instituto politécnico de gestão e tecnologia

MESTRADO EM GESTÃO DA SEGURANÇA E SAÚDE DO TRABALHO

Artificial Intelligence Assisted Documentary Auditing in QHSE Management Systems: Design and Experimental Evaluation of a Prototype Tool

IMEN MNASRI

ISLA

VILA NOVA DE GAIA

MARCH | 2026

INSTITUTO POLITÉCNICO DE GESTÃO E TECNOLOGIA

Artificial Intelligence Assisted Documentary Auditing in QHSE Management Systems: Design and Experimental Evaluation of a Prototype Tool

IMEN MNASRI

Aprovado em 08/05/2026

Composição do Júri

_____

Presidente Prof.a Doutora Emília Quelhas Costa

_____

Arguente Prof. Doutor José Pedro Domingues

_____

Orientador Prof. Doutor Hernâni Veloso Neto

Vila Nova de Gaia
2022

Work submitted in the context of the Applied Project Report, carried out under the supervision of Professor Doctor Hernâni Veloso Neto, and presented to ISLA – Instituto Politécnico de Gestão e Tecnologia de Vila Nova de Gaia for the award of the Master's degree in Occupational Health and Safety Management, in accordance with Order No. 11648/2016 of 29 September, 2nd Series, No. 188.

Acknowledgments

I would like to express my sincere gratitude to all those who contributed directly or indirectly to the completion of this work.

First, I thank my supervisor, **Hernâni Veloso Neto**, for his valuable guidance, careful review, and insightful corrections, which greatly improved the scientific quality and clarity of this project.

Warm thanks also go to **Hachem Abou Zer** for his invaluable assistance with the IT section of the project, particularly for the development and analysis of data, as well as for his methodological explanations that facilitated the understanding of the tools used.

I also wish to acknowledge myself for the dedication, perseverance, and rigor applied throughout this work, which were essential to its completion.

Finally, I would like to thank my friends and colleagues who provided moral support and encouragement throughout this process, helping me to maintain motivation and determination.

This work would not have been possible without all these contributions, and I extend my sincere thanks to everyone involved.

Abstract

This research investigates the potential of artificial intelligence to support and partially automate documentary audits within Quality, Health, Safety, and Environmental (QHSE) management systems aligned with ISO 9001, ISO 14001, and ISO 45001 standards. The main objective of the study was to design and develop a prototype digital platform capable of assisting auditors in the evaluation of documentary compliance through the use of artificial intelligence. To achieve this objective, the research combined a systematic literature review with an applied experimental development approach. First, the requirements of the relevant ISO standards were analyzed in order to extract and formalize documentary compliance criteria. Based on this analysis, a structured digital checklist was designed to serve as the normative framework for document evaluation. A prototype application was then developed integrating an artificial intelligence service capable of analyzing textual content and comparing it with the predefined checklist criteria. The system generates structured audit outputs indicating whether each requirement is classified as compliant, partially compliant, or non-compliant. The methodological evaluation consisted of two complementary phases : an experimental validation protocol designed to assess the system's reproducibility, robustness, and logical consistency under controlled testing conditions, followed by a documentary case study using publicly available organizational documents in order to observe the operational behavior of the system in a realistic context. The results indicate that artificial intelligence can effectively support documentary auditing by improving the efficiency and consistency of compliance verification processes. The prototype demonstrated high reproducibility of decisions under controlled conditions and showed coherent responsiveness when documents were normatively improved. However, the findings also highlight that AI-assisted auditing cannot fully replace professional judgment, as certain compliance interpretations remain dependent on expert evaluation. Overall, the study demonstrates the technical feasibility and practical relevance of integrating artificial intelligence into QHSE documentary auditing as a decision-support tool. The research supports a hybrid intelligence approach in which artificial intelligence enhances the efficiency and structure of audit processes while human expertise remains essential for final compliance validation.

Keywords : Artificial Intelligence ; QHSE Management Systems ; ISO 9001 ; ISO 14001 ; ISO 45001 ; Compliance Audit ; Digitalization.

Resumo

Esta investigação analisa o potencial da inteligência artificial para apoiar e automatizar parcialmente auditorias documentais em sistemas de gestão de Qualidade, Saúde, Segurança e Ambiente (QHSE), alinhados com as normas ISO 9001, ISO 14001 e ISO 45001. O principal objetivo do estudo foi conceber e desenvolver uma plataforma digital protótipo capaz de apoiar auditores na avaliação da conformidade documental através da utilização de inteligência artificial. Para alcançar este objetivo, a investigação combinou uma revisão sistemática da literatura com uma abordagem de desenvolvimento experimental aplicado. Inicialmente, foram analisados os requisitos das normas ISO relevantes com o objetivo de extrair e formalizar critérios de conformidade documental. Com base nesta análise, foi concebida uma checklist digital estruturada que serve como enquadramento normativo para a avaliação dos documentos. Posteriormente, foi desenvolvida uma aplicação protótipo que integra um serviço de inteligência artificial capaz de analisar o conteúdo textual dos documentos e compará-lo com os critérios previamente definidos na checklist. O sistema gera resultados de auditoria estruturados, indicando se cada requisito é classificado como conforme, parcialmente conforme ou não conforme. A avaliação metodológica consistiu em duas fases complementares: um protocolo de validação experimental destinado a avaliar a reprodutibilidade, robustez e consistência lógica do sistema em condições de teste controladas, seguido de um estudo de caso documental baseado em documentos organizacionais publicamente disponíveis, com o objetivo de observar o comportamento operacional do sistema num contexto realista. Os resultados indicam que a inteligência artificial pode apoiar eficazmente auditorias documentais, melhorando a eficiência e a consistência dos processos de verificação de conformidade. O protótipo demonstrou elevada reprodutibilidade das decisões em condições controladas e revelou uma resposta coerente quando os documentos eram melhorados do ponto de vista normativo. No entanto, os resultados também evidenciam que a auditoria assistida por inteligência artificial não pode substituir totalmente o julgamento profissional, uma vez que determinadas interpretações de conformidade continuam a depender da avaliação de especialistas. De forma geral, o estudo demonstra a viabilidade técnica e a relevância prática da integração da inteligência artificial em auditorias documentais QHSE como ferramenta de apoio à decisão. A investigação sustenta um modelo de inteligência híbrida, no qual a inteligência artificial

melhora a eficiência e a estrutura dos processos de auditoria, enquanto a experiência humana permanece essencial para a validação final da conformidade.

Palavras-chave : Inteligência Artificial; Sistemas de Gestão QHSE; ISO 9001; ISO 14001; ISO 45001; Auditoria de Conformidade; Digitalização.

General Index

List of Figures

List of Tables

List of Abbreviations

AI	Artificial Intelligence
AIMS	Artificial Intelligence Management System
API	Application Programming Interface
C	Conform
CSR	Corporate Social Responsibility
DL	Deep Learning
DMS	Document Management System
EDMS	Electronic Document Management System
EMS	Environmental Management System
ERP	Enterprise Resource Planning
ESG	Environmental, Social and Governance
GenAI	Generative Artificial Intelligence
ISO	International Organization for Standardization
LLM	Large Language Model
ML	Machine Learning
NC	Non-Conform
NLP	Natural Language Processing
OCR	Optical Character Recognition
OHS	Occupational Health and Safety
PC	Partially Conform
PDCA	Plan-Do-Check-Act
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QHSE	Quality, Health, Safety and Environment
VS Code	Visual Studio Code
XAI	Explainable Artificial Intelligence

1. Introduction

Over the past decade, organizations operating under Quality, Health, Safety and Environmental (QHSE) management systems have faced increasing normative complexity. International standards such as ISO 9001:2015, ISO 14001:2015 and ISO 45001:2018 require organizations not only to implement compliant processes, but also to produce documented evidence demonstrating structured planning, operational control, performance evaluation and continual improvement. In this regulatory architecture, documented information is central to certification, accountability and risk governance. Document audits therefore represent a critical component of QHSE systems. They verify whether organizational documentation meets normative requirements, reflects operational reality and supports compliance verification. However, document auditing remains a cognitively intensive, time-consuming and interpretative task, highly dependent on auditor expertise and exposed to variability in interpretation, documentation quality and contextual ambiguity. In multi-standard environments, where ISO clauses overlap and interrelate, the complexity of documentary assessment further increases. Simultaneously, the rapid evolution of digital technologies—particularly artificial intelligence (AI), large language models (LLMs), and Natural Language Processing (NLP)—is transforming documentary audit processes. Research indicates that AI-enabled systems and electronic document management platforms offer significant potential to enhance audit efficiency, consistency and traceability (Popara et al., 2023). Nevertheless, automation in regulated environments requires careful consideration of technical, organizational and human constraints, including system integration, professional upskilling, resistance to change and regulatory alignment.

Against this backdrop, this research addresses the following central question :

How can artificial intelligence be used to automate QHSE document audits and improve the efficiency and accuracy of compliance verification ?

Rather than adopting a purely technological perspective, this study examines the conditions under which AI can function as a structured decision-support system within regulated contexts, without compromising normative rigor or professional responsibility.

To answer this question, the study has the following general objective: to explore and evaluate the potential of artificial intelligence for supporting the auditing of QHSE

documents, aiming to enhance efficiency, consistency, and accuracy in compliance verification with ISO standards.

The study also pursues the following specific objectives:

1. To conduct a systematic literature review to identify current approaches, technologies, and existing gaps in AI-based document auditing within the QHSE domain.
2. To analyze QHSE standards (ISO 9001, ISO 14001, ISO 45001) to extract and formalize documentary compliance requirements.
3. To design a structured digital checklist that aligns with normative requirements and supports systematic verification of presented documentation.
4. To implement and test an experimental prototype system allowing users to define audit criteria, upload documents, and manage the audit workflow, without aiming to develop a complete industrial software system.
5. To integrate an AI engine capable of analyzing document content and assessing compliance against the defined checklist.
6. To generate automated audit results, including compliance status (compliant, partially compliant, non-compliant) and AI-supported justifications or comments.
7. To conduct an experimental evaluation protocol and applied case study to assess the functionality, reproducibility, and practical applicability of AI-assisted document auditing.

Rather than adopting a purely technological perspective, this study examines the conditions under which AI can function as a structured decision-support tool within regulated contexts, without compromising normative rigor or professional responsibility.

1.1 Definition of Core Concepts in QHSE Document Auditing

In order to ensure conceptual clarity and facilitate a structured analysis, the key notions underpinning this research are defined below.

QHSE Management Systems

Integrated systems designed to manage quality (ISO 9001), environment (ISO 14001), and occupational health and safety (ISO 45001), requiring structured process control and rigorous documentation (Yew & Goh, 2015).

Documented Information

According to ISO standards, this term encompasses both maintained documentation (e.g., procedures, policies) and retained records used to demonstrate conformity and operational control (ISO 9001:2015; ISO 14001:2015; ISO 45001:2018).

Artificial Intelligence (AI)

The capability of computational systems to perform tasks requiring human-like reasoning, learning, pattern recognition, and decision support, notably through machine learning, NLP, and expert systems (Popara et al., 2023; Noordin et al., 2022).

Audit Process

A systematic and documented evaluation of conformity with established criteria, structured into preparation, execution, and reporting phases as defined by ISO 19011:2018.

Automation

The use of digital systems to perform structured tasks with limited human intervention, aiming to improve speed, consistency and traceability in compliance verification.

Electronic Document Management Systems (EDMS)

Digital platforms that centralize, structure and control documentation necessary for regulatory compliance and audit readiness.

These definitions establish the conceptual foundation necessary for analyzing the interaction between AI technologies and QHSE documentary auditing.

1.2 Technical Context of QHSE

The implementation of ISO-based QHSE management systems relies heavily on rigorous document control. Increasing operational complexity and regulatory expectations require

structured workflows and reliable traceability mechanisms (Balabanov & Davletshin, 2018). In practice, audits frequently begin with an extensive documentary review, including policies, procedures, training records, incident reports and maintenance logs (Shabani et al., 2024). Compliance therefore depends not only on operational execution but also on the availability, accessibility and integrity of documented evidence. ISO standards require continuous process control supported by up-to-date and structured documentation. Empirical studies on integrated management systems highlight the necessity of aligning documentary practices with detailed normative clauses (Yew & Goh, 2015). These constraints generate a growing need for intelligent digital tools capable of supporting structured documentary compliance checks within QHSE systems.

1.3 Justification for the Study

The integration of AI into auditing and compliance processes has attracted increasing academic attention. However, the literature remains fragmented. Existing research often focuses on isolated technological applications while neglecting systemic interactions between automation and professional judgment (Kokina et al., 2025). Persistent challenges include ethical considerations, regulatory constraints, uneven professional preparedness and limited domain-specific validation in highly regulated environments such as QHSE (Wijaya et al., 2025). Moreover, regulatory compliance automation requires formalization of normative requirements into operational representations interpretable by AI systems (Adebayo et al., 2022). Explainability and traceability constitute central requirements in regulatory contexts. Automated systems must produce interpretable and auditable outputs, particularly under emerging governance frameworks such as the European Union AI Act (Winkler et al., 2025; Raji et al., 2020). Consequently, hybrid intelligence models—where AI augments rather than replaces professional expertise—are increasingly recommended (Dellermann et al., 2021). In the specific domain of QHSE documentary auditing, experimentally validated automation frameworks remain scarce. This gap justifies the development and empirical evaluation of structured AI-assisted systems tailored to ISO-based compliance environments.

1.4 Research Design and Structure

This research adopts a structured methodological framework integrating normative analysis, systematic literature review and experimental validation. First, a systematic literature review examines current applications of AI in auditing and regulatory compliance, identifying technological capabilities and structural limitations. Second, a prototype automated document analysis platform is developed to operationalize ISO requirements through structured checklists and AI-driven textual evaluation. The system generates ordinal compliance verdicts (Non-compliant, Partially compliant, Compliant) supported by explicit justifications and documentary evidence excerpts. Four controlled experimental tests assess reproducibility, robustness, sensitivity to predefined non-conformities and responsiveness to documentary corrections. These experiments are complemented by applied case studies using real corporate documents. The dissertation is structured as follows. Chapter 2 presents the systematic literature review. Chapter 3 details the methodological and technical framework of the developed platform. Chapter 4 reports experimental and case study results. Chapter 5 discusses implications, limitations and operational recommendations.

2 Literature Review

2.1 Methodology

To successfully conduct this systematic review, it is essential to adopt a rigorous and structured approach that ensures the quality, relevance, and reproducibility of the results. The chosen methodology is based on recognized academic research protocols that enable the identification, selection, analysis, and synthesis of relevant publications on artificial intelligence applied to document auditing in the QHSE context. A systematic review is a methodical and transparent literature review approach that aims to minimize bias by following explicit and reproducible procedures (Page et al., 2021). This methodological approach aims to ensure a logical and coherent sequence between the different phases of the review, while guaranteeing transparency in the data extraction and filtering process (Page et al., 2021).

2.1.1 PRISMA Strategy

The methodology adopted for this systematic review follows the recommendations of the PRISMA 2020 protocol, which structures systematic reviews into four main stages: identification, screening, eligibility, and inclusion (Page et al., 2021). The PRISMA protocol (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) provides a standardized framework to enhance the clarity and completeness of literature reviews, especially in evidence-based research. During the identification phase, an extensive search was conducted across multiple databases using predefined keywords to capture all relevant publications on artificial intelligence applied to QHSE document auditing. After removing duplicates, articles were screened based on title and abstract to assess their relevance (Kitchenham & Charters, 2007). The full texts of the preselected articles were then assessed using predefined inclusion and exclusion criteria, retaining only those that directly addressed the objectives of the review. This rigorous process ensures the quality and representativeness of the sources used for analysis (Binh, 2025).

2.1.2 Databases

To ensure broad and reliable coverage of the relevant literature, several academic databases were consulted: Scopus, IEEE Xplore, ScienceDirect, SpringerLink, and

Google Scholar. These databases were selected for their relevance to engineering, computer science, and management, offering peer-reviewed publications from major journals and conference proceedings. The search period was limited to the years 2015 to 2025 to include the most recent advancements in artificial intelligence, auditing, and QHSE management (Binh, 2025).

