Research Article

AI and Expert Perspectives: A Comparative Analysis on Training Needs Assessment of Sudurpaschim Province

Ichhya Shrestha[1] and Ashrita Dhital[2]

[1,2]Nepal Administrative Staff College, Jawalakhel, Lalitpur, Nepal

Abstract

Training Need Assessment (TNA) is one of the critical functions of human resource capacity development within the organization. However, the expert-based approaches appear to be comparatively more time-consuming, costly and susceptible to subjective biases. Therefore, this study presents a comparative study between AI-based and expert-based training need assessment within a public institution context, particularly of Sudurpaschim Province, Nepal. In this study, a mixed research approach has been employed to evaluate the quality, time and cost involved in conducting both types of TNA and stakeholders' acceptance towards AI-generated TNA. The quality of both TNA reports has been assessed through Goldstein & Ford model for TNA evaluation, while ISTAM (Intelligent Systems Technology Acceptance Model) has been employed to measure stakeholder acceptability of AI-generated TNA. The findings reveal that the quality of AI-based TNA is nearly comparable with expert-based TNA in organizational analysis. However, it shows significant differences in task/KSA analysis and person analysis. Expert TNA has an overall score of 74 percent, whereas AI TNA scores 67 percent. Furthermore, AI TNA takes 50.71 percent less time and 74.34 percent lower costs, but it achieves moderate stakeholder acceptance (66.57/100). The study recommends a blended approach where AI supports efficiency while human experts provide contextual depth and empirical grounding.

Keywords: Training Needs Assessment, Artificial Intelligence, ISTAM, Goldstein and Ford, Hybrid Approach

Introduction

Training Needs Assessment (TNA) is an important function of human resource development and institutional capacity building to increase organisational effectiveness and productivity (Becker, 1964). Historically, trained personnel including trainers, managers and subject specialists have carried out TNA within the Nepali public sector using both qualitative and quantitative techniques like surveys, performance assessments, and interviews to analyze the results for determining needs. Nevertheless, the process is highly subjective and takes

Corresponding author:
Ichhya Shrestha, Nepal Administrative Staff College, Jawalakhel, Lalitpur, Nepal.
E-mail: ichhya.shrestha@nasc.org.np
© Nepal Administrative Staff College
The Trainer: A Journal of Capacity Development
Volume 1 | Issue 1 | 2026
https://doi.org/10.3126/thetrainer.v1i1.102215

time to adjust to changing demands, and hence, is not able to address the evolving digital skills requirement of the public administration.

Artificial Intelligence (AI) is not a new concept. The theories behind artificial intelligence have been around for the last seven decades, applied by computer scientists such as Alan Turing, Marvin Minsky, and John McCarthy. AI offers the benefit of lower cost of data storage and higher computing power. Hence, many organizations are willing to unlock their value trapped in a huge and vast volume of data through AI. AI is a collection of technologies that enable machines to act with higher levels of intelligence and emulate human capabilities of sense, comprehension and act (NITI Aayog). AI offers a new paradigm to every sector, including training institutions. With the help of technologies such as machine learning, natural language processing, and predictive analysis, artificial intelligence will be able to recognize skill gaps and provide recommendations while minimizing subjective influence (Davenport & Ronanki, 2018). Moreover, the application of artificial intelligence in HR development is growing from recruitment processes, performance appraisal to tailored training (Davenport & Ronanki, 2018).

Nepal's 16th Five-Year Plan (2081/82-2085/86) strategically addresses structural obstacles in Nepal's development by integrating new technologies and management methods grounded in research and development. In this, AI acts as a bridge to achieve transformative change. The 16th Plan has acknowledged the development of AI research centers and the adoption of AI-driven e-governance and digitalization. The plan further prioritizes human resource development to cultivate a skilled workforce for the digital era. Similarly, the National AI Policy is intended to bolster by institutionalizing AI education, vocational training and capacity building programs. The National AI Policy approved by the Government of Nepal in August 2025 is an important milestone towards ethical and inclusive AI (Government of Nepal, 2025). This policy aligns with the Digital Nepal Framework (2019) and 16th Plan's vision of a knowledge-driven economy (Government of Nepal, 2019; National Planning Commission, 2024). However, there are few empirical studies validating AI in public sector applications such as TNA.

The emerging global issue of data colonialism, where developing countries turn out to be passive data providers to AI systems managed by other countries, is a matter of grave concern. According to the findings of PwC (2017) and Harari (2024), nearly 70 percent of AI's projected $15.7 trillion global value by 2030 will concentrate in China and North America, further escalating the divide between data-rich and data-poor countries. Against this backdrop, using AI without strong local governance may increase dependency and inequity instead of empowerment. Thinkers like Harari (2024), Suleyman (2025), and Tegmark (2017) have examined how large language models operate, and caution against treating these systems as autonomous decision makers in stable situations where data are easily available. However, AI falls short and cannot adapt to uncertain situations where there is lack of data. In this regard, A large language model processes text by breaking it into tokens and learning,

from patterns across billions of documents. In this, words and phrases are statistically likely to follow one another. On this basis, it generates a range of plausible continuations and, in effect, "bases sensible actions on those predictions" rather than exercising independent judgement (Suleyman & Bhaskar, 2025). Tegmark (2017) similarly observes that although such models can pass the Turing Test. They discover patterns and relations among words without relating these words to anything in the real world, and they continue to struggle with common sense reasoning, as shown by the Winograd Schema Challenge. This distinction has profound meaning for public administration indicating current LLMs are best understood as decision support tools that generate probabilistic, pattern-based outputs for human review, not as decision makers that reason or act independently. Even so, there are limited studies to prove whether AI can match or excel the efficiency, accuracy, conceptual depth, and practical value of human judgement. This raises critical questions for Nepal's public administration particularly in TNA: Can AI truly understand contextual nuances and human learning needs? Should training decisions in the public sector rely on algorithmic judgement? Therefore, it is imperative for countries like Nepal to conduct research on AI in various domains like TNAs to understand the efficacy of AI.

This study is relevant for several reasons. First, it offers novel empirical research on the comparative study of expert-led and AI-driven TNA within Nepal's public sector context, addressing a critical gap in organizational capacity building. Second, the findings of the study will provide evidence for policymakers and training institutions on the use of AI tools for training needs analysis. AI helps institutions to optimize time, cost and human resources without compromising quality. Third, the study is valuable for the wider discussion about digitalization in public sector HRM practices, providing insights on how AI can augment human capability instead of replacing them. Finally, by examining stakeholder acceptability in AI-generated TNA, the study highlights practical considerations for implementation, ensuring that AI-driven TNAs are actionable and contextually appropriate.The general objective of this study is to conduct a comparative analysis of AI-based and expert-driven TNA of Sudurpaschim Province. Specifically, the study aims to: (1) compare the quality of TNA between expert and AI versions; (2) compare the time and cost required for each approach; and (3) measure stakeholder acceptability of AI-generated TNA.The study is guided by two research questions: How do quality, cost and time differ between the two methods? And how do stakeholders accept AI-driven TNA compared to expert-driven ones?

