinary fields. Researchers can discover mechanical relationships in different fields of knowledge. However, LLMs may have difficulty recalling the correct information when asked outside of the learning content. KGs can be used to solve this problem, by providing a visual graph structure and rich information (Buehler, 2024). As KG technology and LLMs are constantly being improved, more and more researchers have become interested in developing specialised domain knowledge inference tools or building domain expert systems (Liu et al., 2024). Gao et al. (2024) proposed a KG pattern for bridge maintenance based on an LLM. It improves the integration of bridge design and practical inspection insights. To avoid some risks in the project life cycle, their study proposes an LLM framework constrained by flood knowledge and integrated with geographic information systems (GISs). It can also generate accurate information about floods under the constraints of connection points and relationships in the knowledge graph, which can improve the accuracy of disaster prediction (Zhu et al., 2024). The literature shows that project managers are able to more fully understand the needs of stakeholders by integrating different analytical methods. They can make more informed decisions in a dynamic environment.

Research design and methodology

4.1 Research approach

As explained in section 2.2, the objective of this study is to develop and demonstrate an AI-driven systematic framework to enhance stakeholder management by understanding the complex and dynamic nature of stakeholder engagement and deriving actionable insights for project managers.

Case studies are provided to show that the proposed models are conceptually robust, illustrating how they can generate actionable insights for project managers when engaging diverse stakeholder groups. In achieving this, the first study aim is to develop a practical approach for understanding stakeholder complexity in the engagement strategies of mega-infrastructure projects. The approach compares the expected engagement for each stakeholder issue – defined as stakeholder awareness and manually coded by project professionals from historical project reports – with the realised engagement, or stakeholder outcome, evidenced in official records and extracted through a LightRAG-based retrieval pipeline. By examining the alignment or misalignment between stakeholder awareness and stakeholder outcomes on different issues (Kujala et al., 2022), this study operationalises stakeholder complexity through the magnitude and direction of the awareness–outcome gap. This provides project teams with practical insights to adjust engagement strategies and allocate resources more effectively.

In facilitating this research design, three research approach layers were adopted. First, stakeholder issues are specified using an ESG-derived framework (Zhang et al., 2024) to ensure comprehensive coverage of ESG concerns while retaining project specificity. Then, five project management researchers assign 1–10 awareness scores to stakeholder–issue pairs after closely reading historical project records. The scores are then aggregated by issue to form a criterion-level awareness baseline.

Second, a LightRAG pipeline ingests the official corpus, applies delimiter-aware chunking with overlap, embeds chunks and queries, and retrieves issue-relevant evidence at a similarity threshold tuned for precision. The pipeline constrains synthesis to event-mediated associations, and preserves provenance so that every stakeholder involved in an issue can be traced to the source text.

Third, stakeholder engagement awareness (x) and outcome (y) are analysed via scatter plots with a ±0.10 practical-equivalence band around the identity line. Points inside the band indicate proportional conversion; points above signal over-realisation that warrants stewardship rather than more awareness work; points below flag under-realisation and the need for targeted conversion measures (capacity, channels, timing, trust).

Furthermore, to better facilitate stakeholder management in terms of the stakeholder dynamics, the study builds KGs from LightRAG-extracted entities and event contexts, enforcing two modelling rules: only stakeholders and events are permitted as node types, and all stakeholder–stakeholder ties are event-mediated. This design keeps relationships interpretable and auditable, and it supports centrality-based filtering to surface 'key stakeholders' from large graphs without discarding complexity. The full workflow is deployed in a platform for industrial practitioners that unifies document management, KG construction and visualisation, and an LLM-guided chatbot that answers issue-level questions with reference to underlying sources. The whole research approach is evaluated comparatively across three megaprojects that differ in sectoral, regulatory and ownership contexts: Scarborough Energy Project (Australia), HS2 (United Kingdom) and California High-Speed Rail (United States). The overall research methodology workflow is illustrated in Figure 1.

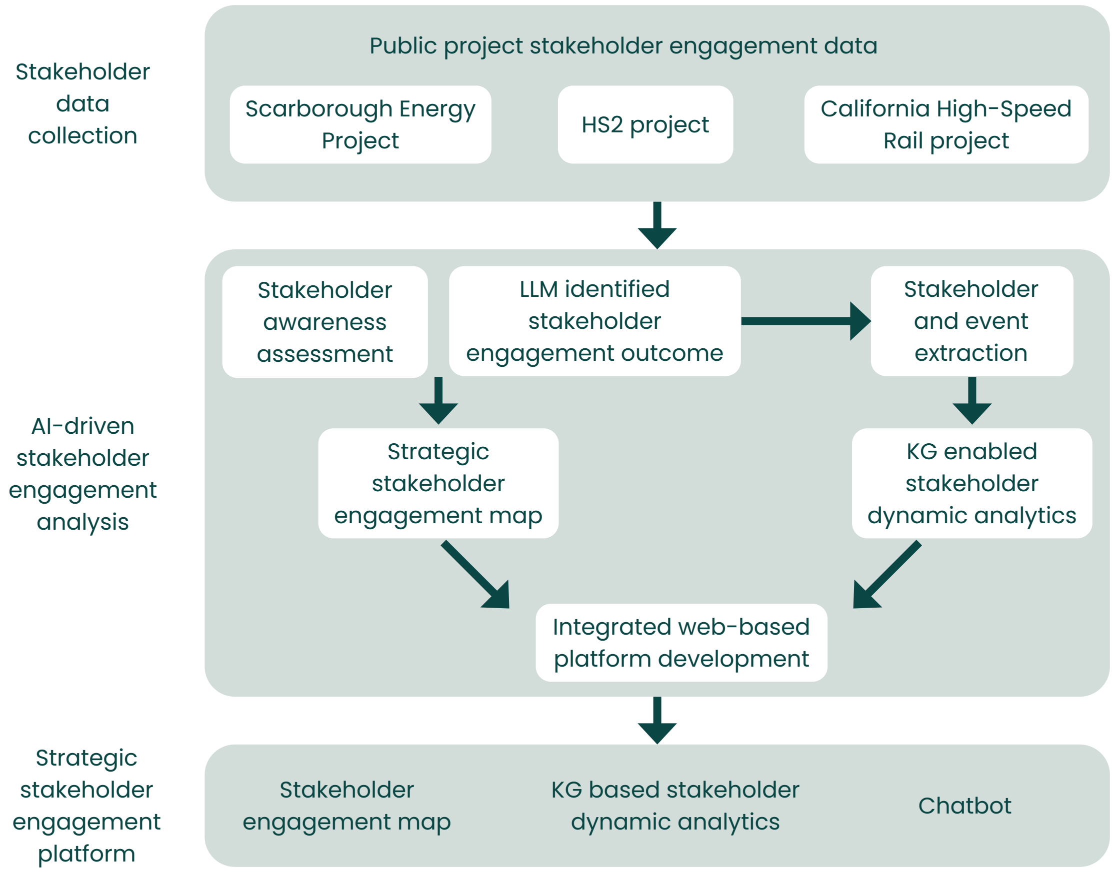

Figure 1: The research methodology workflow

4.2 Data collection methods

The data for this study mainly comes from publicly available stakeholder engagement project documents. The documents collected include:

- project environmental impact assessment
- project proposal
- business plan from the project host organisation
- board meeting minutes and agendas
- contractor and progress report
- community engagement and consultation report
- project annual report.

After sampled reports were collected and the dataset was prepared for the construction of KGs and further analysis, the data was pre-processed into structured tabular data. This process was conducted by five project management researchers manually screening the prepared project record and recording the information needed in a spreadsheet. As per the core of stakeholder theory introduced by Freeman (1984), this process begins with the identification of key stakeholders of the three case projects.

Inspired by previous studies on stakeholder management research in project management (Xue et al., 2023, 2020b), each identified stakeholder was categorised into different interest groups (e.g. government, regulator, council, commercial/contractor, NGO, local/community). Furthermore, scoring metrics were developed for each stakeholder and various related issues. To identify stakeholder-related issues, this study employed existing project-level ESG metrics (Zhang et al., 2024) as a case study, which will be discussed in Chapter 5. Researchers were instructed to manually score the required engagement level for each stakeholder in terms of how this stakeholder could have an impact on this issue from their professional judgement after screening the sampled document. These criteria were assigned a score from 1 to 10, with detailed descriptions shown in Table 1. These manual scores form the issue-level awareness baseline against which LightRAG-derived stakeholder outcomes (realised engagement evidenced in official records) are later compared.

Table 1: Scoring rubric for required stakeholder engagement levels on project issues

Score range	Description
1–2	Negligible effect and little to no need for engagement
3–4	Limited engagement required
5–6	Noticeable effect that calls for short-term engagement
7–8	Substantial change requiring significant engagement effort
9–10	Profound, long-term effects requiring an extensive, sustained engagement strategy

Regarding data for facilitating the LightRAG process, the role of chunking is significant in terms of the chunk's size and overlap (Lyu et al., 2025), which enhances the accuracy of information retrieval by breaking down a large text corpus into segments. All official project reports collected were first converted into machine-readable PDF files and then segmented into smaller chunks. This chunking step ensures that each unit of text is both contextually coherent (e.g. a table section, modelling result or regulatory statement) and computationally manageable for LightRAG processing. Each chunk is tagged with its document metadata in terms of issuer, date, type, page range and project, which makes the retrieval results traceable. LightRAG then embeds both chunks and queries into a vector space and retrieves relevant passages by similarity search.

Downstream, a threshold-based algorithm (Hönen et al., 2023) monitors the frequency and significance of salient signals (recurring keywords, phrases and thematic markers). When a signal exceeds a predefined threshold, the system expands retrieval around that context to initiate deeper analysis. Retrieved evidence is normalised to extract stakeholder events and counts, which are then categorised and aggregated by issue under different criteria and thresholds to support multiple analytical perspectives and sensitivity settings. This integration of LightRAG with thresholding makes the pipeline adaptive to corpus size and heterogeneity, ensuring thorough, automated identification of stakeholder-specific insights while capturing both broad trends and fine-grained details.