2.1.3 Search String

Search queries combined the main concepts using Boolean operators, according to the following formula : ("artificial intelligence" OR "machine learning") AND ("audit" OR "compliance check") AND ("QHSE" OR "ISO 9001" OR "ISO 14001" OR "ISO 45001") AND ("document" OR "documentation"). These expressions were slightly adapted based on the syntax of each database to optimize the relevance of the results obtained (Kitchenham & Charters, 2007).

2.1.4 Inclusion and Exclusion Criteria

Eligibility criteria define the characteristics of the studies included in this systematic review. Included are peer-reviewed articles published in English or French that focus on the application of artificial intelligence to the auditing of QHSE documents, specifically related to ISO 9001, ISO 14001, and ISO 45001 standards (Wijaya et al., 2025). Peer-reviewed articles are scholarly works that have been evaluated by independent experts in the field, ensuring a minimum standard of scientific rigor. Excluded are purely theoretical articles without practical application, publications outside the scope of ISO standards, and non-peer-reviewed documents (technical reports, unpublished theses). This selection ensures the relevance, scientific rigor, and thematic focus of the corpus on the topic under study (Kokina et al., 2025).

2.1.5 Selection and Evaluation Process

The article selection process for this systematic review was conducted in accordance with the PRISMA 2020 protocol (Page et al., 2021), to ensure the rigor, transparency, and reproducibility of the study. A total of 100 articles were initially identified from the selected databases (see section 2.2). After removing duplicates, 90 unique articles were subjected to a first screening phase based on title and abstract. This step allowed for the

exclusion of irrelevant or off-topic publications, reducing the corpus to 80 articles preselected for full-text review, as illustrated in the figure above.

These articles were then independently evaluated by two reviewers according to the predefined inclusion and exclusion criteria (see section 2.4). Any discrepancies were resolved through discussion until a consensus was reached. This consensus-building process helps reduce individual bias and improve inter-rater reliability in study selection (Binh, 2025). At the end of this process, 63 articles were retained for final analysis. It is also worth noting that no contact with the study authors was necessary to retrieve missing information, as all relevant data were available in the publications themselves.

Finally, the methodology deliberately retained a broad range of texts, including diverse approaches and contexts. This choice was made in order to enrich the project's literature review by incorporating a variety of perspectives and analyses, thereby providing a more comprehensive and nuanced understanding of the topic under investigation.

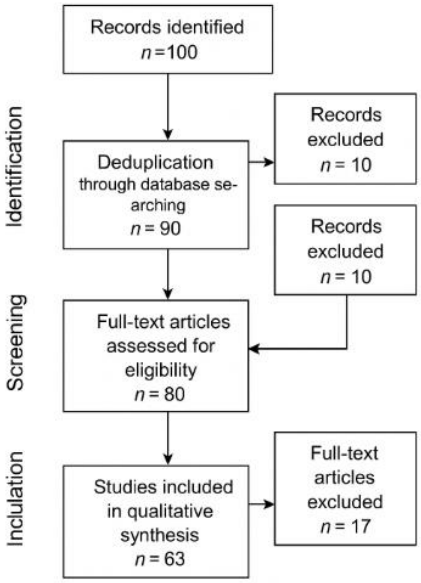


Figure 1 : PRISMA flow diagram

2.2 ISO QHSE Standards Requirements Related to Documented Information

Document management plays a fundamental role in QHSE management systems by ensuring compliance with international normative requirements. ISO 9001, 14001, and 45001 have introduced harmonized frameworks for the control of documented information, covering the creation, control, and retention of key documents essential for QHSE system performance. Understanding these normative requirements is crucial to grasp the challenges of document auditing and the automation of verifications through digital and intelligent tools (Balabanov & Davletshin, 2018).

2.2.1 ISO 9001 / 14001 / 45001 – Key Clauses on Documented Information

ISO 9001:2015, ISO 14001:2015, and ISO 45001:2018 place a central emphasis on the management of documented information, replacing the traditional notions of "procedures" and "records." This change reflects a more integrated and flexible approach to managing evidence and operations under quality, environmental, and safety systems. This unified concept includes both maintained documents (e.g., policies, procedures, operating instructions) and retained records used as evidence of conformity (ISO 9001:2015; ISO 14001:2015; ISO 45001:2018).

- **ISO 9001:2015 – Clause 7.5** states that the organization must create, update, and control the documented information necessary for the effectiveness of the quality management system. It clearly distinguishes between documentation used to conduct activities and that used to prove they were carried out.
- **ISO 14001:2015 – Clauses 7 to 10** require up-to-date documentation on compliance obligations, significant environmental aspects, operational criteria, and responses to environmental nonconformities.
- **ISO 45001:2018 – Clause 7.5** reinforces the need for documented information ensuring compliance with the occupational health and safety management system, including risk assessments, implemented controls, and safety training records.

2.2.2 Types of Commonly Used QHSE Documents

QHSE management requires a variety of formalized documents aimed at both structuring processes and providing objective evidence. According to the above ISO standards, commonly used documents include:

- Quality and environmental manuals,
- Operational procedures and instructions,
- Risk prevention plans,
- Audit and incident reports,
- Safety data sheets (SDS),
- Training, certification, and maintenance records.

These documents serve as both operational guides and compliance records, ensuring transparency and accountability across the QHSE system. This documented information spans the entire PDCA (Plan-Do-Check-Act) cycle — from planning through implementation to verification — and must be presented in a readable, controlled, up-to-date, and accessible format (ISO 9001:2015, Clause 7.5.3) (Balabanov & Davletshin, 2018).

2.2.3 Traceability and Evidence in Audits

Document traceability is a cornerstone of QHSE audits, as it demonstrates that requirements have been understood, implemented, and met. Traceability refers to the ability to track and verify the history, location, or application of data or documents throughout the system's lifecycle. Records are the objective evidence auditors require to validate compliance (ISO 9001:2015, Clause 9.2.2) (Shabani et al., 2024).

In the absence of valid records, the activity is presumed not to have occurred. Furthermore, documented information must be properly controlled, regularly updated, protected from accidental loss, and accessible to authorized individuals, as stipulated in Clauses 7.5.2 and 7.5.3 of the three ISO standards. This requirement is crucial not only for satisfying internal and external audits but also for ensuring the robustness of the QHSE system and its alignment with continuous improvement principles (Yew & Goh, 2015).

2.2.4 Challenges Related to Multiple Formats and Sources

One of the main challenges in managing QHSE information lies in the variety of document formats: office files (PDF, Excel), scanned images, emails, paper reports, etc. This heterogeneity complicates the centralization, analysis, and quick access to key data by auditors (Shabani et al., 2024). Such diversity of formats, often called "data

heterogeneity," can hinder automation and increase the risk of errors or omissions in audit preparation.

As highlighted by Shabani et al. (2024) in their study on safety audits in Zimbabwe, these challenges are exacerbated when documentation is dispersed across departments without a structured document management system. In many cases, companies lack a unified and centralized repository, leading to information loss, difficulty in ensuring compliance, and slower audit processes. In addition, the absence of a clear version control or automatic update system can result in the use of outdated documents, potentially causing nonconformities or critical findings during audits.

To address these constraints, the authors stress the need for integrated digital solutions such as document management systems (DMS) or digital QHSE management platforms, which improve the reliability, traceability, and accessibility of documented information (Balabanov & Davletshin, 2018; Shabani et al., 2024).

2.3 QHSE Document Audit Process

The QHSE document audit process is a key step in ensuring compliance and performance of quality, hygiene, safety, and environmental management systems. It relies on standardized practices and rigorous collection of documentary evidence, thereby ensuring the reliability of findings and the relevance of recommendations (ISO 19011, 2018). Documentary evidence refers to objective records such as procedures, logs, or training certificates that demonstrate the implementation of audited requirements. This process involves successive phases, from preparation to reporting, with constant attention to objectivity, traceability, and continuous improvement.

2.3.1 Audit Phases

The QHSE audit follows a rigorous methodology defined by the international standard ISO 19011:2018, which provides guidelines for auditing management systems. This standard bases the audit process on seven key principles, including the risk-based approach (Clause 4.g), essential for prioritizing critical areas and efficiently allocating audit resources. A risk-based approach is a strategic auditing principle that focuses efforts on areas with the greatest potential for nonconformity or impact on objectives.

- The preparation phase includes defining the audit objectives, scope, criteria, and methods (Clauses 5.5.2–5.5.3), as well as reviewing relevant documents beforehand (Clause 6.3.1). Particular attention is paid to analyzing risks and opportunities related to the audit program (Clause 5.3), which shapes the audit plan within a continuous improvement logic.
- The data collection phase (Clauses 6.4.6 and 6.4.7) is carried out through interviews, observations, and document reviews. The collected data must be objective (Clause 3.8), verifiable, and relevant, then compared against the defined audit criteria (Clause 3.7). Verification enables the identification of nonconformities (Clause 3.21) and opportunities for improvement (Clause 3.10, note 2).
- Finally, the audit report phase (Clause 6.5) formalizes the findings, impartially reflecting deviations, best practices, and recommendations. This documented report becomes a strategic foundation for management review and QHSE performance improvement.

2.3.2 Traditional Tools: Spreadsheets, Paper Forms, Non-AI Platforms

Traditional tools used for QHSE audits—spreadsheets, paper forms, and digital platforms without artificial intelligence—present many limitations. As Xu Yao et al. (2024) point out, these processes are *"labor-intensive and reliant on human expertise,"* exposing audits to human error, lack of transparency, and difficulties in standardization, especially in multi-site or multi-standard environments. These tools are typically static, require manual input, and lack integration with automated monitoring or real-time data sources. The absence of automation hinders early anomaly detection or dynamic report generation. Additionally, manual tracking of regulatory updates or document versions affects the responsiveness and reliability of audits.

2.3.3 Limitations : Manual Workload, Risk of Omissions, Difficult Multi-Site Consolidation

When conducted with traditional methods, QHSE audits face several major obstacles. High administrative workload slows down data collection and analysis. Moreover, the lack of intelligent systems makes it difficult to consolidate results across multiple sites or different frameworks (ISO 9001, 14001, 45001). Multi-site consolidation refers to the

process of aggregating audit data from different physical locations into a coherent and standardized report.

Xu Yao et al. (2024) emphasize the need for **consistency**: without suitable tools, audits are prone to variable interpretations depending on the auditor or audited entity. Similarly, Leocádio, Malheiro, and Reis (2024), cited in their review, call for enhanced technological agility and the upskilling of auditors to meet the growing complexity of audit environments.

2.3.4 Needs: Time Savings, Consistency, Automated Checklist Updates

In light of these challenges, several critical needs emerge:

- Time savings: By using a dynamic risk assessment model, audits can focus on high-impact areas, reducing overall time while increasing the relevance of findings. Xu Yao et al. (2024) report a 24% improvement in operational efficiency in their test scenarios.
- Consistency: Intelligent systems such as “compliance copilots” help standardize interpretations, unify auditor practices, and ensure reliable traceability in handling nonconformities.
- Automatic checklist updates: With information extraction algorithms and self-updating databases, checklists can be refreshed in real-time, integrating new regulatory, normative, or internal requirements.

Audit checklists are structured tools that guide auditors through the verification of specific compliance points; automating them ensures alignment with the latest standards. The integration of artificial intelligence through LLM-based systems (Large Language Models), as described by Xu Yao et al. (2024), addresses these challenges by providing proactive assistance to auditors: semantic analysis, risk prioritization, automatic report generation, best practice extraction, etc. LLM-based systems refer to advanced AI models trained on large volumes of text data, capable of understanding and generating human-like language to assist in tasks such as document analysis, classification, and summarization. These systems transform QHSE document auditing into an agile, predictive, and outcome-oriented process.

2.4 AI Technologies for Document Auditing

The growing use of artificial intelligence in document auditing relies on several complementary technologies that enable the automation of information collection, analysis, and synthesis. These techniques improve audit speed, reliability, and consistency, particularly in complex regulatory environments. The following subsections detail the contributions of NLP, OCR, ML/DL, and expert systems, based exclusively on recent scientific studies (Xu Yao, 2022; Noordin et al., 2022; Hu et al., 2023).

2.4.1 NLP – Entity Extraction, Classification, Text Similarity

Natural Language Processing (NLP) automates named-entity extraction and document classification, facilitating the detection of inconsistencies and compliance verification. NLP is a branch of artificial intelligence focused on enabling machines to understand and interpret human language, both written and spoken (Noordin et al., 2022). Noordin et al. (2022) show that NLP tools significantly reduce document analysis time while improving precision, handling up to one hundred times faster processing of contractual documents while maintaining high-quality identification of critical clauses. This performance gain is particularly useful in highly regulated sectors such as finance.

2.4.2 OCR – Converting Scanned Documents into Exploitable Text

Optical Character Recognition (OCR) converts paper or scanned documents into usable text data, reducing manual data entry. OCR is a technology that enables the conversion of different types of documents, such as scanned paper documents or images, into editable and searchable data (Popara et al., 2022). Popara et al. (2022) demonstrate that integrating OCR into digital audit processes enhances the speed and reliability of information retrieval from large volumes of documents. Zhang (2021) adds that combining OCR with AI enables the automatic detection of anomalies in commercial documents, further improving audit accuracy.

2.4.3 ML / DL – Automated Detection of Non-Conformities

Machine Learning (ML) and Deep Learning (DL) are used to analyze large datasets to identify anomalies often invisible to human analysis. ML is a subset of AI where algorithms learn patterns from data to make decisions or predictions without being

explicitly programmed, while DL is a more complex form of ML using neural networks with multiple layers to model intricate data representations (Xu Yao, 2022). Xu Yao (2022) reports that supervised ML models applied to auditing improve anomaly detection and reduce human errors. Additionally, Noordin et al. (2022) indicate that these technologies increase audit precision by up to 50% and reduce audit risk by 30%, by more quickly identifying inconsistencies.

2.4.4 Expert Systems – Explicit Business-Rule Engines

Expert systems rely on explicitly defined logical rule engines to model domain knowledge. These are AI systems that emulate decision-making abilities of human experts by applying a set of predefined rules to data (Hu et al., 2023). Hu et al. (2023) present a model based on multiple fuzzy rules for corporate audit governance, enabling the formalization of conformity criteria with transparency. These systems facilitate decision traceability and are well-suited to regulated environments where clarity and precision of rules are critical, notably in triggering alerts for financial thresholds or unobserved procedures.

2.4.5 Comparison: Advantages and Limitations

According to Noordin et al. (2022), AI ensures uniformity in the application of audit criteria, while accelerating data processing and reducing human bias. However, Xu Yao (2022) highlights that the lack of explainability in machine learning models limits their adoption in regulated contexts where traceability is essential. Moreover, Iwuanyanwu et al. (2023) identify as adoption barriers data quality, system complexity, and the need for auditor training.

In response to these limitations, Explainable Artificial Intelligence (XAI) is increasingly seen as a middle ground. It makes algorithmic decisions understandable to human users without compromising performance. Zhang, Cho, and Vasarhelyi (2022) explain that techniques such as LIME (Local Interpretable Model-agnostic Explanations) and SHAP (SHapley Additive exPlanations) help identify which variables most influenced a model's decision. These tools are particularly useful in automated audits where standards demand that decisions be justified and traceable. XAI thus supports greater AI adoption in

regulated environments by reconciling algorithmic innovation with transparency requirements (Zhang et al., 2022).

2.5 Industrial Configurations of AI in QHSE Systems

Recent research indicates that the industrial deployment of artificial intelligence in auditing and compliance environments does not occur as a uniform technological shift but rather as a progressive configuration of specific functional modules embedded within organizational systems (Kokina et al., 2025). Studies on AI adoption in auditing emphasize that implementation is typically driven by operational needs, regulatory constraints, and existing information system architectures rather than by full-system replacement strategies (Wijaya et al., 2025). Furthermore, the alignment between AI capabilities and regulatory requirements requires structured mapping between technological functions and normative obligations, particularly in standards-driven environments (Imperial et al., 2025). In regulated domains, governance mechanisms and lifecycle controls are increasingly considered prerequisites for industrial AI deployment (ISO/IEC 42001, 2023; Winkler et al., 2025). These observations frame AI integration in QHSE systems as a configuration problem involving functional scope, system architecture, and compliance governance rather than purely algorithmic performance.