Literature Review

Artificial Intelligence and Its Landscape in Nepal

Albaladejo-González et al. (2024) claimed that AI is advancing in training and assessment. However, AI is not fully implementable because most of the ideas are in the research stage. Chen (2022) suggested that AI systems in training can shift the stages of knowledge base construction, surveys, training structuring and feedback processes. In the context of training, large language models could process large volumes of words and translate and

summarize them, a capacity which was exclusive to human skills. Furthermore, machine learning models can predict the kind of skills needed in the future (Bessen, 2019). However, Suleyman and Bhaskar (2025) suggested that AI should not replace human contact or ability, but the safety of AI should be continually enacted. Tegmark (2017) asserts that though AI can pass the Turing Test, it does not clearly understand context. Also, AI is not very good at common sense tasks which is proven through Winograd Schema Challenge.

AI technology in Nepal is still at an early stage. According to the AI Readiness Index 2024, Nepal is ranked 150 out of 188 countries. The National Artificial Intelligence Policy 2082 established artificial intelligence as a computer system that can reason, engage in decision-making, problem-solving, and learning like humans. The adoption of National AI Policy 2082, Digital Nepal Vision 2023 and the Sixteenth Five Year Plan 2081/82-2085/86 hints at Nepal's commitment toward incorporating AI across governance structures, public service delivery including human resource development (Government of Nepal, 2025; National Planning Commission, 2024). The policy makes provisions to establish key institutions including AI Supervision Council, National AI Center, and the AI Regulatory Authority and prioritizes sectorial implementation of AI-enhanced public services and for HR upskilling. However, the policy efforts do not align with empirical findings of AI in a provincial context. As per a study by Mahat (2024) in Ramechhap district, despite being positive about the potential of AI, the public servants are concerned about infrastructure readiness and digital literacy.

On the contrary, the AI policy landscape is geopolitically divided. Considering this, Nepal must develop its own regulatory strategy that incorporates international best practices and be agile enough to accommodate local requirements (Kushwaha et al., 2025). Without proper local governance structures that empirically test AI technologies, Nepal would largely remain trapped in perpetual dependence rather than enabling its own indigenous capacity (Malhotra, 2021).More broadly, recent research on AI adoption across public organizations points to organizational readiness, resources, skills, and governance arrangements as key determinants of successful implementation, rather than the technical sophistication of the AI system alone (Neumann et al., 2024). This reinforces the argument that Nepal's institutional and regulatory preparedness, not just AI availability, will determine whether public sector applications such as TNA can be responsibly scaled.

Training Needs Assessment: Theory and Practice

Training Needs Assessment is defined as the process of determining the discrepancy between the current and desired level of performance and finding ways to bridge the gap by educational intervention (McGehee & Thayer, 1961). Generally, Training Needs Assessment involves three layers of analysis: organizational, task, and individual (Goldstein & Ford, 2002). Traditional TNA processes, guided by an expert, have strong reliance on personal judgement, contextual understanding and tacit knowledge while being highly prone to bias and inconsistency (Brown, 2002). On the other hand, AI could be used to discover the

trainee characteristics and tailor learning programs based on their unique KSAs, eliminating human bias (Maity, 2019).

AI's comparative weakness in task/KSA and person analysis, established later in this study's findings, is consistent with the nature of tacit knowledge itself. Knowledge that experts hold but rarely document explicitly. The knowledge on structured task inventories, direct observation of current behavior, and iterative interaction cannot be documented explicitly in the text alone. Task and KSA analysis require judgement about the frequency, importance and difficulty of tasks that are difficult to extract from documentation. Similarly, person analysis depends on contextual intelligence, intuition, behavioral assessment and access to primary performance data, appraisal records and supervisor input which are also rarely fully captured in secondary sources. Because organizational culture and informal institutional cues are similarly absent from most written records, AI systems that rely on document synthesis are structurally better positioned for organizational level analysis than for the field level, person centered stages of the Goldstein and Ford model.

The theoretical foundation of this research is based on the Training Need Evaluation Process Model developed by Goldstein and Ford (2002). This model comprises five stages including organizational analysis, requirement analysis, task and KSA analysis, person analysis, and design and evaluation input. Organizational analysis refers to assessing organizational goals, resources, climate, and constraints. Requirement analysis involves defining target jobs and boundaries, while task and KSA analysis focus on identifying tasks and determining the necessary knowledge, skills and abilities for each task. Person analysis identifies trainees and their needs, and the final stage of design and evaluation input reviews existing programs and aligns selection with training. In this study, this model served as the baseline in designing instruments for measuring the quality of TNAs.

The Goldstein and Ford (2002) framework was selected for four reasons. It is theoretically foundational, representing the most widely cited refinement of McGehee and Thayer's (1961) three-level organization-task-person (O-T-P) model for over six decades. It is comprehensive yet diagnostic, enabling level-specific comparison between AI-generated and expert-generated reports rather than a single aggregate score, which aligns with the study's findings. It is evaluative rather than descriptive, treating needs assessment as the foundational input to training design and evaluation, which aligns directly with the study's objective of comparing final TNA report quality. Lastly, it offers institutional fit, as it is the framework already embedded in the TNA practice of the Nepal Administrative Staff College (NASC), which ensure that evaluation criteria were meaningful to practitioners and consistent with the institution's own standards.

Prompt Engineering for AI-Based TNA

Generative AI refers to systems capable of producing new text, images or other content. LLMs are a class of generative AI trained on very large volumes of text to identify statistical

patterns between words and tokens. When an input is given to LLM, it predicts and generates the most statistically plausible continuation rather than retrieving a stored answer or reasoning independently about the request (Suleyman & Bhaskar, 2025; Tegmark, 2017). NotebookLM and Claude, introduced below, are both LLM based tools used in this study strictly as decision support instruments for drafting and structuring the AI-generated TNA report.

In this study, the prompt for generating the AI-supported TNA report was engineered following established best practices. Google's Gemini prompting guidebook (2024) has simplified effective prompt design around four core elements: persona, task, context, and format. We followed these practices by assigning a clear expert role, providing detailed contextual framing specific to Sudurpaschim Province, using structured instructions with delimiters, specifying output format matching NASC's TNA templates, emphasizing evidence-based claims with source disclosure, and incorporating safeguards against hallucination. The prompt was further refined through iterative testing and expert feedback from different IT professionals. In this study, two tools were combined for a specific reason. NotebookLM was used first because it is source-grounded by design, restricting its output to the documents and links it is given, which significantly reduces hallucination. A comparative evaluation of ChatGPT, Gemini and NotebookLM on a 300-page reporting task found hallucination rates of roughly 40 percent for ChatGPT and Gemini compared with only 13 percent for NotebookLM (Hagar et al., 2025). However, NotebookLM's source grounded output was comparatively concise and lacked the analytical depth required for a full TNA report. Its output was therefore used as structured input for Claude. Claude is a model recognized for stronger reasoning, more coding-like structured output, effective handling of long context inputs and a comparatively lower tendency toward hallucination (Anthropic,2025).