4.3 Data analysis methods

The analytical stack comprises four elements: construction of the x-axis (awareness), construction of the y-axis (outcome), alignment visualisation and modelling, and graph-based exploration. Awareness totals originate from the expert rubric described in Table 1. Outcome scores are generated programmatically by the LightRAG pipeline. For each issue, a query is formulated for LLMs with a concise definition and key characteristics, so that retrieval favours issue-relevant passages over generic mentions.

The system operates in hybrid mode, first deriving keywords and patterns from the definition, then routing the query through both vector and graph indexes. Stakeholders and event-centred relationships are returned with their supporting chunks. A constrained synthesis step produces a structured table that binds each distinct stakeholder to a specific event passage and the source chunk identifiers used. The criterion-level outcome is computed as the cardinality of unique, evidenced stakeholders meeting the threshold for that issue. Because raw counts depend on corpus size and documentation practice, values are min–max normalised to the unit interval [0,1] to enable comparison across issues and cases. The resulting awareness–outcome dataset is reproducible and audit-ready: every point on the plot resolves to citable text.

In previous stakeholder management literature, several studies reported that the 'dark side' of stakeholder engagement may occur due to misalignment or ignorance of strategies in different contexts (Abosag et al., 2016; Kujala et al., 2022). To present the 'alignment' in the case study, it is examined with issue-level scatter plots (e.g. Figure 2), which adopt a ±0.10 practical-equivalence band around the identity line $y = x$. Points inside the band indicate proportional conversion of awareness into participation and suggest that the prevailing engagement strategy design is efficient. Points above $y = x + 0.10$ indicate over-realisation, where realised engagement exceeds expectation; such cases typically call for consolidation and channel co-ordination rather than additional awareness spend. Points below $y = x - 0.10$ indicate under-realisation, where intended engagement has not converted into documented participation; these are prioritised for targeted interventions that reduce the cost of engagement (capacity-building, channel redesign, sequencing with decision gates, trust routines).

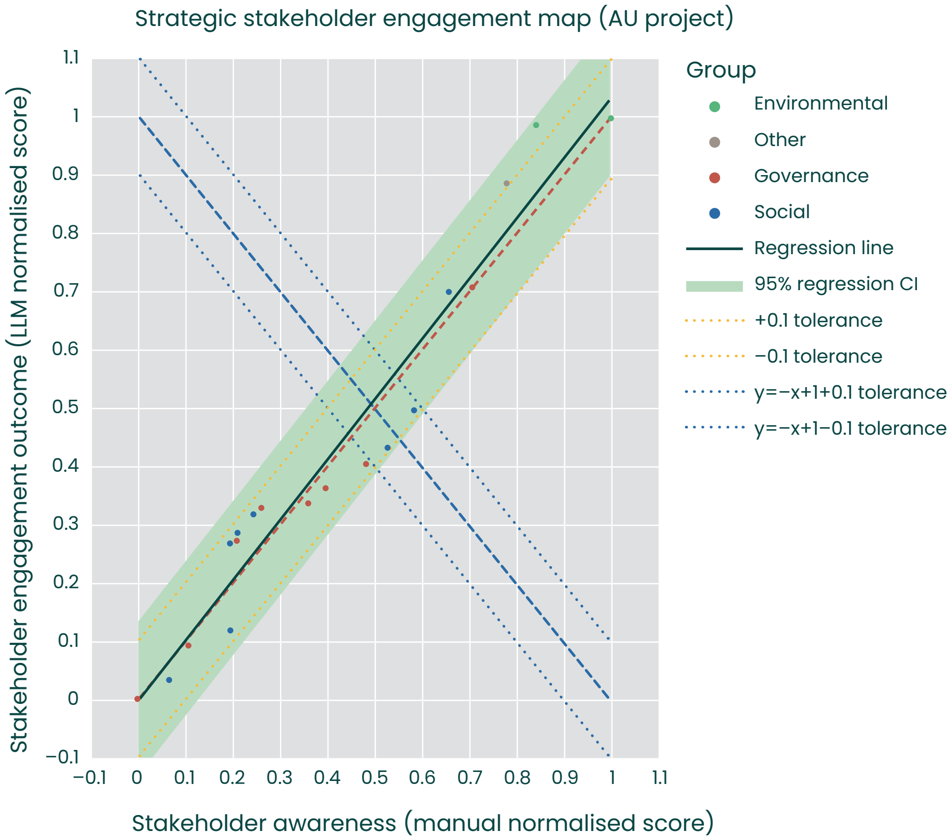

Figure 2: Visualisation of comparison of manually scoring and LLM scoring of stakeholder issues (Australia project)

To further support industrial practitioners in understanding stakeholder dynamics, KGs can be constructed from LightRAG outputs. The graphs permit only two node types: stakeholders and events. Each node is allowed to connect with another node type only. In another words, only stakeholder-event edges are allowed in a graph. This means each graph only contains stakeholder-event connections, guaranteeing that every relationship is linked to a specific event and is completely traceable. To keep graphs clean and consistent, any entities that might have different names or representations are converted into a single, standardised name. This graph-building and analysis capability is part of a user-friendly web platform that follows a three-step research process:

- a knowledge base for ingestion and KG building
- an interactive network visualiser for exploration and export
- an LLM-powered chatbot that answers specific questions about issues, providing answers linked directly back to the original source data for verification.

Because the settings and safety measures used for analysis are consistent across both the research pipeline and the web platform, any figures or charts created for a final report can be instantly regenerated by the user with the exact same parameters. This consistency significantly enhances reproducibility for both researchers and practitioners.

Findings

5.1 Overview of the case studies

To test the capability of LLM-driven techniques for analysing stakeholder engagement outcomes, the project-level ESG metrics developed by Zhang et al. (2024) were adopted as a guideline for defining and developing stakeholder issues. Three mega-infrastructure projects were selected: the Scarborough Energy Project in Australia, the High Speed Two (HS2) project in the UK, and the California High-Speed Rail project in the US.

The Scarborough Energy Project, led by Woodside Energy Ltd (Woodside), is a major liquefied natural gas (LNG) development located approximately 375 km off the coast of Karratha, Western Australia, in the Carnarvon Basin. The project was proposed in 2020 by Woodside, the leading Australian petroleum exploration and production company, and the first cargo was targeted in 2023 (Woodside Energy, 2020). From a sustainability perspective, the Scarborough gas reservoir is characterised by an exceptionally low CO_2 content of less than 0.1%, and design efficiencies make it one of the lowest carbon intensity LNG sources for the Asian export market (Woodside Energy, 2025). Socially, the project emphasises community development, aboriginal participation and local employment. By 2024, it had awarded over A$190m in contracts to Karratha-based businesses and A$104m to aboriginal businesses across Australia. The project also initiated partnerships with local educational and training organisations, supporting workforce development and engagement with aboriginal communities (Woodside Energy, 2024). Extensive cultural heritage assessments, both archaeological and ethnographic, have also been conducted in collaboration with related local Aboriginal organisations.

The UK's HS2 project is a government-led transport public infrastructure project intended to improve rail connectivity between London and the north of England (High Speed Two Ltd, 2017). The scale and complexity of the HS2 project, coupled with its controversial nature due to environmental and property displacement concerns, make it an ideal case for studying stakeholder engagement in public infrastructure projects. Stakeholders have been actively involved in consultations regarding route alignments, environmental mitigation measures and community compensation schemes, underscoring the importance of stakeholder influence in shaping project outcomes.

Within the United States, California High-Speed Rail is one of the most ambitious and revolutionary infrastructure projects. The project, which was approved by California voters in 2008 under Proposition 1A, aims to connect major metropolitan centres such as San Francisco and Los Angeles with an electrified rail system spanning 800 miles. Eventually, it will also connect Sacramento and San Diego. The California High-Speed Rail Authority is carrying out the project using a phased delivery model. The first operating segment will span 171 miles in the Central Valley of California between Merced and Bakersfield (California High-Speed Rail Authority, 2015). The Central Valley segment has served as the testing ground for construction strategies, stakeholder co-ordination and funding mechanisms. As of 2023, the Authority had created over 12,000 jobs, cleared 422 miles environmentally and received $3.3bn in new federal funding – signalling renewed federal support for high-speed rail in the US (California High-Speed Rail Authority, 2024).

Environmental, social and governance (ESG) frameworks have attracted extensive research and industrial interest in recent years. They are comprehensive frameworks that enable business practice and performance to be evaluated across a range of sustainability and ethical issues (Adewumi et al., 2024; Gillan et al., 2021). For example, environmental performance measures a company's energy consumption, waste generation, natural resource utilisation, and impact on ecosystems and habitats. Social performance encompasses stakeholder engagement, community relations, diversity and inclusion initiatives, and health and safety practices. Governance evaluates a company's policies, practices and procedures that guide business decisions while ensuring legal compliance and meeting obligations (Adewumi et al., 2024). Reber et al. (2022) suggest that ESG reporting offers a more comprehensive perspective on organisational performance and can improve the quality of corporate information, enhance alignment with societal expectations regarding accountability and transparency, support sustainability objectives, and enhance an organisation's reputation and public image.

From a project management perspective, the importance of ESG reporting extends beyond organisational benefits. The Project Management Institute emphasises that actively reporting and sharing ESG initiative outcomes can increase the visibility of a project's impact, create buy-in for broader ESG visions and ensure future ESG projects achieve maximum value (PMI, 2023). In line with this, recent studies have emphasised the need for effective stakeholder engagement in addressing ESG-related issues in projects (Derakhshan et al., 2019; Zhang et al., 2024; Zuofa & Ochieng, 2016). Building on these insights, this study adopts the project-level ESG framework developed by Zhang et al. (2024) as a basis for comparing stakeholder awareness (expert-judged expectations of engagement) with stakeholder outcomes (realised engagement evidenced in project disclosures), thereby providing a new lens for analysing stakeholder complexity through ESG-related project issues.

