5 / 2026

REPORT

Large Language Models in Analytical Processes

Insights from NOKUT's Thematic Analyses of the Third Periodic Review Cycle

2026

NOKUT

NOKUT – The Norwegian Agency for Quality Assurance in Education – is a government administrative body under the Ministry of Education and Research.

NOKUT has its own board and operates independently in professional matters related to accreditation, supervision, and evaluations to assess the quality of higher education and tertiary vocational education. In addition, NOKUT carries out administrative tasks on behalf of the Ministry.

The purpose of NOKUT's activities is to ensure and promote quality in higher education and tertiary vocational education, and to promote quality development that ensures a high international standard in the educational programmes offered by the institutions. NOKUT's purpose and responsibilities are established in the Universities and University Colleges Act and the Vocational Education Act.

NOKUT aims to ensure that society has confidence in the quality of Norwegian higher education and tertiary vocational education. Through its work, NOKUT seeks to support institutions in their quality assurance efforts.

NOKUT uses expert committees in accreditation, supervision, evaluations, and projects.

You can read more about our work at nokut.no.

Title	Large Language Models in Analytical Processes
Author(s)	Gustavo Guajardo, Jon Lanestedt, Philipp Emanuel Leivestad Friedrich
Date	15.01.2026
Report no.	5-2026
ISSN-no	1892-1604

Preface

NOKUT's third round of periodic reviews, conducted between 2017 and 2024, has resulted in a total of 48 reviews and corresponding review reports. This has given us the opportunity to examine overarching patterns, shared challenges, and examples of good practice in the institutions' systematic quality work over the past eight years.

In the process of organising, analysing and validating the data and findings, we have received valuable support from an internal reference group at NOKUT. The group has consisted of colleagues with solid experience from the third round of periodic reviews: Ane B. Lillehammer, Eva Refsdal, Frøydis Maurtvedt, Hedvig Maria Bergem, and Hege Brodahl, as well as Gustavo Guajardo, who has contributed with specialised expertise in the use of artificial intelligence in the analytical work, and Sumera Majid for proofreading and language editing.

The reports aim to support the further development of systematic quality work in the sector. We hope the findings will provide a useful basis for reflection, learning, and continued quality enhancement.

Table of Content

Abstract

This report explores how a large language model, specifically *NotebookLM*, developed by *Google DeepMind,* can serve as a support tool in analytical processes related to external quality assurance in higher education. The analysis draws on 48 institutional review reports from NOKUT's third round of periodic reviews of higher education institutions in Norway. These reports form the database for the study, with a particular focus on common challenges and examples of good practice in quality assurance and quality enhancement. By using *NotebookLM* to analyse the reports, we gained insight into the model's ability to identify key challenges and effective practices in institutional quality work. Compared to manual analysis, *NotebookLM* demonstrates strengths in processing large volumes of data and detecting patterns. However, manual analysis offers a deeper understanding of context and nuance. The combination of AI-supported and manual analysis thus emerges as the most robust and comprehensive approach. The findings indicate that while *NotebookLM* shows promise as an analytical support tool, its effective use depends on structured data, well-formulated prompts, proficiency in AI tools, subject-matter expertise, and human oversight to ensure quality and accuracy.

1 Introduction

The purpose of this report is to explore how large language models can be used to analyse data from the third round of NOKUT's periodic reviews of higher education institutions in Norway (2017-2024), and how this compares to traditional manual analysis methods.[1] The analysis draws exclusively on publicly available data in the form of NOKUT's official review reports, which do not contain personally identifiable information. The main report addresses three research questions, the third of which is:

- ***How can NOKUT use artificial intelligence to analyse the data from the third round of periodic reviews?***

This report does not present a comprehensive review of the findings generated by the language model. Instead, it reflects on the experience of using a large language model as an analytical tool in the evaluation of institutional review reports. It forms part of a broader thematic analysis consisting of three interconnected reports (see Table 1). The overarching aim of this analysis is to further develop NOKUT's methodology for periodic reviews, support institutions' internal quality efforts, and provide authorities, the sector, and the broader public with deeper insight into challenges and development opportunities in the internal quality assurance of higher education.

Given the rapid pace of technological advancement, it is important to examine how emerging tools, such as large language models (hereafter referred to as "models" or "artificial intelligence, (AI)"), can contribute to more efficient and improved analysis and evaluation processes, particularly for external quality assurance bodies. We also assume that such models may offer new perspectives on quality work in higher education.

We believe that this report will be particularly relevant for other quality assurance agencies considering the use of AI in their analytical processes. For a fuller understanding of systematic quality work at Norwegian universities and university colleges, we recommend reading the two additional reports that form part of this thematic analysis.

Table 1 Overview of Report Structure

NOKUTs Thematic Analysis	
Report A	Thematic Analysis of NOKUT's Third Periodic Review Cycle – Main Report
Report B	Thematic Analysis of NOKUT's Third Periodic Review Cycle – University Colleges with Accredited Study Programmes and No Self-Accreditation Rights
Report C	**Large Language Models in Analytical Processes – Insights from NOKUT's Thematic Analyses of the Third Periodic Review Cycle**

[1] We have used the licensed version of *Microsoft Copilot* to proofread the reports in this thematic analysis, both to support high linguistic quality and to evaluate this functionality in practice.

