Digital Transformation and Society

Beyond technical capability: how institutional resistance shapes generative AI exposure in civil and environmental engineering occupations

Received 24 January 2026
Revised 13 June 2026
18 June 2026
14 August 2026
24 August 2026
Accepted 26 August 2026

Yuxuan Chai
Department of Civil and Environmental Engineering, University of Strathclyde, Glasgow, UK

Abstract

Purpose – This article introduces the Integrated Sociotechnical Exposure Framework (ISEF) to assess generative AI exposure in civil and environmental engineering by separating technical AI suitability (TA) from institutional resistance (IR). The contribution is an operational measurement framework, not a claim that institutional constraint is a new concept.

Design/methodology/approach – The framework is applied to 674 O*NET tasks across 28 occupations. Task-variable pairs are scored by three large language models under a fixed rubric and then aggregated to occupation level. Exposure is modelled with a multiplicative form, and sensitivity checks compare additive, min-rule, and geometric alternatives. A small external practitioner survey is used as a calibration check rather than as task-level validation.

Findings – A large group of occupations score well on technical AI suitability but stay heavily constrained by liability, compliance, and sign-off requirements. The cleaned practitioner survey points in the same direction: across 15 task vignettes, higher perceived institutional constraint is strongly associated with lower adjusted AI exposure.

Research limitations/implications – The scoring still leans on LLMs, and the practitioner sample is small. Because no same-task expert panel is available in the present revision, the results should be read as a transparent first-stage index. The nine-response practitioner exercise is preliminary calibration, not validation of the occupation-level index or its constructed adjusted-score analogue. The natural next step is a larger same-task expert panel paired with adoption-linked field evidence.

Practical implications – For engineering organisations, adoption strategy is better anchored in accountable workflows, auditability, and professional review processes than in model capability on its own.

Originality/value – The paper pulls governance constraints into the exposure measure itself, and uses that lens to explain why technical-only indices tend to overstate near-term generative AI uptake in regulated engineering work. Its measurement logic is potentially transferable to other regulated professions, but any such application requires domain-specific specification and calibration of the institutional-resistance construct.

Keywords Generative AI, Institutional resistance, Civil engineering, Occupational exposure, Sociotechnical systems, Professional regulation

Paper type Research article

Introduction

Civil and environmental engineering is a good place to take the phrase “AI exposure” seriously. Generative systems can now draft reports, produce code, and assist analytical work at a level that would have looked implausible only a few years ago (OpenAI, 2023; Bubeck

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The author is a self-funded PhD researcher at a UK university.

Conflict of interest: The author declares no competing interests. The author is responsible for the study design, analysis, interpretation, and manuscript content.

Digital Transformation and Society
Emerald Publishing Limited
e-ISSN: 2755-077X
p-ISSN: 2755-0761
DOI 10.1108/DTS-01-2026-0039

et al., 2023). Engineering decisions still pass through licensure, liability, and review chains that were built long before these tools existed. Whether AI can produce a particular output, and whether that output can enter an accountable workflow, are two different questions; much of the current confusion around generative AI in engineering comes from treating them as one.

The exposure literature has built strong instruments for the first question. Task-content measures identified codifiable work, routine procedures, and pattern-based activity as the natural unit of analysis (Autor, 2015; Frey & Osborne, 2017). That foundation remains useful for regulated engineering, but it is incomplete. A task may read as technically suitable for AI assistance and still sit inside a sign-off, audit, or liability structure that keeps AI-generated outputs out of the accountable record (Jasanoff, 2016).

This paper develops the Integrated Sociotechnical Exposure Framework (ISEF) in response. ISEF is a task-level framework for civil and environmental engineering that scores technical AI suitability and institutional resistance through the same rubric-based protocol, so neither side gets reduced to background. Its novelty is practical and methodological: it turns a familiar sociotechnical insight into a reproducible occupation-level measurement design.

Research questions. Three questions guide the analysis:

(1) How can technical AI suitability and institutional resistance be operationalised separately for civil and environmental engineering tasks?

(2) How do occupation-level exposure patterns change when institutional resistance is incorporated into technical exposure measurement?

(3) How sensitive are the resulting rankings and groupings to modelling choices, including functional form and clustering assumptions?

The multiplicative form developed below treats mandatory sign-off and similar requirements as binding limits on workable exposure, not as small adjustments at the margin. The analysis covers 28 engineering occupations and 674 O*NET task statements (Peterson *et al.*, 2001). Throughout, exposure is used as a descriptive indicator of where technical suitability meets institutional permission, and it is kept separate from claims about adoption timing or labour displacement.

Literature review

This section positions the framework in four bodies of work: task-based exposure measurement, professional regulation, sociotechnical governance, and digital transformation in engineering workflows. The review is not used to claim that institutions have been ignored in prior scholarship. The point is narrower: existing literatures explain why adoption is institutionally mediated, while the present paper turns that insight into an occupation-level measurement protocol.

Task-based AI exposure measurement

Task-based exposure measures have a long lineage in labour economics. The early generation linked routine task content to substitution risk and set the analytical template later work would extend (Autor, 2015; Frey & Osborne, 2017). Labour-market studies then connected task displacement to job polarisation and wage inequality, while also showing why occupation-level averages can hide task-level variation (Goos, Manning, & Salomons, 2014; Arntz, Gregory, & Zierahn, 2017; Acemoglu & Restrepo, 2022). Subsequent studies narrowed the frame further, moving from whole occupations to the specific capabilities a machine-learning system can perform (Brynjolfsson, Mitchell, & Rock, 2018; Webb, 2020). Reviews of AI, labour, and productivity also stress that exposure measures need to be separated from realised firm-level adoption (Raj & Seamans, 2019). Generative AI has revived that agenda: recent indices ask whether language models can perform or assist with occupational tasks

(Eloundou, Manning, Mishkin, & Rock, 2023; Felten, Raj, & Seamans, 2021; Tolan *et al.*, 2021), while experimental and field evidence shows that gains are task-contingent rather than uniform (Noy & Zhang, 2023; Dell'Acqua *et al.*, 2023). These measures handle technical fit competently, but they remain thinner on the institutional side of adoption, which is the part of the problem this paper focuses on.

ISEF keeps task-level measurement but separates two dimensions that exposure scores typically fold together. Technical AI suitability (TA) asks whether generative AI can plausibly support the task. Institutional resistance (IR) asks whether responsibility, compliance duties, data governance, or organisational accountability make practical use harder. In engineering, the second question often matters as much as the first: a useful draft or calculation still has to clear review before it becomes actionable evidence.

Professional regulation and sociotechnical governance

Institutional accounts help explain why technical fit alone tends to undersell adoption frictions. Abbott (1988) argues that professional jurisdictions are sustained through expertise, credentialing, and control over what counts as legitimate practice; Freidson (2001) reads that control as a durable institutional logic, not a transitional friction. The implication for digital tools is sharper in Susskind and Susskind (2015): technology can redistribute the tasks professionals carry out without redistributing authority, sign-off, or liability. STS and organisational scholarship add the safety-critical angle. Jasanoff (2004) treats science and technology as co-produced with social order, and Jasanoff (2016) extends the argument to how risk technologies become acceptable—through standards, review procedures, and accountability arrangements, alongside any efficiency case. Sociomaterial and algorithmic-work research makes the same point at organisational scale: digital systems alter work through routines, documentation, and control relations rather than through technical capability alone (Leonardi, 2012; Kellogg, Valentine, & Christin, 2020).

Civil and environmental engineering sits at the sharp end of this argument because outputs are tied to public safety, environmental compliance, procurement rules, and professional sign-off. AI can help with drafting, checking, simulation support, and documentation, but the institutional layer still decides how those outputs are reviewed, recorded, and accepted. AI governance scholarship is useful here because it treats accountability, transparency, risk classification, and auditability as design conditions rather than optional afterthoughts (Calo, 2017; Floridi & Cowls, 2019; Raji *et al.*, 2020; Veale & Borgesius, 2021; Weidinger *et al.*, 2022). ISEF scores this institutional layer at task level using the same rubric-based logic applied to technical suitability, so the two sides of the problem stay comparable.