2.5.1 Typical Functional Scopes (Planning, Reporting, Document Control, Monitoring)

Industrial implementations of artificial intelligence in QHSE systems are generally structured around discrete operational functions rather than comprehensive end-to-end automation (Kokina et al., 2025). The literature indicates that deployment strategies tend to target modular processes where efficiency gains and standardization can be achieved without fundamentally restructuring the entire management system. Within audit planning and resource allocation, AI-supported systems are described as facilitating task coordination and optimization through data-driven allocation mechanisms, contributing to reductions in administrative workload and improvements in operational efficiency (Wijaya et al., 2025; Kokina et al., 2025). These configurations illustrate how AI can optimize preparatory phases of the audit process while preserving the central role of professional evaluative judgment. In reporting contexts, Large Language Models (LLMs) and NLP-based systems are reported to support automated drafting of audit reports and

structured summarization of findings. Such applications reduce manual documentation effort and enhance consistency in reporting formats, particularly in multi-site or high-frequency audit environments (Xu Yao et al., 2024; Kokina et al., 2025). The emphasis remains on augmenting documentation workflows rather than substituting analytical oversight. Concerning document control and compliance verification, AI systems combining NLP techniques and rule-based logic are used to extract regulatory requirements and verify the presence of documented information within organizational records (Arora et al., 2015; Haataja et al., 2021). These approaches align with ISO requirements related to documented information and traceability, especially where structured mapping between normative clauses and internal documentation is required (Adebayo et al., 2022). In monitoring and anomaly detection, machine learning and deep learning models are described as supporting the identification of irregularities and risk indicators within safety and compliance systems (Xu Yao, 2022). In industrial QHSE platforms, AI-driven analytics, including predictive trend extraction and automated pattern recognition, enable more proactive identification of hazardous conditions and emerging compliance risks (Shabani et al., 2024). Across these operational configurations, the literature consistently frames AI as an assistive technology aimed at enhancing efficiency, consistency, and standardization, while maintaining human judgment as a central component of compliance evaluation and decision-making processes (Dellermann et al., 2021).

2.5.2 Integration Models (Standalone vs ERP/DMS Integrated; Data Flows; Governance)

The literature distinguishes two principal industrial integration configurations in the deployment of AI within compliance and auditing environments. One configuration consists of standalone analytical systems operating as independent modules that process uploaded documents or extracted datasets without deep embedding into enterprise information systems (Kokina et al., 2025). Although this architecture facilitates relatively rapid deployment and operational flexibility, it may constrain interoperability and limit continuous or real-time data exchange across organizational infrastructures. A second configuration involves the integration of AI functionalities within Enterprise Resource Planning (ERP) or Document Management Systems (DMS), enabling continuous data flows between operational records and compliance-related modules (Balabanov &

Davletshin, 2018). However, such integration frequently encounters interoperability challenges, particularly when legacy systems and complex permission structures are involved (Shabani et al., 2024). Within these configurations, structured data governance emerges as a central requirement. The use of standardized data schemas enhances traceability and auditability of AI-generated outputs, thereby supporting alignment with ISO compliance requirements (Kvalito, 2024). More broadly, aligning AI systems with regulatory standards necessitates computational mapping between normative frameworks and technological capabilities (Imperial et al., 2025). From a governance standpoint, ISO/IEC 42001:2023 introduces requirements concerning AI lifecycle management, transparency, and risk control, reinforcing the need for structured oversight mechanisms in regulated domains (ISO 42001, 2023; Winkler et al., 2025).

2.5.3 Recurring Structural Constraints in Industrial Deployment

Although several studies report efficiency gains associated with AI-supported auditing and compliance processes, the literature also suggests that large-scale industrial deployment may depend on a range of structural and organizational conditions (Kokina et al., 2025). These conditions appear to extend beyond algorithmic performance and relate more broadly to the systemic context in which AI tools are embedded. A frequently discussed constraint concerns the integration of AI modules into existing enterprise infrastructures. Legacy ERP systems, heterogeneous data environments, and fragmented information architectures may complicate interoperability and limit seamless data exchange (Balabanov & Davletshin, 2018; Shabani et al., 2024). In regulated industrial settings, AI applications are often introduced within pre-established governance structures and documentation workflows, which may influence their operational effectiveness. Organizational readiness is also described as a relevant factor in the literature. The implementation of AI-supported auditing tools may require adjustments in internal processes, clarification of roles, and the development of digital competencies among professionals (Wijaya et al., 2025). In this perspective, AI adoption can be interpreted not merely as a technological upgrade, but as part of a broader socio-technical transformation that may necessitate structured change management approaches. In addition, several authors underline the potential importance of governance mechanisms when AI systems are deployed in compliance-sensitive environments. Hybrid intelligence frameworks tend to conceptualize AI as augmenting rather than replacing human

expertise, suggesting that supervisory oversight may remain necessary for evaluative and interpretative tasks (Dellermann et al., 2021). From this viewpoint, industrial deployment could require clearly defined accountability arrangements and explicit allocation of decision-making responsibilities. Taken together, these recurring themes in the literature indicate that the industrialization of AI in QHSE systems may be influenced by organizational maturity, infrastructural compatibility, and governance alignment, rather than being determined solely by advances in machine learning techniques.

2.6 Automated Generation of Normative Checklists

Automating the generation of normative checklists represents a significant advance in improving the quality and compliance of audits and control processes. This automation relies on advanced NLP techniques to extract and structure regulatory requirements, then adapt these checklists to the specific context of the audited organization (Arora et al., 2015; Cutting-Decelle et al., 2018). Normative checklists are structured lists based on standards or regulations used to verify compliance and ensure that all required criteria are met during audits (Arora et al., 2015).

2.6.1 Clause Extraction (NLP Parsing / Linguistic Patterns)

Automated extraction of normative clauses relies on NLP models capable of recognizing linguistic patterns specific to regulatory requirements. Arora et al. (2015) demonstrate that by casting specification templates as NLP patterns, it is possible to automate conformity verification to these templates in a robust and precise manner. This method enables rapid identification of relevant clauses in complex regulatory documents, thereby reducing manual effort and associated errors.

2.6.2 Mapping : Linking Standards to Internal Documents

Effectively applying a standard like ISO in a company requires linking the standard's written requirements to the organization's internal documents (procedures, reports, forms, etc.). Cutting-Decelle et al. (2018) proposed a method using ontologies intelligent dictionaries to understand technical terms in standards. Ontologies are structured representations of knowledge that define concepts and their relationships within a domain, allowing software to interpret and connect complex information accurately

(Cutting-Decelle et al., 2018). This approach allows software to automatically detect key normative concepts (e.g., “preventive maintenance” or “production planning”) and associate them with best practices already in place. This creates clear links between normative requirements and actual organizational practices, facilitating compliance and tailored checklists.

2.6.3 Personalization : Adapting Checklists to Context

Standard-based checklists often need customization for real-world application. For instance, in the medical field, Wang et al. (2020) developed a tool that uses AI to automatically read PDF clinical reports and generate a personalized checklist compliant with the CONSORT standard (used for clinical trials). The tool reads documents as an expert would, identifies critical sections (methods, results, etc.), and generates a list of items to check. Personalization here means tailoring generic checklists to fit the specific needs and context of the audited entity, improving relevance and effectiveness (Wang et al., 2020). This proves it is possible to automate the creation of context-specific checklists, enhancing accuracy and saving professionals time.

2.6.4 Academic Prototype : Semi-Automated Checklist Generation

In an academic research context, Wang et al. (2015) developed a prototype called UMTG capable of automatically generating verification scenarios (or test cases) from simple natural language task descriptions. This means that, from a text describing how a process should operate, the system can infer all the verification steps, including some those humans might overlook. The system successfully automated 95% of the described cases and even proposed new ones, demonstrating that AI can help create complete and reliable checklists even in complex domains such as engineering or auditing.

2.7 Technical Limitations and Implementation Challenges

The adoption of artificial intelligence in the QHSE (Quality, Health, Safety, Environment) field presents clear advantages but also faces several technical and regulatory challenges. According to Winkler et al. (2025), one of the main challenges is the explainability of AI systems, i.e., their ability to justify their decisions. This criterion is essential in sensitive domains like QHSE, where every decision must be clearly and understandably justified. However, the AI Act (European regulation on AI) only proposes

general rules without specifying concrete methods to evaluate explainability. The AI Act is a European Union regulation aiming to ensure that AI systems are trustworthy, transparent, and respect fundamental rights, especially in sensitive fields like QHSE (Winkler et al., 2025). This makes it difficult to use AI in regulated contexts such as QHSE, which require precise justifications.

2.7.1 Data : Lack of Annotated Corpora

AI models learn from annotated data, meaning documents already analyzed and categorized by experts. In legal and regulatory fields, such datasets are extremely rare. Zin et al. (2024) show that, due to the lack of manually annotated data, some models are trained on automatically generated content (e.g., using GPT-3). This improves short-term performance but raises concerns about quality and reliability. Constant et al. (2025) confirm that existing corpora are often incomplete or imprecise, limiting the reasoning capabilities of AI systems. For instance, their AnnoCaseLaw project had to be manually enriched with 471 legal cases annotated by experts to fill this gap. These findings apply directly to the QHSE sector, where no openly accessible annotated databases currently exist. This limitation hinders AI progress in this field.

2.7.2 Complex Normative Language

QHSE documents, such as ISO standards, are often written in legalistic language that is hard for machines to interpret. Guha et al. (2023) explain that regulatory texts pose a dual challenge : they must be legally accurate while remaining readable for analytical tools. However, they are often ambiguous, and divergent interpretations are not always well captured. The E-NER study (2022) also notes that while many legal texts are available, few are annotated in a way that is useful for AI. For example, Legal-BERT, a legal AI model, was trained on a dataset that is not publicly accessible. Normative language refers to formal, rule-based text such as laws or standards that specify obligations or requirements, often complex and ambiguous for automated interpretation (Guha et al., 2023). This lack of clarity and accessible data hampers the effectiveness of AI tools in properly handling QHSE normative documents. The evaluation of robustness aims to determine whether a system produces stable outputs when confronted with formal modifications that do not alter the intended meaning. From a measurement perspective, when multiple evaluations must be compared and the scale is ordinal, Krippendorff’s

alpha is a recognized reliability coefficient, applicable to a variable number of raters and to different levels of measurement, including the ordinal level (Krippendorff, 2011). In natural language processing, robustness is also examined through behavioral testing and controlled perturbation approaches. (Ribeiro et al.,2020) propose CheckList as a general behavioral testing methodology designed to uncover systematic model failures. Similarly, (Jia and Liang, 2017) demonstrate that inserting distractor sentences can significantly degrade the performance of systems without altering the correct expected answer, thereby highlighting the importance of invariance testing.

2.7.3 Explainability : Justifying a Decision to the Manager

In a QHSE context, it is crucial that managers understand why an AI has determined whether a document is compliant. Winkler et al. (2025) highlight that the rules proposed by the AI Act remain vague and are hard to translate into concrete indicators that tools can integrate. Additionally, the *Survey on Explainable AI and Law* (2023) notes that legal reasoning methods are varied and complex. To be credible, AI systems must tailor their explanations to the specific context. This means systems must not only give answers but also clearly state the rule used, the level of certainty, and the reasoning followed. Currently, achieving this level of explainability remains technically challenging. Explainability is the capacity of AI systems to provide understandable and transparent reasons for their outputs, which is essential for trust and accountability in regulated domains (Winkler et al., 2025). The reproducibility of verdicts constitutes a central dimension of the reliability of an evaluation system, particularly when it is used as a decision-support tool. In measurement science, agreement between two categorical evaluations is commonly quantified using Cohen's κ coefficient, which adjusts the observed agreement for the agreement expected by chance (Cohen, 1960). This coefficient is widely employed to assess reliability when two independent evaluations must be compared, for example in concordance or test–retest studies. The interpretation of κ may rely on commonly used methodological benchmarks, such as the scale proposed by Landis and Koch (1977), which classifies the strength of agreement into interpretable categories ranging from slight to almost perfect.

2.7.4 Security & Confidentiality

Using AI in sensitive fields like quality, health, safety, and environment raises major concerns around data confidentiality. As noted in *Challenges and Considerations in Annotating Legal Data* (2024), "annotating legal texts involves confidentiality risks; sensitive data (contracts, decisions) require rigorous security protocols before any sharing or processing." This requirement is especially critical in QHSE, where the data involved often include incident reports, internal audits, or confidential HR information. Any AI solution must therefore natively integrate data protection mechanisms during both the training and usage phases. Data confidentiality refers to protecting sensitive information from unauthorized access or disclosure, crucial when handling personal or proprietary QHSE data (Challenges and Considerations in Annotating Legal Data, 2024).

2.7.5 AI Compliance : Governance & Standards (ISO 42001)

Implementing AI in regulated contexts also requires strong governance and compliance practices. The ISO/IEC 42001:2023 standard addresses this need by specifying "requirements for establishing, implementing, maintaining, and continually improving an AI Management System (AIMS), ensuring responsible and effective use of AI" (ISO 42001, 2023). It specifically addresses AI lifecycle management, process transparency, and risk control. According to KPMG (2024), "ISO/IEC 42001 provides a structured framework for AI governance: risk management, impact assessment, lifecycle management, and third-party oversight." ISO/IEC 42001 is an international standard that defines guidelines and best practices for managing AI systems responsibly and transparently within organizations. Adopting this standard is a strategic opportunity to enhance the credibility and robustness of AI tools in the QHSE field.

2.8 Comparative Analysis: Standards vs. AI Capabilities

Aligning normative requirements, particularly those from ISO standards, with the current capabilities of artificial intelligence (AI) technologies is a central issue for ensuring effective operational compliance. Imperial, Jones, and Madabushi (2025) emphasize: "We assess the criticality levels of different standards across domains and sectors and complement them by grading the current compliance capabilities of state-of-the-art GenAI models. … We argue that aligning GenAI with standards through computational

methods can help strengthen regulatory and operational compliance. Generative AI (GenAI) refers to AI models capable of producing new content, such as text or images, based on learned patterns from large datasets (Imperial et al., 2025). This approach highlights the importance of standardizing generative AI models to meet regulatory requirements and operationalize compliance.

2.8.1 Overview of Key ISO Requirements : AI Coverage vs. Critical Points

Recent research shows that current AI technologies can meet certain ISO standard requirements, particularly those related to document traceability. Adebayo, Sow, and Bulut (2022) demonstrate that AI techniques can be used to automatically analyze regulatory standards. Their approach allows normative requirements to be linked with technical controls from the early stages of system design. This involves technologies such as Optical Character Recognition (OCR) and automated mapping, which facilitate the alignment of regulatory documents with ISO requirements (Adebayo et al., 2022). Furthermore, Kvalito (2024–2025) emphasizes the importance of using standardized data schemas. These formats ensure a clear audit trail of the data used to train or modify an AI model. This supports ISO-compliant traceability of both data and AI-driven decisions. Data schemas refer to structured formats that organize data consistently to ensure interoperability, traceability, and compliance with standards like ISO (Kvalito, 2024).

2.8.2 Technological Gaps : Identifying Areas Unaddressed by Current AI

Despite these advances, several key aspects of ISO standards remain poorly covered by current AI tools. Ojewale et al. (2024) note that while many tools assist in evaluating AI systems, they fall short of ensuring true accountability during audits. Specifically, AI struggles to assess qualitative elements such as safety culture, human interpretation, or organizational context—dimensions that demand nuanced and often subjective understanding. The evaluation of a system's ability to detect deviations (non-conformities) can be formalized as a classification problem, for which the literature provides standardized metrics derived from the confusion matrix. Recall (sensitivity) measures the proportion of actual positive cases that are correctly identified, whereas precision measures the proportion of predicted positive cases that are correct. The F1-score synthesizes the trade-off between precision and recall. These metrics constitute widely accepted standards for the evaluation of automated decision systems and

classification models (Powers, 2011). Their use avoids reliance on arbitrary thresholds and enables performance to be discussed in terms of the balance between under-detection (false negatives) and over-detection (false positives), which is particularly relevant in contexts where classification errors carry operational consequences. Safety culture encompasses the shared values, beliefs, and practices related to safety within an organization, which are difficult to quantify and analyze using AI (Ojewale et al., 2024). Organizational context refers to the unique circumstances and environment of a company that influence how regulations apply, requiring human judgment beyond AI's current capabilities (Nannini et al., 2023).