2 Evidence Base and Analysis Process

2.1 Evidence base

The primary data source for this analysis consists of review reports from NOKUT's third periodic reviews cycle. A total of 48 reports, amounting to approximately 1,800 pages and covering the period from 2017 to 2024, were reviewed using a large language model to analyse how institutions describe and demonstrate their quality assurance efforts in the reports.

Each of the nine review requirements is typically assessed over two to three pages within each report.[2] For each review, NOKUT appoints an expert committee responsible for the assessments presented in them. These assessments are based on documentation submitted by the institutions as part of the review process.

In this third round, NOKUT required documentation at both the institutional and programme levels. The institutional-level documentation was intended to provide an overarching view of how the institution systematically ensures and enhances the quality of its study programmes. By contrast, the programme-level documentation was intended to illustrate how quality assurance is actually practised within individual study programmes and to offer concrete examples of how the institution's quality system is applied.

In the reviews, particular emphasis is placed on how well the institution's quality assurance practices comply with the review requirements, both in terms of formal procedures and their practical implementation.

A complete list of institutions and their corresponding review reports is provided in Appendix 2.

2.2 Use of the AI Model *NotebookLM*

The AI-assisted analysis of the review reports was conducted using *NotebookLM*, a model developed by *Google DeepMind* for contextual information analysis and insight generation. *NotebookLM* is built on *Google's Gemini* model and is designed to help users analyse large volumes of text-based content. It functions as an AI-powered assistant for working with complex document collections, enabling users to extract key information, summarise content, and identify connections across multiple sources.

NotebookLM operates by analysing and extracting information solely from the sources provided by the user. For the model to respond to questions, it must first be granted access to a defined dataset, such as uploaded documents, links to websites, or video transcripts. The model processes this material and bases its responses strictly on the content within these sources. Unlike general-purpose language models such as *ChatGPT-4o*, *Gemini*, or *Claude*, *NotebookLM* does not access external databases or the open internet. This means

[2] See Appendix 1 for further details on the wording of the requirements for systematic quality assurance work.

that although the model is built on a *Retrieval-Augmented Generation* (RAG) framework, unlike many other RAG solutions, *NotebookLM* does not integrate a large, external knowledge base. This gives the user complete control over the information the model uses to generate answers. For example, if asked “How do you boil an egg?”, *NotebookLM* will only be able to respond if that information is present in the uploaded material. If not, it will refrain from generating an answer based on assumptions or general knowledge. This design significantly reduces the likelihood of the model producing fabricated or misleading responses, commonly referred to as "hallucinations" in AI.

An additional strength of *NotebookLM* is that it includes references to the sources from which its answers are drawn. This built-in citation feature supports transparency and facilitates the quality assurance of the model’s outputs.

Although some versions of the *Gemini* model support context windows of up to one million tokens[3] (enabling, in theory, the simultaneous processing of vast amounts of text) this does not mean that *NotebookLM* analyses entire documents all at once. As noted earlier, the tool uses a *Retrieval-Augmented Generation* (RAG) approach, retrieving and processing only the most relevant chunks of text as needed. This allows for faster and more targeted responses.

While this method enables the handling of large datasets, its accuracy can be compromised if the documents are poorly structured, overly repetitive, or contain conflicting information. According to Google Support (2025), *NotebookLM* performs most effectively when working with a total document volume of 100 to 200 pages, provided that the content is well-organised and internally consistent.

Nevertheless, errors may occur. The model does not "understand" content in the same way a human does; instead, it generates responses based on patterns in the text. As a result, it may occasionally misinterpret, oversimplify, or incorrectly merge information, leading to imprecise or misleading conclusions.

Therefore, human quality assurance remains essential, particularly in professional or high-stakes contexts where accuracy and nuance are critical.

2.3 Methodological Development and Systematisation of the Dataset

A key prerequisite for generating meaningful and relevant results with models like *NotebookLM* is the formulation of the input, commonly referred to as a *prompt*, and hereafter also called an *instruction*.

[3] Tokens are small pieces of text that language models use to understand and process text. A token can be a word, part of a word or punctuation. For example, the sentence ‘This is a test.’ is divided into approximately 5 tokens. Models have a limit to how many tokens they can handle simultaneously (this is called the ‘context window’). While some models support up to 1 million tokens, tools like *NotebookLM* only retrieve relevant snippets of text rather than reading everything at once.

A prompt is the directive given to a language model to elicit specific types of responses or analytical outputs. For example, when analysing review reports, an instruction might ask the model to identify how an institution describes its quality assurance practices, how student feedback is integrated into those practices, or how feedback loops are structured. The way an instruction is phrased significantly influences the model's output in terms of relevance, depth, and nuance (Brown et al., 2020; Reynolds & McDonell, 2021).

The term *instructional design* (or prompt design) refers to the systematic effort to develop and refine such inputs. In a scientific analysis context, where precision, reliability, and interpretability are essential, instructional design becomes a critical methodological component. Prior studies have shown that even minor variations in wording can lead to substantial differences in the quality and relevance of model-generated responses (Zhou et al., 2022). This is especially true in the analysis of complex, context-rich texts such as review reports, where well-crafted instructions can enable the extraction of insights that would otherwise require manual coding and interpretation.