Generative AI in engineering workflows

Engineering research has tracked the spread of BIM, digital twins, machine learning, reservoir modelling, and other data-driven tools in infrastructure and subsurface work (Bock, 2015; Sacks, Eastman, Lee, & Teicholz, 2018; Mohaghegh, 2017; Azimi, Eslamlou, & Pekcan, 2020). Generative AI adds a particular complication on top of this trajectory: outputs can read as competent while remaining hard to verify, attribute, or certify. Foundation-model and generative-AI reviews make this verification problem especially salient because general-purpose models can be repurposed across domains faster than institutional controls can be standardised (Bommasani *et al.*, 2021; Dwivedi *et al.*, 2023). That gap carries weight when a flawed design note, permit draft, or environmental assessment has legal and safety consequences.

This distinction matters for Digital Transformation and Society because generative AI is not simply another productivity tool inserted into an existing workflow. In regulated engineering, model outputs become useful only when they can be documented, checked, attributed, insured, and signed off. A measurement framework that scores technical suitability without also scoring those governance conditions will tend to overstate the near-term exposure of safety-critical professional work.

Positioning of the present study

Exposure is therefore treated here as a sociotechnical construct: a description of where task suitability and institutional constraint meet, rather than a forecast of job loss or a snapshot of current firm-level uptake. The originality of ISEF lies in operationalisation. It combines a task-level rubric, a multi-model scoring protocol, and a gating specification that can be recalculated by other researchers. The empirical contribution remains descriptive, and the LLM-based scores should be read as a first-stage measurement exercise—one that can be inspected, challenged, and extended with stronger practitioner validation.

Methods

Sample and data

The study examines 28 civil and environmental engineering occupations from the O*NET database (version 28.0) (Peterson *et al.*, 2001). The final sample consists of 674 unique tasks across managers, engineers, technologists/technicians, and related professionals. Table 1 provides an overview of the sample structure.

Measurement protocol

Technical and institutional factors are evaluated using the Generative AI Impact Index (GAII), which consists of 12 variables (seven TA; five IR). Table 2 provides a quick reference, while the full definitions and rubric are available in Supplementary Material B. The rubric is written out in full so that scoring decisions can be checked line by line.

The variable list was built through an iterative review of three sources: task-based AI exposure studies, sociotechnical and professional-regulation concepts, and civil engineering workflow characteristics. The TA variables capture recurring forms of task suitability in prior

Table 1. Sample structure by occupation category

Category	Occupations	Tasks
Managers	4	84
Engineers	10	248
Technologists/Technicians	7	169
Scientists & Others	7	173
Total	*28*	*674*

Source(s): O*NET 28.0

Table 2. ISEF variable quick reference

Variable	Type	Intuition
Cognitive/Data Analysis	TA	Structured reasoning, statistical modelling, data interpretation
Technological Integration	TA	Digital workflows, IoT/BIM coordination
AI Readiness	TA	Rule-based, procedural tasks suitable for generative AI assistance
Automation Potential	TA	General suitability for hybrid human-AI execution
Collaboration	TA	Teamwork, stakeholder coordination
Innovation/Synergy	TA	Pattern discovery, creative synthesis
Domain Expertise	TA	Specialised engineering knowledge and judgement
Compliance/Responsibility	IR	Legal liability, regulatory sign-off requirements
Industry Barriers	IR	Professional gatekeeping, institutional inertia
Algorithm Limitation	IR	Tasks requiring emotional intelligence, negotiation
Data Security/Privacy	IR	Sensitive data, confidentiality constraints
Sociotechnical Challenges	IR	Organisational politics, public trust issues

exposure measures, including data analysis, structured procedures, digital integration, and augmentation potential. The IR variables capture barriers repeatedly identified in professional and safety-critical domains, including legal responsibility, privacy, organisational trust, and sectoral gatekeeping. The list is not claimed to be exhaustive; it is a parsimonious operationalisation designed for transparent scoring and future recalibration. Domain Expertise is treated consistently as a TA variable because specialised knowledge can make a task more amenable to AI-assisted retrieval, drafting, or checking when appropriate expert oversight remains in place. Institutional accountability for that expert judgement is captured separately by Compliance/Responsibility and related IR variables.

The analysis measures *exposure*: the joint degree to which tasks are technically suitable and institutionally permissible. It does not forecast replacement or job loss.

The task–variable mapping protocol is set out in Supplementary Material A. The complete measurement protocol, defined in Supplementary Material B, documents every scoring rule used in the analysis. For each task, the initial mapping was conducted with a focus on the specific context of civil engineering. This conservative protocol involved reading task descriptions and identifying relevant variables supported by the text. The mapping was checked for internal consistency before model scoring, but it should not be confused with independent task-level expert validation. Supporting diagnostic tables and figures are provided in Supplementary Material C.

Scoring protocol

Task-variable pairs were evaluated using three large language model systems under fixed settings and a prompt template: ChatGPT-5, DeepSeek-R1, and Qwen-3. The same prompt, rubric, and aggregation code can be rerun by other researchers without proprietary tuning.

Design considerations for the scoring process included:

(1) Three independent evaluation runs per model to mitigate session effects.

(2) A single prompt template requiring a 0–10 score with a brief justification tied to the task text.

(3) No post-hoc cleaning of outliers to maintain a realistic assessment of variance.

(4) Clearing of session history between runs to ensure independent measurements.

Consistency was monitored across runs and stayed within the limits acceptable for a descriptive measurement exercise. Inter-model agreement is useful but cannot stand in for external validity. The three systems likely share training data, alignment pressures, and shared assumptions about professional work, so their consensus is treated here as a scalable first-stage signal rather than a substitute for domain-expert scoring. In response to reviewer concerns about circularity, a small practitioner survey was added during the revision period to check whether the TA/IR split survives outside the model-scoring loop; respondents rated 15 task vignettes from the questionnaire on AI assistance potential and institutional constraints. This survey is reported as calibration evidence only. It does not estimate expert–LLM agreement on the same tasks, and the manuscript therefore avoids presenting LLM consensus as independent validation.

Aggregation and analysis

The aggregation of task-level evaluations into occupation-level indices follows a structured procedure:

Stage 1: Consensus Scores. A mean is computed across all systems and runs for each task-variable pair (Equation 1):

$$\bar{s}_{tv} = \frac{1}{M \times R} \sum_{m=1}^{M} \sum_{r=1}^{R} s_{tv}^{(m,r)}$$

where $s_{tv}^{(m,r)}$ is the score for task t and variable v from system m and run r, with $M = 3$ systems and $R = 3$ runs per system.

Stage 2: Variable Aggregation. Evaluations are aggregated to the variable level using O*NET importance weights (Equation 2):

$$S_{vo} = \frac{\sum_{t \in o, t \to v} w_{to} \cdot \bar{s}_{tv}}{\sum_{t \in o, t \to v} w_{to}}$$

where w_{to} is the normalised importance weight for task t in occupation o, with $\sum_{t \in o} w_{to} = 1$.

Stage 3: Index Construction. TA and IR are computed as unweighted means of their constituent variables and rescaled to a 0–100 range (Equation 3):

$$TA_o = 100 \cdot \frac{1}{|\mathrm{TA}_o|} \sum_{v \in \mathrm{TA}_o} \frac{S_{vo}}{10}, IR_o = 100 \cdot \frac{1}{|\mathrm{IR}_o|} \sum_{v \in \mathrm{IR}_o} \frac{S_{vo}}{10}$$

Exposure is then calculated through the multiplicative specification (Equation 4), incorporating the institutional context:

$$\mathrm{Exposure}_o = TA_o \times \left(1 - \frac{IR_o}{100}\right)$$

The multiplicative specification is used as the primary model because it represents institutional resistance as a partial gate on technical capability. The overall ISEF procedure is visualised in Figure 1. Exploratory K-means diagnostics are then applied to the standardised occupation-level scores to identify broad patterns within the data; the resulting labels are not treated as validated occupational strata.

Cluster analysis

K-means clustering is applied to standardised occupation-level scores as a diagnostic exercise. Solutions for $k = 2$ through $k = 10$ are evaluated using the elbow method, silhouette scores, and gap statistics (Tibshirani, Walther, & Hastie, 2001). With only 28 occupations, clustering carries clear risks: small shifts in input weighting can move borderline cases between groups. The main presentation reports the original $k = 5$ diagnostic for transparency, but interpretation is deliberately collapsed to broad lower-, middle-, and higher-exposure patterns rather than discrete occupational strata. Supporting diagnostic tables and figures are provided in Supplementary Material C.