5.2 Stakeholder complexity in the case studies

Beyond the plot chart in Figure 2, all plot charts of three case study projects are shown in Figure 3. In the region to the left of the yellow tolerance lines ($y = x + 0.1$), the scatter plot indicates the scenario where the actual stakeholder engagement exceeds the expert-assessed awareness levels. This suggests that stakeholders, despite lower perceived consciousness regarding project-related activities, exhibit a higher degree of engagement than anticipated. Within the central region bounded by the yellow tolerance lines ($y = x \pm 0.1$) in the HS2 project, the plot reflects a balanced relationship between expert-assessed awareness and actual outcome in stakeholder engagement. This area represents an optimal tolerance zone where the input–output ratio demonstrates minimal deviation, suggesting an efficient and sustainable engagement strategy. The alignment of data points in this region indicates that the project team's efforts to cultivate awareness are proportionate to the observed engagement levels, supporting a model of effective resource utilisation and stakeholder management. This equilibrium is critical for maintaining project momentum without overextending efforts on awareness campaigns. To the right of the yellow tolerance lines ($y = x - 0.1$), the scatter plot reveals a situation where expert-assessed awareness significantly exceeds actual stakeholder engagement. This indicates a high level of input directed towards raising stakeholder consciousness. However, the resulting level of engagement was lower than the team had expected. A single data point, representing resource sharing capability, resides in this quadrant. This discrepancy suggests that despite substantial efforts to enhance awareness, external or contextual factors may be limiting stakeholder involvement. This scenario warrants a reevaluation of engagement strategies, potentially focusing on addressing barriers to engagement rather than increasing awareness inputs.

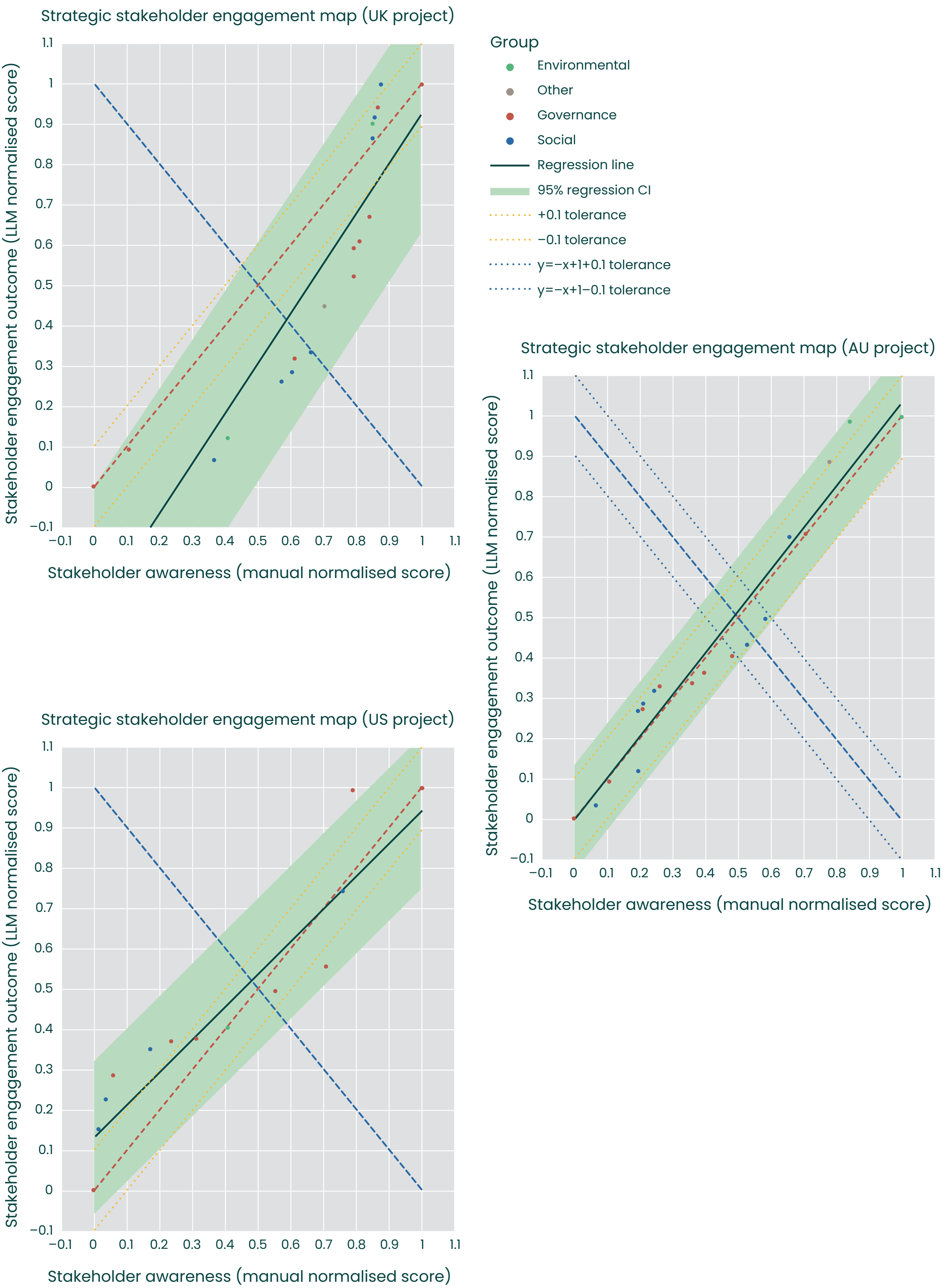

Figure 3: Stakeholder awareness–outcome scatter plots for the three case studies

5.3 AI-driven stakeholder dynamics analytic platform

After the comparison and analysis of stakeholder awareness and outcome had been completed, KGs were constructed and visualised. Additionally, to better facilitate the practice in stakeholder management for industrial practitioners, a platform integrating KG construction and an LLM chatbot was developed. The user interface for the platform is shown in Figure 4. There are two main functions within this stakeholder management platform: Knowledge Base and Chat. All of the completed coding was integrated within this platform, which includes KG construction, KG-based stakeholder network visualisation and a chatbot that responds to user queries based on the stakeholder knowledge graph.

The platform is designed to make our research pipeline usable by practitioners who must interrogate large volumes of official project documentation and translate insights into engagement decisions. It unifies three capabilities behind a simple web interface:

- a knowledge base, where users upload and manage project document sets and trigger KG construction
- an interactive KG visualiser that renders stakeholders (organisations) and events (documented actions or decisions) as a network
- an LLM-powered chatbot that answers issue-level questions using the KGs and the source texts.

Although the research design is organised around two primary functions – Knowledge Base and Chat – we also expose a Charts workspace to view alignment figures (scatter plots and heatmaps) generated in the analysis stage, so that analysts can move from computation to interpretation in one place.

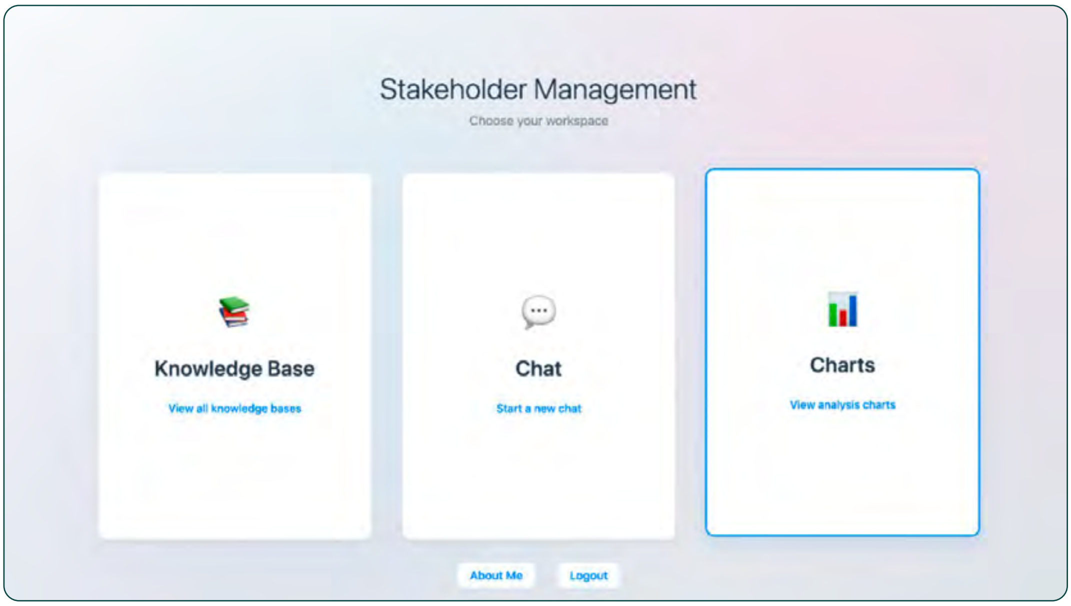

Figure 4: The user interface for the stakeholder management platform

The platform is backed by a LightRAG implementation, and uses the same parameters and safeguards described in the methodology (e.g. hybrid retrieval, similarity thresholding and provenance-preserving chunking). Within the Knowledge Base function (Figure 5), users can create a new project repository and upload files (in .txt, .pdf, .doc or .docx format). Upon upload, the ingestion service detects file encodings, converts documents into machine-readable text and segments them into context-coherent chunks using a token-aware function with overlap. This segmentation respects semantic split markers (e.g. headings) when available and otherwise falls back to sliding windows of up to around 1,024 tokens with a 128-token overlap. Each chunk is recorded with a unique identifier, content, token length and order index.