In the following sections, we describe how *NotebookLM* was used to analyse the review reports and how instruction design was integrated into the development of the analytical method. We outline the different types of instructions applied, ranging from open-ended, exploratory prompts to structured, task-specific questions, and discuss how these variations influenced the insights that could be extracted.

2.3.1 Exploration phase

At the beginning of the analysis process, we chose to start with short and specific questions rather than longer and more comprehensive instructions. This approach aimed to minimise the risk of generating inaccurate or fabricated information, while increasing the likelihood that the model would retrieve relevant and accurate responses based on the dataset.

During this phase, we identified several challenges that needed to be addressed before proceeding further. First, we found that the volume of material (48 review reports totalling approximately 1,800 pages) was initially too large for the model to process effectively in a single operation. One reason for this is the structural variation among the reports. Although all reports follow a common framework with nearly identical chapter headings, they were written by different expert committees and NOKUT case officers. These subtle differences introduced inconsistencies that posed challenges for the model, particularly when similar information appeared in varying locations across different reports. This made it more difficult for *NotebookLM* to locate and consolidate the relevant content needed to answer specific instructions. This finding aligns with limitations identified by Google itself, which recommends using *NotebookLM* with a moderate number of pages and documents to ensure accurate and relevant outputs.

Another challenge we encountered was inconsistency in how institutions were referred to. Some were listed by their full names (e.g., *University of Oslo*), while others were referred to by abbreviations (e.g., *UiO*). This inconsistency extended to file names as well, some included the full institutional name, while others used only abbreviations. As a result, the model occasionally became "confused," misapplying abbreviations or inventing ones that did not exist, which led to inaccuracies in the output.

2.3.2 Simplifying and Systematising the Dataset

In light of the challenges identified during the exploratory phase, it became evident that further structuring of the review reports was necessary. Specifically, there was a need for greater consistency in chapter titles and section headings across all reports. Although the content was generally comparable, key themes (such as feedback loops in institutional quality work) were sometimes addressed under different review requirements in different reports. This variation stemmed from the context-dependent nature of such discussions, which naturally vary across institutions. However, it made it difficult to direct the model to the most relevant sections and to compare similar content across reports.

To address this, we adopted a strategy that broke down the dataset in a way that made it easier for the model to identify and compare relevant information across the reports. To support this effort, we also consulted another language model, *ChatGPT-4o*, to explore how to optimise *NotebookLM*'s performance and improve the reliability of its responses. One suggested solution was to organise the reports into separate folders, referred to as "Notebooks" in *NotebookLM*, based on a chosen thematic structure, and to pose the same questions within each folder.

We decided to group the reports by institutional accreditation level, using this as the primary categorisation principle. NOKUT distinguishes between four types of institutions:

1. universities
2. specialised Universities
3. university colleges with institutional accreditation
4. university colleges with accredited study programmes, but without institutional accreditation

Accordingly, we created four corresponding folders in *NotebookLM*[4], each containing the reports relevant to one of these categories.

Although we later analysed each report individually, this folder-based structure helped to reduce the contextual scope the model had to process at any given time, thereby lowering the risk of misinterpretation or error. For instance, this structure made it easier for the model to identify recurring challenges and shared features within each institutional group. It also facilitated comparison between groups, as the analysis was performed sequentially rather than across the entire dataset simultaneously.

In the literature, such strategies are commonly referred to as chunking, whereby larger volumes of data are divided into smaller, manageable units in order to reduce context load

[4] For more information about the different institutional categories and accreditation levels, see reports A and B as referenced on page 3 of this report.

and support the language model's analysis. In this project, the segmentation occurred primarily at the document and context level, through thematic grouping of reports into separate folders, rather than by splitting individual texts into smaller text fragments (Liu et al. 2023). The purpose of this segmentation was therefore not to analyse the reports at a group level, but to improve the model's working conditions.

Based on these considerations, we developed a training document that included two specific instructions (see Box 1). For each institution and its corresponding review report, we instructed the model to extract the following information:

- institutional category
- accreditation rights
- content related to the requirements in § 2-1 (2) of the *Regulations Relating to the Quality of Education* (with accompanying instruction)
- content related to the requirements in Section 4-1 (3) of the *Regulations Relating to the Supervision and Control of Higher Education* (with accompanying instruction)

We chose to focus on the requirements in Section 2-1 (2) and Section 4-1 (3) because the analyses in Reports A and B showed that these were the two areas where institutions most frequently encountered challenges in the third periodic review cycle. Accordingly, the model was instructed to prioritise these sections to identify overarching and structural challenges in institutional quality assurance work.

In addition to the review reports, we included two supplementary documents in the dataset to support *NotebookLM*'s analysis. The first was NOKUT's *Guide to Systematic Quality Work*, published in 2022 ("Supporting Document 1"). This guide presents NOKUT's interpretation of the regulatory requirements for universities and university colleges, as well as general guidance on the types of documentation typically expected during a review (NOKUT, 2022).

The second document ("Supporting Document 2") was developed specifically for this analysis. It provides a systematic overview of the official names and abbreviations of all higher education institutions in Norway, ensuring greater consistency and helping the model avoid confusion when referencing institutions.