2.8.3 Hybrid Model : Combining AI with Human Expertise

In response to these limitations, several publications recommend a hybrid model rather than positioning AI as a replacement for human experts. The Knapsack Report (2024) underlines that AI does not replace humans but rather redefines auditors' roles toward higher-value activities such as strategic analysis and continuous improvement. While AI automates repetitive tasks, human expertise remains essential to understand complex situations, interpret standards, and make context-aware decisions. A study published by MDPI (2023–2024) supports this position : AI can reduce the need for human auditors in certain tasks, but improper use may compromise audit quality. Therefore, AI use must be carefully managed, and human experts should always be involved in critical audit phases. A hybrid model integrates AI automation with human judgment to optimize efficiency while maintaining reliability and contextual understanding in compliance auditing (Knapsack Report, 2024). When the objective is to compare scores before and after an intervention on the same units, statistical literature recommends the use of paired tests. In the case of ordinal data without an assumption of normality, the Wilcoxon signed-rank test constitutes a standard non-parametric method for comparing two paired conditions (Wilcoxon, 1945). This test assesses whether the distribution of the before–after differences is centered around zero or whether a systematic tendency toward improvement exists. Its use is particularly appropriate when system outputs are expressed on a low-cardinality ordinal scale (e.g., 0/1/2), as is the case for conformity verdicts in documentary auditing.

2.9 Recommendations & Review Conclusion

This review emphasizes the necessity of aligning generative AI (GenAI) models with regulatory and operational standards to reinforce compliance. Imperial, Jones, and Madabushi (2025) state that “aligning GenAI with standards through computational methods can help strengthen regulatory and operational compliance. [...] The gap between regulation and automation invites new research agendas for hybrid systems and formalized evaluation benchmarks.” This framework highlights current automation limitations and future directions.

2.9.1 Future Research Directions : Reference Frameworks, Ontologies, Public Datasets

To fully leverage hybrid systems combining AI and human intelligence, Dellermann et al. (2021) stress the importance of "new shared knowledge structures such as ontologies and open annotated datasets,” which “provide the necessary formalism for collaboration between AI agents and human experts.” In the QHSE context, Wang, Ye, and Pan (2025) emphasize “the urgent need for domain-specific datasets in HSE contexts and the development of structured regulatory reasoning frameworks.” These references justify the creation of shared frameworks, QHSE ontologies, and annotated public corpora, which are essential for the reliability and efficiency of AI systems in this domain.

2.9.2 Implications for My Project : Which Technologies to Prioritize

Haataja, Hyrynsalmi, and Leppänen (2021) show that Natural Language Processing (NLP) is particularly effective “in extracting requirements and rules from unstructured regulatory documents,” while “rule-based expert systems are more effective for enforcing deterministic logic and domain constraints.” Rule-based expert systems apply predefined logical rules to make deterministic decisions, ideal for enforcing clear compliance criteria (Dellermann et al., 2021). Dellermann et al. (2021) thus recommend designing systems that assign judgment-intensive tasks to human experts while allowing AI to handle repetitive or pattern-based processing. This supports prioritizing NLP technologies combined with expert systems in your QHSE platform to maximize relevance and robustness.

2.9.3 Summary Answer to the Research Question & Next Steps

The research question is: How can artificial intelligence be used to automate the audit of QHSE documents and improve the efficiency and accuracy of compliance verification? Recent research shows that AI can effectively automate certain auditing tasks, particularly through Natural Language Processing (NLP) and expert systems. Haataja, Hyrynsalmi, and Leppänen (2021) demonstrate that NLP is well-suited to extract requirements from unstructured regulatory documents, while rule-based systems are more effective in applying the deterministic logic of normative domains. However, Wang et al. (2025) caution that large language models (LLMs) have limitations when it comes to tasks requiring contextual judgment or formal compliance logic, thus necessitating human intervention or the use of structured rule engines. Dellermann et al. (2021) recommend designing hybrid systems where AI handles repetitive tasks, while human experts address complex decisions. Finally, Raji et al. (2020) emphasize the importance of iterative validation and human oversight throughout the lifecycle of automated audit tools to ensure their reliability. These findings support a hybrid approach that combines AI and human expertise according to the nature of the requirements. The next key step in the project is to validate this approach in real-world use cases to concretely assess its robustness and benefits.

3 Materials and Methods

3.1 Methodological positioning and link with the literature review

This chapter presents the methodology implemented to operationally apply the results of the systematic literature review. This review identified the normative requirements for Quality, Health, Safety, and Environment (QHSE) document audits, the artificial intelligence technologies suitable for analyzing normative documents, and the main limitations associated with automating this type of process. The review's conclusions highlighted four major methodological challenges: (1) the need to translate normative requirements into operational representations usable by an automated system, (2) the reliability and reproducibility of decisions produced by artificial intelligence models, (3) robustness to linguistic and structural variations in documents, and (4) the requirement for a hybrid approach that ensures human oversight in normative interpretation. The methodology developed in this chapter was designed as a direct response to these challenges, ensuring explicit consistency between the theoretical results of the review and their experimental implementation. Based on these results, the methodology adopted relies on an applied and experimental approach, structured around two complementary components. The first component consists of evaluating the behavior of an automated document audit application using predefined experimental tests, designed to analyze its reliability, robustness, normative consistency, and the stability of the results produced. The second component is based on conducting a document case study, grounded in the analysis of online QHSE documents, to illustrate the practical use of the application in a context closely resembling professional practice.

The methodology is structured around three main steps:

1. The selection and translation of the requirements of ISO 9001, ISO 14001 and ISO 45001 standards into auditable checklists;
2. The integration of these checklists into an automated document audit application based on artificial intelligence;
3. The application was evaluated using experimental tests, supplemented by a documentary case study.

The objective of this chapter is to precisely describe the methodological choices, tools used, and procedures applied to evaluate the capacity of artificial intelligence to automate, in a controlled manner, the QHSE document audit. The theoretical and conceptual elements already developed in the literature review are not repeated in this chapter, which focuses exclusively on the methodological implementation of the work.

3.2 Normative framework and design principles for checklists

This section describes the methodology adopted to build the checklists used by the automated document audit application. It constitutes the operational link between the normative requirements from ISO 9001, ISO 14001 and ISO 45001 standards, and their use in a document analysis system based on artificial intelligence.

The objective of this step is to transform a general, intentionally flexible and adaptable normative framework into a structured set of verifiable documentary criteria compatible with automated analysis. The approach follows a rigorous selection of relevant requirements, their translation into explicit audit questions, the identification of documents to be analyzed, and the structuring of checklists according to both normative and documentary logic.

3.2.1 Selection of relevant normative requirements

The selection of normative requirements included in the final checklist was carried out using a structured methodological approach, designed to make the requirements of ISO 9001, ISO 14001, and ISO 45001 usable within the framework of automated document analysis. This step is based on the principle that only requirements that can be assessed from explicit documentary evidence can be reliably automated.

Initially, all the requirements of the three standards were analyzed to identify those explicitly referring to the production, updating, maintenance, or control of documented information. This analysis allowed us to exclude requirements based primarily on field observations, behaviors, operational practices, or in-depth contextual interpretations. The remaining requirements were then filtered according to their documentary verifiability. Only those that could be assessed using formalized documentary evidence were retained. This evidence includes, in particular, written policies, documented procedures, records, plans, reports, or tracking charts, considered observable and usable by an automated

analysis application. Based on this, the final checklist primarily covers requirements related to:

- To the formalization of the QHSE policy and objectives;
- To the mastery of documents and records;
- To the formalization of processes, responsibilities and working methods;
- Documentary evidence associated with planning, monitoring, performance evaluation and continuous improvement;
- Documentation requirements related to internal audits, non-conformity management and management review.

When requirements from different standards had similar objectives or comparable content, they were harmonized to limit redundancies and simplify analysis. This approach made it possible to create a single, consistent framework compatible with automated assessment, while respecting the specific characteristics of each standard.

3.2.2 Method for translating ISO requirements into audit questions

The selected normative requirements were then translated into audit questions, forming the operational interface between the normative framework and the automated document audit application. The objective of this step is to transform general requirements into concrete criteria that can be reliably assessed using documents. The translation method adopted is based on a systematic and reproducible approach. Each requirement was analyzed individually to identify elements that could be verified solely through documentary evidence. This analysis aims to isolate the components of the requirement that are likely to be observed in a document, regardless of the specific organizational context. The requirements were then reformulated as audit questions geared towards document verification. The general normative formulations were translated into explicit questions focusing on the existence, availability, or formalization of documented information. Elements involving human judgment, qualitative assessment, or contextual interpretation were deliberately excluded to limit evaluation bias. Particular attention was paid to the linguistic standardization of the questions. The wording was simplified and harmonized to ensure clarity, consistency, and stable interpretation by the application—essential for reproducible results. For each question, the expected documentary evidence

was explicitly defined to guide the automated analysis toward observable and relevant elements. Artificial intelligence assisted in the formulation of the questions as a tool for drafting and linguistic standardization, without intervening in the normative interpretation. Each question then underwent manual review and validation to ensure its alignment with ISO standards requirements.

3.2.3 Identification of documents to be analyzed and operationalization of documentary requirements

ISO standards deliberately formulate their documentation requirements in a generic manner, without imposing specific named documents or formats. In the context of an automated document audit, this characteristic necessitates a specific methodological step to identify the concrete documents to be analyzed. In this work, this step consists of transforming the document categories defined by the standards into operational documents that can be effectively analyzed by the application. The selected requirements were first grouped according to their documentary nature, distinguishing in particular:

- documents to be maintained (policies, procedures, plans);
- Information to be kept in the form of records or evidence.

These categories were then translated into a list of concrete documents commonly found in QHSE management systems and accessible in digital format. The selected documents correspond to organizational instances of normative categories, chosen based on their functional role as evidence of compliance, rather than their exact title. This operationalization makes it possible to clearly define the units of analysis used by the application and to stabilize the document scope of the automated audit, while taking into account the diversity of document practices between organizations.

3.2.4 Structuring checklists

The checklists were structured to ensure consistency between the organization of the normative framework and the constraints of automated document analysis. The approach adopted is based on a document-centric logic, in which documents constitute the primary evidence of conformity. The checklist is organized by document type, with each document representing an independent unit of analysis for the application. For each document, the questions are grouped according to the applicable normative framework

(ISO 9001, ISO 14001, and/or ISO 45001), allowing the analysis to be adapted to the actual scope of the organization under study. In addition, the checklists follow the harmonized structure of ISO standards, based on the major categories of the management system:

A. Context of the organization;
B. Leadership;
C. Planning;
D. Support;
E. Operation ;
F. Performance evaluation;
G. Improvement.

The documents to be analyzed are assigned to these categories according to their functional role within the management system. This combined organization makes it possible to produce structured, readable, and usable results, facilitating both the interpretation of the results and the identification of areas for improvement.

3.3 Development of the automated document auditing application

The development of the application is part of an experimental approach aimed at exploring the feasibility and limitations of automating Quality, Health, Safety, and Environment (QHSE) document audits using artificial intelligence. The objective is not to design a certifiable tool or a system intended for industrial deployment, but rather to develop a functional prototype that allows the evaluation, within a controlled framework, of the actual contribution of artificial intelligence to document compliance analysis. The application is designed as an audit support tool, focused on verifying the presence or absence of explicit documentary elements with respect to predefined normative requirements. It enables the analysis of artificial intelligence behavior when confronted with structured documentary criteria, the observation of the consistency and stability of the results produced, and the conduct of both experimental tests and a documentary case study based on documents accessible online. Within the framework of this work, limited technical assistance was sought to provide methodological guidance in the development of the prototype, particularly to clarify certain implementation steps related to the use of computational tools and artificial intelligence technologies. This assistance was strictly

advisory in nature and did not involve the direct development of the system. The design of the prototype, the configuration of the analysis framework, the integration of the checklists, and the execution of the experimental tests were carried out as part of this research work. This methodological approach ensures consistency between the development of the application, the creation of checklists based on ISO 9001, ISO 14001, and ISO 45001 standards, and the research objectives defined in this study. It also frames the use of artificial intelligence within a clearly defined scope compatible with the rigorous requirements inherent in QHSE document auditing.

3.3.1 Development environment and functional objectives

The automated document auditing application was implemented using Python, chosen not for its advanced software development capabilities, but as a tool to leverage existing artificial intelligence functionalities for processing text documents. Python is widely used in applied research projects integrating data analysis and artificial intelligence tools, making it suitable for an experimental approach outside the strict domain of computer science. The work was carried out using the Visual Studio Code (VS Code) environment, which served as a functional interface for running the program, integrating the necessary parameters, and monitoring the progress of the analyses. The use of this environment was practical and operational, enabling the execution and testing of the application under controlled conditions, without requiring extensive software development.The prototype was developed using existing Python modules and structured scripts, adapted and parameterized specifically for the experimental needs. The GPT code was then adjusted to meet the project's specific objectives. This approach allowed the work to focus on the methodological and functional aspects of the document audit, rather than the programming itself. The adjustments primarily concerned the analysis parameters, the integration of the standardized checklist, and the definition of the tool's expected behavior. This approach reflects a logic of reasoned use of existing digital tools, in which artificial intelligence is used as a technical support for a clearly defined QHSE objective. Human oversight remains central, particularly for selecting the standardized framework, structuring the checklists, and interpreting the results produced by the application.

3.3.2 Integration of the Artificial Intelligence Application Programming Interface (API)

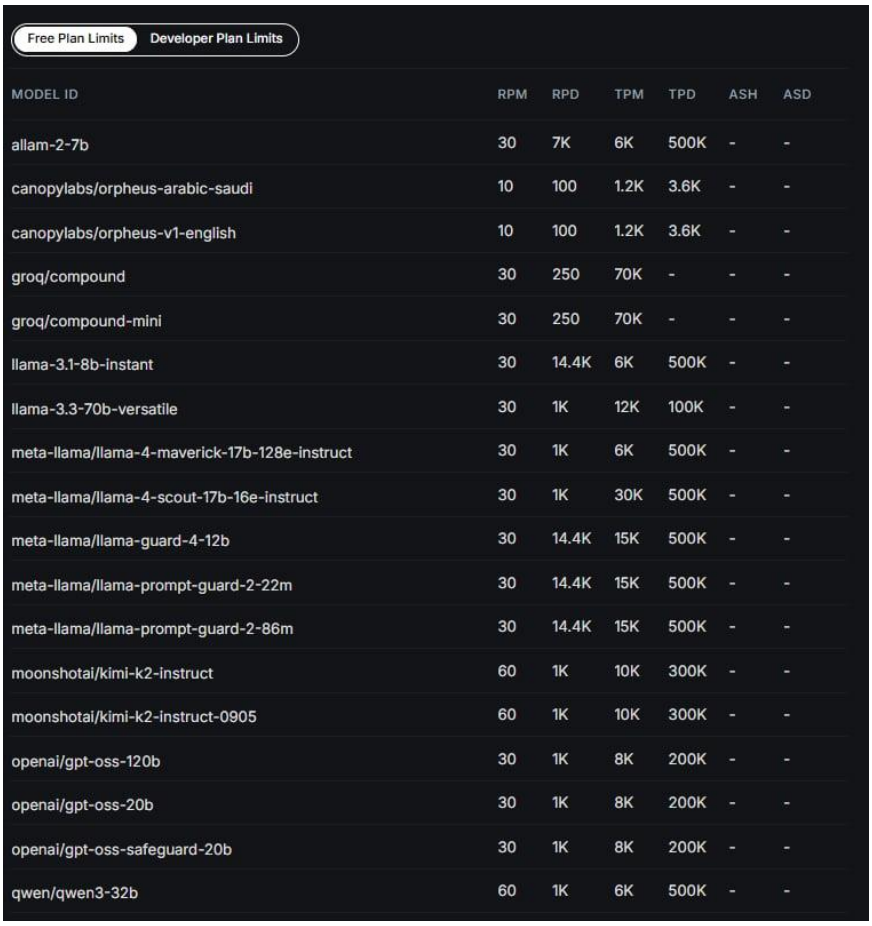
Automatic document analysis relies on the use of an Application Programming Interface (API) provided by the Groq platform, which allows access to artificial intelligence models specialized in text analysis. In this work, the API serves as a means of accessing an artificial intelligence service, and not as a software development tool in itself. Integrating this API is based on a standardized procedure, accessible to non-technical users. It requires creating a user account on the Groq platform, followed by the generation of a personal authentication key. This key is integrated into the application to authorize the secure exchange of data between the developed tool and the artificial intelligence service, in accordance with the authentication protocols defined by the platform. Communication between the application and the service is ensured via a dedicated connection address (base URL) provided by the platform. This address directs requests to the selected language model without requiring any intervention on the underlying technical infrastructure. The application uses the standard connection code recommended by the platform, ensuring compliant use of the service and correct management of call parameters. From a methodological point of view, this integration allows for a clear separation of:

- The documents analyzed;
- The normative framework, materialized by checklists;
- The analysis engine, outsourced via the artificial intelligence service.