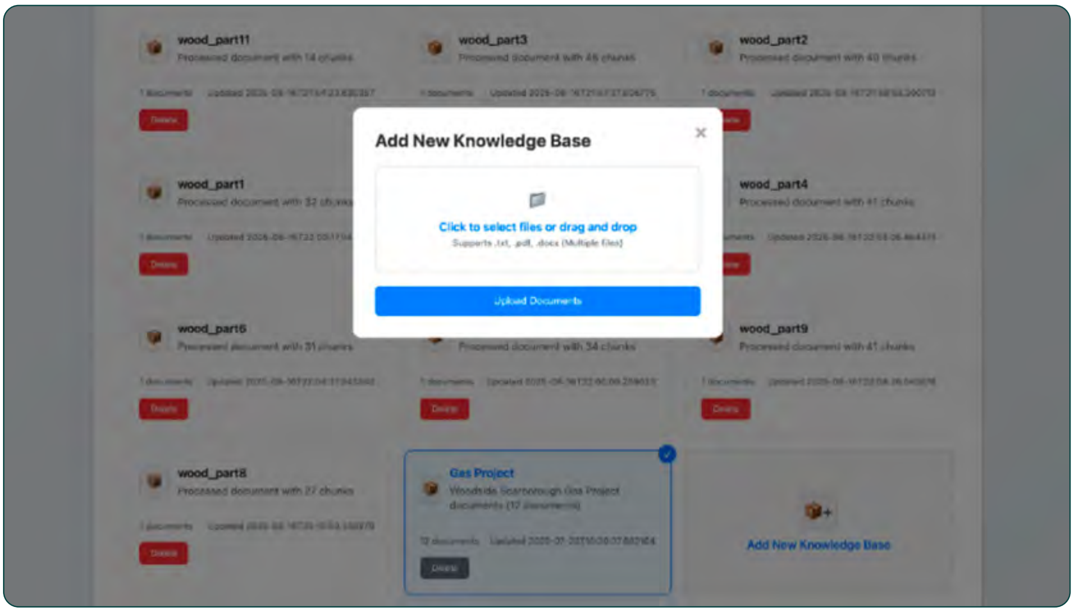

Figure 5: The user interface for the Knowledge Base function

These design choices make retrieval both efficient and auditable because every fact later surfaced by the chatbot can be traced to one or more original chunks in the corpus. After the file is uploaded, a visualised KG will be generated, representing the stakeholder network within the project (Figure 6). The platform executes an LLM-guided entity-event extractor to build the KGs, with two strict rules guiding this process:

- Only two entity types are allowed: stakeholder and event.
- Direct stakeholder-to-stakeholder edges are disallowed, and all connections must be event-mediated (stakeholder ↔ event ↔ stakeholder).

In parallel, the system constructs vector indexes for entities and relationships, including their names, descriptions and keywords. This dual index supports both local (entity-centric) and global (relationship-centric) retrieval downstream.

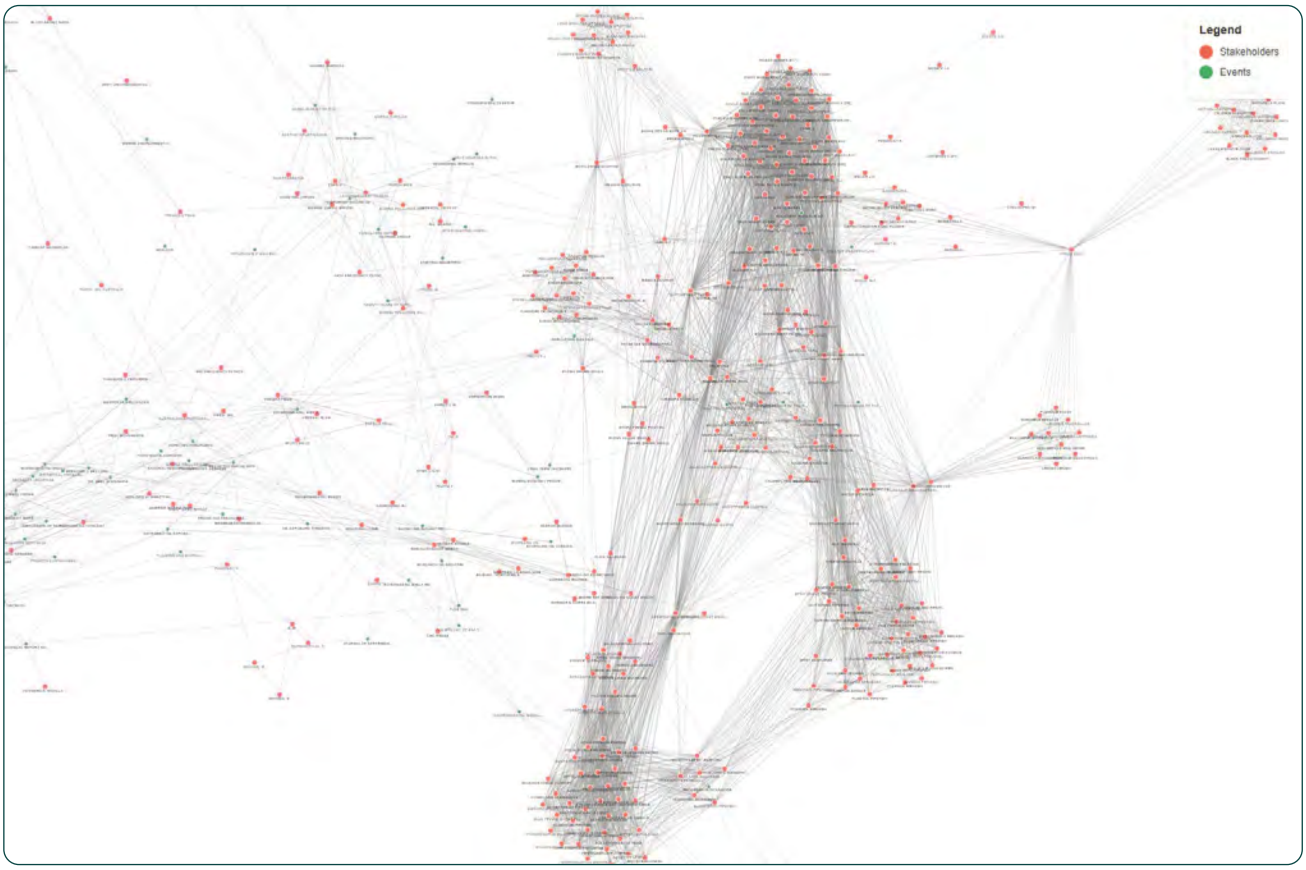

Figure 6: A visualised KG of a stakeholder network

Discussion and analysis

6.1 Complexity in stakeholder engagement

Stakeholder complexity has been widely defined in the literature as the dense interdependencies among stakeholders and the issues they are connected to throughout the project life cycle (K. Y. Mok et al., 2017a; Xue et al., 2023). In responding effectively to stakeholder complexity, project managers are required to choose fit-for-purpose engagement strategies at the issue level. However, misalignments often arise because of perception gaps between how stakeholders view their role and how project professionals anticipate it (Kujala et al., 2022), and from the way stakeholder influence shifts dynamically across the long span of megaprojects (Gonzalez-Porras et al., 2021). These challenges make it essential for project teams to continuously adjust their strategies, using feedback from earlier engagement practices to adjust their approach.

Furthermore, due to the development and advancement of LLMs, there is a growing trend for LLM applications to expand beyond trivial conversation to complex data analytics (Oh et al., 2024). In practice, management teams increasingly deploy LLMs to process large, heterogeneous evidence bases and to inform data-driven decisions (Rao et al., 2023). Yet, within project management, and specifically in stakeholder engagement research, LLM applications remain relatively underexplored (Aaltonen et al., 2024). This study addresses that gap by examining how LLMs can be leveraged to analyse stakeholder engagement through historical official project disclosures, and by comparing their outputs with professional judgement to detect where alignment or misfit occurs between anticipated and realised engagement.

Regarding the professional's judgement in stakeholder engagement, according to previous literature, efforts were made to shed light on stakeholder expectations, stakeholder perception and public consciousness of stakeholders (Bahadorestani et al., 2020; Haldar et al., 2021; Liu, 2022). But these constructs do not precisely capture the issue-level, prescriptive assessment needed here. We operationalise stakeholder awareness as the project team's informed appraisal of how intensively specific stakeholders ought to be engaged on a given issue. By contrasting this awareness with LLM-evidenced outcomes, we identify gaps and translate them into adjustments to engagement strategy.

In organisational literature, the concept of 'fit' has been identified as critical in understanding and improving organisational performance (Bundy et al., 2018; Drazin & Van de Ven, 1985; Hofer & Schendel, 1978). Combined with stakeholder theory, literature has also shed light on the 'fit' between corporation and stakeholder to explain co-operative behaviours and enhance corporate survival (Bundy et al., 2018; Tantalo & Priem, 2016). Furthermore, Bundy et al. (2018) demonstrate that the quality of the relationship between organisation and stakeholder relies on the extent of fit or misfit across two dimensions: value congruence and strategic complementarity. Their results show that high congruence on both dimensions produces co-operation, partial alignment produces compromise, and misalignment on both generates conflict. Importantly, they highlight that this fit is dynamic, shifting as stakeholder values, organisational strategies and external conditions evolve. In this study, the awareness–outcome gap functions as an empirical signal of stakeholder engagement alignment or misalignment (fit or misfit). Based on a historical stakeholder-related project report, project professionals may assume high awareness and design engagement strategies accordingly, but if project records show weak or absent participation, the result mirrors Bundy et al.'s (2018) notion of misfit. That is, where stakeholder engagement awareness is high, but the outcome is evidenced to be low, this misalignment indicates an under-engagement with stakeholders, where further efforts and resources are needed for allocation.

According to the scatter plots in Figure 3, the misalignment between the project professional's judgement and actual outcome in stakeholder engagement strategies can be observed across all three projects and different stakeholder issues. For example, in the scatter plot for the US project (California High-Speed Rail), a single point representing the social factor 'Long-term Commitment' is observed to the left of the tolerance zone. This suggests that despite a lower expert-assessed awareness of this aspect, the actual stakeholder engagement evidence is notably high. This may indicate an over-engagement scenario, where the project team may consider allocating less effort to engagement regarding this issue.

The slopes of the case-level regressions (outcome on awareness) also provide a succinct gauge of portfolio alignment. A slope below 1 indicates that increases in expected engagement are not matched one-for-one by realised engagement (portfolio-level under-realisation), whereas a slope above 1 implies slight amplification in what is documented relative to expectations. The UK coefficient of 0.82 suggests that, on average, the outcome grows more slowly than awareness. The US coefficient of 1.13 indicates modest amplification overall, even though several criteria are under-realised and call for targeted fixes. The Australian coefficient of 1.03 is near-linear, consistent with the concentration of points inside the tolerance band and selective over-realisation on environmental issues. Together with the scatter plots, these coefficients offer a clear reading: the UK portfolio requires attention to conversion losses; the US portfolio mixes balanced issues with a sizeable under-realised cluster; the Australia portfolio is broadly proportionate with pockets that merit stewardship rather than additional awareness.