One feature of the $k = 5$ diagnostic needs a direct word. The singleton diagnostic category contains only one occupation, Engineering Teachers, Postsecondary. The most plausible reading is structural: O*NET task wording for teaching roles is heavy on instruction, curriculum design, and student assessment, which sits awkwardly against the design-and-permit vocabulary that dominates the rest of the sample. The singleton is therefore retained only as a transparent diagnostic artefact. It is not interpreted as a cluster and is not used to support substantive claims. Bootstrap stability is also only moderate in the wider solution; the discussion therefore stays at the level of broad exposure bands and avoids fine-grained cluster identities.

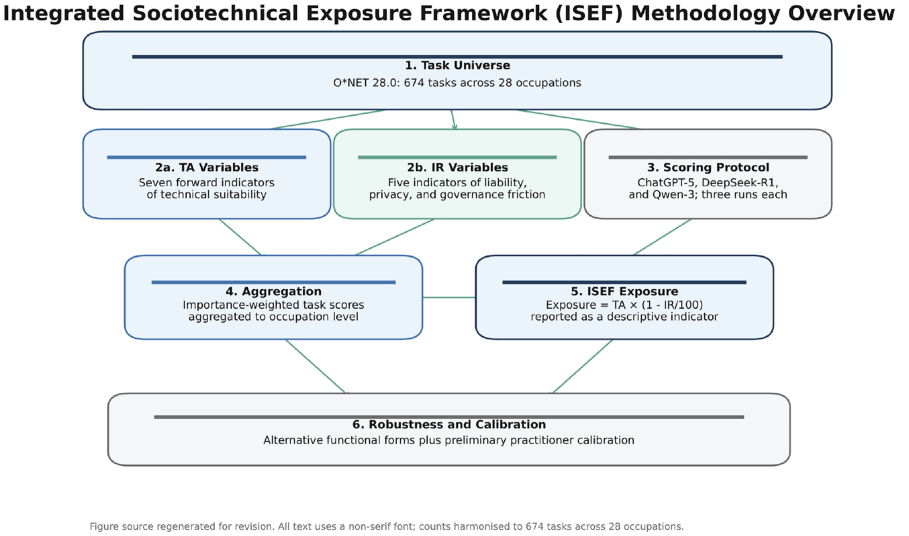

Figure 1. Complete ISEF workflow for the 674-task, 28-occupation sample: task universe; TA and IR indicators; three-model rubric scoring; occupation-level aggregation; ISEF exposure calculation; and functional-form sensitivity plus preliminary practitioner calibration. Domain Expertise is treated as a TA variable, while accountability and sign-off barriers are captured through IR variables

Results

Measurement reliability

Table 3 reports agreement across the three systems. On occupation-level averages, Cronbach's α and the ICC both sit around 0.93, which is enough for the descriptive purposes of the paper. Kendall's W tells a less tidy story. Its mean is 0.87, but the range stretches from 0.09 to 0.99, and that spread deserves a serious look before any operational use of the index.

Most occupations sit near the top of the W distribution, where the three systems agree on both the average scores and the within-occupation rank order of tasks. The very low W values are concentrated in a small group of compliance-heavy roles whose task statements lean on words such as "inspect," "certify," "ensure," and "approve." For these occupations the systems converged on the overall exposure level but disagreed on which sign-off and licensing tasks were the most binding inside the bundle. Two implications follow. First, fine-grained rank ordering is the weakest part of the analysis, and the affected occupations should be flagged for targeted human re-rating before the index is used to inform decisions on individuals or specific workflows. Second, the high mean W cannot be read as independent validation. Where three LLMs trained on overlapping corpora interpret legalistic phrasing in similar ways, the agreement may sit inside the models' shared priors as much as inside the task content. We treat the W extremes as a flag for human review rather than as background noise.

Table 3. Cross-model reliability statistics (occupation-level averages)

Statistic	Mean value	Range
Cronbach's α	0.93	0.67–0.99
ICC(2,1)	0.93	0.67–0.99
Kendall's W	0.87	0.09–0.99

Heterogeneity in exposure (expectation 1)

Figure 2 reports ordered ISEF exposure scores for the 28 occupations under the multiplicative specification. Table 4 summarises the exploratory diagnostic groups, and Table 5 reports the occupation-level scores. Scores range from 8.1 to 15.7 on the 0–100 realised exposure index. That observed range is narrow on purpose: once institutional resistance is multiplied in, the safety-critical character of the whole domain compresses exposure even where technical AI suitability is high. Exact ordered scores are retained for transparency and reproducibility, but the analysis does not treat adjacent ranks as substantively distinct.

Civil Engineers (11.3) and Transportation Engineers (11.2), for example, sit too close to treat as meaningfully different given the measurement noise documented above. The interpretive weight sits in the TA/IR decomposition and in broad bands: a small group of occupations combine high TA with especially high IR, others sit on lower IR floors, and another set is moderate on both axes. The diagnostic summaries in Table 4 and Figure 3 carry the same caveat—the singleton diagnostic group and only moderate bootstrap stability mean the labels are visual aids, not discrete categories. Most visible variation comes from differences in institutional resistance, with technical capability moving less.

Technical suitability under institutional constraint

Splitting TA and IR brings out a pattern that gets blurred when both are folded into one score. Among occupations with similarly high technical potential, the ones with stronger institutional resistance end up with noticeably lower exposure. The whole sample is regulated work, which keeps the gap muted, but the pattern is still readable in the numbers.

Mining & Geological Engineers and Petroleum Engineers make the point concrete. Both sit above 72 on TA. The Mining role pairs that with very high IR (88.8) and ends up at the bottom of the exposure ladder (8.1); the Petroleum role sits on lower resistance (78.5) and lands at 15.7. The comparison is not meant to suggest that one occupation is "more exposed" in any absolute sense. The point is that two technically similar roles get pulled in opposite directions by their institutional surroundings.

This is the same gap that technical-only exposure indices tend to read over. A task can look well suited to AI assistance and still sit inside a workflow where a licensed engineer's signature is the final binding step.

Divergence from technical indices

To see how the rankings diverge from technical-only measures, the relationships between exposure, TA, and IR are examined across the 28 occupations. Exposure correlates strongly and negatively with institutional resistance ($r \approx -0.88$); its correlation with technical potential is small and slightly negative ($r \approx -0.26$). These are descriptive diagnostics, not causal estimates, but they point to a clear pattern within this regulated domain: institutional resistance does most of the work in shaping the final exposure scores.

The same pattern helps explain why technical-only indices can mis-rank occupations whenever IR is both high and unevenly distributed. In civil engineering, governance and liability constraints look more consequential for near-term exposure than marginal differences in model capability. Occupation-level TA, IR, and exposure profiles are reported in Figure 4; variable-level model means are reported in Table 6.

Heatmap patterns

Figures 5 and 6 show technical-side and institutional-side heatmaps across the 28 occupations. The heatmaps use a consistent DTS-style blue-to-deep-blue scale and place occupation names on the bottom axis for readability. Reading the two side by side makes one feature obvious: institutional resistance is rarely spread evenly across a role. In most occupations it concentrates