Box 1. Structure of the Training Document

Institution category: University, Specialised University, university college with institutional accreditation, or university college without institutional accreditation

- *NotebookLM's response:* […]

Accreditation rights: Detailed description of the institution's accreditation status and authority

- *NotebookLM's response:* […]

Requirement § 2-1 (2) of the *Regulations on Quality in Higher Education*, with accompanying instruction:
Use the document *nokuts-veileder-til-institusjonenes-systematiske-kvalitetsarbeid_august-2022* as a basis for understanding the content of § 2-1 (2). Based on this requirement, what are the main challenges mentioned in the reports related to fulfilling § 2-1 (2)?

- *NotebookLM's response:* […]

Requirement § 4-1 (3) of the *Regulations Relating to the Supervision and Control of Higher Education*, with accompanying instruction:
Use the document *nokuts-veileder-til-institusjonenes-systematiske-kvalitetsarbeid_august-2022* as a basis for understanding the content of § 4-1 (3). Based on this, what are the main challenges identified in the reports concerning compliance with § 4-1 (3)?

- *NotebookLM's response: […]*

2.3.3 Development of Instructions (Prompt)

Based on the responses generated by *NotebookLM* from the initial training document, particularly in response to the two instructions related to § 2-1 (2) of the *Regulations Relating to the Quality of Education* and § 4-1 (3) of the *Regulations Relating to the Supervision and Control of Higher Education*, we proceeded to further develop and refine the instructions with the assistance of *ChatGPT-4o*.

We informed *ChatGPT-4o* that we were working with *NotebookLM*. Since it is already familiar with this model, no additional explanation was needed. The prompt development process took the form of a dialogue, in which we provided *ChatGPT-4o* with information about the dataset available to *NotebookLM*, as well as the specific types of insights we were aiming to extract from the documents.

One general recommendation from *ChatGPT-4o* was to give *NotebookLM* as much clarity and structure as possible when formulating instructions. In particular, when referring to regulatory requirements, it advised us to either (a) reference the location of the relevant content in the dataset, such as *Supporting Document 1,* or (b) summarise or quote the relevant requirement directly within the instruction.

An early technical challenge we encountered with *NotebookLM* was related to language. Although the model supports multiple languages, it initially returned responses in English, despite the dataset and instructions being entirely in Norwegian. To address this, we began each instruction with the phrase *"Svar på norsk"* ("Answer in Norwegian"). This effectively resolved the issue. More recently, we've observed that N*otebookLM* no longer needs this additional prompt and correctly responds in Norwegian by default.

In the next phase, we analysed the updated training document, which now included NotebookLM's responses, using a new set of instructions to conduct a more comprehensive and comparative analysis across institutions. To facilitate this, we shared the updated document with *ChatGPT-4o* and informed the model that we wanted to deepen the analysis. We asked it to suggest two to four new instructions aimed at identifying common challenges, structural weaknesses, and examples of good practice in institutional quality work.

ChatGPT-4o initially proposed four suggestions. One of them was deemed unsuitable, as it assumed the reports included self-reported challenges from the institutions, whereas the reports only contained assessments made by the expert committees and NOKUT. After clarifying this with *ChatGPT-4o*, it proposed three revised instructions. In the end, we selected two instructions from the original suggestions and two from the revised set. These four final instructions focused on identifying shared challenges, structural weaknesses, and examples of good practice in the institutions' quality assurance work (see Table 2).

Table 2 Final Instructions

No.	Theme	Question
1	**Challenges related to § 2-1 (2)**	What are the most recurring challenges institutions face in fulfilling the requirement in § 2-1 (2)? Are there differences across institutional categories?
2	**Challenges related to § 4-1 (3)**	What are the most recurring challenges institutions face in fulfilling the requirement in § 4-1 (3)? Are there differences across institutional categories?
3	**Common challenges and structural weaknesses**	What challenges recur regardless of institutional category, and what might this indicate about shared structural weaknesses in quality assurance work?
4	**Good practices in quality assurance in higher education**	Based on the analysis of the sources, what characterises good quality assurance practices at higher education institutions? What signs and indicators suggest that an institution is working systematically and purposefully with quality, both in its educational offerings and in its organisational culture?