6.2 AI-driven stakeholder dynamic analytics

The organisation's capabilities to possess and absorb knowledge are widely recognised as a critical foundation of competitive advantage (Cohen & Levinthal, 1990; Mueller et al., 2011). In project management research, knowledge management is one of the key success factors where an organisation's knowledge base can be produced and utilised (Cope et al., 2006; Gasik, 2011). With the advancement of digital transformation, a web-based system for a project knowledge management platform has been developed and applied in the industry (Elwakil & Zayed, 2018; Iheukwumere-Esotu & Yunusa-Kaltungo, 2022; Kivrak et al., 2008). Despite these advances, however, relatively little attention has been paid to the development of knowledge management platforms specifically oriented towards stakeholder management. This gap is significant, given that stakeholder engagement constitutes one of the most knowledge-intensive aspects of project work, involving complex networks of actors, shifting expectations, and contested issues that require systematic capture, organisation and analysis. Extending knowledge management practices into stakeholder management would allow project organisations not only to store historical records of interactions but also to transform these records into actionable intelligence for future engagement strategies.

The platform developed in this study functions as a knowledge management tool for stakeholder management, extending beyond the passive storage of past project documents. It is capable of automatically pre-processing project data and generating structured outputs for knowledge graph construction (Figure 6). To capture stakeholder dynamics over time, the platform can also generate multiple network snapshots across different project phases in response to user prompts. Examples of the stakeholder networks generated for distinct construction phases are shown in Figure 7.

In addition, the platform allows users to interactively explore a knowledge graph by clicking on individual nodes, which reveals detailed information about the selected stakeholder or event. This interactive functionality provides project practitioners with immediate access to contextual data – such as stakeholders' roles, associated issues and linked events – thereby offering a richer understanding of how stakeholders are embedded within the network. As illustrated in Figure 8, such dynamic interaction moves beyond static stakeholder maps by enabling practitioners to trace specific connections, uncover patterns of influence and identify clusters of activity that may otherwise remain hidden. By linking visualised nodes directly to underlying project records, the platform could enhance interpretability and evidence-based decision-making in stakeholder engagement strategies.

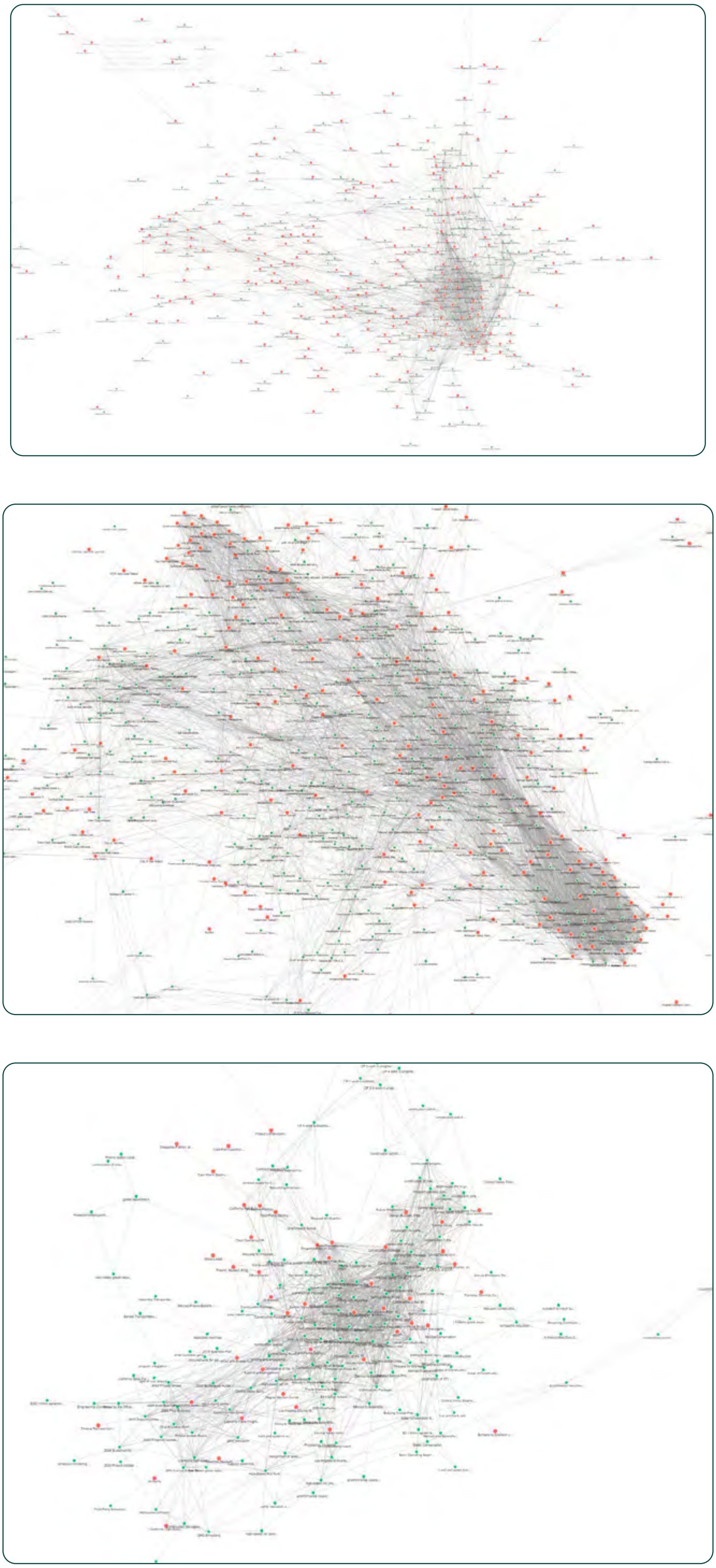

Figure 7: Stakeholder networks during different construction phases (US project)

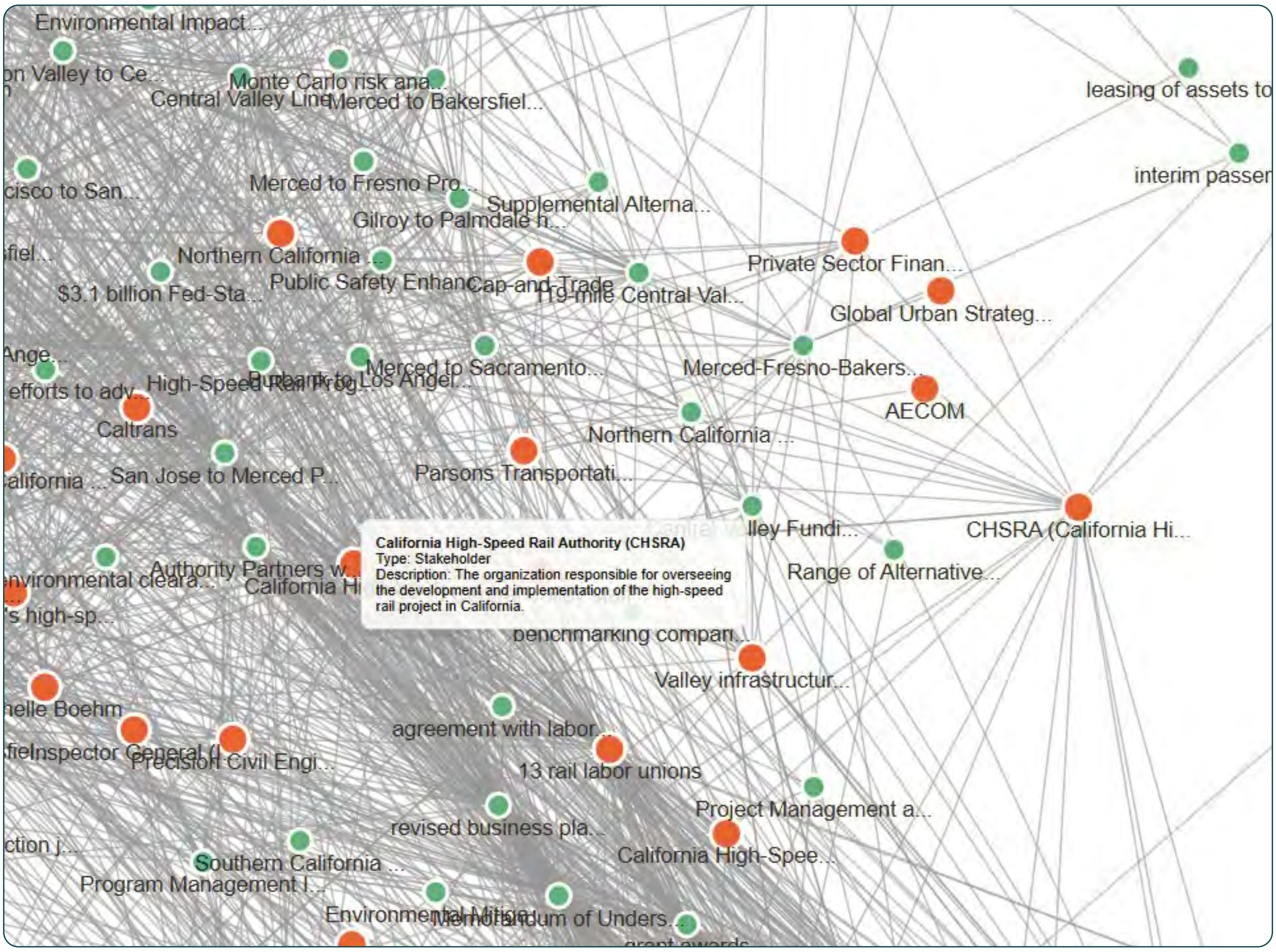

Figure 8: Stakeholder network node description

The first chatbot, ELIZA, was introduced by Weizenbaum (1966). It was developed based on a language model and could respond to users by identifying keywords with given rules. Regarding the input processing and response generation methods, chatbots can be categorised into rule-based, retrieval-based and generative models (Adamopoulou & Moussiades, 2020). A rule-based model, such as ELIZA and some online chatbots, is developed on a predefined fixed set of rules or decision trees, and thus responds to conversations by recognising the input text without creating new texts (Adamopoulou & Moussiades, 2020; Yang et al., 2024). A retrieval-based model draws from a predefined knowledge base of possible responses to select the most relevant answer (Yang et al., 2024). The generative model provides answers based on both current and previous users' messages, typically powered by advanced neural networks, which could have better contextual understanding and engage in more open-ended conversation (Adamopoulou & Moussiades, 2020).