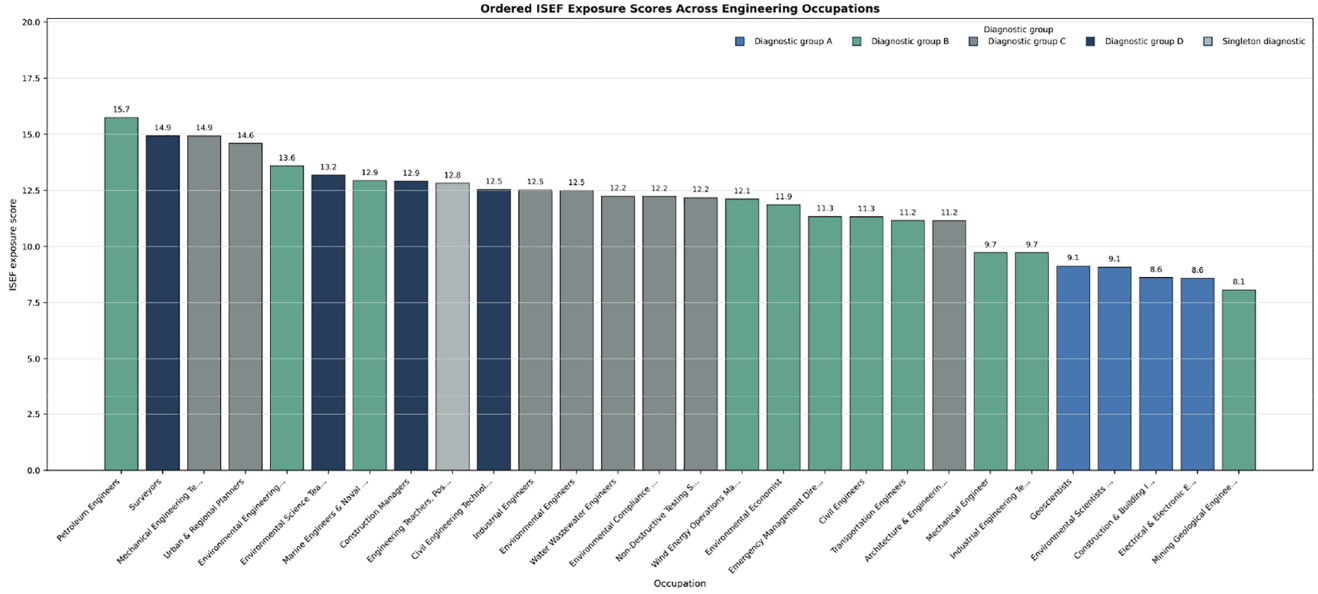

Figure 2. Ordered ISEF exposure scores across 28 engineering occupations. Colours indicate exploratory diagnostic groupings only, and labels report one-decimal exposure scores; adjacent ranks should not be interpreted as substantively distinct

Digital Transformation and Society

Table 4. Exploratory diagnostic group summary statistics (exposure)

Diagnostic group	*N*	Mean	Median	SD	Representative occupations
Diagnostic group A	4	8.85	8.85	0.29	Geoscientists; Environmental Scientists and Specialists; Construction and Building Inspectors; Electrical and Electronic Engineering Technologists and Technicians
Diagnostic group B	11	11.60	11.33	2.08	Petroleum Engineers; Environmental Engineering Technologists and Technicians; Marine Engineers and Naval Architects; Wind Energy Operations Managers; Environmental Economist
Diagnostic group C	8	12.79	12.36	1.29	Mechanical Engineering Technologists; Urban and Regional Planners; Industrial Engineers; Environmental Engineers; Water Wastewater Engineers
Diagnostic group D	4	13.39	13.04	1.06	Surveyors; Environmental Science Teachers, Postsecondary; Construction Managers; Civil Engineering Technologists and Technicians
Singleton diagnostic	1	12.82	12.82	0.00	Engineering Teachers, Postsecondary

Note(s): Summary statistics are calculated from the unrounded occupation-level exposure scores. Table 5 rounds occupation exposures to one decimal place, so manual recomputation from the displayed values may differ slightly in the second decimal. These are exploratory diagnostics from the $k = 5$ solution, not validated occupational strata. The singleton diagnostic is retained only to make the instability of the solution visible

in a small cluster of legally sensitive tasks—permits, signoffs, safety checks, and audit-facing documentation—and those few tasks tend to drive the IR score for the whole occupation.

The comparison also has implications for how exposure is interpreted. Where institutional resistance is concentrated in a limited set of accountability-related tasks, changes in governance arrangements and signing responsibilities may have a greater effect on occupational exposure than incremental improvements in the underlying AI tools. This suggests that exposure cannot be assessed from technical capability alone, particularly for tasks involving formal approval, safety assurance, and regulatory accountability.

Validation checks. Several checks were run on the scoring protocol, the functional form, and the plausibility of the IR rubric. These checks are deliberately separated from validation claims. The practitioner survey helps calibrate the TA/IR split across 15 concrete task vignettes, the FAR-clause mapping illustrates why institutional resistance belongs in the construct, and the benchmarking exercise checks the technical component against prior exposure measures. None of these exercises is treated as a substitute for a same-task expert panel.

Functional Form Justification. The multiplicative specification (Equation 4) treats institutional resistance as a gate, not a marginal offset. In high-stakes engineering work, a licensed engineer's signature requirement can suppress the practical exposure of a design task that otherwise looks highly suitable for AI assistance. Exposure therefore falls towards zero as resistance approaches its upper bound. The logic is theoretical, not statistically identified, which is why alternative specifications are tested below.

Sensitivity Analysis. Occupational scores were compared across four functional forms: additive mean exposure ($(TA + [100 - IR])/2$), multiplicative exposure ($TA \times [1 - IR/100]$), a min-rule specification ($\min(TA, 100 - IR)$), and an equal-weight geometric mean of technical potential and realised institutional permission. Table 7 and Figure 7 summarise the diagnostics. The additive mean produces a noticeably different fine-grained ordering from the multiplicative model (Spearman $\rho = 0.44$), while the min-rule is more closely aligned but still changes several ranks (Spearman $\rho = 0.80$). The equal-weight geometric mean is rank-equivalent to the multiplicative model (Spearman $\rho = 1.00$), because both are monotonic

Table 5. ISEF occupation scores (multiplicative exposure)

Occupation	SOC	TA	IR	Realised	Exposure	Diagnostic
Petroleum Engineers	17-2171.00	73.2	78.5	21.5	15.7	Diagnostic group B
Surveyors	17-1022.00	65.4	77.2	22.8	14.9	Diagnostic group D
Mechanical Engineering Technologists	17-3027.00	60.3	75.3	24.7	14.9	Diagnostic group C
Urban and Regional Planners	19-3051.00	58.6	75.1	24.9	14.6	Diagnostic group C
Environmental Engineering Technologists and Technicians	17-3025.00	71.3	80.9	19.1	13.6	Diagnostic group B
Environmental Science Teachers, Postsecondary	25-1053.00	63.5	79.2	20.8	13.2	Diagnostic group D
Marine Engineers and Naval Architects	17-2121.00	70.2	81.6	18.4	12.9	Diagnostic group B
Construction Managers	11-9021.00	69.0	81.3	18.7	12.9	Diagnostic group D
Engineering Teachers, Postsecondary	25-1032.00	59.6	78.5	21.5	12.8	Singleton diagnostic
Civil Engineering Technologists and Technicians	17-3022.00	66.6	81.2	18.8	12.5	Diagnostic group D
Industrial Engineers	17-2112.00	56.3	77.8	22.2	12.5	Diagnostic group C
Environmental Engineers	17-2081.00	55.1	77.3	22.7	12.5	Diagnostic group C
Water Wastewater Engineers	17-2051.02	55.7	78.0	22.0	12.2	Diagnostic group C
Environmental Compliance Inspectors	13-1041.01	56.3	78.3	21.7	12.2	Diagnostic group C
Non-Destructive Testing Specialists	17-3029.01	54.5	77.7	22.3	12.2	Diagnostic group C
Wind Energy Operations Managers	11-9199.09	65.5	81.5	18.5	12.1	Diagnostic group B
Environmental Economist	19-3011.01	69.0	82.8	17.2	11.9	Diagnostic group B
Emergency Management Director	11-9161.00	69.3	83.7	16.3	11.3	Diagnostic group B
Civil Engineers	17-2051.00	62.8	82.0	18.0	11.3	Diagnostic group B
Transportation Engineers	17-2051.01	68.1	83.6	16.4	11.2	Diagnostic group B
Architecture and Engineering Management	11-9041.00	54.3	79.4	20.6	11.2	Diagnostic group C
Mechanical Engineer	17-2141.00	69.2	85.9	14.1	9.7	Diagnostic group B
Industrial Engineering Technologists and Technicians	17-3026.00	72.3	86.5	13.5	9.7	Diagnostic group B
Geoscientists	19-2042.00	70.8	87.1	12.9	9.1	Diagnostic group A
Environmental Scientists and Specialists	19-2041.00	64.1	85.8	14.2	9.1	Diagnostic group A
Construction and Building Inspectors	47-4011.00	71.2	87.9	12.1	8.6	Diagnostic group A
Electrical and Electronic Engineering Technologists and Technicians	17-3023.00	67.5	87.3	12.7	8.6	Diagnostic group A
Mining Geological Engineers	17-2151.00	72.1	88.8	11.2	8.1	Diagnostic group B