Therefore, making it better suited to producing a detailed, traceable final report. NotebookLM thus provided grounded raw synthesis, while Claude provided the structuring and analytical elaboration. The final refined prompt was applied across multiple AI models: NotebookLM was used initially due to its tendency to remain strictly within provided sources; this output was then fed into Claude Opus 4.7 for more detailed, structured analysis. This was substantiated by a study that found NotebookLM produced hallucinations in 13 percent of outputs compared to approximately 40 percent for ChatGPT and Gemini (Hagar et al., 2025).The complete engineered prompt, including its data handling and traceability rules, is provided in Appendix A for reproducibility.

Stakeholder Acceptability: ISTAM Framework.

Technology Acceptance Model (TAM) reasons that acceptance is influenced by perceived usefulness and ease of use, which then influence attitudes and intention to use (Davis, 1989). Unlike conventional IT technologies, AI systems tend to generate complicated

and ambiguous output, raising "black box" concerns. Vorm & Combs (2022) proposed the Intelligent Systems Technology Acceptance Model (ISTAM) which is an extension of TAM by incorporating transparency and trust. According to ISTAM, transparency increases perceptions of ease of use and usefulness and trust, which in turn strengthens acceptance. The five constructs for assessing stakeholder acceptability in this study were: perceived ease of use, transparency, trust, perceived usefulness, and behavioral intention.

In public institutions specifically, acceptability depends not only on usefulness and trust but also on legitimacy and governance expectations. Perceived legitimacy is threatened when algorithmic systems weaken participation, procedural fairness, transparency, accountability and equity of outcomes (Grimmelikhuijsen et al., 2022). The perceived fairness, accountability and transparency are strongly linked to trust and adoption of algorithmic services more broadly (Shin & Park, 2019).

Despite growing global interest in applying AI in Human Resource Management and TNA, several important research gaps persist. No study has yet compared the quality of AI-generated TNA reports with expert-prepared reports within Nepal's public sector. There is limited empirical evidence on stakeholder acceptance of AI-generated TNA recommendations. Although AI has been applied in skill-gap analysis globally, its effectiveness in resource-constrained public institutions remains largely unexplored. The theoretical limitations discussed above show AI's pattern based rather than reasoning-based operation, and its structural weakness in eliciting tacit knowledge gap. This study later reports in task/KSA and person analysis, suggesting the gap is not incidental, but a predictable consequence of how these systems process information. Finally, little guidance exists on how AI can be meaningfully integrated into existing TNA workflows without diminishing the contextual judgment that human experts provide. This study addresses these gaps through a systematic comparative analysis grounded in Nepal's emerging policy context.

Conceptual Framework

The conceptual framework of this study positions the TNA as the independent variable (traditional expert-driven approach or the AI-driven approach) on the same organizational datasets. The dependent variables are quality, time and cost for comparing two approaches. Firstly, the quality of TNA is assessed in relation to the Goldstein and Ford model of Training Need Evaluation process that consists of organizational analysis, requirement analysis, task and KSA analysis, person analysis and input design and evaluation of training. Secondly, efficiency is assessed by the time (in hours) and cost (in NPR) taken to complete each approach. Both reports are scored and results are compared. The evaluative variable is stakeholder acceptability specific to AI condition only. It assesses the willingness of managers and employees to endorse, rely on and implement AI-generated TNA. This is measured through perceived usefulness, ease of use, transparency and behavioral intention drawing on TAM, UTAUT and ISTAM frameworks (Davis, 1989; Venkatesh et al., 2003; Vorm & Combs, 2022).

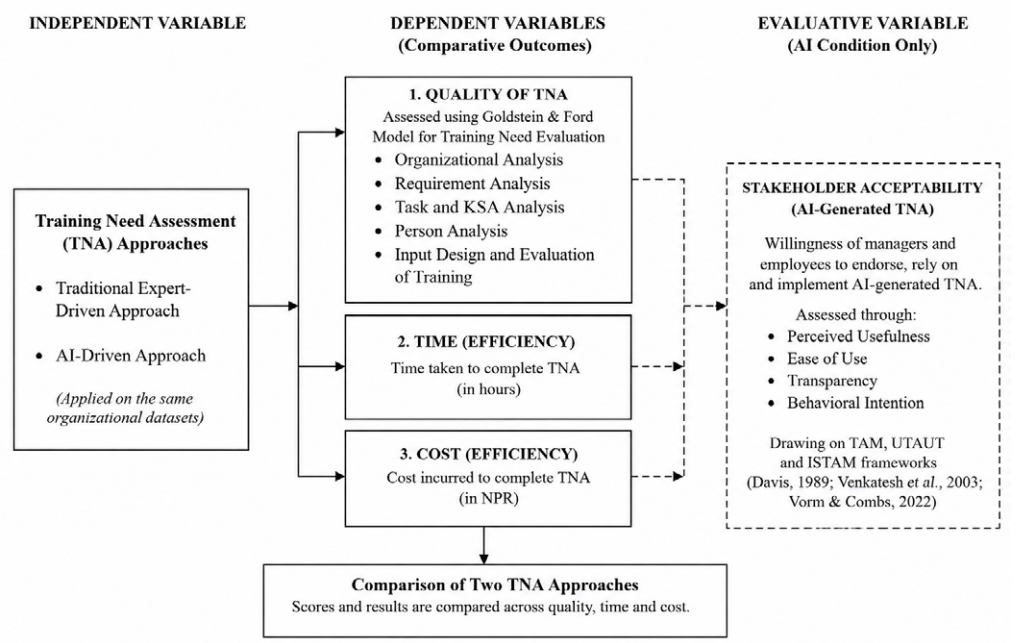


Figure 1: Conceptual framework of the study

Methods

The study follows a mixed-method design. Specifically, the study adopts a convergent parallel mixed methods design (Creswell & Plano Clark, 2018), in which quantitative and qualitative responses were collected simultaneously, analysed independently, and merged at the interpretation stage. The rationale for combining both approaches is methodological triangulation (Denzin, 1978; Creswell & Plano Clark, 2018): A purely quantitative comparison of scores would have reported that the expert report scored higher in task/KSA analysis without explaining the substantive reasons stakeholders gave for that gap; a purely qualitative account would have captured those reasons but could not have established the magnitude of the time (50.71%) and cost (74.34%) savings or the reliability of the measurement instruments. By merging both, the study achieves breadth and depth. The quantitative strand establishes the size and direction of differences, while the qualitative strand contextualizes and explains them, with points of convergence between the two approaches strengthening the overall confidence in the findings (for example, efficiency emerging as the dominant theme in both ISTAM scores and interview comments).

Since NASC conducted various expert-driven TNAs, a comprehensive TNA report with access to raw data served as the baseline. The same dataset including the format and layout were then used to generate the AI-assisted TNA. By keeping the dataset constant, the research design ensures that differences in outcomes can be attributed primarily to the

method of analysis (expert vs. AI) rather than variations in data sources. Both reports were then evaluated for quality using the same blind rated instrument, and the AI report alone was additionally assessed for stakeholder acceptability.