Within the project management field, there are a few existing research and industrial practices of using a chatbot system supported by an LLM in enhancing project design conversations. Regarding the effectiveness of communication, and compared with traditional surveys, chatbots can be more effective and preferable (Dortheimer et al., 2024; Te Pas et al., 2020). Chatbot interactions might also be seen as more enjoyable, which could lead to increased engagement with diverse communities and greater stakeholder participation (Dortheimer et al., 2024). For this project, a chatbot was developed as the core interaction surface for industrial users (Figure 9).

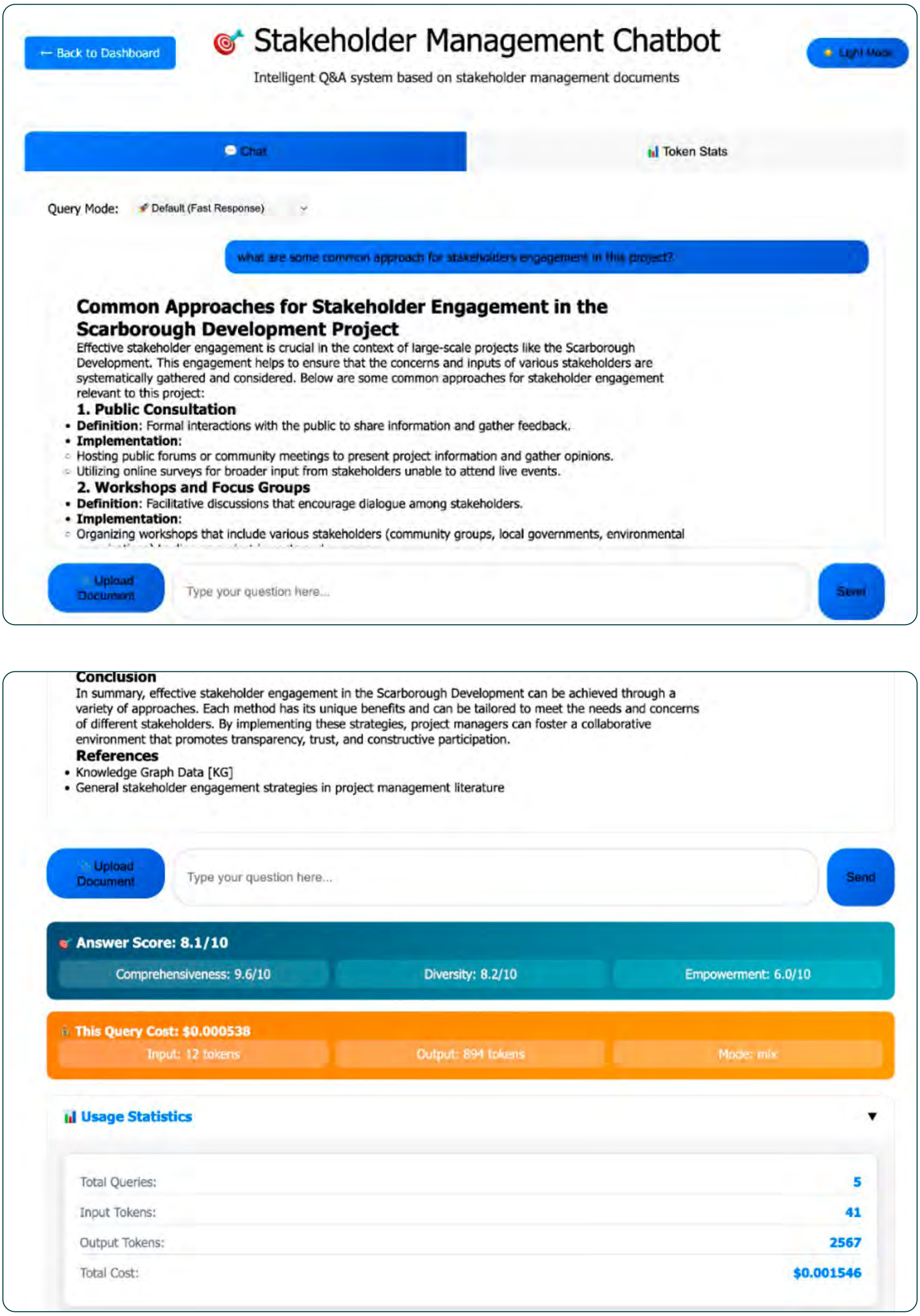

Figure 9: Stakeholder management chatbot

For instance, in the case of HS2, the constructed knowledge graph identified 2,435 stakeholders connected through 3,697 documented events. While this comprehensive mapping provides a rich representation of stakeholder involvement, the sheer scale of nodes and relationships can overwhelm users seeking to conduct targeted analyses. To address this challenge, the platform allows users to specify a limited number of stakeholders to be highlighted as ‘key stakeholders’ for their investigation. These key actors are filtered and prioritised according to their centrality measures within the KG network, ensuring that the most influential stakeholders are surfaced for closer examination while maintaining analytical efficiency. Figure 10 illustrates an example query posed to the LLM: ‘show me the top 20 stakeholders’.

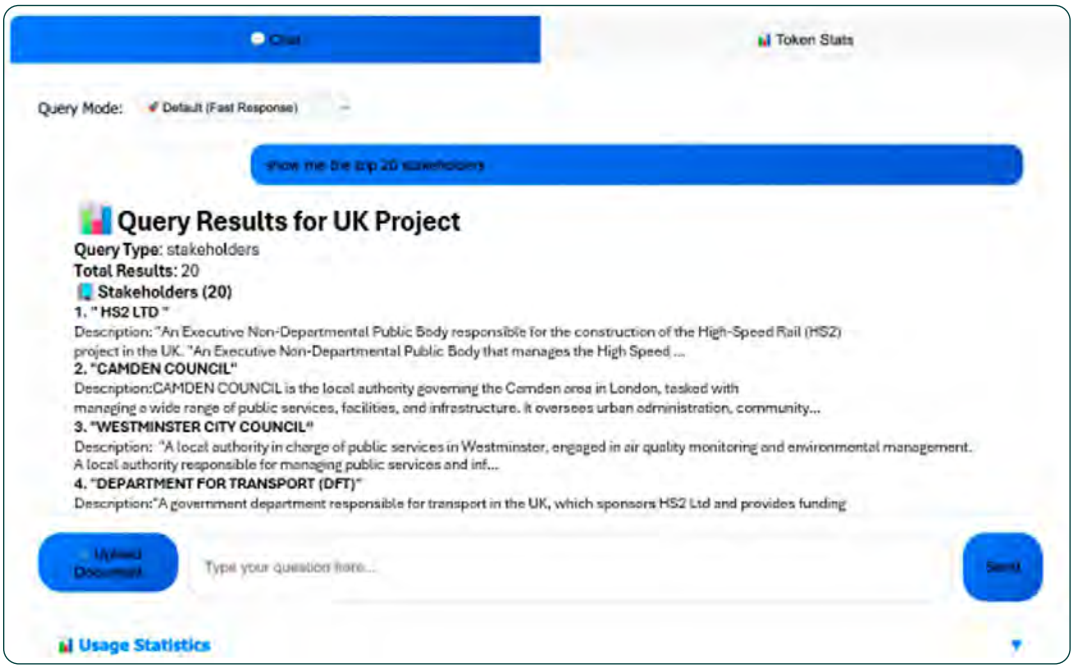

Figure 10: Snapshot of the chatbot’s list of key stakeholders

Conclusion

7.1 Summary of the research

This study presents an AI-driven framework and web-based platform designed to enhance automated stakeholder engagement by addressing the complexity and dynamic nature of large infrastructure projects. The proposed framework integrates large language models (LLMs) with retrieval-augmented analysis to extract stakeholder events, relationships and outcomes from large corpora of official project documents. Case studies were provided to show that the proposed models are conceptually robust, illustrating how they can generate actionable insights for project managers when engaging diverse stakeholder groups.

In addressing stakeholder complexity, this study operated at the level of stakeholder issues – rather than solely at the level of actors – and the research has introduced a practical framework for comparing stakeholder awareness, as assessed by project professionals, with stakeholder outcome, as evidenced in official project disclosures processed through a LightRAG pipeline. The magnitude and direction of the awareness–outcome gap has been used as a direct indicator of alignment or misalignment, thereby providing project teams with a structured means of diagnosing when engagement strategies are proportionate, when they are over-realised and when they fall short. The comparative analysis of three emblematic case studies – Australia's Scarborough Energy Project, the UK's High Speed Two (HS2) project and the California High-Speed Rail project – demonstrated the versatility of this approach. Each case revealed distinctive patterns of alignment and misalignment that reflect both the institutional environments and the types of issues at stake. The cross-case analysis produced distinct yet interpretable profiles.

To capture stakeholder dynamics more fully, the study also develops knowledge graphs (KGs) that map stakeholders and their event-mediated interactions over time. Importantly, these analytical capabilities are integrated into a web-based platform, which combines corpus ingestion, automated evidence retrieval, awareness–outcome alignment charts, KG visualisation and an LLM-enhanced chatbot for provenance-linked Q and A. Together, the framework and the tools demonstrate how AI techniques – particularly LLMs and KGs – can transform fragmented records into actionable intelligence, offering practitioners a scalable and transparent way to navigate the socio-technical complexity of megaprojects and to recalibrate engagement strategies in real time.

7.2 Contributions and key takeaways

This study contributes to debates on big data in stakeholder management. It demonstrates the value of official project documents as a structured, reliable data source. Based on the strategic stakeholder engagement map, and by reading the awareness–outcome gap, the research operationalises stakeholder complexity as a measurable construct that points directly to managerial action. Technically, the application of LightRAG and KGs addresses the reliability issues often found in standard AI tools; by ensuring all insights are strictly traceable to source evidence, the framework supports the high standards of accountability required in project governance. The resulting web-based platform effectively bridges the gap between raw archival data and immediate decision support, turning static records into active diagnostic tools.