Note(s): TA, IR, Realised, and Exposure are shown to one decimal place. Exposure is calculated from the unrounded TA and IR scores using $TA \times (1 - IR/100)$ before final rounding, so recomputation from the displayed one-decimal TA and IR values may differ by 0.1 in a small number of rows. The final column reports exploratory diagnostic labels only; adjacent ranks and group labels should not be read as stable occupational strata

functions of $TA \times (100 - IR)$. These results strengthen the main caution of the paper: broad exposure bands are more defensible than exact occupational ranks. The occupations that move most across specifications include Industrial Engineering Technologists and Technicians, Water Wastewater Engineers, Environmental Engineers, and Non-Destructive Testing

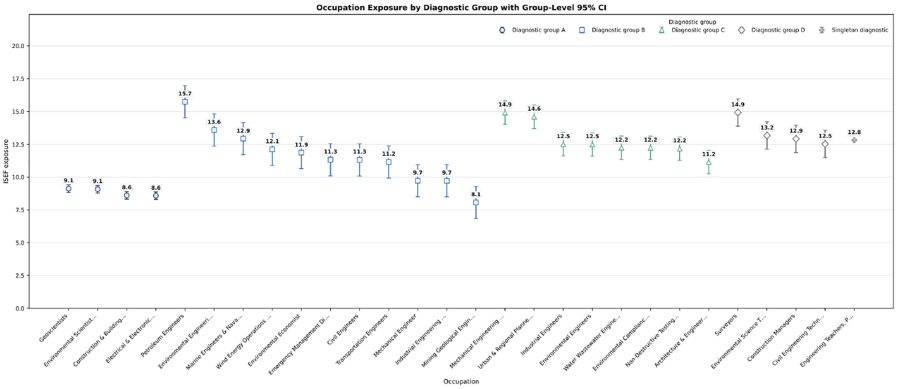

Figure 3. Occupation exposure by exploratory diagnostic group with group-level 95% confidence intervals. Points and intervals are descriptive diagnostics, not validated occupational strata

Specialists. Weighted and unweighted O*NET aggregation produce nearly identical multiplicative rankings (Spearman $\rho = 0.998$).

Illustrative Anchoring via Contract Clauses. Supplementary Material D maps three representative US federal procurement workflows to the institutional-resistance rubric. Engineering work in these settings is governed by inspection, audit, records, and acceptance clauses such as FAR 52.246-4 (Inspection of Services) and FAR 52.215-2 (Audit and Records). Three contracts cannot validate the index. They are useful for showing why liability, auditability, and review requirements belong inside the IR construct in the first place.

Cross-Index Benchmarking. TA scores were benchmarked against established technical exposure indices (Eloundou *et al.*, 2023; Felten *et al.*, 2021). TA here means technical suitability for generative AI assistance, with no claim of full task automation. The correlation with prior "Exposure" scores is strong ($r \approx 0.72$), so the technical side of ISEF tracks existing capability measures closely. Final ISEF exposure diverges from those rankings once IR is applied; the divergence is part of the contribution, since governance acts as a gate on what counts as workable exposure. Table 8 summarises the comparison.

Preliminary Practitioner Calibration. A short practitioner survey was fielded during the revision period as preliminary calibration, not validation, because a fair criticism of the main scoring exercise is that LLMs are being asked to grade work LLMs may one day do. The questionnaire in question.png presented 15 concrete civil, environmental, and adjacent engineering task vignettes, including drilling-plan design, environmental permit preparation, site-plan compliance review, legal boundary descriptions, notices of violation, FEA stress analysis, and building-inspection sign-off. For each vignette, respondents scored two independent 0–10 dimensions. D1 asked how much generative AI could technically assist the task assuming no legal, liability, or sign-off restrictions. D2 asked how much actual sign-off requirements, professional liability, regulatory compliance, and audit traceability limit the use of AI-generated outputs.

The raw Google Forms export contained 11 anonymous records. Two were removed because they did not contain the full set of D1/D2 task ratings and professional-judgement confirmations. The cleaned file therefore retains 9 complete responses across all 15 vignettes. The retained panel includes academic/researcher, mid-level engineer, senior engineer, and government/regulator roles; 6 of the 9 respondents reported holding a PE licence. Table 9 reports the vignette-level means after these exclusions. This preliminary calibration does not validate the 674-task scoring item by item, the occupation-level ISEF index, or the constructed calibration-only adjusted score; no expert–LLM task-level correlation is reported. What it

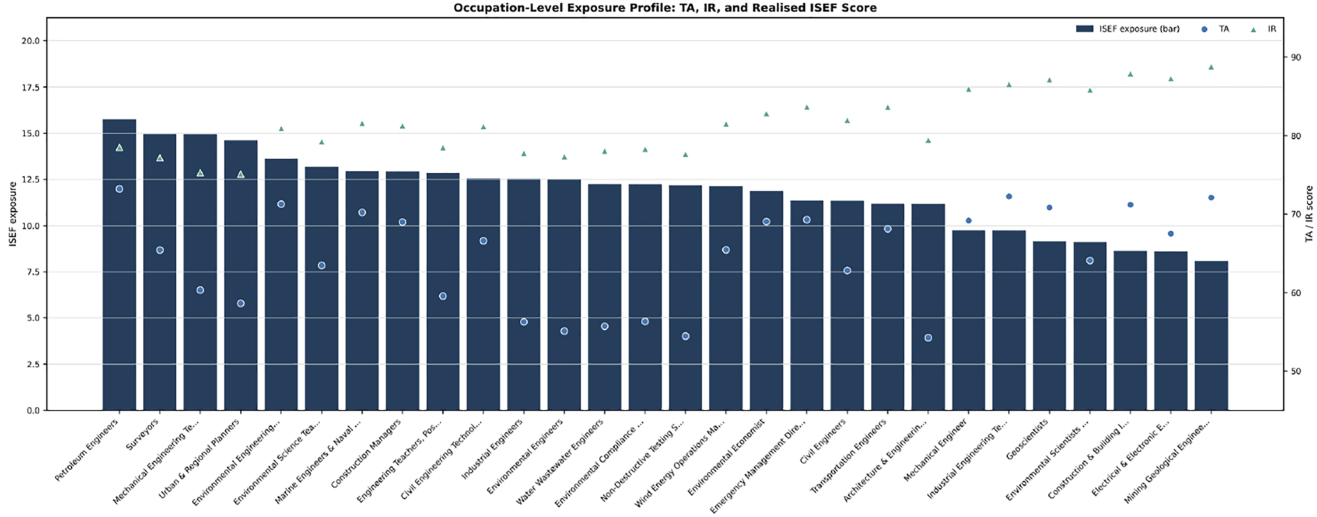

Figure 4. Occupation-level profiles of technical AI suitability (TA), institutional resistance (IR), and realised ISEF exposure

Table 6. ISEF variable mean scores by model

Variable	Dir.	ChatGPT-5	DeepSeek	Qwen-3	Consensus
Cognitive/Data Analysis	Fwd	6.87	7.39	8.02	7.38
Technological Integration	Fwd	6.81	7.49	7.85	7.40
AI Readiness	Fwd	7.67	8.13	8.38	8.05
Automation Potential	Fwd	7.01	7.51	7.75	7.46
Collaboration	Fwd	3.06	4.58	5.80	4.37
Innovation/Synergy	Fwd	5.30	5.68	6.63	5.83
Domain Expertise	Fwd	3.92	5.04	6.00	4.96
Compliance/Responsibility	Barrier	8.32	8.21	8.76	8.38
Industry Barriers	Barrier	7.45	7.59	7.53	7.57
Algorithm Limitation	Barrier	8.50	8.37	8.79	8.54
Data Security/Privacy	Barrier	7.24	7.07	7.52	7.27
Sociotechnical Challenges	Barrier	7.93	7.69	8.18	7.91

checks is narrower: whether practitioners draw a similar broad distinction between technical usefulness and institutional permissibility when judging realistic engineering tasks.