Population, Sample, and Participants

The population of this study includes all TNAs that NASC completed during the two fiscal years 2080/81 and 2081/82 for various government institutions in Nepal. From this population, the TNA report of Sudurpaschim Province prepared by the experts was purposely selected as a sample, where a thorough expert-driven TNA had been done. Additionally, the report was comprehensive and had access to raw data for further analysis. The participants of expert TNA involved 46 government officials holding positions above 7th levels taken from different ministries of Sudurpashchim Province. To assess the quality of both the expert-driven and AI-generated reports, a panel of independent evaluators, including stakeholders and academics, was formed. Three external evaluators provided unbiased assessments of both reports. Additionally, 15 participants from various provincial ministries reviewed the overall quality of the two reports, while 17 stakeholders evaluated the acceptability of the AI-generated report.

Data Collection Methods

The study employed two main data collection methods. First, secondary data collection was carried out through the existing expert-driven TNA reports prepared by NASC and the underlying raw data. These documents served as input for generating the AI-driven TNA. Second, primary data for comparing report qUality was collected through a structured workshop-based approach. Reports were anonymized and labeled as Report A and Report B. Participants were given materials two to three days prior to the workshop for thorough review and were then briefed on evaluation criteria. They were divided into groups of two members each for structured discussion and evaluation. For assessing acceptability of the AI-generated report, stakeholders were approached individually, briefed on study objectives, given the AI report, and asked to evaluate using a structured sheet within two to three days.

Instruments

The main research tool was the TNA Evaluation Scoring Sheet, which was based on Goldstein and Ford's Model of Training Need Evaluation Process and comprised five essential dimensions: organizational analysis, requirement analysis, task and KSA analysis, person analysis, and design and evaluation input. A 5-point rating system was used for each dimension, with 1 denoting "strongly disagree" and 5 denoting "strongly agree." The study used Vorm and Combs's ISTAM tool, which measures perceived ease of use, transparency, trust, perceived utility, and behavioral intention on a 5-point scale, to assess stakeholder acceptability of the AI-generated TNA. Additionally, brief feedback surveys and semi-structured interviews were employed to record qualitative remarks. Three NASC faculty with TNA expertise participated in a pilot test to improve the assessment framework and questionnaire.

Data Analysis

Both quantitative and qualitative methods were used in the study. In quantitative analysis, time and cost data were presented descriptively in percentages and the average scores provided for the reports were compared. Descriptive statistics like means and percentages were used to examine survey results. Qualitative data from interviews was analyzed thematically to compare quality and capture stakeholder acceptability. When comparing AI and expert-driven TNA, this mixed-methods approach offers quantifiable and interpretive insights.

Data Validity and Reliability

Content validity was established by designing the instrument in consultation with TNA NASC experts. Furthermore, the prompt for generating the AI TNA went through several rounds of interactions and was finalized in consultation with three IT professionals consisting of a software engineer with six years of experience in the field of AI, a NASC IT specialist and an IT professional, ensuring content validity. In addition, internal validity was strengthened by applying the same dataset for both TNAs and through blind evaluation by three independent raters using the same NASC report template. Reports were anonymized and labeled only as Report A and Report B with randomized order, ensuring evaluators remained blind to report type. Because each report can differ in tone or phrasing, both reports were also reformatted into the identical NASC report template and section structure before evaluation, so that formatting and stylistic cues could not signal report origin to the raters.

The Cronbach's alpha for Expert TNA, across all 25 TNA quality items, is 0.91, indicating that respondents answered consistently across all the questions. Compared to Expert TNA, AI-generated TNA scores slightly higher with the reliability score of 0.95, showing stronger construct-level reliability. Similarly, for the acceptability assessment, Cronbach's alpha of 0.87 indicates good internal consistency. This shows that respondents answered the acceptability items in a coherent and consistent manner, reflecting a reliable measurement of constructs. Thus, the findings can be considered methodologically credible and trustworthy. Because a Cronbach’s alpha this high can also reflect item redundancy or overlap between related constructs rather than only true reliability. It is recommended that future studies apply confirmatory factor analysis and assess discriminant validity before drawing casual conclusions from similarly high internal consistency values. Regarding ethics, all personal identifiers were removed from datasets before AI processing, data was used strictly for research purposes, and informed consent was sought from relevant departments, employees and experts. Also, AI-assisted tools were used for language refinement.

Ethical Consideration

Regarding ethics, all personal identifiers were removed from datasets before AI processing, data was used strictly for research purposes, and informed consent was sought from relevant departments, employees and experts. Also, AI-assisted tools were used for language refinement.

Results and Discussion

Time and Cost Efficiency

The expert TNA required 41 working days (984 hours), incurring a total expenditure of NPR 787,692 for the service fee, transportation, and allowance for four people. In contrast, the AI-enabled TNA process was done in only 20 days and 5 hours (485 hours), with an associated total expenditure of NPR 202,087. This includes the cost of subscribing to the AI system and travel for fieldwork. Using the AI-enabled TNA would be 50.71 percent quicker and 74.34 percent cheaper, calculated as:

$$\textit{Time Saved} = (984 - 485) / 984 \times 100 = 50.71\%$$

$$\textit{Cost Saved} = (787{,}692 - 202{,}087) / 787{,}692 \times 100 = 74.34\%$$

Table 1: TNA Cost and Time Comparison — Expert and AI-Based

Item	Expert TNA	AI-Based TNA
Time Breakdown		
Questionnaire / Prompt design	10 days	10 days
Data collection / Data preparationvw	14 days (7 field + 7 tablework)	5 days
Data cleaning / Data feeding	7 days	5 hours
AI generation time	—	3 days
Analysis, report writing / Editing & refinement	10 days	2 days
Total days	41 days	20 days + 5 hrs
Total hours	984 hrs	485 hrs (↓ 50.71%)
Days allocated (as per proposed plan)	67.5 days	—
Cost Breakdown (NPR)		
Professional Fee		
Professional fee calculation	67.5/120 × 1,051,600	—
Professional fee total (NPR)	591,525	—
AI Subscription (AI-Based TNA only)		
Claude + GPT subscriptions ($40 × 148)	—	5,920
Travel & Logistics (4 team members)		
Transport (NPR)	137,667	137,667
Daily allowance (NPR)	58,500	58,500
TOTAL COST (NPR)	787,692	202,087 (↓ 74.34%)

Quality Comparison: Goldstein and Ford Training Need Evaluation Model

The total score for the Expert TNA (Report A) and the AI TNA (Report B) were 74 percent and 67 percent, respectively, representing a quality gap of 7 percent. Across Goldstein and Ford's five phases of TNA analysis, the comparative scores between the two reports were: organizational analysis showed near parity (Expert: 14.53; AI: 14.00); requirement analysis had a moderate gap (19.07 vs. 16.53); task/KSA analysis had the largest gap (30.53 vs. 24.93); person analysis had a moderate difference (18.07 vs. 15.20); and finally, design and evaluation input had the closest difference (10.67 vs. 9.60).

Organizational analysis: near parity (14.53 vs. 14.00)

Goldstein and Ford position organizational analysis as the foundation of TNA, examining organizational goals, resources, climate and constraints. The near-identical scores suggest that AI's capacity to synthesize secondary sources such as strategy documents, annual reports and policy texts is functionally equivalent to expert judgment at this stage, which suggests that organizational analysis is largely a desk based, document-driven activity. However, experts can still check organizational context through informal cues and stakeholder conversations that secondary sources alone do not capture.