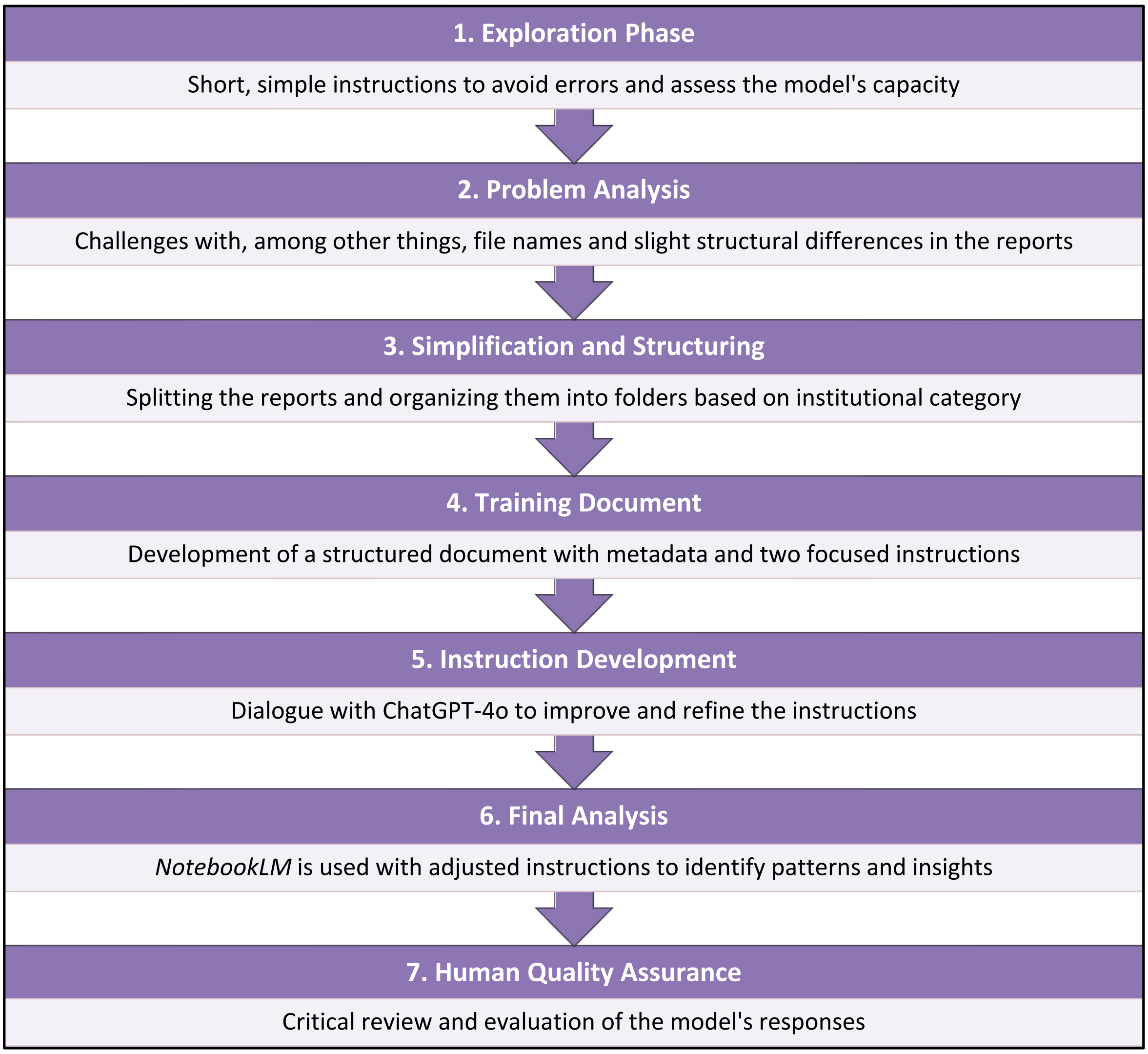
The instructions presented in Table 2 were the final ones used in the analysis process, partly because, at that stage, we were satisfied with the quality of the responses generated by *NotebookLM*. However, this does not mean that the responses were considered final or ready for publication without further refinement and professional judgment.

Although we will not go into the findings in detail here, it is worth noting that the model identified several recurring themes related to institutional quality work. These included, for example, challenges in systematising quality assurance efforts and unclear areas of responsibility and reporting structures. The model also highlighted that examples of good practice were often characterised by the strategic use of quality data, clearly defined roles and responsibilities, strong anchoring in institutional leadership and governing bodies, and integration of quality work with core areas such as teaching, research, and the learning environment.

These findings illustrated the model's usefulness in detecting patterns across large volumes of text. At the same time, they reinforced the continued importance of human expertise for interpretation, contextualisation, and ensuring analytical accuracy and professional relevance.

Figure 1 below summarises the entire process of developing the analysis and systematising the dataset for training *NotebookLM*, including the various phases we undertook.

Figure 1 Analysis Development and Systematization of the Data Basis



3 Discussion and Methodological Reflections

Our experience with *NotebookLM* demonstrates that the model can support the identification of insights, recognition of patterns, and articulation of tacit knowledge. At the same time, this type of workplaces significant demands on structure, methodology, and professional judgment. In this section, we offer reflections on the differences and similarities between manual and AI-assisted analysis and discuss five key factors that are critical to the effective use of artificial intelligence in analytical processes. These factors illustrate both the potential and the limitations of AI tools when applied to complex qualitative assessments. They include structure and data quality, standardisation and professional judgment. the role and value of the model in the analysis process, skills required for responsible AI use, and ethical considerations and trust in AI-supported work.

3.1 Manual vs. AI-Assisted Analysis – Key Reflections

Both the manual analyses and those supported by AI revealed many of the same challenges in the institutions' quality assurance practices. However, we also observed meaningful differences in what each approach contributes. Manual analysis offers a deeper understanding of context, process, and intent. It enables the interpretation of linguistic nuance, assessment of institutional maturity, and insight into how quality systems function in practice. In many cases, we have engaged directly with the institutions involved, experiencing the cases firsthand, not just through the lens of the review reports. This level of contextual knowledge cannot be replicated by an external actor or, in this case, a machine which lacks experience-based understanding.