On that limited question, the pattern is consistent with the framework. Technical assistance is rated highest for course-material development (Engineering Teacher, mean D1 = 8.7), permit application preparation (Environmental Engineer, D1 = 7.2), and radiographic-image interpretation (NDT Specialist, D1 = 7.0). Institutional constraints are rated highest for notices of violation (Environmental Inspector, mean D2 = 8.3), completed building-inspection sign-off (Construction Inspector, D2 = 8.3), facility compliance inspection (Environmental Engineer, D2 = 8.0), and legal boundary descriptions (Surveyor, D2 = 7.7). The constructed calibration-only adjusted score (AI Help $\times(1 - \text{Barriers}/10)$) is highest for course-material development (6.4) and lowest for building-inspection sign-off (0.2) and notices of violation (0.4); it is a descriptive summary of these nine responses, not a validated exposure measure. Across the 15 task means, AI assistance potential and institutional constraints are strongly negatively associated ($r = -0.93$; $p < 0.001$), while AI assistance potential and the calibration-only adjusted score are strongly positively associated ($r = 0.93$; $p < 0.001$). The survey therefore supports the core TA/IR distinction, while still falling short of the larger same-task expert panel needed for full validation.

Reliability and Dispersion. Within-system variation averaged 0.73 on a 0–10 scale, and 89% of task-variable pairs showed low dispersion. Even so, a small set of occupations showed weaker cross-model rank agreement, which is why the cluster results are treated as exploratory. The silhouette analysis in Figure 8 supports the use of groupings for description, but not as stable occupational classes.

Discussion

Reading exposure as a sociotechnical indicator

Civil and environmental engineering is governed at least as much by review chains as by technical skill. Licensure, liability, and regulated sign-off decide what counts as acceptable evidence, who is allowed to issue it, and who carries the risk when something goes wrong. ISEF starts from that working reality and keeps it inside the measurement instead of treating it as background.

Several occupations in the sample combine high technical potential with substantial institutional resistance. The resulting exposure scores stay inside a moderate band, so “high exposure” here is a relative position within a regulated profession; it is not a prediction of displacement. The plausible entry points for generative AI in this domain sit at the edges of the

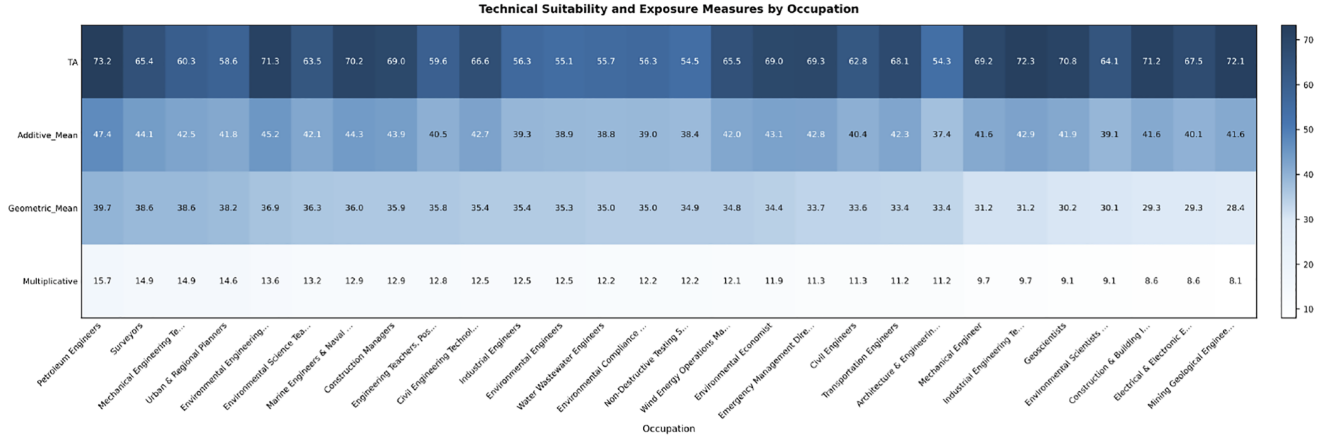

Figure 5. Technical-side heatmap across the 28 occupations

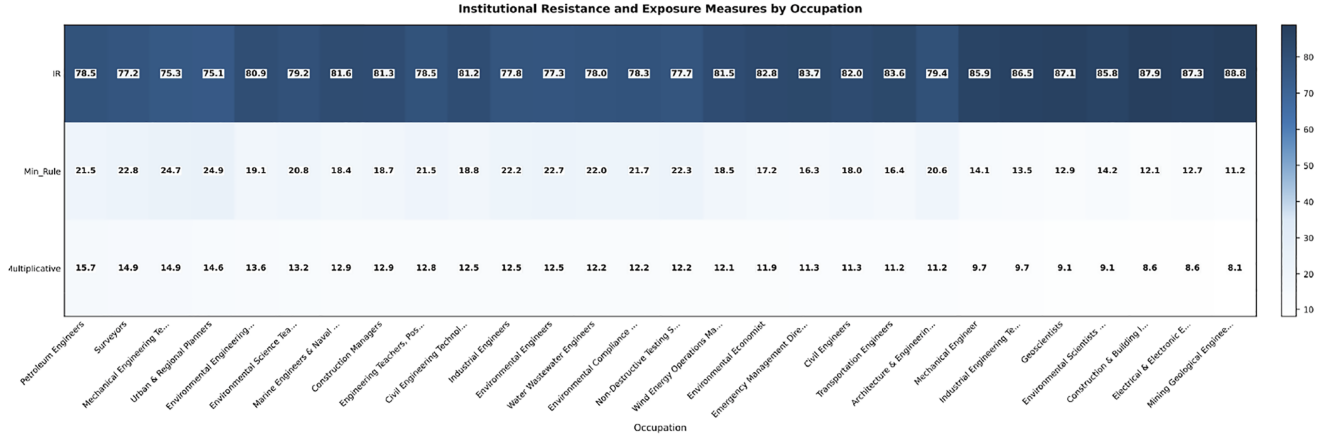

Figure 6. Institutional-side heatmap of institutional resistance (IR), min-rule exposure, and multiplicative exposure across the 28 occupations. The redundant realised-permission row is omitted because it is a direct linear transformation of IR

Table 7. Functional-form sensitivity checks

Specification	Rank correlation	Interpretation
Multiplicative	Primary model	*TA×(1-IR/100)*; treats institutional resistance as a partial gate on technical suitability
Additive mean	Spearman $\rho = 0.44$	Equivalent in ranking to *TA+(100-IR)* after rescaling. It gives less force to high institutional resistance than the multiplicative model
Min-rule	Spearman $\rho = 0.80$	Uses the lower of TA and realised permission. It compresses high-resistance occupations and supports broad-band rather than fine-rank interpretation
Geometric mean	Spearman $\rho = 1.00$	Equal-weight geometric mean is rank-equivalent to the multiplicative model because both are monotonic functions of *TA×(100-IR)*
O*NET weighting check	Spearman $\rho = 0.998$	Weighted and unweighted aggregation produce nearly identical rankings

Note(s): Additive mean, min-rule, and geometric mean are computed from the same occupation-level TA and IR scores as the primary multiplicative model. Full occupation-level scores are provided in the replication repository

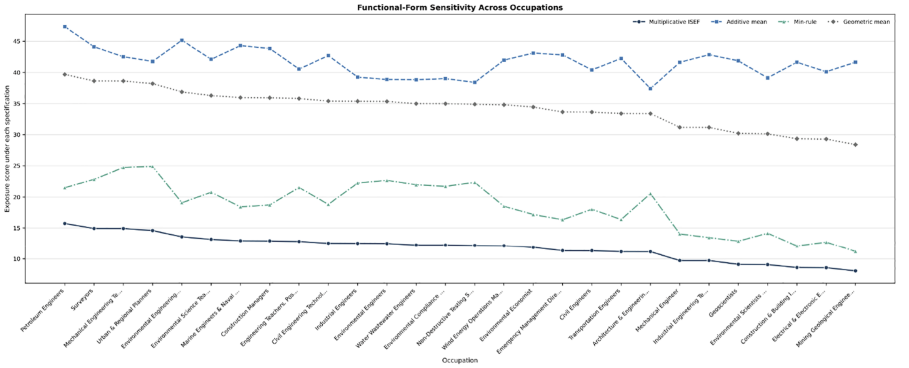

Figure 7. Functional-form sensitivity across multiplicative, additive, min-rule, and equal-weight geometric-mean occupation-level exposure specifications

Table 8. Cross-index benchmarking of the technical AI suitability component

Measure	Benchmark evidence	Interpretation
Technical AI Suitability (TA)	Pearson correlation with existing technical AI exposure indices: $r \approx 0.72$	The TA component aligns with prior technical-only exposure measures. This supports the technical side of the rubric without claiming realised adoption
ISEF Exposure	Descriptive divergence from technical-only rankings	Divergence is expected because ISEF applies institutional resistance as a gating factor. The difference is part of the sociotechnical argument, not a failure of the TA measure