Requirement analysis: moderate gap (19.07 vs. 16.53)

This stage defines the target job and its boundaries the entry point into the analytical cascade. The expert's advantage here is consequential rather than merely numerical meaning if the job scope is wrongly bounded, every downstream decision about tasks and KSAs is built on a flawed foundation. AI's lower score likely reflects its inability to probe ambiguities, since it cannot ask why a role is structured a certain way or detect inconsistencies between a formal job description and actual practice, whereas experts navigate this through interviews, observation and iterative clarification.

Task and KSA analysis: largest gap (30.53 vs. 24.93*)*

This is the technical heart of the Goldstein and Ford model and the point of greatest divergence. The expert's higher score reflects an ability AI cannot easily replicate conducting structured task inventories, observing incumbent behavior, and eliciting tacit knowledge through interaction. Deriving KSAs particularly distinguishing what should be selected for versus trained requires judgement about the frequency, importance and difficulty of tasks that is difficult to extract from text alone. AI can competently map documented tasks, but it struggles with tasks that are rarely written down, contextually variable, or visible only through direct observation. This gap represents the most significant practical concern for treating AI-generated TNA as a replacement rather than a supplement to expert judgement.

Person analysis: moderate gap (18.07 vs. 15.20)

Person analysis answers Goldstein and Ford's two operational questions: who needs training, and what kind. This requires real performance data, appraisal records, supervisor

assessments and competency ratings, ideally gathered at the individual level. The expert's advantage reflects access to primary data sources and the ability to triangulate across them. AI can infer performance gaps from aggregated or publicly available information but cannot evaluate the performance of specific employees, a limitation more about data access than analytical ability, one that could narrow if an AI system were integrated with internal HR data rather than limited to external sources.

Design and evaluation input: closest gap (10.67 vs. 9.60)

This stage involves reviewing existing programs, designing evaluation frameworks, and aligning selection with training systems. Both approaches performed comparably, reviewing existing training programs, designing assessment and performance appraisal tools, and aligning selection and training systems by identifying which KSAs should be hired for versus trained. Neither scored particularly high in absolute terms, suggesting this phase may have been less thoroughly addressed by both approaches.

The findings reveal that the effectiveness of AI decreases as the model moves from the broader organizational analysis (document-driven approach) to a narrowed-down task/KSA and personal analysis (field-level/personal approach). AI is strong in secondary synthesis but weak in personal/primary observations. This reinforces Kahneman & Klein's (2009) claim that experts often take a long time, need many resources, and are exposed to evaluators' cognitive biases.In terms of practical significance, a 7 percent gap on a 25-item scoring instrument is not trivial. It is concentrated almost entirely in the two stages task/KSA and person analysis that Goldstein and Ford treat as the technical and operational core of a TNA. This means the gap is operationally meaningful in the sense that an AI-generated report can be trusted more readily for organization level framing than for the specific, actionable task and person level recommendations that ultimately drive training design. This pattern is broadly consistent with evidence from other domains. For instance, a validity study of AI-based clinical skills assessment found that AI could match trained human raters on structured, well defined rating tasks, but such parity is more attainable in narrow, clearly bounded assessments than in the open-ended, context dependent judgments required at the task/KSA and person levels of TNA (Johnsson et al., 2023).

Qualitative Analysis for Comparing TNA Quality

Based on study participants' feedback, the Expert TNA appeared to be descriptive and comprehensive. It conducted a holistic assessment of the situation by addressing the gaps in knowledge, skills, attitude, as well as behavioral aspects including the presence of "hakim" mentality. Moreover, this approach was specific enough to identify the differences between the entry-level and in-service officials. Majority of participants perceived it as objective, practical, and representative of the province-specific context. Nevertheless, the participants raised concerns on the absence of monitoring and evaluation mechanism, coverage of multiple levels of government, poor link between the training needs and performance, limited sample size (46 participants out of 88 local governments and 9 districts), potential

social desirability bias from self-assessment, as well as the underrepresentation of induction training for 8th level officials.

The AI-generated TNA was considered concise and well-structured, showing the use of triangulation consisting of direct evidence, secondary sources, and inference. The TNA showed an effective connection between training practices at both national and international levels with SPARTA. It also shows that its analysis was in line with the pertinent constitutional and legislative acts. One of the key strengths of the AI-TNA was its clarity and prioritization framework based on frequency, severity, strategic significance, and feasibility. However, the priority matrix assigned equal weight to all four parameters without any explanation. Besides this, the TNA did not have the wider contextual outlook of the expert-generated report and relied heavily on documentation over first-hand sources such as quantitative standards, competencies and performance indicators.

Overall, out of 15 study respondents, 8 preferred the Expert TNA, 5 remained neutral, and only 2 preferred the AI-generated TNA. Those preferring the expert report found it more practically relevant and implementable due to its practical grounding. Supporters of the AI report appreciated its clarity, simplicity, and structured presentation.

Stakeholder Acceptability: ISTAM Results The average overall score for AI-based TNA acceptability was 66.57 out of 100, indicating moderate and conditional acceptance. Across the four ISTAM constructs, usefulness scored highest (mean 3.62), followed by behavioral intention (mean 3.33), transparency (mean 3.26), and ease of use (mean 3.09).

Usefulness was driven primarily by efficiency: time saving scored 4.00 and effort saving scored 3.71, both the highest items in the entire study. However, items related to AI decision quality (3.35) and accurate identification of training needs (3.41) remained only moderate, suggesting stakeholders primarily value efficiency gains rather than fully trusting the quality of AI-generated findings.

Behavioral intention reflected cautious and consistent support across all three items (intent to support = 3.35, accept as input = 3.29, willingness to recommend = 3.35). Transparency revealed a critical explainability gap: stakeholders could verify outputs (3.41) but could not clearly see the reasoning process behind them (3.12) or whether limitations were adequately disclosed (3.24). For instance, in the quality comparison, the AI-generated TNA's prioritization matrix, that ranked training needs based on four key factors: frequency, feasibility, severity, and strategic importance, was highly appreciated by the evaluators for its structured and systematic presentation. However, concerns emerged regarding the transparency of the prioritization process, as all four factors appeared to be assigned equal weight without any clear explanation, rationale, or underlying logic. This lack of clarity created "black box" concerns among stakeholders, making it difficult for them to understand how the final prioritization scores were generated. As a result, issues related to trust,

interpretability, and stakeholder acceptance were raised, despite the perceived usefulness of the matrix itself.

Ease of use was the most concerning construct (mean 3.09). Within it, the item measuring whether the AI-supported TNA fits into the workflow without extra effort or support scored just 2.41, the only item in the entire survey to fall below the 3.0 midpoint signaling active disagreement rather than mere neutrality. Because stakeholders could still use and interpret the AI outputs reasonably well (use = 3.29, interpret = 3.47), the low score is better read as a friction point about embedding AI outputs into existing institutional systems than as a rejection of the outputs themselves. Similarly, a related item on whether recommendations fit the organization's policy and systems scored a modest 3.18. This pattern points less to a technological limitation of the AI tools than to organizational readiness, institutional processes, reporting formats and decision-making systems. The study is consistent with public servants' documented concerns about infrastructure readiness and digital literacy elsewhere in Nepal (Mahat, 2024) The strategic implications are that human judgment is increasingly necessary, particularly when organizations grow more complex with AI adoption (Goldfarb & Lindsay, 2022).