NotebookLM, on the other hand, has shown clear strengths in structuring, summarising, and identifying patterns across large volumes of text. It effectively extracts explicit elements from documents, links findings to relevant regulatory requirements, and organises indicators of good quality practice. The model has been particularly helpful in highlighting common challenges across institutions and suggesting generalisable good practices. This has enhanced our ability to see the bigger picture and to systematise findings in a way that would be difficult to achieve manually within the same timeframe. Furthermore, the use of AI prompted greater awareness of the need for improved data structuring and consistency, an issue we explore in more detail later in this chapter. At the same time, we acknowledge that, like humans, the model can sometimes draw rigid or oversimplified conclusions, particularly in relation to causality or implicit factors that require nuanced, professional interpretation.

In our view, the greatest potential lies in combining manual and AI-assisted analysis. Using AI as a support tool, to identify patterns, structure large datasets, and propose analytical categories, can free up time and resources for deeper, human-led interpretation and evaluation. In this way, AI serves not as a replacement, but as a powerful supplement that enhances both the quality and efficiency of the overall analytical process (see also Davison et al., 2024).

3.2 Key Aspects of the Analysis Process

3.2.1 Structure and Data Quality

NotebookLM was generally able to identify relevant information in the review reports. However, as expected, the precision of its responses was highly dependent on the quality and structure of the underlying data, as well as on the design of the prompts. While the reports themselves are technically sound, some information was unclearly phrased, dispersed across multiple sections, or difficult for the model to access due to inconsistencies in structure from one report to another.

A key technical consideration when using large language models is how documents are broken down into smaller text segments to facilitate processing (Hitch, 2024; Nguyen-Trung, 2025). This process occurs automatically and typically does not account for thematic coherence. As a result, fragmented, unstructured, or highly contextual information may be misinterpreted or overlooked by the model. In our experience, shorter, well-structured texts with clear headings, consistent formatting, and thematic markers produce significantly better outcomes. Such structuring supports both human and AI-based analysis across multiple documents.

3.2.2 Standardisation and Professional Judgement

In NOKUT's review work, both standardisation and professional judgement play complementary roles. Variation in how case officers and expert committees assess and document institutional quality work highlights the need for greater standardisation. Increasing consistency across reports would not only support comparability but also create a stronger methodological foundation for AI tools like *NotebookLM* to function accurately and consistently.

At the same time, we believe standardisation should not come at the expense of professional discretion. The third cycle of review showed that flexibility and sensitivity to institutional context can coexist with methodological rigour. A clear conceptual framework and standardised terminology can strengthen professional judgement rather than limit it. Importantly, language models such as *NotebookLM* place increased demands on structured, consistent input (Nguyen-Trung, 2025), a requirement that, in turn, can serve as an incentive to further develop both methodological clarity and academic freedom in tandem.

3.2.3 The Role of AI Models: Assistants with Analytical Value

NotebookLM demonstrated an ability to "read between the lines" and synthesise holistic insights that were not always explicitly stated in the reports. For instance, when asked in Prompt 3 to identify common challenges and structural weaknesses, the model was able to draw on information from across the dataset, recognise thematic patterns, and infer possible root causes.

These capabilities highlight the potential of AI models as analytical assistants - tools that can support, enhance, and structure the analysis process but not replace human expertise or professional judgement (Hitch, 2024). That said, the model also exhibited notable limitations. In several instances, it confused institution names and abbreviations, which led to inaccurate or misleading conclusions. This underscores the need for careful human oversight, not only during the interpretation of results, but also in the preparation and structuring of data.

3.2.4 Skills Development for AI Use

Effective use of large language models requires new competencies for analysts and case officers. These go beyond technical proficiency and include a deeper understanding of AI's capabilities, limitations, and appropriate applications in analytical work. In the long term, AI tools may streamline workflows and automate certain tasks, thereby freeing up time for more in-depth, human-led analysis. However, this requires intentional skills development and a clear strategy for integrating AI into existing work processes.

Our experience also suggests benefits in combining multiple AI models. For example, using *ChatGPT-4o* to help refine and test prompts for use in *NotebookLM* allowed us to better leverage each model's strengths. This raises methodological questions around division of labour, transparency, and control: Which model plays which role, and how do we document their interaction? Ensuring consistency and traceability becomes especially important when different models, each with distinct capacities, context windows, and interpretive tendencies, are used in tandem. This multi-model approach opens new methodological possibilities but demands critical reflection, documentation, and awareness at all stages of the process.

3.2.5 Ethical Considerations and Trust in AI-Generated Analyses

The use of AI in analytical work naturally raises questions about trust, responsibility, and professional integrity (Davison, 2024; Hitch, 2024; Nguyen-Trung, 2025). How much confidence can we place in the model's responses? How do we ensure that its outputs are valid, reliable, and professionally sound? We believe that AI-generated results must always be reviewed and validated by professionals with relevant expertise. Even when the model performs well, human judgement remains essential, particularly for contextual interpretation, relevance assessment, and placing findings within a broader analytical framework.

A related ethical consideration concerns data protection and rights associated with data use. Although AI tools may make data access feel easier, the same ethical and academic standards that apply to manual analysis must also guide AI-supported work. In this project, we limited our dataset to NOKUT's publicly available review reports, which contains no personally identifiable information.