This structural separation ensures that the application remains focused on QHSE document auditing, while the technical aspects related to the internal functioning of the model are delegated to the platform.

3.3.3 Choice of language model and operational constraints

The Groq platform provides a defined set of large language models, whose characteristics differ in terms of processing capacity, execution speed, and usage constraints. Figure 1 shows all the models accessible during application development.

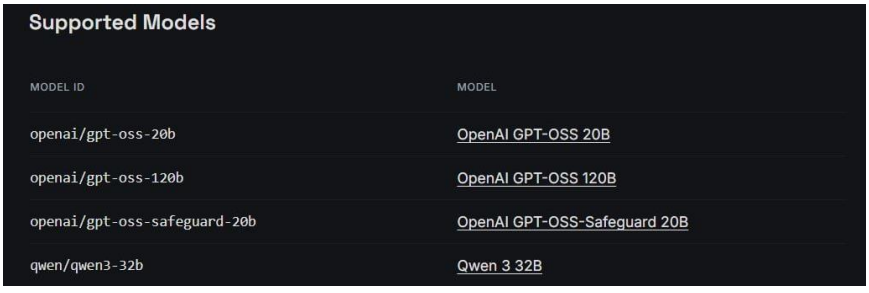


Free Plan Limits | Developer Plan Limits

MODEL ID	RPM	RPD	TPM	TPD	ASH	ASD
allam-2-7b	30	7K	6K	500K	-	-
canopylabs/orpheus-arabic-saudi	10	100	1.2K	3.6K	-	-
canopylabs/orpheus-v1-english	10	100	1.2K	3.6K	-	-
groq/compound	30	250	70K	-	-	-
groq/compound-mini	30	250	70K	-	-	-
llama-3.1-8b-instant	30	14.4K	6K	500K	-	-
llama-3.3-70b-versatile	30	1K	12K	100K	-	-
meta-llama/llama-4-maverick-17b-128e-instruct	30	1K	6K	500K	-	-
meta-llama/llama-4-scout-17b-16e-instruct	30	1K	30K	500K	-	-
meta-llama/llama-guard-4-12b	30	14.4K	15K	500K	-	-
meta-llama/llama-prompt-guard-2-22m	30	14.4K	15K	500K	-	-
meta-llama/llama-prompt-guard-2-86m	30	14.4K	15K	500K	-	-
moonshotai/kimi-k2-instruct	60	1K	10K	300K	-	-
moonshotai/kimi-k2-instruct-0905	60	1K	10K	300K	-	-
openai/gpt-oss-120b	30	1K	8K	200K	-	-
openai/gpt-oss-20b	30	1K	8K	200K	-	-
openai/gpt-oss-safeguard-20b	30	1K	8K	200K	-	-
qwen/qwen3-32b	60	1K	6K	500K	-	-

Figure 2 - Language models available on the Groq platform at the time of application development

Among these models, only some met the methodological requirements of a QHSE document analysis, particularly due to their ability to handle large volumes of text, maintain semantic consistency across structured normative content, and produce stable responses within a systematic evaluation framework. Figure 3 highlights the models offering the most suitable overall performance based on criteria combining contextual analysis capabilities, processing efficiency, and economic constraints of use.



Supported Models

MODEL ID	MODEL
openai/gpt-oss-20b	OpenAI GPT-OSS 20B
openai/gpt-oss-120b	OpenAI GPT-OSS 120B
openai/gpt-oss-safeguard-20b	OpenAI GPT-OSS-Safeguard 20B
qwen/qwen3-32b	Qwen 3 32B

Figure 3 - Subset of models identified as relevant for QHSE document analysis

Following this comparative analysis, the openai/gpt-oss-120b model was selected. This choice was based on its ability to handle a large textual context, an essential element for reviewing voluminous normative documents, as well as the quality of its semantic understanding of technical and regulatory formulations. Furthermore, the usage limitations associated with this model (RPM, RPD, and TPM) appeared compatible with the project's experimental needs, allowing for stable and reproducible operation during the testing and case study phases. The selection process was thus based on clearly defined methodological and operational criteria: • the ability to process large volumes of text; • the quality of the semantic interpretation of normative content; • compatibility with a structured evaluation approach focused on document verification.

The use of a large language model is consistent with the work identified in the literature review, which highlights the relevance of natural language processing (NLP) technologies for extracting and analyzing normative requirements. Unlike an expert system based solely on explicit rules, this approach allows for the management of linguistic variability in institutional documents. Furthermore, the lack of annotated corpora specific to the QHSE domain led to the rejection of a supervised learning approach. The choice of a pre-trained model thus aims to reconcile semantic flexibility with experimental feasibility. These parameters were integrated from the design phase to guarantee experimental consistency and the reproducibility of the results obtained.

3.3.4 Functional settings and role assigned to artificial intelligence

To ensure that the use of artificial intelligence aligns with the project objectives and the requirements of the QHSE document audit, explicit functional settings were implemented during application initialization. These settings aim to define the behavior of the language model and limit its scope to strictly normative document analysis. An explicit role of "QHSE auditor" was thus assigned to the artificial intelligence. This role contextualizes the model's responses and guides it toward a document analysis approach focused on verifying document compliance with predefined requirements. The goal is not to simulate full human judgment, but to guide the AI toward a structured and reproducible analysis. The context provided to the model includes references to ISO 9001, ISO 14001, and ISO 45001 standards, used as a general framework. These standards define the scope of the document audit, while the operational assessment relies exclusively on checklists

developed beforehand. This separation allows for human control over the definition of the normative requirements.

3.3.5 Document management and checklist usage

The application analyzes QHSE documents submitted digitally using a structured checklist based on ISO standards requirements. This checklist serves as the central operational reference for the document analysis system and guides the entire conformity assessment process. The checklist is integrated as a structured table, exported in Portable Document Format (PDF) to ensure a readable and standardized presentation. For each document analyzed, the text content is extracted and then compared to the questions defined in the checklist. Artificial intelligence is used to identify the presence or absence of relevant documented information, taking into account possible variations in wording. To facilitate this comparison, the application uses an embedding model as a tool for semantic text organization. This mechanism allows for the comparison of different formulations expressing the same documentary information, without being limited to a word-for-word match. It helps improve the consistency of the analysis without interfering with the normative assessment. The application does not implement supervised learning models, complex automatic non-conformity detection, or rule-based expert systems. The decision was made to limit itself to structured document analysis to ensure the traceability, reproducibility, and explainability of the results. Artificial intelligence is used exclusively as a tool for analyzing and structuring information, without any autonomous normative judgment.Although the literature mentions the use of ontologies to structure normative requirements, this approach was not adopted in this experimental work. Building a comprehensive QHSE ontology would require a collaborative, multi-expert process extending beyond the project's timeframe. The structured checklist provides an intermediate formalization, ensuring normative traceability without complicating the experimental architecture.

3.4 Experimental Evaluation Protocol

This section presents the experimental protocol implemented to evaluate the functionality and relevance of the automated document audit application developed within the framework of this work. The evaluation is based on an applied experimental approach, designed to analyze the application's behavior in response to various document situations

representative of real-world QSE audit contexts. The protocol was designed with methodological rigor and reproducibility in mind, while taking into account the lack of a specific normative framework dedicated to the evaluation of document audit tools based on artificial intelligence.

3.4.1 General objectives of the evaluation

The overall objective of the experimental evaluation is to assess the application's ability to produce consistent, reliable, and usable results within the context of a QSE (Quality, Safety, and Environment) document audit. It aims to evaluate whether the tool can provide relevant support for document analysis without replacing the auditor's professional judgment. More specifically, the evaluation aims to:

- Analyze the temporal reproducibility of the results produced by the application;
- Evaluate the robustness of the system in the face of variations in the form and structure of documents;
- Measure the application's sensitivity to detecting obvious non-conformities;
- Verify the logical consistency of the verdicts in the event of targeted document improvement.

These objectives guided the definition and organization of the experimental tests presented in the following sections.

3.4.2 Experimental conditions

The experimental conditions described above constitute the common operational framework for all the evaluations carried out. They guarantee the homogeneity of the analyses and ensure methodological consistency between the different tests implemented. On this basis, the experimental tests were structured to examine each of the dimensions defined in the evaluation objectives separately.

3.4.2.1 Experimental Environment

The tests were conducted using a functional version of the automated document audit application, operating under controlled conditions. The application was run exclusively in automated analysis mode, without human intervention in the interpretation or

correction of results, to evaluate its intrinsic capabilities. Analysis parameters were kept constant throughout all tests, ensuring the comparability of the results.

3.4.2.2 Documentary corpus

The corpus of documents used in the experimental evaluation consists of twelve online QSE/OHS documents. The selection was made to cover a diversity of document configurations representative of various organizational contexts. The selected documents exhibit notable differences in terms of:

- Editorial structure and level of formalization;
- Degree of completeness of information;
- Dominant normative orientation (quality, environment or occupational health and safety).

This diversity aims to replicate conditions similar to those encountered during a real-world document audit, characterized by frequent heterogeneity in the quality, clarity, and structure of the documents analyzed. To ensure consistency between the studied corpus and the document typology derived from the audit checklist, the documents were renamed according to the categories selected for the experimental analysis. The following table presents the correspondence between the original titles and their designation within the study.

Table 1 - Correspondence between the original titles of the documents and their designation in the checklist

Document No.	**Original title of the document**	**Designation used in the checklist**
D1	Risk Management – Continuous Improvement	Continuous Improvement Plan
D2	Environmental Emergency Response Management Plan for Northern Territory Mining Operations Pty Ltd	Environmental Emergency Plan
D3	Pilot Group Environmental Manual	Environmental Manual
D4	NQA Management Systems Surveillance (Remote) Process Audit Report	Internal Audit Reports (ISO 9001)
D5	Procedure for the Management of Nonconformities and Corrective Actions	Management Procedure – Nonconformities and Corrective Actions

Document No.	Original title of the document	Designation used in the checklist
D6	Joint Occupational Health & Safety Committee Meeting Minutes	Minutes of Occupational Health and Safety Management Review
D7	Standard Operating Procedures for Occupational Health and Safety	Occupational Health and Safety Operational Procedures
D8	Inditex's Group Occupational Health and Safety Policy	Occupational Health and Safety Policy
D9	Rohini College of Engineering and Technology Organization Chart	Organization Chart
D10	Operational Planning and Control – Group Capital Division	Planning and Operational Control Procedure
D11	EHQMS System Manual – TM Steels	Quality Manual
D12	Risk and Opportunity Register – Master Sheet	Register of Environmental Risks and Opportunities

3.4.2.3 Verdict scale

In the application studied, the analysis is performed question by question, using a standardized checklist. Each question receives an independent verdict, without automatic aggregation at the document level. The unit of analysis used for all experimental tests is therefore the question, not the document. The verdicts assigned to each question are based on a three-level ordinal scale:

- 0: Non-compliant (NC)
- 1: Partially compliant (PC)
- 2: Compliant (C)

This scale allows for a graded assessment of conformity, in line with document audit practices.

3.4.3 Detailed description and justification of experimental tests

The experimental evaluation of the application is based on four complementary tests designed to examine:

- The temporal stability of verdicts,

- Robustness in the face of variations in form without semantic change,
- The ability to detect expected non-conformities,
- The consistency of verdicts after targeted documentary improvement.

In the absence of a specific normative framework dedicated to evaluating document auditing tools based on artificial intelligence, the experimental protocol was structured using established metrics from measurement science (random-corrected agreement coefficients), classification system evaluation (recall, precision, F1 score), and nonparametric statistics for paired comparisons (Wilcoxon signed-rank test). This approach grounds the evaluation in recognized methodological foundations comparable to those used in the evaluation of automated systems. In accordance with general recommendations for evaluating AI systems, the approach combines reliability indicators and classification performance metrics to produce comparable, reproducible, and interpretable results within an academic context.

The unit of analysis used is the question (and not the document), with each question receiving an ordinal verdict (0 = NC; 1 = PC; 2 = C). The selected indicators are based on coefficients recognized in measurement science and systems evaluation, including chance-corrected agreement coefficients (κ and α) and standard performance metrics (recall, precision, F1).

3.4.3.1 Test 1 – Reproducibility of results (temporal reliability)

This test aims to evaluate the stability of the verdicts produced by the application when the same document corpus is analyzed at two distinct times, under strictly identical conditions. It follows a test-retest reliability approach, commonly used in psychometrics and the evaluation of measurement instruments to assess the consistency of an analytical system over time. The objective is to determine whether the tool provides consistent decisions regardless of temporal fluctuations or potential internal variations in the model.

A. Methodological Foundation

The verdicts assigned by the application are expressed on a three-level ordinal scale:

- 0: Non-compliant (NC)

- 1: Partially compliant (PC)
- 2: Compliant (C)

Given the ordinal nature of this scale, reproducibility assessment cannot be limited to a simple crude agreement rate. An indicator that takes into account:

- From the agreement expected by chance;
- The distance between ordinal categories.

B. Experimental Protocol

1. Creation of a corpus composed of 12 QSE/SST documents.
2. First automated analysis of the corpus (Run 1) and recording, for each activated question, of the assigned verdict (0/1/2).
3. New analysis of the same corpus (Run 2), after a delay of between 48 hours and 7 days.
4. Strict maintenance of experimental conditions :
    - No changes to the documents.
    - No adjustments to the analysis parameters were made.
    - Same version of the model.
5. Question-by-question matching of the verdicts obtained during the two analyses.

The unit of analysis corresponds to each question activated in the checklist.

C. Main indicator: Cohen's weighted κ coefficient

Reproducibility is measured using Cohen's weighted kappa coefficient.

Unlike simple κ, weighted κ allows:

- To correct the expected agreement by chance;
- To integrate the ordinal structure of the data;
- To weigh disagreements differently depending on their magnitude.

Thus, a disagreement between Non-compliant and Partially compliant is considered less serious than a disagreement between Non-compliant and Compliant.

The coefficient is defined by:

$$\kappa_w = \frac{P_o - P_e}{1 - P_e}$$

Or :

- Po: proportion of observed weighted agreement
- Pe: proportion of weighted agreement expected by chance

Quadratic weightings were chosen in order to increase the penalty for extreme disagreements.

The value of κ varies between −1 and 1:

- κ = 1: perfect agreement
- κ = 0: agreement equivalent to random
- κ < 0: agreement less than chance

D. Additional descriptive indicator

For descriptive purposes, the gross agreement rate (Percent Agreement – PA) is also reported:

$$PA = \frac{\text{Number of identical verdicts}}{\text{Total number of verdicts compared}}$$

However, this rate is not used as a primary indicator because it does not correct for the effect of chance.