Qualitatively, stakeholders accepted AI-based TNA primarily for its time-saving capability, efficiency, and data-driven analysis. Main reasons for non-acceptance included concerns of trust in AI, that its policies and processes lack clarity, that recommendations appear generic, and that AI fails to capture regional differences such as geographical, social, cultural, economic, and legal variations.

Interpreting the Quality Gap

The 7 percent quality gap (74% vs. 67%) indicates that the participants favored expert TNA more than AI TNA. This conclusion emphasizes Wilson and Daugherty's (2018) assertion that even though AI typically has an upper hand in recognizing patterns and predicting future outcomes, it falls behind humans when it comes to contextual analysis and ethical decision-making. Likewise, the near parity scores in organizational analysis is consistent with a review on Flevy by Robinson (n.d), which depicts AI performing well on document-driven, desk-based tasks. Likewise, the considerable gap observed in task and KSA analysis, which Goldstein and Ford (2002) consider as the technical heart of TNA, reveals that AI cannot understand structured task inventories, behavior, and tacit knowledge from text alone.

Harari's (2024) philosophical caution about "alien intelligence" is especially relevant in this context. AI may produce convincing results without genuine understanding of underlying processes involved, thereby creating what Duncan (2026) terms as "workslop". The report generated by AI may look well-crafted, however, is devoid of substance and context required for decision-making. Substantiating this, the study implies that expert-generated TNA was able to recognize the attitudinal and behavioral factors, like the "hakim mentality,"

illustrating contextual understanding can only be achieved through human analysis, which AI cannot analyze simply based on documents and website access.

Efficiency Gains and Their Implications

The 50.71 percent time and 74.34 percent cost savings in AI-generated TNA represents substantial efficiency gains which uphold Yan et al.'s (2024) assertion of AI's potential to enhance learning through personalized assistance, quick feedback, and novel assessment techniques. Considering Nepal's resource-strained public sector, such gains have a notable impact on the organizations that are limited in budget and time frames. These results align well with Maity (2019), signifying AI overcoming lengthy and tiresome TNA procedures. However, cost reductions should be considered along with quality results, as lower cost does not mean much unless the produced TNA gives clear recommendations. The stakeholders' moderate quality trust (M = 3.35 for decision quality) suggests that efficiency gains are recognized but do not necessarily translate into confidence in AI-generated findings.

Stakeholder Acceptability and the ISTAM Framework

The ISTAM findings show a pattern consistent with the framework's notion. Usefulness, specifically efficiency gains, is the primary driver of acceptance. Yet the transparency gap where outputs can be seen but the reasoning process behind them cannot be fully understood represents the classic explainability problem identified by Vorm and Combs (2022). The stakeholders could verify what the AI produced (M = 3.41) but remained uncertain about how and why it was produced, since the weakest transparency items concerned the availability of information to review the process (M = 3.12) and the disclosure of limitations (M = 3.24). This is not merely a technical shortfall but a matter of trust. Without visibility into the reasoning path, stakeholders cannot independently judge whether a recommendation follows sound logic or reflects an unexamined assumption, which is precisely the concern that surfaced around the equally weighted prioritization matrix. According to Grimmelikhuijsen et al. (2022), legitimacy of algorithms as tools of decision-making is compromised when systems reduce participation, procedural fairness, and accountability just like stakeholders from Sudurpaschim Province have emphasized. Related evidence on algorithmic services more broadly finds that perceived fairness, accountability and transparency are closely linked to trust and to willingness to adopt algorithmic tools (Shin & Park, 2019). Workflow integration emerged as the most important factor (M = 2.41) amongst the aspects of acceptability. This denotes that the issue is not whether the stakeholder considers AI useful per se, but whether AI output fits within established organizational workflows such as institutional processes, reporting formats and decision-making systems.Digital literacy and institutional readiness compound this challenge. The provincial level public servants elsewhere in Nepal have expressed similar concerns about infrastructure readiness and digital literacy when adopting AI tools (Mahat, 2024), and NASC's own established reporting practices and templates were not originally designed to accommodate AI-generated input, which likely contributes to the friction stakeholders reported. In that regard, it augments the results presented by Neumann

et al. (2024) about the role of organizational readiness, resources, skills, and governance arrangements.

A high internal consistency values (Cronbach's Alpha) observed in this study, while indicative of strong reliability, may also reflect potential item redundancy or multicollinearity among related constructs. Future researchers are cautioned to conduct confirmatory factor analysis and assess discriminant validity indices (e.g., Average Variance Extracted) to rule out multicollinearity before drawing causal inferences from similar measurement models.

Beyond acceptability, the use of AI to support public-sector decision-making raises broader questions of governance, accountability and ethics that extend past the individual stakeholder's trust. Because algorithmic legitimacy in government depends on preserving participation, procedural fairness, accountability and equitable outcomes (Grimmelikhuijsen et al., 2022; Shin & Park, 2019), institutions adopting AI-supported TNA should treat explainability as governance requirements. Every AI-generated recommendation that feeds into a training or budgetary decision should be traceable and subject to accountable human review, consistent with this study's finding that stakeholders accept AI output more readily when they trust the process that produced it.

Complementary Role of AI in TNA

The triangulation of quantitative ratings and qualitative comments shows a clear pattern suggesting a blended approach. An expert-based TNA brings the analytical rigor associated with primary research findings. The AI-based TNA, on the other hand, provides better policy alignment and justifications for addressing governments where policy alignment entails institutional weight. According to the Hybrid Intelligence Framework proposed by Dellermann et al. (2019), a combination of AI and human judgment leads to better results as opposed to either approach independently.This is consistent with related human-in-the-loop evidence from clinical skills assessment, where AI-based scoring showed strong agreement with trained human raters on structured tasks but performed best when paired with, rather than substituted for, expert judgement (Johnsson et al., 2023). A pattern this study's own findings reproduce at the organizational level (where AI performs comparably to experts) versus the task/KSA and person levels where expert judgement remains distinctly stronger.

Delimitations of the Study

The delimiting factors of the research are also crucial to address. The raw data used for the AI analysis was first noted in diaries by the expert team. This data had certain nuances that could only be interpreted through firsthand field experience. Likewise, comparisons drawn were limited to selected chapters of the expert's report. Also, prompt engineering required significant intervention and skill, translating to results not easily replicable without comparable competence and guidance. These factors highlight why an AI-only technique remains inadequate.

Conclusion and Implications

The comparison between expert and AI-based TNA of Sudurpaschim Province is a novel study in Nepal's public sector context. The key findings of this study are that expert TNA outperforms AI-generated TNA in terms of overall quality (74% compared to 67%), particularly in the stages of task/KSA and person analysis. However, AI achieves near parity scores in document-driven organizational analysis while saving 50.71 percent in time and 74.34 percent in cost compared to expert-led TNA. The score of 66.57 out of 100 implies moderate and conditional acceptance of TNA generated by AI. The findings reveal that usefulness driven by efficiency in time and cost are the strongest drivers for accepting AI-generated TNA. However, the lowest mean score of 2.41 in integrating AI into the workflow remains the greatest challenge in acceptance. Likewise, the transparency reveals a critical explainability gap where stakeholders can verify output but cannot understand the reasoning behind them, indicating the importance of human judgement and contextual intelligence particularly in field-level analysis.