It is also important to be aware of the environmental aspects of AI technology. AI systems require large amounts of electricity and energy (Verdecchia et al., 2023), which naturally raises questions about whether the resource consumption is proportionate to the added value AI actually provides, for example in analytical work. Although we are still in an exploratory phase regarding the use of AI (examining which types of tasks the technology is suited for such as analysis, summarisation, or language editing), we believe that its use must be guided by both environmental considerations and assessments of actual utility. In addition, it is necessary to be aware of technology dependency as a potential vulnerability, particularly in relation to the responsible use of AI. An increasing reliance on such tools may weaken independent professional judgment and critical reflection. For now, we accept the premise that AI as a working tool should be explored, but there should always be a cost-benefit assessment underlying its use.

Ultimately, the use of AI must be grounded in open dialogue about values and responsibility. This applies to how NOKUT uses such tools, and also to how institutions and expert committees may choose to integrate them into their work. Transparency, technological literacy, and respect for professional autonomy are key. As language models become more advanced, it is increasingly important to preserve the human role in assessment processes, ensuring that it is people, not machines, who make the final judgement.

4 Conclusion and Recommendations

The use of *NotebookLM* in this thematic analysis has shown that large language models can deliver accurate, reflective, and occasionally surprising insights, particularly when provided with favourable working conditions, including well-structured data and precisely formulated instructions. The model demonstrates an ability to abstract from individual wording and extract overarching patterns. However, it is also highly sensitive to small variations in prompt phrasing, which places significant demands on prompt design.

One key finding is the importance of a well-prepared and structured dataset. *NotebookLM* processes information in discrete units of information, meaning it analyses segmented portions of text independently. When information is dispersed, unclearly written, or inconsistently formatted, there is a higher risk that relevant insights may be missed. Our experience confirms that shorter, clearer, and more targeted texts consistently produce better results.

Another important lesson was the impact of uniform terminology and consistent file naming. Inconsistencies in abbreviations and file names caused confusion for the model and required manual correction. This demonstrates how the use of AI models can also have a beneficial *disciplinary effect* on data organisation and structure, pushing users toward improved consistency and clarity.

A particularly interesting aspect of *NotebookLM* is its so-called *emergent abilities* - unexpected capabilities that arise in complex analytical tasks. These can be both beneficial and challenging. In the context of analysing review reports, where information can be fragmented and cross-cutting analysis is essential, the model showed potential, but also revealed limitations, such as misinterpreting institutional abbreviations, confusing information sources, or generating interpretations that a human case officer would not necessarily make.

For AI models like *NotebookLM* to be used responsibly and effectively, human oversight is essential. Case officers with subject-matter expertise must guide and quality-assure the analysis process. Since language models do not "understand" content in a human sense but operate based on statistical patterns, professional judgment remains crucial for ensuring analytical relevance and accuracy.

At the same time, it is important to recognise that artificial intelligence is evolving rapidly. Even within the scope of this project (2025), we observed significant advancements in model capabilities, functionality, and available tools. It is therefore reasonable to expect that the potential for AI-supported analytical work will continue to grow in the near future.

This also implies that we at NOKUT are open to testing and adopting models other than *NotebookLM* should these prove more appropriate. The use of AI should therefore be grounded in clear internal guidelines governing when and how such tools may and should be used, particularly with regard to privacy, data integrity, IT security, transparency, accountability, and sustainability, and that risk assessments of relevant tools are conducted prior to their adoption.

In closing, we contend that AI models have the potential to strengthen and streamline 'traditional' analyses, but that this presupposes a deliberate and systematic effort in both preparation and implementation. Without strategic governance and competence development in this area, the use of AI in analytical work may contribute to ambiguity and misinterpretation rather than to qualitatively superior insights and decision-making foundations.

References

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, *33*, 1877-1901

Davison, R.M., Chughtai, H., Nielsen, P., Marabelli, M., Iannacci, F., van Offenbeek, M., Tarafdar, M., Trenz, M., Techatassanasoontorn, A.A., Díaz Andrade, A. and Panteli, N. (2024). The ethics of using generative AI for qualitative data analysis. *Inf Syst J, 34*: 1433-1439

Google Support (2025). https://support.google.com/notebooklm?sjid=12491595086394054000-EU#topic=16164070, retrieved May 15th, 2025

Hitch D. (2024). Artificial Intelligence Augmented Qualitative Analysis: The Way of the Future? *Qual Health Res. 2024 Jun; 34*(7):595-606.

IEEE Computer Society Team (2025). AI and Resource Consumption: Examining the Environmental Impact. Tilgjengelig på nett: https://www.computer.org/publications/tech-news/research/ai-and-the-environment?utm_source=chatgpt.com, 15. Juni 2025

Liu, N. F., Lin, K., Hewitt, J., Paranjape, A., Bevilacqua, M., Petroni, F., & Liang, P. (2023). Lost in the middle: How language models use long contexts. *arXiv preprint arXiv:2307.03172*.

Nguyen-Trung, K. (2025). ChatGPT in thematic analysis: Can AI become a research assistant in qualitative research?. *Qual Quant* (2025).

NOKUT 2022. VEILEDNING. Systematisk kvalitetsarbeid i høyere utdanning: *Veiledning til kravene i universitets- og høyskoleloven, studiekvalitetsforskriften og studietilsynsforskriften* . https://www.nokut.no/siteassets/tilsyn-og-kvalitetsarbeid/nokuts-veileder-til-institusjonenes-systematiske-kvalitetsarbeid_august-2022.pdf

Reynolds, L., & McDonell, K. (2021). Prompt programming for large language models: Beyond the few-shot paradigm. In *Extended abstracts of the 2021 CHI conference on human factors in computing systems* (pp. 1-7)

Verdecchia, R., Sallou, J., & Cruz, L. (2023). A systematic review of Green AI. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *13*(4), e1507.