E. Interpretation

The interpretation of the κ values follows the grid conventionally used in methodological research:

Table 2 - Value of K and the lever of agreemment

me	Level of agreement
$\kappa < 0$	lower than random agreement
0.00–0.20	Light chord
0.21–0.40	Passable agreement
0.41–0.60	Moderate agreement
0.61–0.80	Substantive agreement
0.81–1.00	Near-perfect agreement

A high value of weighted κ suggests strong temporal stability of the application. However, this indicator should be interpreted taking into account:

- From the distribution of categories,
- Regarding the sample size,
- And the application context (automated document audit).

This initial test allows us to assess the intrinsic reliability of the system. High reproducibility is a necessary condition for the tool's credibility, but it is not sufficient to demonstrate its overall performance, which will be examined in subsequent tests focusing on robustness, sensitivity, and causal consistency.

3.4.3.2 Test 2 – Robustness to variations in document form

This test aims to evaluate the application's ability to maintain stable verdicts when the analyzed text undergoes formal modifications that do not alter its relevant normative content. The objective is to verify the system's semantic invariance, that is, its ability to base its decisions on the normative substance of the document rather than on stylistic, syntactic, or editorial variations. This test therefore examines the model's external robustness in the face of controlled textual perturbations.

A. Methodological Foundation

Natural language processing models can exhibit excessive sensitivity to superficial variations in text (reformulation, addition of irrelevant content, sentence restructuring). In the context of normative document auditing, such sensitivity would constitute a major methodological bias, likely to produce unstable verdicts without any real change to the substance.

Given:

- The presence of three evaluations per unit of analysis (O, R, N),
- An ordinal scale with three levels (0 = NC, 1 = PC, 2 = C),

An indicator suitable for multiple comparisons and ordinal data is required.

B. Experimental Protocol

1. Selection of five predominantly narrative documents (D3, D6, D7, D11, D12), deemed suitable for reformulations without normative alteration.
2. Generating three versions for each document:
    - **O**:original version
    - **R**: revised version (≈ 20–30% of the content modified, without normative changes)
    - **N**: original version enriched with a neutral paragraph without normative scope
3. Manual verification that versions R and N do not modify:
    - The regulatory requirements,
    - Responsibilities,
    - The commitments,
    - Nor the evidence.
4. Automated analysis of the three versions under strictly identical conditions (same model, same parameters).
5. Question-by-question matching of the verdicts attributed to the three versions.

The unit of analysis corresponds to each question activated in the checklist.

C. Main indicator: Krippendorff ordinal alpha (α)

Robustness is measured using Krippendorff's alpha for ordinal data.

This indicator has several advantages:

- Applicable to more than two assessments;
- Adapted to ordinal data;
- Robust against unbalanced distributions;

- Based on an explicit measure of disagreement.

Alpha is defined as:

$$\alpha = 1 - \frac{D_o}{D_e}$$

Or :

- Do: disagreement observed
- From: disagreement expected by chance

The value of α varies between −1 and 1:

- α = 1: perfect agreement
- α = 0: agreement equivalent to random
- α < 0: systematic disagreement

D. Calculation of the observed disagreement (Do)

For each question, the three versions (O, R, N) generate three possible comparisons:

- O–R
- O–N
- R–N

The disagreement is calculated using an ordinal quadratic distance:

$$\delta_{ij} = (v_i - v_j)^2$$

Or :

- Vi and vj correspond to the numerical values of the verdicts (0, 1, 2).

This weighting allows for:

- To further penalize extreme deviations (NC ↔ C),
- To consider deviations of a certain level as less serious.

The observed disagreement corresponds to the weighted average of distances across all comparisons.

E. Calculation of the expected disagreement (De)

The expected disagreement is estimated from the overall distribution of the categories observed across all analyzed versions. It represents the level of disagreement theoretically expected if the verdicts were assigned randomly according to this marginal distribution.

F. Additional descriptive indicators

In addition to alpha, three descriptive indicators are reported:

- Percentage of perfect stability (O = R = N)
- Proportion of variations limited to a single level
- Proportion of extreme variations (difference of two levels)

These indicators make it possible to characterize the nature of the divergences and to identify any critical instabilities.

J. Interpretation

The interpretation of α values follows the thresholds generally accepted in reliability analysis:

Table 3 - value of alpha and the level of agreement

Value of α	Level of agreement
$\alpha \geq 0.80$	High reliability
$0.67 \leq \alpha < 0.80$	Acceptable reliability
$\alpha < 0.67$	Insufficient reliability

A high value of α suggests that formal changes do not substantially affect verdicts. However, statistical stability does not necessarily guarantee normative validity; it only indicates resistance to controlled textual variation.

3.4.3.3 Test 3 – Sensitivity to the detection of expected non-conformities

This test aims to evaluate the application's ability to correctly identify obvious, predefined documentary non-conformities and to avoid the erroneous assignment of "compliant"

verdicts when normative requirements are absent. It examines the system's discriminatory capacity in the face of intentionally identified documentary defects.

A. Methodological Foundation

In a normative audit context, the most critical risk is not the over-detection of non-conformities, but the issuance of false conformity verdicts, which may lead to incorrect decisions.

For this reason, the expected non-conformity is defined as a positive class in the performance analysis.

This methodological choice allows us to:

- To focus the evaluation on the system's ability to detect real deficiencies;
- To estimate the risk of false conforms (type II errors in this application context).

The analysis is based on the construction of a binary confusion matrix derived from the initial scale (0/1/2), grouping the categories in a way that is consistent with the objective of the test.

B. Experimental Protocol

1. Selection of three documents presenting documentary gaps identified as manifest (D2, D5, D10).
2. Prior definition, for each question concerned, of the expected verdict, based on an independent normative analysis.
3. Binary recoding for performance analysis:
    - Positive class: non-compliant with expectations
    - Negative class: another verdict
4. Automated document analysis.
5. Construction of a confusion matrix comparing the expected verdict and the obtained verdict.

The unit of analysis corresponds to each question for which an expected non-conformity has been defined.

C. Performance Indicators

The following indicators are calculated from the confusion matrix:

$$Recall = \frac{TP}{TP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Or :

- **TP**: non-conformities correctly detected
- **FN**: undetected non-conformities (false conformities)
- **FP**: Conformities incorrectly classified as non-compliant

The F1-score synthesizes the compromise between detection capability and alert accuracy.

D. Interpretation

- A low recall indicates a high number of false positives (FP), which constitutes a critical audit risk.
- Low accuracy indicates a tendency towards over-detection (high FP).
- A high F1 score indicates a satisfactory balance between sensitivity and accuracy.

The analysis favors a comparative and contextual interpretation rather than the application of fixed thresholds, taking into account:

- Due to the limited sample size,
- The exploratory nature of the test,
- And the specific nature of the normative domain.

3.4.3.4 Test 4 – Consistency before/after document correction

The test aims to verify that the targeted addition of explicit normative elements in a document leads to a consistent improvement in the verdicts assigned by the application.

The added elements include, in particular:

- The appointment of a person in charge,
- The description of an action,
- Setting a deadline,
- The identification of evidence or a recording.

The objective is to assess the functional causal validity of the system: a substantial normative change must produce a consistent adjustment of the verdict.

A. Methodological Foundation

Unlike Tests 1 and 2, which assess system stability, this test examines its responsiveness to a real and controlled improvement in document content.

It is part of an experimental logic of the pre-test / post-test type on paired units.

Since the verdicts are expressed on an ordinal scale (0 = NC, 1 = PC, 2 = C), the analysis requires a test:

- Adapted to ordinal data,
- Applicable to paired observations,
- Not assuming normality.

B. Experimental Protocol

1. Selection of two documents presenting initial documentary gaps (D5 and D12).
2. Initial automated analysis and recording of verdicts (the “before” phase).
3. Targeted document modification consisting of the addition of explicit normative elements.
4. Manual verification that the changes only concern the targeted requirements.
5. New automated analysis under strictly identical conditions.

6. Question-by-question matching of before/after verdicts, limited to questions directly impacted by the changes.

The unit of analysis corresponds to each question addressed by the documentary intervention.

C. Statistical test : Wilcoxon signed-rank

The analysis is based on the non-parametric Wilcoxon signed-rank test.

This test is appropriate because :

- The data is ordinal ;
- The observations are matched ;
- The distribution of differences is not assumed to be normal.

The test examines whether the median difference between the verdicts before and after modification is significantly greater than zero.

D. Statistical Hypotheses

- H_0: no significant difference (median difference = 0)
- H_1: significant improvement (median difference > 0)

The test is one-sided, as only improvement is expected theoretically.

E. Reported Indicators

The following elements are presented :

- Value of the Wilcoxon statistic,
- Associated p-value,
- Effect size (r), calculated when the sample size allows,
- Descriptive percentage of questions :
    - Improved,
    - Unchanged
    - Degraded.

The effect size is calculated according to:

$$r = \frac{Z}{\sqrt{N}}$$

Or :

- Z corresponds to the normalized statistic of the test,
- N corresponds to the number of paired observations.

F. Interpretation

- $p < 0.05$: statistically significant improvement in verdicts after documentary correction.
- $p \geq \mathbf{0.05}$: absence of a statistically demonstrated effect, requiring further discussion.

However, the interpretation is not based solely on statistical significance. The distribution of observed changes and effect size are also taken into account.

A statistically significant improvement suggests that the system adjusts its verdicts consistently in response to an explicit normative correction.

3.4.4 Statistical Computation Procedure and Technical Assistance

The statistical evaluation of the experimental tests was conducted through a structured procedure ensuring methodological transparency and separation between analytical design and computational execution. The selection of the statistical indicators (weighted Cohen's κ, Krippendorff's α, recall, precision, F1-score, and Wilcoxon signed-rank test) was based on an independent review of methodological literature concerning measurement reliability and classification system evaluation. The automated application developed in this study does not perform statistical calculations. It generates ordinal verdicts (0 = NC, 1 = PC, 2 = C) for each activated checklist question. These results were exported in tabular format and constituted the raw dataset used for statistical analysis. The computation of the statistical coefficients was carried out separately using computational tools, with technical assistance from a computer engineer to ensure the correct implementation of the formulas and the appropriate parameterization of the

statistical procedures. All formulas, parameters, and data structures used for the calculations are explicitly described in this methodology section, ensuring traceability and reproducibility by researchers possessing equivalent statistical and computational competencies.

3.5 Documentary Case Study

Following the experimental validation phase of the application, a documentary case study was conducted to examine its behavior in a real-world environment. This step aimed to compare the tool with institutional documents actually published by organizations, in order to observe its operation in a practical context and to assess its ability to process heterogeneous, unstructured content specifically designed for a normative audit.

3.5.1 Objective of the case study

The case study aims to analyze the behavior of the automated audit application when applied to real documents from companies with publicly available documentation. Unlike the previously presented experimental tests, this phase does not seek to validate the algorithm technically. The tests assessed the system's reproducibility, robustness, and logical consistency. The case study complements this, examining its operational applicability within a concrete organizational context. It is an applied, non-certification study. Its objective is not to evaluate the actual compliance of the studied companies with ISO standards, but rather to observe how the tool functions when faced with institutional documents written for organizational communication purposes, rather than solely for regulatory purposes.

3.5.2 Case selection and documentary scope

The selection of the cases studied was part of a reasoned and intentional approach, excluding any logic of random sampling. The objective was not to achieve statistical representativeness, but to create a relevant and diverse body of documentation, capable of testing the application in different organizational contexts and assessing its robustness in the face of varied sectoral configurations.

Four international companies were selected: Jan De Nul, a Belgian group specializing in maritime works, dredging, and offshore infrastructure; Mota-Engil, a major player in the international construction and infrastructure sector; TDF, a French operator of broadcasting and telecommunications infrastructure; and Total Energies, an integrated energy group operating in the oil, gas, renewable energy, and electricity sectors. These organizations have formalized governance structures and documented management systems, making them suitable for in-depth normative analysis.

The choice of these companies is based on several cumulative criteria: the availability of publicly accessible institutional documents; the existence of policies, reports and procedures explicitly linked to Quality, Health, Safety and Environment (QHSE) management systems; as well as sectoral diversity (maritime infrastructure, construction, telecommunications, energy), allowing the application to be exposed to distinct organizational environments in terms of risks, regulatory issues and interested parties.

The scope of the documentation was strictly defined. Only documents with structured content explicitly linked to ISO 9001, ISO 14001, or ISO 45001 standards were included. Documents with an exclusively promotional or financial purpose, or lacking usable organizational content in light of the standards requirements, were excluded. This approach ensures the methodological consistency of the corpus while maintaining sufficient diversity to assess the applicability and transferability of the analysis tool in heterogeneous document configurations.

3.5.3 Association “Company Document – Document Checklist”

Each company document selected for the case study was associated with a single type of auditable document from the application's integrated checklist. This methodological decision aims to maintain analytical consistency and avoid any artificial overlap of regulatory requirements. The association is based on an analysis of the document's primary purpose. Thus, depending on whether it is a policy, an action plan, an operational procedure, a contextual analysis, or a document similar to a register, the document was linked to the most relevant corresponding checklist item.

It is important to note that no document was associated with multiple categories of the checklist simultaneously. This choice ensures a clear unity of analysis and avoids an expansive interpretation that could compromise the methodological rigor of the study.

The following tables present, for each of the companies studied, the correspondences retained.

Table 4 - Correspondence between company documents and audit checklist categories

Business	Company document	Exact match in the checklist	Reference framework
Jan De Nul	Code of Conduct	Register of interested parties	ISO 9001 – Context of the organization
Jan De Nul	QHSSE Policy Statement	Environmental policy	ISO 14001 – Leadership
Jan De Nul	Energy Action Plan & Report	Environmental Action Plan	ISO 14001 – Planning
Jan De Nul	Annual Report	Environmental context analysis (global report)	ISO 14001 – Context of the organization
Mota-Engil	Code of Ethics and Professional Conduct	Register of Interested Parties	ISO 9001 – Context of the organization
Mota-Engil	Social Responsibility Policy	Register of Interested Parties	ISO 45001 – Context of the organization
Mota-Engil	SHEQ Policy (11/21/2023 – Rev 00)	Occupational Health and Safety Policy	ISO 45001 – Leadership
TDF	Hygiene, Safety and Environment Charter (2021)	Organization Chart, Job Descriptions, or Occupational Health and Safety Responsibility Matrix	ISO 45001 – Leadership
TDF	Health, Safety and Environment Policy (2021)	Occupational Health and Safety Policy	ISO 45001 – Leadership
TDF	TDF Group Code of Ethics	Register of Interested Parties	ISO 9001 – Context of the organization
TDF	Corporate Social Responsibility Policy	Environmental context analysis (global report)	ISO 14001 – Context of the organization
TotalEnergies	Safety, Environment and Quality Charter	Organization Chart, Job Descriptions, or Occupational Health and Safety Responsibility Matrix	ISO 45001 – Leadership
TotalEnergies	The 12 Golden Rules	Occupational Health and Safety Operational Procedures	ISO 45001 – Operation

Business	Company document	Exact match in the checklist	Reference framework
TotalEnergies	Code of Conduct	Register of Interested Parties	ISO 9001 – Context of the organization
TotalEnergies	Sustainability & Climate 2024 Progress Report	Environmental context analysis (global report)	ISO 14001 – Context of the organization

3.5.4 Automated analysis procedure

The document analysis was conducted according to a standardized protocol, ensuring the reproducibility of the approach and the consistency of the processing of the cases studied. First, the corresponding document type was selected within the application, based on its association with the checklist. This selection enabled the automatic activation of questions related to the chosen document category and the associated normative framework. The company document was then imported into the analysis interface. The application then performed automated content processing, comparing the document text to the requirements formulated in the questions of the activated checklist. For each question, the tool assigned a verdict from among three categories: compliant, partially compliant, or non-compliant. This classification is based on the identification of textual elements deemed relevant to the normative requirement being examined.

In addition to the verdict, the application generates a concise explanatory sentence justifying the decision. This statement aims to clarify the reasoning behind the assigned classification, establishing a direct link between the checklist requirement and the content actually detected in the analyzed document. Finally, the text excerpts considered as evidence are presented to ensure the traceability of the analysis and allow for human verification of the results.