The study offers five evidence-based recommendations. First, AI-generated TNA should disclose the methodology used in gathering, analyzing, and interpreting the data used for the AI output. Second, a human-in-the-loop approach should be adopted as a formal design approach for TNA with expert validation as a required step. A clearer implementation framework consistent with the stage level findings above can be framed as:

- AI-supported tasks: document review, literature and policy synthesis, data organization, preliminary analysis, report drafting, formatting, and prioritizing training needs.
- Human-led tasks: field observations, interviews, task/KSA analysis, person analysis, contextual interpretation, stakeholder consultation, validation of AI outputs, and final decision-making.

Third, AI results should be localized to Sudurpaschim Province's context by using locally verified, organization-specific data aligned with the country's rules, regulations, and frameworks. Fourth, governments should plan to enhance capacity building of users to use AI as well as build clear AI policies for broader acceptance. Fifth, AI-assisted TNA is more suitable as complementary input to experts' decision-making rather than the standalone basis for determining training needs.

Since the study is based on one province of the country, a key limitation of this study is that the findings of the study cannot be generalized across Nepal's diverse public sector contexts. Similarly, the AI output was also heavily dependent on the availability and quality of raw data and on the prompt engineering skills. The findings are further bound by the specific AI models used NotebookLM for source-grounded synthesis and Claude for structured elaboration with prompt engineering expertise applied to combine them. Because LLMs are updated frequently and their hallucination and reasoning characteristics vary across models

and versions (Hagar et al., 2025), the magnitude of the quality gap and the time and cost savings reported here should be treated as specific to this model pairing and this point in time and may not hold identically for other LLMs or future model versions. Because the study is based on a single provincial case, the extent to which these findings transfer to other contexts warrants caution. Sudurpaschim Province's own sample of 46 respondents was already acknowledged as limited relative to its 88 local governments and 9 districts. The same logic likely applies across provinces and federal agencies with comparable resource constraints, dataset availability and TNA practices to Sudurpaschim Province may see similar patterns, but provinces or agencies with denser primary data, more mature digital infrastructure or different institutional workflows could see a narrower or wider quality gap and different acceptability results. Broader implementation would therefore be most appropriate where a comparable expert baseline dataset exists for validation and where institutions have already begun building the digital capability and workflow readiness this study finds lacking in Sudurpaschim Province. However, the methodology and prompting techniques could set the tone for comparable study in the future.

Further research should expand across multiple contexts, institutions and different provinces, investigate the integration of AI with real-time HR data systems, develop standardized AI governance frameworks for TNA in Nepal's public sector along with guidelines for implementation More broadly, the hybrid AI and expert approach recommended here offers a concrete mechanism for operationalizing Nepal's National AI Policy 2082 with assigning AI to document review, synthesis and drafting while reserving field level analysis and final judgement for experts. The institutions can therefore pursue the Policy's aims of efficient, ethical and transparent AI enhanced public service delivery without displacing the contextual and evidence-based human resource development that Nepal's public sector requires. In this regard, the study's contribution extends beyond TNA itself to demonstrate, empirically, one workable model for responsibly integrating AI into public-sector governance and evidence based HRD more broadly. As Nepal continues to implement its National AI Policy 2082 and advance toward a knowledge-based economy, the responsible and evidence-based integration of AI into public sector capacity building will be essential for sustainable human resource development.

Finally, future research should also investigate under what conditions a hybrid AI-expert model can close the quality gap in task/KSA and person analysis while preserving the time and cost efficiencies demonstrated in this study. Three subsequent questions follow: How can human-in-the-loop configurations be designed to compensate for AI's limitations in field-dependent analysis? How can explainability and workflow integration gaps be addressed to strengthen stakeholder trust and institutional adoption? And do the patterns observed in Sudurpaschim Province hold when replicated across other provinces, institutions, and TNA datasets? Answering these questions would shift the focus from whether AI can match expert judgement to how the two can be most effectively combined in resource-constrained public institutions.

Data Availability Statement

The data supporting the findings of this study are available from the corresponding author upon reasonable request, subject to confidentiality, anonymity, and ethical requirements. Personal identifiers were removed from the datasets before AI processing, as described in the Methodology section.

Declaration of Interest

The authors declare that they have no competing interests relevant to this article.

Funding Statement

This research was supported by a faculty internal research grant from the Nepal Administrative Staff College.

Declaration of Generative AI Use

Generative AI tools were used for generating and structuring the AI-based Training Needs Assessment (TNA) report as part of the research process, and for language refinement of the manuscript. The authors reviewed, verified, and took full responsibility for the final content of the article.

References

Albaladejo-González, M., Gómez Mármol, F., & Ruipérez-Valiente, J. A. (2024). Artificial intelligence to support the training and assessment of professionals: A systematic literature review. *ACM Computing Surveys, 57*(3), 1–29. https://doi.org/10.1145/3699712

Anthropic. (2025). *Claude 4 system card*. https://assets.anthropic.com/m/6c940a1b69ed6a1c/original/Claude-4-System-Card.pdf

Becker, G. S. (1964). *Human capital: A theoretical and empirical analysis*. University of Chicago Press.

Bessen, J. (2019). *AI and jobs: The role of demand* (NBER Working Paper No. 24235). National Bureau of Economic Research. https://doi.org/10.3386/w24235

Brown, J. (2002). Training needs assessment: A must for developing an effective training program. *Public Personnel Management, 31*(4), 569–578. https://doi.org/10.1177/009102600203100412

Chen, Z. (2022). Artificial intelligence-virtual trainer: Innovative didactics aimed at personalized training needs. *Journal of the Knowledge Economy, 14*(2), 2007–2025. https://doi.org/10.1007/s13132-022-00985-0

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.

Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard Business Review, 96*(1), 108–116.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340. https://doi.org/10.2307/249008

Dellermann, D., Ebel, P., Söllner, M., & Leimeister, J. M. (2019). Hybrid intelligence. *Business & Information Systems Engineering, 61*(5), 637–643. https://doi.org/10.1007/s12599-019-00595-2

Denzin, N. K. (1978). *The research act: A theoretical introduction to sociological methods* (2nd ed.). McGraw-Hill.

Duncan, D. S. (2026, February 3). How do workers develop good judgment in the AI era? *Harvard Business Review*. https://hbr.org/2026/02/how-do-workers-develop-good-judgment-in-the-ai-era

Goldfarb, A., & Lindsay, J. R. (2022). Prediction and judgment: Why artificial intelligence increases the importance of humans in war. *International Security, 46*(3), 7–50. https://doi.org/10.1162/isec_a_00425

Goldstein, I. L., & Ford, J. K. (2002). *Training in organizations: Needs assessment, development, and evaluation* (4th ed.). Wadsworth/Thomson Learning.