Note(s): This benchmark is used as a concurrent-validity diagnostic for technical AI suitability. It does not validate realised adoption. Final ISEF exposure intentionally diverges from technical-only indices once institutional resistance is included

Table 9. Preliminary practitioner calibration survey after removing incomplete responses

Task	Survey vignette	*N*	AI help	Barriers	Calibration-only adjusted score
T01	Petroleum Engineer – Design drilling plans	9	5.2	5.3	2.5
T02	Petroleum Engineer – Analyze geological data to determine drilling locations	9	6.0	4.9	3.1
T03	Environmental Engineer – Prepare environmental permit applications	9	7.2	4.6	3.7
T04	Environmental Engineer – Inspect facilities for environmental compliance	9	4.2	8.0	1.1
T05	Mining Engineer – Prepare technical reports for regulatory agencies	9	6.0	5.4	2.6
T06	Geoscientist – Analyze soil, rock, and groundwater samples	9	5.2	5.2	2.9
T07	Civil Engineer – Review site plans for building code compliance	9	5.0	7.2	1.0
T08	Surveyor – Prepare legal property boundary descriptions	9	2.7	7.7	0.9
T09	Construction Manager – Review and approve contractor change orders	9	6.3	5.8	2.9
T10	Water/Wastewater Engineer – Design water treatment processes	9	5.3	5.8	2.5
T11	Environmental Inspector – Issue notices of violation	9	2.6	8.3	0.4
T12	NDT Specialist – Interpret radiographic images to identify defects	9	7.0	4.7	3.8
T13	Engineering Teacher – Develop course syllabi and materials	9	8.7	2.7	6.4
T14	Mechanical Engineer – Perform stress analysis using FEA software	9	4.6	6.7	1.7
T15	Construction Inspector – Sign off on completed building inspections	9	1.4	8.3	0.2

Note(s): Values are means on a 0–10 scale from 9 complete responses after excluding 2 incomplete records from the raw Google Forms CSV. The 15 vignettes and dimension wording are taken from question.pdf. D1 asks how much generative AI can technically assist the task assuming no legal, liability, or sign-off restrictions. D2 asks how much actual sign-off, liability, compliance, and audit-traceability requirements limit use of AI-generated outputs. A response was retained only when all 15 D1/D2 paired ratings were present and each task-level professional-judgement confirmation was marked yes. The calibration-only adjusted score is computed at respondent-task level as AI Help $\times(1 - \text{Barriers}/10)$ and then averaged. It is a descriptive calibration summary, not a validated exposure measure

workflow—drafting, checking, summarising, and supporting documentation—where outputs can be folded into review routines that already exist.

ISEF does not claim to capture every driver of adoption. Firm strategy, procurement culture, informal professional norms, and individual project leadership all matter and sit largely outside the index. What the framework adds is a way of keeping governance conditions in the same analytical frame as technical suitability, which task-only exposure measures struggle to do.

When technical fit meets accountability limits

The roles that look most open to AI assistance—structured design calculations, simulation support, technical documentation—are also the ones tightly coupled to public safety and infrastructure reliability. The same coupling keeps institutional resistance high in precisely the places where technical scores are high.

There is a forward-looking edge to this observation. As generative models become more capable, the consequences of an undetected error scale with them. If AI-assisted outputs feed into safety-critical work, oversight requirements may tighten before they relax. Institutional

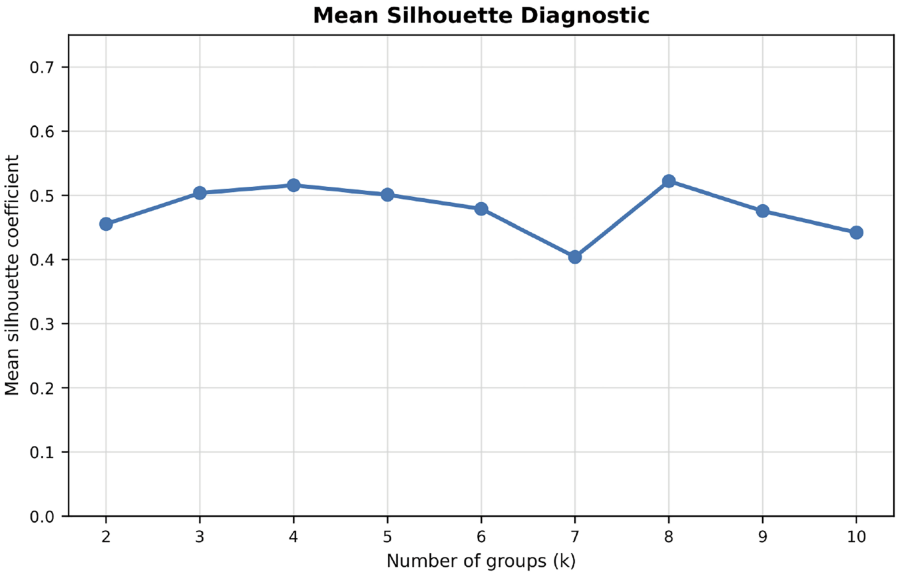

Figure 8. Silhouette analysis across candidate K-means solutions. The reported groupings are exploratory diagnostics, not validated occupational strata

resistance can therefore be a durable feature of the profession, not just a temporary lag behind the technology.

The picture is also less uniform than the occupation-level numbers suggest. Project governance, signing authority, and how individual firms allocate work between licensed and unlicensed staff almost certainly matter below the occupation level. Differences across licensure and liability regimes will also pull the same occupation in different directions across jurisdictions.

Potential applicability beyond engineering

ISEF is designed around a general regulated-work principle: technical capability becomes practically relevant only after it passes domain-specific accountability, documentation, privacy, and authority requirements. That principle may be useful in professions such as healthcare, law, accountancy, architecture, and regulated utilities, where AI-assisted outputs may be technically plausible but cannot be adopted without professional oversight. Transfer is not automatic. Each domain would need its own task universe, regulatory map, IR indicators, weights, and practitioner calibration; for example, clinical governance and patient confidentiality are not interchangeable with engineering sign-off and infrastructure liability. ISEF should therefore be treated as a portable measurement architecture rather than a ready-made cross-profession score.

Updating ISEF as capabilities and regulation change

Both sides of the framework are time dependent. Technical suitability can change as models gain tool use, multimodal reasoning, reliability controls, or domain-specific evaluation evidence. Institutional resistance can also change as regulators issue guidance, insurers revise coverage, clients amend procurement requirements, and professional bodies clarify acceptable review and sign-off practices. A future longitudinal implementation should version the TA and IR rubrics, retain dated regulatory evidence, re-score a fixed task panel at planned intervals,

and reconvene a larger practitioner panel when material changes occur. Such updates would allow observed changes in exposure to be separated from changes in the measurement rule itself.

What this means in practice

For workforce planning, exposure is most defensible as a descriptive baseline; it should not be read as a direct estimate of job-loss risk. In regulated roles, AI use is likely to settle first on documentation and analytical support, where outputs can be slotted into audit trails and existing oversight routines.

For governance, the framework surfaces a tension between AI capability and existing liability arrangements that will not dissolve quickly. The concrete levers sit exactly where that tension lives: AI-specific clauses in engineering procurement contracts, professional indemnity guidance for AI-assisted deliverables, audit-trail requirements for generated outputs, and clearer rules on whether a professional engineer's sign-off can rely on AI-supported calculations or documentation. These are not marginal adjustments to an exposure score; they decide which parts of a workflow are institutionally permissible in the first place.

For training and professional development, the heterogeneous exposure profiles imply differentiated needs. Some roles benefit from stronger emphasis on integrating and supervising generative tools; others need more weight on competencies tied to accountability, public safety, and document defensibility.

Limitations

A few limitations deserve to be named openly rather than left to the framing. LLM scoring showed strong agreement across the three systems, and that agreement may partly reflect shared training priors, especially in how legalistic phrasing is read. The practitioner survey adds an external check, but it is not the same as asking engineers to re-score the same O*NET tasks. Nine complete respondents provide preliminary calibration only; they cannot support population-level claims, task-level validation, or validation of the constructed calibration-only adjusted score. A larger, purposively stratified expert panel should re-score the same O*NET tasks sampled from high-, medium-, and low-exposure occupations, report inter-rater agreement and uncertainty, and test whether calibration patterns replicate across professional roles and jurisdictions. This design would allow direct expert–LLM correlations to be estimated.