Key takeaways:

- **Monitor the stakeholder engagement strategy execution gap**: Treat the issue-level awareness–outcome gap in the strategic stakeholder engagement map as the core indicator of engagement effectiveness in better responding to stakeholder complexity.
- **Monitor dynamic networks**: Use KGs to identify key stakeholders and event pathways, distinguishing who exerts influence and at which points of the project life cycle.
- **Manage the project dataset**: Use the platform as a knowledge management tool, capable of breaking down historical project reports into structured, reusable evidence and storing it for future analysis.
- **Utilise the platform**: Interact with the LLM-driven chatbot to query stakeholder issues and receive provenance-linked answers, enhancing both transparency and usability for practitioners.

7.3 Areas for future research

While this study advances both methodological and practical understanding of stakeholder complexity, several avenues remain for further development.

First, the analysis in this report is based primarily on consolidated corpora for each case. Future work should explore time-sliced datasets and KGs to capture how alignment changes across project phases.

Second, the reliance on official project disclosures, while strengthening transparency and auditability, inevitably leaves out informal or undocumented interactions. Future research could test the integration of complementary data sources – such as structured public submissions, parliamentary transcripts or moderated community feedback – under carefully designed reliability checks to enrich the evidence base.

Third, the platform developed here currently serves as a diagnostic tool. A valuable next step would be to extend it into a decision-support system, allowing practitioners to simulate alternative engagement strategies and anticipate how these might shift awareness–outcome alignment on specific issues.

Finally, further testing is needed across different sectors and governance regimes. Applying the methodology to energy transition projects, urban regeneration initiatives or projects in emerging economies could reveal boundary conditions and refine parameters (e.g. tolerance bands or similarity thresholds) for different institutional contexts.

For further enquiries regarding this research, the underlying AI framework or potential collaboration opportunities, please contact the research lead, Dr Jin Xue, Lecturer at the School of Project Management, University of Sydney.

Email: jin.xue@sydney.edu.au

Website: Dr Jin Xue (sydney.edu.au)

References

Aaltonen, K. et al. (2015). Stakeholder dynamics during the project front-end: The case of nuclear waste repository projects. *Project Management Journal*, *46*(6), 15–41. doi.org/10.1002/pmj.21549

Aaltonen, K. et al. (2024). Stakeholder engagement: Theoretical and methodological directions for project scholarship. *International Journal of Project Management*, *42*(7), 102649. doi.org/10.1016/j.ijproman.2024.102649

Abosag, I. et al. (2016). What is dark about the dark-side of business relationships? *Industrial Marketing Management*, *55*, 5–9.

Adamopoulou, E. & Moussiades, L. (2020). An overview of chatbot technology. In *IFIP International Conference on Artificial Intelligence Applications and Innovations* (pp. 373-383). Springer International Publishing.

Adewumi, A. S. et al. (2024). Sustainability assessment frameworks for delivering Environmental, Social, and Governance (ESG) targets: A case of Building Research Establishment Environmental Assessment Method (BREEAM) UK New Construction. *Corporate Social Responsibility and Environmental Management*. doi.org/10.1002/csr.2768

Bahadorestani, A. et al. (2020). Novel approach to satisfying stakeholders in megaprojects: Balancing mutual values. *Journal of Management in Engineering*, *36*(2), 04019047. doi.org/10.1061/(ASCE)ME.1943-5479.0000734

Bao, Q. et al. (2024). Hierarchical construction and application of machining domain knowledge graph based on as-fabricated information model. *Advanced Engineering Informatics*, *62*, 102638. doi.org/10.1016/j.aei.2024.102638

Barraket, J. (2020). The role of intermediaries in social innovation: The case of social procurement in Australia. *Journal of Social Entrepreneurship*, *11*(2), 194–214.

Branco, M. C. & Rodrigues, L. L. (2008). Social responsibility disclosure: A study of proxies for the public visibility of Portuguese banks. *The British Accounting Review*, *40*(2), 161–181.

Buehler, M. J. (2024). Generative retrieval-augmented ontologic graph and multiagent strategies for interpretive large language model-based materials design. *ACS Engineering Au*, *4*(2), 241–277. doi.org/10.1021/acsengineeringau.3c00058

Bundy, J. et al. (2018). Organization–stakeholder fit: A dynamic theory of cooperation, compromise, and conflict between an organization and its stakeholders. *Strategic Management Journal*, *39*(2), 476–501. doi.org/10.1002/smj.2736

California High-Speed Rail Authority (2015). *Project Update Report to the California State Legislature*. hsr.ca.gov/wp-content/uploads/2024/01/SB1029-ProjectUpdate-FINAL-022715-A11Y.pdf

California High-Speed Rail Authority (2024). *2024 Business Plan*. hsr.ca.gov/wp-content/uploads/2024/05/2024-Business-Plan-FINAL.pdf

Castelblanco, G. & Guevara, J. (2024). Stakeholder dynamics: Rethinking roles and responsibilities in user-pay transport PPP projects. *Journal of Construction Engineering And Management*, *150*(8), 05024009. doi.org/doi:10.1061/JCEMD4.COENG-14477

Chen, X. et al. (2020). A review: Knowledge reasoning over knowledge graph. *Expert Systems with Applications*, *141*, 112948. doi.org/10.1016/j.eswa.2019.112948

Chung, K. S. K. et al. (2023). Response strategies for community stakeholder engagement on social media: A case study of a large infrastructure project. *International Journal of Project Management*, *41*(5), 102495. doi.org/10.1016/j.ijproman.2023.102495

Cohen, W. M. & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, *35*(1), 128–152.

Cope III, R. F. et al. (2006). Enhancing project management with knowledge management principles. *Proceedings of Allied Academies International Conference*. Academy of Management Information and Decision Sciences.

Cravero, C. (2017). Socially responsible public procurement and set-asides: A comparative analysis of the US, Canada and the EU. *Arctic Review on Law and Politics*, *8*(1), 174–192.

De Schepper, S. et al. (2014). Stakeholder dynamics and responsibilities in public–private partnerships: A mixed experience. *International Journal of Project Management*, *32*(7), 1210–1222. doi.org/10.1016/j.ijproman.2014.01.006

Derakhshan, R. et al. (2019). Project governance and stakeholders: A literature review. *International Journal of Project Management*, *37*(1), 98–116.

Dortheimer, J. et al. (2024). Evaluating large-language-model chatbots to engage communities in large-scale design projects. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, *38*, e4, Article e4. doi.org/10.1017/S0890060424000027

Drazin, R. & Van de Ven, A. H. (1985). Alternative forms of fit in contingency theory. *Administrative Science Quarterly*, *30*(4), 514–539.

El Baz, J. & Ruel, S. (2024). Achieving social performance through digitalization and supply chain resilience in the COVID-19 disruption era: An empirical examination based on a stakeholder dynamic capabilities view. *Technological Forecasting and Social Change*, *201*, 123209. doi.org/10.1016/j.techfore.2024.123209

Elwakil, E. & Zayed, T. (2018). Construction productivity fuzzy knowledge base management system. *Canadian Journal of Civil Engineering*, *45*(5), 329–338.

Ford, G. et al. (2024). Simplifying complexity? On quality decision-making and nonconformance outcomes of megaprojects. *IEEE Transactions on Engineering Management*, *71*, 5443–5454.

Freeman, R. E. (1984). *Strategic Management: A Stakeholder Approach*. Cambridge University Press.

Gao, Y. et al. (2024). Exploring bridge maintenance knowledge graph by leveraging GrapshSAGE and text encoding. *Automation in Construction*, *166*, 105634. doi.org/https://doi.org/10.1016/j.autcon.2024.105634

Gasik, S. (2011). A model of project knowledge management. *Project Management Journal*, *42*(3), 23–44. doi.org/10.1002/pmj.20239

Gillan, S. L. et al. (2021). Firms and social responsibility: A review of ESG and CSR research in corporate finance. *Journal of Corporate Finance*, *66*, 101889.

Gonzalez-Porras, L. et al. (2021). Understanding stakeholder influence: Lessons from a controversial megaproject. *International Journal of Human Resources Development and Management*, *21*(2–3), 191–213. doi.org/10.1504/ijhrdm.2021.116923

Haldar, K. et al. (2021). Institutional challenges and stakeholder perception towards planned water reuse in peri-urban agriculture of the Bengal Delta. *Journal of Environmental Management*, *283*, 111974. doi.org/10.1016/j.jenvman.2021.111974

Hassan, O. A. (2018). The impact of voluntary environmental disclosure on firm value: Does organizational visibility play a mediation role? *Business Strategy and the Environment*, *27*(8), 1569–1582.

High Speed Two Ltd (2017). *High Speed Two Phase One Information Paper*. assets.publishing.service.gov.uk/media/5a82e50eed915d74e62382ab/A1_-_Development_of_the_HS2_Proposed_Scheme_v1.6.pdf

Hofer, C. W. & Schendel, D. (1978). *Strategy Formulation: Analytical Concepts*. West Publishing Company.

Hong, W.-T. et al. (2024). A data-driven conceptual framework for understanding the nature of hazards in railway accidents. *Transport Policy*, *152*, 102–117. doi.org/10.1016/j.tranpol.2024.05.007

Hönen, J. et al. (2023). Threshold-based algorithms for an online rolling horizon framework under uncertainty – with an application to energy management. arXiv preprint, arXiv:2311.11307.

Iheukwumere-Esotu, L. O. & Yunusa-Kaltungo, A. (2022). Development of an interactive web-based knowledge management platform for major maintenance activities: Case study of cement manufacturing system. *Sustainability*, *14*(17), 11041. mdpi.com/2071-1050/14/17/11041

Jiang, H. et al. (2015). Public-opinion sentiment analysis for large hydro projects. *Journal of Construction Engineering and Management*, *142*. doi.org/10.1061/(ASCE)CO.1943-7862.0001039

Jiang, S. et al. (2024). Product innovation design approach driven by implicit relationship completion via patent knowledge graph. *Advanced Engineering Informatics*, *61*, 102530. doi.org/10.1016/j.aei.2024.102530

Kivrak, S. et al. (2008). Capturing knowledge in construction projects: Knowledge platform for contractors. *Journal of Management in Engineering*, *24*(2), 87–95. doi.org/doi:10.1061/(ASCE)0742-597X(2008)24:2(87)

Kujala, J. et al. (2022). Stakeholder engagement: Past, present, and future. *Business & Society*, *61*(5), 1136–1196. doi.org/10.1177/00076503211066595

Lai, X. et al. (2024). Kansei engineering for the intelligent connected vehicle functions: An online and offline data mining approach. *Advanced Engineering Informatics*, *61*. doi.org/10.1016/j.aei.2024.102467

Li, Y. et al. (2018). The impact of environmental, social, and governance disclosure on firm value: The role of CEO power. *The British Accounting Review*, *50*(1), 60–75.