3.5.5 Structuring the results

The results generated by the application were systematically recorded for each document analyzed. The results were presented in structured tables, question by question. For each activated question, the table displays the wording of the requirement from the checklist, the assigned verdict (compliant, partially compliant, or non-compliant), and the explanatory sentence generated by the application. This structure allows for a consistent analytical reading of the results and facilitates comparative analysis between the different

cases studied. A cross-sectional summary was then produced to identify general trends observed across the analyzed companies, without conducting a comprehensive normative assessment of their compliance.

3.6 Methodological limitations

The study has several limitations related to the nature of the document corpus and its scope. The documents analyzed are exclusively from public sources and do not necessarily correspond to the internal documents used by organizations within their certified management systems. Consequently, certain elements required by ISO standards, such as formal registers, detailed evaluation matrices, or internal operational plans, are not accessible within the scope of this research. Furthermore, the application studied is a tool to assist in document analysis. It does not replace a certification audit conducted by a qualified auditor and does not incorporate the on-site observation, interview, or contextual verification dimensions inherent in actual audits. Finally, the scope of the case study is intentionally limited to a small number of companies and documents, in accordance with a thorough, academic-level qualitative approach.

4 Results

4.1 Presentation of analytical tools

The document analysis conducted in this study relies on two complementary instruments: (i) a structured, standardized checklist and (ii) an analysis platform enabling the systematic processing of company documents according to selected requirements. The integration of these tools aims to ensure a consistent, traceable, and reproducible assessment of document content, thereby limiting the variability in interpretation typically associated with manual auditing. The checklist provides the reference framework: it formalizes, in the form of verifiable questions, the expectations from the ISO 9001, ISO 14001 and ISO 45001 standards. The platform, for its part, operationalizes this framework by supporting the analysis process (selection of the relevant checklist, reading of the document, assignment of a verdict and generation of a textual justification), which then allows a standardized presentation of the results in the form of tables. In this chapter, the objective is not to detail the design of the tool (developed in the Materials and Methods section), but to present the instruments as they are used to produce the verdicts reported in the following sections.

4.1.1 Structure of the assessment checklist (Appendix A)

The checklist used is built on a multi-reference framework integrating three management systems: quality (ISO 9001), environment (ISO 14001) and occupational health and safety (ISO 45001). It is organized by major clauses (A to G) in accordance with the architecture of the standards, covering in particular: organizational context, leadership, planning, support, operational activities/performance, performance evaluation and improvement.

Within each clause, the checklist is structured by families of expected documents (e.g., SWOT analysis, PESTEL study, register of interested parties, management review minutes, policy, action plan, audit reports, non-conformities and corrective action procedure, etc.). For each document family, a series of operational questions verifies the presence of explicitly required normative elements, such as:

- Document identification (availability, dating, versioning/traceability);

- Required content (presence of expected sections: internal/external issues, interested parties, obligations, objectives, indicators, etc.);
- Implementation methods (methodology described, sources/consultations, scope, update frequency);
- Management dimension (responsible/owner, actions, deadlines, follow-up, evidence/record).

The evaluation from the checklist results, for each activated question, in a three-level ordinal verdict (Compliant, Partially Compliant, Non-Compliant), accompanied by a textual justification generated during the analysis. This structure allows for the aggregation of results by document, clause, and reference framework, while still maintaining a detailed, question-by-question analysis.

To preserve the readability of the results chapter, the complete checklist (all sections, document types and associated questions for ISO 9001, ISO 14001 and ISO 45001) is presented in Annex 1. In this chapter, it is used as a reference tool for the production and organization of results tables.

4.1.2 Presentation of the analysis platform (Appendix B)

The production of the results presented in this chapter relies on the use of a dedicated automated document analysis application, developed within the framework of this research. This platform provides the operational support for the experiment: it enables the activation of standardized checklists, the import of organizational documents, the execution of automated processing, and the structured delivery of results. The principles of technical architecture and methodological choices relating to natural language processing were presented in the Materials and Methods section. This section is limited to describing the interface as it is used to generate the results presented in Chapter 4.

A. General interface architecture and usage sequence

The platform's interface is organized around a centralized workspace that allows for the analysis of one document at a time. It includes a side panel dedicated to selecting normative modules, a main area displaying activated questions and generated verdicts, and a top bar grouping functions for managing files, settings, and exporting results.

The ISO 9001:2015, ISO 14001:2015, and ISO 45001:2018 standards are structured by clauses and document types. The user selects the module corresponding to the document being analyzed (for example: Context Analysis, Register of Interested Parties, Occupational Health and Safety Policy, etc.), which determines the activation of the relevant normative questions.

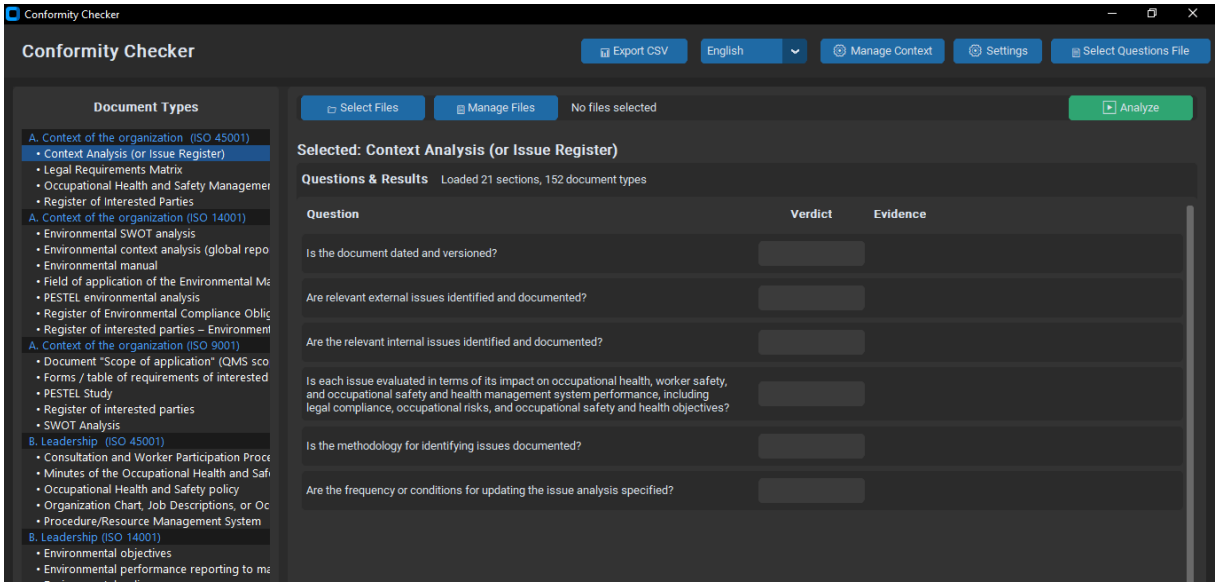


Figure 4 - Main interface of the document analysis platform

B. Operational analysis sequence

The platform presents the results in an analytical table organized according to the structure of the activated checklist. The display follows a question-by-question logic, ensuring a comprehensive overview of the assessed requirements.

For each item, the interface systematically displays:

- The question from the checklist, corresponding to the operational formalization of the normative requirement;
- The verdict awarded, expressed according to the ordinal scale used in the study (Compliant, Partially compliant, Non-compliant);
- An explanatory sentencejustifying the evaluation in light of the analyzed content;
- An obvious fact, corresponding to a textual extract from the evaluated document, allowing precise identification of the passage on which the verdict was based.

This standardized reporting—question, verdict, justification, and evidence—ensures the comparability of results between documents and between companies. It also allows for subsequent statistical analysis of the verdicts, while maintaining an explicitly documented qualitative basis.The complete interface of the analysis platform, including the various steps of checklist selection, document import and results display, is presented in detail in Appendix B.

C. Management of normative frameworks

The platform includes a normative repository management module that allows for the import and selective activation of ISO texts used as an analytical context. Applicable standards can be activated or deactivated depending on the scope of the evaluation. This feature ensures the consistency of the interpretative framework used when processing documents. It also contributes to the traceability of the experimental setup by explicitly identifying the reference framework used for each analysis.

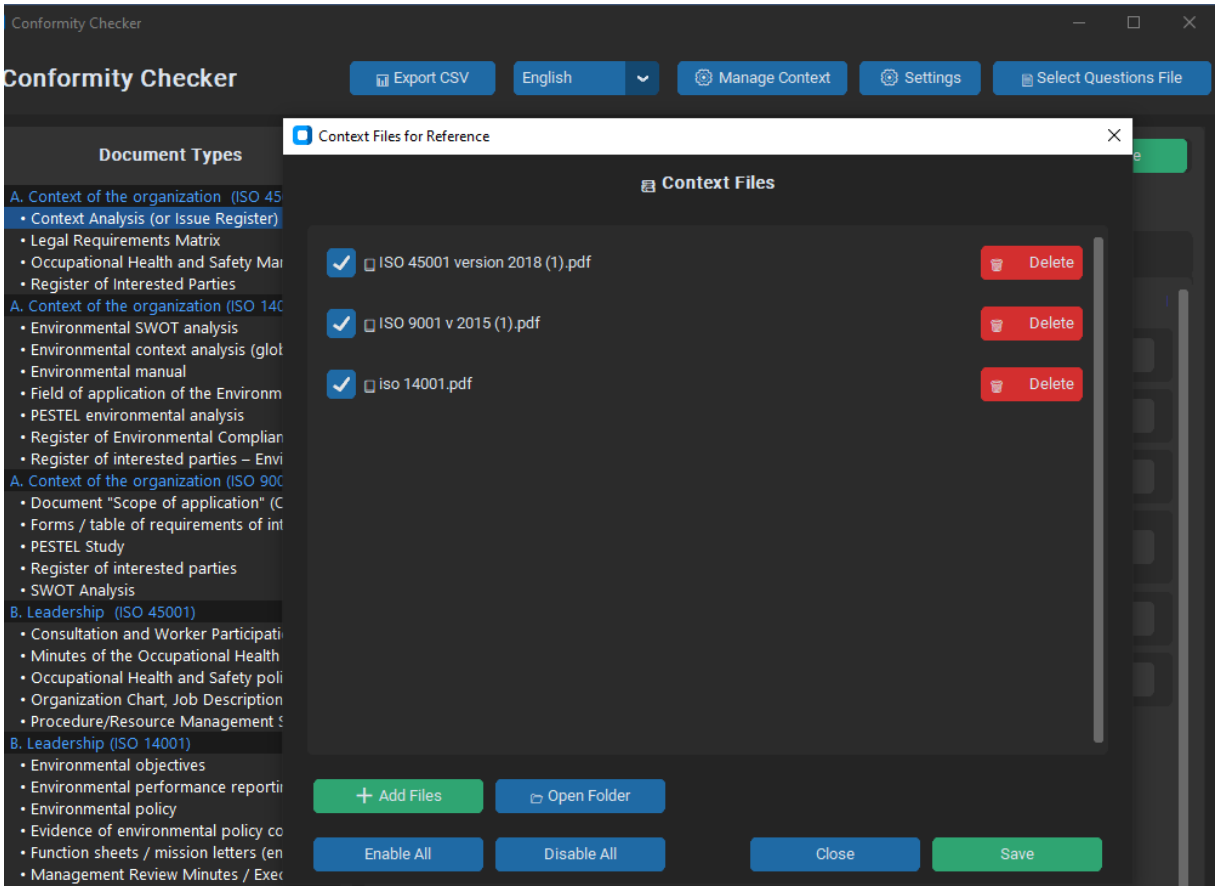


Figure 5 - Context file management interface (ISO repositories activation).

D. Technical configuration of the analysis environment

A configuration module allows the definition of execution parameters, including the selection of the language model and the registration of API parameters. This step ensures the stability and reproducibility of the analyses performed within the experimental framework.

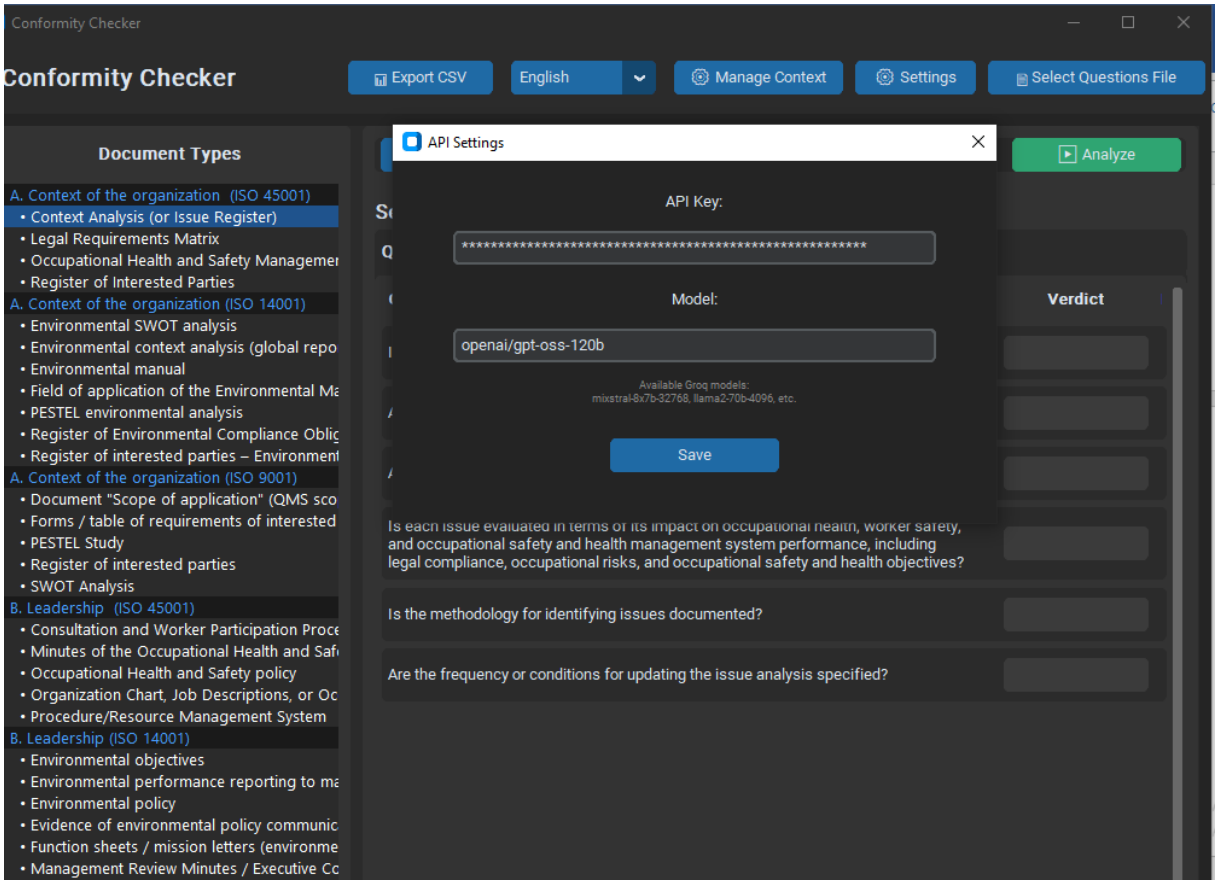


Figure 6 - Technical parameterization module (model configuration and execution environment).

4.2 Results of experimental tests

This section presents in a structured manner the results obtained from the four experimental tests conducted to evaluate the reliability, robustness and decision-making consistency of the automated document analysis application.

Each test examines a specific dimension of the tool's operation:

- Test 1: temporal stability of verdicts;

- Test 2: robustness in the face of formal variations in the text;
- Test 3: ability to detect expected non-conformities;
- Test 4: consistency of adjustments following targeted document correction.

All assessments are based on the checklist integrated into the application, the full version of which is presented in Appendix A. This checklist constitutes the analytical framework used for assigning verdicts.