Institutional resistance is also, openly, a constructed measure. Different rubric items, weightings, or functional forms would shift the resulting exposure profile. The current reading is a cross-section of present-day conditions; it does not track how procurement standards, professional indemnity, or sectoral regulation may move under generative AI. Following exposure through time, and linking it to observed adoption inside real projects, is work for the next study.

Conclusion

ISEF was built to look at generative AI exposure in civil and environmental engineering through two lenses at once: what a task is technically suited to, and what institutional rules will let through. The aim is descriptive. The framework reports where suitability and governance constraints meet; it does not try to time adoption or forecast job loss.

Applied to 674 tasks across 28 occupations, the analysis surfaces a pattern that is easy to miss when only the technical side is measured. Many roles read as open to AI assistance, yet the exposure score stays inside a moderate band once liability, compliance, and sign-off requirements are added to the picture. Indices that leave the institutional side outside the measurement will tend to misread how generative AI is likely to enter regulated engineering work in the near term.

The more useful question that follows is set at the workflow level, not the occupation level. Which parts of a typical project become institutionally permissible for AI support, under what review arrangements, and with what liability conditions? The same question can be tested in other regulated professions only after their institutional conditions have been specified and calibrated. The released data, scoring rubric, and code are intended as a starting point for that kind of follow-up: larger same-task expert panels, adoption-linked case studies, longitudinal rubric updates, and project-level tracing of where review and responsibility actually shift.

Ethics statement:
The main analysis uses publicly available occupational task data. The revision also reports an anonymous practitioner calibration survey that collected professional judgements without personal identifiers or sensitive personal data. Ethical approval was not required under the author's institutional understanding for this minimal-risk anonymous consultation.

Data access and replication:
The occupation-level scores, functional-form sensitivity outputs, preliminary practitioner-calibration files, questionnaire, and figure-generation code supporting the reported results will be made available in a public repository upon publication (or are available from the author upon request). The complete task-level model outputs and model-response logs are not retained in the current archive.

Acknowledgments
I sincerely thank my supervisor, Dr Stephen Suryasentana, for his guidance and support throughout the research. I am also grateful to all participants who contributed to the calibration. Finally, I thank the reviewers for their detailed and constructive feedback across several rounds of view.

Supplementary material
The supplementary material for this article can be found online:

References

Abbott, A. (1988). *The system of professions: An essay on the division of expert labor*. University of Chicago Press.

Acemoglu, D., & Restrepo, P. (2022). Tasks, automation, and the rise in US wage inequality. *Econometrica*, *90*(5), 1973–2016. doi: 10.3982/ecta19815.

Arntz, M., Gregory, T., & Zierahn, U. (2017). Revisiting the risk of automation. *Economics Letters*, *159*, 157–160. doi: 10.1016/j.econlet.2017.07.001.

Autor, D. H. (2015). Why are there still so many jobs? The history and future of workplace automation. *Journal of Economic Perspectives*, *29*(3), 3–30. doi: 10.1257/jep.29.3.3.

Azimi, M., Eslamlou, A. D., & Pekcan, G. (2020). Data-driven structural health monitoring and damage detection through deep learning: State-of-the-art review. *Sensors*, *20*(10), 2778. doi: 10.3390/s20102778.

Bock, T. (2015). The future of construction automation: Technological disruption and the upcoming ubiquity of robotics. *Automation in Construction*, *59*, 113–121. doi: 10.1016/j.autcon.2015.07.022.

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... and Liang, P. (2021). On the opportunities and risks of foundation models. arXiv:2108.07258.

Brynjolfsson, E., Mitchell, T., & Rock, D. (2018). What can machines learn?. In *AEA Papers and Proceedings* (Vol. 108, pp. 43–47).

Bubeck, S., Chadrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., ... and Zhang, Y. (2023). Sparks of artificial general intelligence: Early experiments with GPT-4. arXiv:2303.12712.

Calo, R. (2017). Artificial intelligence policy: A primer and roadmap. *UC Davis Law Review, 51*(2), 399–435.

Dell'Acqua, F., McFowland III, E., Mollick, E., Lifshitz-Assaf, H., Kellogg, K. C., Rajendran, S., ... Lakhani, K. R. (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of artificial intelligence on knowledge worker productivity and quality. Harvard Business School Working Paper, 24-013.

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., ... Wright, R. (2023). 'So what if ChatGPT wrote it?' Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management, 71*, 102642. doi: 10.1016/j.ijinfomgt.2023.102642.

Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2023). GPTs are GPTs: An early look at the labor market impact potential of large language models. arXiv:2303.10130.

Felten, E., Raj, M., & Seamans, R. (2021). Occupational exposure to artificial intelligence. *Strategic Management Journal, 42*(12), 2195–2217.

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review, 1*(1). doi: 10.1162/99608f92.8cd550d1.

Freidson, E. (2001). *Professionalism: The third logic*. University of Chicago Press.

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation?. *Technological Forecasting and Social Change, 114*, 254–280. doi: 10.1016/j.techfore.2016.08.019.

Goos, M., Manning, A., & Salomons, A. (2014). Explaining job polarization. *American Economic Review, 104*(8), 2509–2526.

Jasanoff, S. (2004). *States of knowledge: The Co-production of science and social order*. Routledge.

Jasanoff, S. (2016). *The ethics of invention: Technology and the human future*. W. W. Norton.

Kellogg, K. C., Valentine, M. A., & Christin, A. (2020). Algorithms at work: The new contested terrain of control. *Academy of Management Annals, 14*(1), 366–410. doi: 10.5465/annals.2018.0174.

Leonardi, P. M. (2012). Materiality, sociomateriality, and socio-technical systems: What do these terms mean? How are they related? Do we need them?. In P. M. Leonardi, B. A. Nardi, & J. Kallinikos (Eds.), *Materiality and Organizing: Social Interaction in a Technological World* (pp. 25–48). Oxford University Press.

Mohaghegh, S. D. (2017). *Data-driven reservoir modeling*. Society of Petroleum Engineers.

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science, 381*(6654), 187–192. doi: 10.1126/science.adh2586.

OpenAI (2023). GPT-4 technical report. arXiv:2303.08774.

Peterson, N. G., Mumford, M. D., Borman, W. C., Jeanneret, P. R., Fleishman, E. A., Levin, K. Y., ... and Dye, D. M. (2001). Understanding work using the occupational information network (O*NET): Implications for practice and research. *Personnel Psychology, 54*(2), 451–492. doi: 10.1111/j.1744-6570.2001.tb00100.x.

Raj, M., & Seamans, R. (2019). Artificial intelligence, labor, productivity, and the need for firm-level data. In A. Agrawal, J. Gans, & A. Goldfarb (Eds.), *The Economics of Artificial Intelligence: An Agenda* (pp. 553–566). University of Chicago Press.

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., ... Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 33–44). ACM.

Sacks, R., Eastman, C., Lee, G., & Teicholz, P. (2018). *BIM handbook: A guide to building information modeling for owners, designers, engineers, contractors, and facility managers* (3rd edition). Wiley.

Susskind, R., & Susskind, D. (2015). *The future of the professions: How technology will transform the work of human experts*. Oxford University Press.

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters via the gap statistic. *JRSS-B, 63*(2), 411–423. doi: 10.1111/1467-9868.00293.

Tolan, S., Pesole, A., Martinez-Plumed, F., Fernandez-Macias, E., Hernandez-Orallo, J., & Gomez, E. (2021). Measuring the occupational impact of AI: Tasks, cognitive abilities and AI benchmarks. *Journal of Artificial Intelligence Research, 71*, 191–236. doi: 10.1613/jair.1.12647.

Veale, M., & Borgesius, F. Z. (2021). Demystifying the draft EU AI Act. *Computer Law Review International, 22*(4), 97–112.

Webb, M. (2020). The impact of artificial intelligence on the labor market. SSRN Working Paper No. 3482150.

Weidinger, L., Uesato, J., Rauh, M., Griffin, C., Huang, P.-S., Mellor, J., … Gabriel, I. (2022). Taxonomy of risks posed by language models. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 214–229). ACM.

Corresponding author
Yuxuan Chai can be contacted at: yuxuan.chai@strath.ac.uk