Liu, H. (2022). Combating unethical producer behavior: The value of traceability in produce supply chains. *International Journal of Production Economics*, *244*, 108374. doi.org/10.1016/j.ijpe.2021.108374

Liu, P. et al. (2024). Joint knowledge graph and large language model for fault diagnosis and its application in aviation assembly. *IEEE Transactions on Industrial Informatics*, *PP*, 1–10. doi.org/10.1109/TII.2024.3366977

Loosemore, M. (2016). Social procurement in UK construction projects. *International Journal of Project Management*, *34*(2), 133–144.

Loosemore, M. & Reid, S. (2019). The social procurement practices of tier-one construction contractors in Australia. *Construction Management and Economics*, *37*(4), 183–200.

Luo, L. et al. (2018). Stakeholder-associated supply chain risks and their interactions in a prefabricated building project in Hong Kong. *Journal of Management in Engineering*, *35*(2), 05018015. doi.org/10.1061/(ASCE)ME.1943-5479.0000675

Lyu, Y. et al. (2025). Crud-rag: A comprehensive Chinese benchmark for retrieval-augmented generation of large language models. *ACM Transactions on Information Systems*, *43*(2), 1–32.

Mitchell, R. K. et al. (1997). Toward a theory of stakeholder identification and salience: Defining the principle of who and what really counts. *The Academy of Management Review*, *22*(4), 853–886. doi.org/10.2307/259247

Mok, K. et al. (2017). Addressing stakeholder complexity and major pitfalls in large cultural building projects. *International Journal of Project Management*, *35*. doi.org/10.1016/j.ijproman.2016.12.009

Mok, K. Y. et al. (2015). Stakeholder management studies in mega construction projects: A review and future directions. *International Journal of Project Management*, *33*(2), 446–457. doi.org/10.1016/j.ijproman.2014.08.007

Mok, K. Y. et al. (2017). A network theory-based analysis of stakeholder issues and their interrelationships in large construction projects: A case study. *International Journal of Construction Management*, *17*(3), 210–227.

Montalbán-Domingo, L. et al. (2019). Social sustainability in delivery and procurement of public construction contracts. *Journal of Management in Engineering*, *35*(2), 04018065.

Moussa, A. & El-Dakhakhni, W. (2022). Managing interdependence-induced systemic risks in infrastructure projects. *Journal of Management in Engineering*, *38*(5), 04022048. doi.org/doi:10.1061/(ASCE)ME.1943-5479.0001071

Mueller, J. et al. (2011). Virtual worlds as knowledge management platform – a practice-perspective. *Information Systems Journal*, *21*(6), 479–501. doi.org/10.1111/j.1365-2575.2010.00366.x

Oh, J. et al. (2024). Better think with tables: Tabular structures enhance LLM comprehension for data-analytics requests. arXiv preprint, arXiv:2412.17189.

Olander, S. (2007). Stakeholder impact analysis in construction project management. *Construction Management and Economics*, *25*(3), 277–287. doi.org/10.1080/01446190600879125

Omotayo, T. S. et al. (2024). Systems thinking interplay between project complexities, stakeholder engagement, and social dynamics roles in influencing construction project outcomes. *Sage Open*, *14*(2), 21582440241255872.

PMI (2023). *Measuring the Impact of ESG Initiatives*. pmi.org/learning/thought-leadership/esg/measuring-the-impact-of-esg-initiatives

Qin, S. et al. (2024). Intelligent design and optimization system for shear wall structures based on large language models and generative artificial intelligence. *Journal of Building Engineering*, *95*, 109996. doi.org/10.1016/j.jobe.2024.109996

Rao, A. et al. (2023). Assessing the utility of ChatGPT throughout the entire clinical workflow: Development and usability study. *Journal of Medical Internet Research*, *25*, e48659. doi.org/10.2196/48659

Reber, B. et al. (2022). ESG disclosure and idiosyncratic risk in initial public offerings. *Journal of Business Ethics*, *179*(3), 867–886.

Rowley, T. J. (1997). Moving beyond dyadic ties: A network theory of stakeholder influences. *The Academy of Management Review*, *22*(4), 887–910. doi.org/10.2307/259248

Shen, G. Q. & Xue, J. (2021). Managing stakeholder dynamics and complexity in mega infrastructure projects. *Frontiers of Engineering Management*, *8*(1), 148–150. doi.org/10.1007/s42524-020-0149-6

Tang, L. et al. (2017). Social media data analytics for the U.S. construction industry: Preliminary study on Twitter. *Journal of Management in Engineering*, *33*. doi.org/10.1061/(ASCE)ME.1943-5479.0000554

Tantalo, C. & Priem, R. L. (2016). Value creation through stakeholder synergy. *Strategic Management Journal*, *37*(2), 314–329.

Te Pas, M. E. et al. (2020). User experience of a chatbot questionnaire versus a regular computer questionnaire: Prospective comparative study. *JMIR Medical Informatics*, *8*(12), e21982.

Thirunavukarasu, A. J. et al. (2023). Large language models in medicine. *Nature Medicine*, *29*(8), 1930–1940. doi.org/10.1038/s41591-023-02448-8

Weizenbaum, J. (1966). ELIZA – a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, *9*(1), 36–45.

Wen, P. et al. (2023). Systematic knowledge modeling and extraction methods for manufacturing process planning based on knowledge graph. *Advanced Engineering Informatics*, *58*, 102172. doi.org/10.1016/j.aei.2023.102172

Woodside Energy (2020). *Scarborough Offshore Project Proposal*. woodside.com/docs/default-source/our-business---documents-and-files/burrup-hub---documents-and-files/scarborough---documents-and-files/scarborough-offshore-project-proposal.pdf?sfvrsn=1f542782_2

Woodside Energy (2024). *Scarborough Energy Project Community Development Report*. woodside.com/docs/default-source/our-business---documents-and-files/scarborough/scarborough-energy-project-2024-community-development.pdf?sfvrsn=98354d04_3

Woodside Energy (2025). *Scarborough Energy Project Factsheet*. woodside.com/docs/default-source/our-business---documents-and-files/scarborough/scarborough-energy-project-fact-sheet---.pdf?sfvrsn=24fc8d2c_17

Xiao, H. & Hao, S. (2023). Public participation in infrastructure projects: An integrative review and prospects for the future research. *Engineering, Construction and Architectural Management*, *30*(2), 456–477. doi.org/10.1108/ECAM-06-2021-0495

Xue, J. et al. (2020a). Dynamic analysis on public concerns in Hong Kong-Zhuhai-Macao Bridge: A topic modeling approach. *Construction Research Congress*. doi.org/10.1061/9780784482889.018

Xue, J. et al. (2020b). Dynamic stakeholder-associated topic modeling on public concerns in megainfrastructure projects: Case of Hong Kong–Zhuhai–Macao Bridge. *Journal of Management in Engineering*, *36*(6), 04020078. doi.org/doi:10.1061/(ASCE)ME.1943-5479.0000845

Xue, J. et al. (2020c). Mapping the knowledge domain of stakeholder perspective studies in construction projects: A bibliometric approach. *International Journal of Project Management*. doi.org/10.1016/j.ijproman.2020.07.007

Xue, J. et al. (2023). Evolution modeling of stakeholder performance on relationship management in the dynamic and complex environments of megaprojects. *Engineering, Construction and Architectural Management*, *30*(4), 1536–1557. doi.org/10.1108/ECAM-06-2021-0504

Xue, J. et al. (2025). Modeling stakeholder resilience in relationship management in a changing environment of megaprojects. *IEEE Transactions on Engineering Management*, *72*, 730–750. doi.org/10.1109/TEM.2025.3538073

Yang, S. et al. (2024). When design workshops meet chatbots: Meaningful participation at scale? *International Journal of Architectural Computing*, *22*(2), 216–237. doi.org/10.1177/14780771241253440

Yang, Y. & Han, J. (2023). Digital transformation, financing constraints, and corporate environmental, social, and governance performance. *Corporate Social Responsibility and Environmental Management*, *30*(6), 3189–3202.

Zhang, R. et al. (2024). Down to Earth: Implementing project-level ESG metrics in Chinese AEC firms' practices. *Journal of Management in Engineering*, *40*(5), 04024035. doi.org/doi:10.1061/JMENEA.MEENG-5941

Zhang, S. et al. (2015). PPP application in infrastructure development in China: Institutional analysis and implications. *International Journal of Project Management*, *33*(3), 497–509.

Zhou, B. et al. (2024). CausalKGPT: Industrial structure causal knowledge-enhanced large language model for cause analysis of quality problems in aerospace product manufacturing. *Advanced Engineering Informatics*, *59*, 102333. doi.org/10.1016/j.aei.2023.102333

Zhou, Z. et al. (2021). Online public opinion analysis on infrastructure megaprojects: Toward an analytical framework. *Journal of Management in Engineering*, *37*, 04020105. doi.org/10.1061/(ASCE)ME.1943-5479.0000874

Zhu, J. et al. (2024). A flood knowledge-constrained large language model interactable with GIS: Enhancing public risk perception of floods. *International Journal of Geographical Information Science*, *38*(4), 603–625. doi.org/10.1080/13658816.2024.2306167

Zuofa, T. & Ochieng, E. (2016). Sustainability in construction project delivery: A study of experienced project managers in Nigeria. *Project Management Journal*, *47*(6), 44–55.

We are the only chartered membership organisation for the project profession

Association for Project Management
Ibis House, Regent Park
Summerleys Road
Princes Risborough
Bucks HP27 9LE
0845 458 1944
apm.org.uk

@APMProjectMgmt
@AssociationForProjectManagement
@APMProjectMgmt
@AssociationForProjectManagement

Association for Project Management is incorporated by Royal Charter RC000890 and a registered charity No. 1171112. Principal office as shown.