Zhou, C., He, J., Ma, X., Berg-Kirkpatrick, T., & Neubig, G. (2022). Prompt consistency for zero-shot task generalization. *arXiv preprint arXiv:2205.00049*.

Appendix

Appendix 1: Requirements for Systematic Quality Work in Round Three

Section	Legal Text	Legal Source
§ 1-6	Universities and university colleges shall establish and maintain a satisfactory internal system of quality assurance to ensure and further enhance the quality of education. Student evaluations shall form an integral part of the quality assurance system.	*The Universities and University Colleges Act*
§ 4-3 (5)	The institution's work on the learning environment shall be documented and shall form part of the institution's internal quality assurance system pursuant to § 1-6.	*The Universities and University Colleges Act*
§ 2-1 (2)	The institutions shall conduct periodic evaluations of their study programmes. Representatives from working life or society at large, students, and relevant external experts shall take part in the evaluations. The evaluation results shall be made public.	*Regulations Relating to the Quality of Education*
§ 4-1 (1)	The institution's quality assurance work shall be anchored in a strategy and shall cover all essential areas of significance for the quality of students' learning outcomes.	*Regulations Relating to the Supervision and Control of Higher Education*
§ 4-1 (2)	The quality assurance work shall be anchored in the institution's board and management at all levels. Through its quality assurance efforts, the institution shall contribute to fostering a culture of quality among staff and students.	*Regulations Relating to the Supervision and Control of Higher Education*
§ 4-1 (3)	The institution shall have arrangements in place to systematically ensure that all study programmes meet the requirements set out in § 3-1 to 3-3 of the Regulations concerning Quality Assurance and Quality Development in Higher Education and Tertiary Vocational Education, as well as Chapter 2 of these Regulations	*Regulations Relating to the Supervision and Control of Higher Education*
§ 4-1 (4)	The institution shall systematically collect information from relevant sources in order to assess the quality of all study programmes.	*Regulations Relating to the Supervision and Control of Higher Education*
§ 4-1 (5)	Knowledge gained from the quality assurance work shall be used to develop the quality of the study programmes and to identify any deficiencies in quality. Deficient quality shall be rectified within a reasonable time.	*Regulations Relating to the Supervision and Control of Higher Education*
§ 4-1 (6)	The results of the quality assurance work shall form part of the knowledge base for the assessment and strategic development of the institution's overall portfolio of study programmes.	*Regulations Relating to the Supervision and Control of Higher Education*

Appendix 2. Overview of Institutions in the Third Round

Group	Institution	Review Period
Pilot Project	The Oslo School of Architecture and Design (AHO) MF Norwegian School of Theology, Religion and Society (MF) NHH Norwegian School of Economics (NHH) Norwegian Academy of Music (NMH) VID Specialized University (VID)	2017–2018
Group 1	Inland Norway University of Applied Sciences (HINN) Western Norway University of Applied Sciences (HVL) Nord University (NORD) University of South-Eastern Norway (USN)	2018/2019
Group 2	Norwegian University of Life Sciences (NMBU) Oslo Metropolitan University (OsloMet) University of Agder (UiA) University of Stavanger (UiS)	2019/2020
Group 3	Molde University College (HiMolde) Volda University College HVO) Østfold University College (HiØ) Kristiania University of Applied Sciences (HK) Sámi University of Applied Sciences (SH/SA)	2019/2020
Group 4	Norwegian Business School (BI) Oslo National Academy of the Arts (KHiO) Norwegian School of Sport Sciences (NIH)	Autumn 2020 –Spring 2021
Group 5	Ansgar University College (AHS) Fjellhaug International University College (FIH) Lovisenberg Diaconal University College (LDH) NLA University College (NLA)	Spring 2021 –Autumn 2021
Group 6	Queen Maud University College (DMMH) The Norwegian Defence University College (FHS) Norwegian School of Leadership and Theology (HLT) The Norwegian Police University College (PHS)	Autumn 2021 –Spring 2022
Group 7	Norwegian University of Science and Technology (NTNU) University of Bergen (UiB) University of Oslo (UiO) The Arctic University of Norway (UiT)	Spring 2022 –Autumn 2022
Group 8	Atlantis Medical College (AMH) The Barratt Due Institute of Music (HBD) Bergen School of Architecture (BAS) Oslo New University College (ONH) Norwegian University College of Green Development (HGUt) The University College of Vocational Education (HØFY) University College of Dance Art (HFDK) The University College of Norwegian Correctional Service (KRUS) Lillehammer Institute of Music Production and Industries (LIMPI)	Autumn 2022 –Autumn 2024

Noroff University College (NUC)
The Norwegian Institute of Children's Books (NBI)
The Norwegian Gestalt Institute AS (NGI)
Bergen Writing Academy (SKA)
NSKI University College (NSKI)
Rudolf Steiner University College

NOKUT

DRAMMENSVEIEN 288 | P.O. BOX 578,1327 LYSAKER | NOKUT.NO