Google. (2024, October). *Prompting guide 101*. https://services.google.com/fh/files/misc/gemini_for_workspace_prompt_guide_october_2024_digital_final.pdf

Government of Nepal. (2025). *National AI policy 2082*. Ministry of Communications and Information Technology. https://aiassociationnepal.org/wp-content/uploads/2025/08/AI_Policy_ko51psc.pdf

Government of Nepal. (2019). *Digital framework Nepal 2019*. Author.

Grimmelikhuijsen, S., & Meijer, A. (2022). Legitimacy of algorithmic decision-making: Six threats and the need for a calibrated institutional response. *Perspectives on Public Management and Governance, 5*(3), 232–242. https://doi.org/10.1093/ppmgov/gvac008

Hagar, N., Agustianto, W., & Diakopoulos, N. (2025). *Not wrong, but untrue: LLM overconfidence in document-based queries*. https://arxiv.org/pdf/2509.25498

Harari, Y. N. (2024). *Nexus: A brief history of information networks from the Stone Age to AI*. Random House.

Johnsson, V., Herling, L., Tolsgaard, M. G., Petersen, O. B., Søndergaard, M. B., Kulasegaram, K., Sundberg, K., & Tiblad, E. (2023). Validity evidence supporting clinical skills assessment by artificial intelligence compared with trained clinician raters. *Medical Education, 58*(1), 105–117. https://doi.org/10.1111/medu.15190

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist, 64*(6), 515–526. https://doi.org/10.1037/a0016755

Kushwaha, M. K., Neupane, R., & Khanal, A. (2025). *Artificial intelligence and diplomacy: A case for Nepal's AI diplomacy*. CESIF.

Mahat, D. (2024). Readiness for artificial intelligence integration in government services: Perspectives from Ramechhap district employees. *Madani Multidisciplinary Journal, 4*(4), 486–495. https://doi.org/10.55927/mudima.v4i4.8435

Maity, S. (2019). Identifying opportunities for artificial intelligence in the evolution of training and development practices. *Journal of Management Development, 38*(8), 651–663. https://doi.org/10.1108/JMD-03-2019-0069

Malhotra, R. (2021). *Artificial intelligence and the future of power: 5 battlegrounds*. Rupa Publications India.

McGehee, W., & Thayer, P. W. (1961). *Training in business and industry*. John Wiley & Sons.

National Planning Commission. (2024, June). *16th five-year plan (2081/82–2085/86 BS)*. Government of Nepal. https://npc.gov.np/content/6461/sixteen-plan--fy-fy-fy-fythting-2081-81-8285-85-86-/

Neumann, O., Guirguis, K., & Steiner, R. (2024). Exploring artificial intelligence adoption in public organizations: A comparative case study. *Public Management Review, 26*(1), 114–141. https://doi.org/10.1080/14719037.2022.2048685

PwC. (2017). *Sizing the prize: What's the real value of AI for your business and how can you capitalize?* PwC Global Artificial Intelligence Study: Exploiting the AI revolution. https://www.pwc.co.nz/insights-and-publications/2023-publications/artificial-intelligence-study.html

Robinson, J. (n.d.). *How is AI technology being leveraged to automate and enhance the TNA process?* Flevy. https://flevy.com/topic/tna/question/optimizing-tna-process-ai-technology-a-deep-dive

Shin, D., & Park, Y. J. (2019). Role of fairness, accountability, and transparency in algorithmic affordance. *Computers in Human Behavior, 98*, 277–284. https://doi.org/10.1016/j.chb.2019.04.019

Suleyman, M., & Bhaskar, M. (2025). *The coming wave*. Vintage.

Tegmark, M. (2017). *Life 3.0*. Penguin Random House UK.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27*(3), 425–478. https://doi.org/10.2307/30036540

Vorm, E. S., & Combs, D. J. Y. (2022). Integrating transparency, trust, and acceptance: The Intelligent Systems Technology Acceptance Model (ISTAM). *International Journal of Human-Computer Interaction, 38*(18–20), 1828–1845. https://doi.org/10.1080/10447318.2022.2070107

Wilson, H. J., & Daugherty, P. R. (2018). Collaborative intelligence: Humans and AI are joining forces. *Harvard Business Review, 96*(4), 114–123.

Yan, L., Greiff, S., Teuber, Z., & Gašević, D. (2024). Promises and challenges of generative artificial intelligence for human learning. *Nature Human Behaviour, 8*(10), 1839–1850. https://doi.org/10.1038/s41562-024-02004-5

Appendix A: AI Prompting Protocol

The NASC report was divided into three sections, and each section was assigned to AI as an independent task. For each task, the AI was first provided with the corresponding section of an existing NASC report as a reference to understand its format, structure, and writing style. Specifically, Section 1- Background was used as the reference for Task 1- Section 1 specific links, resources, Section 2- Contextual Studies for Task 2- Section 2 specific links and resources, and Section 3- Analysis for Task 3-section 3 specific links and resources. Based on this understanding and the provided prompt, the AI then generated the respective section of the report. Then each generated section was compiled into one single report. This was done because when a single report was fed to AI.

Role: You are acting as a specialist in Training Need Assessment (TNA). Your task is to generate a formal TNA report for Sudurpaschim province, Nepal, from the provided resources.

Context: This report is intended for provincial government stakeholders. It must replicate the formal tone and structure of the NASC TNA report template.

Language: The entire report must be written in Formal Nepali.

Resources: Both qualitative and quantitative data collected from interviews and surveys, along with secondary data sources such as website links, books, constitutional laws and regulations, are provided for analysis (The resources here were :the format of the NASC report, Task, Links to the websites for contents, Good Governance (Management and Operation) Act, 2008 (2064 B.S.),Local Government Operation Act, 2017 (2074 B.S.), Constitution of Nepal, Quantative data provided by the experts for their TNA study. All these resources and websites links were solely based on resources that expert used for their TNA study. Besides these, no other resources were used)

Here are data handling and traceability rules that you must follow to generate Training Need Assessment report.

1. Source Fidelity: You should only the sources provided to you. Do NOT fabricate statistics, quotes, or historical facts. Do not hallucinate.
2. Data Gaps: If specific data for a table or section is missing from the uploaded sources, write: "[DATA GAP: Insufficient data available]".
3. Traceability: You must link every finding to its source. You can use the phrases to indicate that it is based on the survey results or interview respondents.

4. Reasoning: In the findings you must distinguish between Direct Evidence (e.g., a specific percentage of respondents reporting a need), Inference (e.g., what a finding suggests about an underlying gap), and Triangulation (e.g., where multiple data sources converge on the same conclusion).
5. Prioritization Logic: You must rank training needs in the report based on frequency, severity, strategic importance, and feasibility.

I have provided you the links and documents (resources), a format of TNA (file name: eg Task 1, Task 2, Task 3). Draft a TNA report based on the command, resources, and Task 1 format.

Workflow: The compiled resources were first processed in NotebookLM to produce a source-grounded synthesis, which was then used as structured input for Claude to produce the final, more detailed and traceable report, consistent with the two-stage process described in the Prompt Engineering section.