The results are presented using a dual approach, both descriptive and statistical. A detailed interpretation and critical analysis of the findings will be the subject of a separate discussion chapter. The verdicts are coded as follows: 0 = Non-compliant; 1 = Partially compliant; 2 = Compliant.

4.2.1 Test 1 – Reproducibility of results

The reproducibility test covered the entire document corpus, representing a total of 88 questions activated across the 12 documents analyzed. The complete table of results obtained during Run 1 and Run 2 is presented in Appendix C.

A. Gross agreement rate

A comparison of the verdicts awarded in the two successive analyses shows:

- 77 identical verdicts,
- 11 discrepancies,
- No direct extreme shift between the categories "Non-compliant" (0) and "Compliant" (2).

The gross agreement rate (Percent Agreement) is therefore:

$$PA = \frac{77}{88} = 0{,}875$$

This represents an observed agreement rate of 87.5%.

B. Marginal distributions

The distributions of the categories for each analysis are as follows:

Table 5 - Marginal Distributions

Category	Run 1	Run 2
NC (0)	21	23
PC (1)	21	18
C (2)	46	47
Total	88	88

The proportions observed are relatively close between the two executions, reflecting an overall stability of the classificatory trends.

C. Weighted Cohen's coefficient κ

Given the ordinal nature of the scale (0/1/2), the quadratic weighted Cohen's coefficient was calculated.

The values obtained are as follows:

- Weighted observed agreement: $P_{o,w}=0.96875$
- Weighted expected agreement: $P_{e,w}=0.64954$

The weighted coefficient κ is therefore:

$\kappa_w=0.911$

D. Level of agreement

According to the usual interpretation framework:

- 0.61–0.80 → substantial agreement
- 0.81–1.00 → near perfect agreement

The value obtained ($\kappa = 0.911$) corresponds to a near-perfect level of agreement.

E. Nature of the observed discrepancies

The qualitative examination of the 11 discrepancies highlights:

- Limited adjustments of a single level (NC ↔ PC or PC ↔ C),
- The absence of a direct switch between extreme categories (NC ↔ C),
- No systematic instability on any particular document.

The observed differences therefore remain moderate on the ordinal scale.

F. Summary of Test 1 Results

The reproducibility test highlights:

- A high gross agreement rate (87.5%),
- A weighted κ coefficient indicating near-perfect agreement (0.911),
- A stable marginal distribution between the two executions,
- Limited and not extreme differences.

These results demonstrate a high temporal stability of the verdicts produced by the application under the experimental conditions considered.

4.2.2 Test 2 – Robustness to variations in document form

The robustness test aims to assess the stability of verdicts when the normative content of a document remains unchanged but its wording is modified, either by partial reformulation or by adding a neutral paragraph devoid of normative scope.The detailed results of this test are presented in Appendix D.

The analysis covered 5 documents, representing a total of 39 activated questions, each evaluated according to three versions:

- O: original version
- R: reformulated version
- N: enhanced version of a neutral text

Each question thus has three verdicts on a three-level ordinal scale (0 = NC; 1 = PC; 2 = C), 117 judgments in total.

A. Preliminary descriptive analysis

Of the 39 questions analyzed:

- 24 questions (61.5%) show perfect stability(O = R = N);
- 12 questions (30.8%) show variations of only one level(for example PC → C or NC → PC);
- 3 questions (7.7%) show an extreme variation of two levels(simultaneous presence of categories 0 and 2).

Thus, 92.3% of the questions show a maximum variation of only one level, which indicates an overall high stability on the ordinal scale.

B. Calculation of Krippendorff's alpha (ordinal data)

In order to assess robustness in a statistically rigorous manner, inter-evaluation reliability was measured using Krippendorff's alpha (α), adapted to ordinal data and multiple comparisons.

Alpha is defined by:

$$\alpha = 1 - \frac{D_o}{D_e}$$

Or :

- Do represents the observed disagreement.
- De represents the disagreement expected by chance.

Observed disagreement (Do)

For each question, three comparisons are possible (O–R, O–N, R–N), namely:

39 × 3 = 117 pairs

The distance used is quadratic ordinal:

$$\delta_{ij} = (v_i - v_j)^2$$

Distribution of observed distances:

- 88 pairs with distance 0
- 24 pairs with distance 1
- 5 pairs with a distance of 4

The average observed disagreement is therefore:

$$D_o = \frac{(24 \times 1) + (5 \times 4)}{117} = \frac{44}{117} = 0{,}376$$

Expected disagreement (De)

The overall distribution of the 117 judgments is as follows:

- 0 (NC): 40 occurrences
- 1 (PC): 28 occurrences
- 2 (C): 49 occurrences

From these proportions and applying Krippendorff's standard formula for ordinal data with root mean square distance, the expected disagreement is: De = 1.346

C. Final calculation of alpha

$$\alpha = 1 - \frac{0{,}376}{1{,}346} = 1 - 0{,}279 = 0{,}721$$

The value obtained is therefore: $\alpha \approx 0.72$

D. Reliability level

According to generally accepted thresholds:

- $\alpha \geq 0.80$: high reliability
- $0.67 \leq \alpha < 0.80$: acceptable reliability
- $\alpha < 0.67$: insufficient reliability

The value obtained ($\alpha \approx 0.72$) corresponds to an acceptable level of reliability.

E. Summary of Test 2 Results

The robustness test highlights:

- A majority of perfect stability (61.5%),
- A large majority of variations were limited to a single level (92.3% of cases),
- The observed disagreement is small compared to the disagreement expected by chance.
- A Krippendorff alpha indicating acceptable reliability (0.72).

These results reflect an overall stability of the verdicts in the face of controlled formal variations in the text, with however a measurable sensitivity to editorial changes.

4.2.3 Test 3 – Sensitivity to the detection of expected non-conformities

Test 3 aimed to evaluate the application's ability to detect obvious nonconformities defined a priori, as well as to avoid the erroneous assignment of "compliant" verdicts when normative requirements were absent. The analysis covered 25 questions from documents D2, D5, and D10. Among these, 3 nonconformities were expected a priori and constituted the positive category. The detailed table of verdicts is presented in Appendix E.

A. Confusion Matrix

Considering the category "Non-compliant" (0) as a positive detection and 1 and 2 grouped together as a negative class, the resulting confusion matrix is as follows:

- TP (True Positives)= 2
- FN (False Negatives)= 1
- FP (False Positives)= 2
- TN (True Negatives)= 20

B. Performance Indicators

The results obtained are:

- Recall = 0.67 (66.7%)
- Precision = 0.50 (50%)
- F1 score = 0.57

C. Summary of Test 3 Results

The system:

- Detects two-thirds of expected non-conformities,
- Generates two false positives out of 22 negative cases.
- It presents a moderate overall compromise between detection and accuracy.

These results reflect a partial ability to detect obvious non-conformities under the chosen experimental conditions.

4.2.4 Test 4 – Logical consistency before/after document correction

This test aims to assess whether targeted document correction (adding explicit normative elements such as responsible party, action, deadline, and evidence) leads to a consistent improvement in verdicts for the directly affected questions. In accordance with the methodological protocol, the statistical analysis focuses exclusively on questions explicitly linked to a corrected nonconformity (LINK TO NON-CONFORMITY = yes). Non-targeted questions are presented for descriptive purposes but are not included in the statistical analysis. The detailed table of verdicts before and after correction is presented in Appendix F.

A. Relevant questions

In total, 8 questions were identified as being directly impacted by the document corrections:

- D5: Q2, Q7, Q10
- D12: Q1, Q2, Q4, Q5, Q9

For each of these questions:

- Initial verdict (v1) = 0 (Not compliant)
- Verdict after correction (v2) = 2 (Compliant)
- Difference :

$di = v2 - v1 = +2$

B. Statistical analysis – Wilcoxon test (one-tailed)

Since the verdicts are expressed on an ordinal scale and compared before/after on the same units, the comparison is based on the Wilcoxon signed-rank test.

Hypotheses:

- H_0: median of differences = 0
- H_1: median of differences > 0

All observed differences being strictly positive and identical:

di = + 2 for i = 1..8

In the event of a tie, ranks 1 to 8 are replaced by their average rank (4.5).

The sum of the positive ranks is therefore: W+ = 36

The sum of the negative ranks is: W− = 0

C. exact p-value

In the case where all differences are oriented in the same direction (n = 8), the exact probability under the null hypothesis is:

$$p = \left(\frac{1}{2}\right)^8 = 0{,}0039$$

This value is below the significance threshold of 0.05.

D. Test result

The null hypothesis is rejected. The results indicate a statistically significant improvement in verdicts after targeted document correction.

E. Summary of Test 4 Results

The test reveals:

- A systematic improvement of the 8 targeted questions,

- The complete absence of degradation,
- Perfect directional consistency of adjustments,
- Strong statistical significance ($p \approx 0.0039$).

These results reflect a consistent responsiveness of the system to the normatively relevant corrections introduced.

4.3 Results of the documentary case study

This section presents the results obtained from applying the automated analysis tool to real-world business documents. The objective is to evaluate the application's ability to interpret an existing document and produce a structured verdict in accordance with the relevant normative requirements.

4.3.1 Case 1 – Jan De Nul

The case study of Jan De Nul is based on the automated analysis of four separate organizational documents. Each document was evaluated against a pre-activated, standardized checklist. This section presents only the results generated by the application.

A. Document 1 – Code of Conduct

The first document analyzed for the company Jan de Nul was the Code of Conduct. The evaluation was conducted using the "Register of Interested Parties" checklist (ISO 9001 – Organizational Context), activated to examine the identification of relevant interested parties, the consideration of their requirements, and the consistency of their integration into the quality management system. A total of five questions were used and automatically analyzed by the application.

Table 6 – doc 1 : Jan de Nul

No.	Checklist question	Verdict	Concise explanatory sentence for the application
1	Is the register available and dated?	Non-compliant	The document discusses values and conduct but contains no reference to a formal register being available or dated.

No.	Checklist question	Verdict	Concise explanatory sentence for the application
2	Does the document list the categories of interested parties (customers, employees, suppliers, authorities, etc.)?	Partially compliant	Several stakeholder groups are mentioned, but they are not structured as a formal categorized register.
3	For each party listed, does the document indicate the requirements/expectations?	Compliant	The document specifies expectations and behavioral requirements applicable to the identified parties.
4	Does the document identify the source of each requirement (contract, law, customer request, etc.)?	Non-compliant	General obligations are mentioned without explicitly linking each requirement to a specific source.
5	Does the document specify the frequency/method of monitoring the requirements of interested parties (who monitors, when)?	Non-compliant	No responsible party, frequency, or monitoring method is formally defined.

B. Document 2 – QHSE Policy Statement

The second document concerns the QHSE Policy Statement. The analysis was carried out using the "Environmental Policy" checklist (ISO 14001 – Leadership), which assesses the environmental policy's compliance with the normative requirements, particularly regarding management commitment, pollution prevention, environmental protection, and continuous improvement. Nine questions were activated and evaluated within this framework.

Table 7 - doc 2 : Jan de Nul

No.	Checklist question	Verdict	Concise explanatory sentence for the application
1	Is the document identified as an official environmental policy, dated and versioned?	Compliant	The document includes a formal policy number, revision (12.00) and a revision date, satisfying identification and versioning requirements.
2	Is the policy appropriate to the purpose and context of the organization?	Compliant	The policy explicitly states its purpose and aligns with the organization's mission and governance.

No.	Checklist question	Verdict	Concise explanatory sentence for the application
3	Does the policy take into account the nature, size and environmental impacts of activities, products and services?	Partially compliant	The policy mentions general environmental commitments but does not explicitly detail the nature and size of impacts.
4	Does the policy provide a framework for setting environmental objectives?	Compliant	The policy states it serves as a guide and basis for establishing environmental objectives and processes.
5	Does it contain a commitment to protect the environment (eg, pollution prevention)?	Compliant	The document explicitly commits to preventing environmental nuisances and reducing pollution and climate impact.
6	Does it contain a commitment to comply with compliance obligations?	Compliant	The policy clearly states commitment to comply with applicable legal and regulatory requirements.
7	Does it contain a commitment to continuous improvement of the organization and environmental performance?	Compliant	The document explicitly refers to continuous improvement within governance and strategy.
8	Is the policy formalized in the form of documented information?	Compliant	The document is issued under formal document control with identification and revision tracking.
9	Is the policy written in an understandable and exploitable way?	Non-compliant	The excerpts do not explicitly demonstrate that the policy is structured as understandable and exploitable documented information.

C. Document 3 – Energy Management Action Plan

The third document examined is the Energy Management Action Plan. The assessment is based on the "Environmental Action Plan" checklist (ISO 14001 – Planning), designed to verify the formalization of environmental objectives, action planning, allocation of responsibilities, and the definition of monitoring indicators and deadlines. Eight questions were selected and submitted to automated analysis.

Table 8 - doc 3 : Jan de Nul

No.	Checklist question	Verdict	Concise explanatory sentence for the application
1	Is the document identified, dated, and versioned?	Compliant	The document control section provides a document number, revision (0.0), and revision date (12/2024).
2	Are actions defined to address significant environmental aspects?	Compliant	Structured actions derived from energy audits are described and monitored.
3	Are actions defined to deal with compliance obligations?	Compliant	Internal and external audits, updates, and management reviews are explicitly mentioned.
4	Are actions defined to address the identified risks and opportunities?	Partially compliant	Corrective actions are described, but they are not explicitly linked to a formal risk and opportunity register.
5	Are the actions integrated into the Environmental Management System processes or into the business processes?	Partially compliant	Actions are integrated into the CO_2 management system, but broader business process integration is not explicitly stated.
6	Does the document specify how the effectiveness of the actions will be evaluated?	Compliant	Effectiveness is evaluated through semi-annual updates, audits, and management review.
7	Are technological options taken into account?	Partially compliant	The document discusses introducing sustainable solutions but does not explicitly reference technological option analysis.
8	Are financial, operational, and commercial constraints considered?	Non-compliant	No explicit reference to financial, operational, or commercial constraints is provided.

D. Document 4 – Annual Report

The fourth document analyzed was the 2017 Annual Report. The review was conducted using the "Environmental Context Analysis (Global Report)" checklist (ISO 14001 – Organizational Context), designed to assess the consideration of internal and external issues, interested parties, and factors likely to influence environmental performance. Eight questions from this checklist were activated and evaluated by the application.

Table 9 - doc 4 : Jan de Nul

No.	Checklist question	Verdict	Concise explanatory sentence for the application
1	Is the document identified, dated, and versioned?	Non-compliant	The excerpts analyzed do not include explicit identification, dating, or version information.
2	Are the internal environmental issues clearly identified?	Non-compliant	The document does not explicitly list or structure internal environmental issues.
3	Are the external environmental issues clearly identified?	Compliant	External environmental problems such as polluted marine sediments and heavy metals are explicitly described.
4	Are the environmental conditions affected by the organization described?	Compliant	Activities such as dredging, soil depollution, and recycling are clearly presented.
5	Are the environmental conditions that are likely to affect the organization described?	Partially compliant	Some site-specific environmental conditions are mentioned without comprehensive organization-wide coverage.
6	Are the issues analyzed in relation to the expected results of the Environmental Management System?	Non-compliant	No explicit analysis linking environmental issues to expected EMS results is provided.
7	Is the method for context analysis explained?	Non-compliant	No formal method of context analysis is described in the exceptions.
8	Is a review or periodic update planned?	Non-compliant	The document does not explicitly state that a periodic review or update is planned.

E. Factual Summary of Case 1 – Jan De Nul

In total, the analysis of the four documents involved:

- 5 requirements (Document 1)
- 9 requirements (Document 2)
- 8 requirements (Document 3)
- 8 requirements (Document 4)

That makes 30 requirements evaluated.

The overall distribution of verdicts is as follows:

- Compliant: 14
- Partially compliant: 6
- Non-compliant: 10

These results correspond exactly to the verdicts generated by the application for the documents and checklists activated within the framework of this case study.

4.3.2 Case 2 – Mota-Engil

A. Document 1 – Code of Ethics and Professional Conduct

The first document analyzed for Mota-Engil was the Code of Ethics and Professional Conduct. The evaluation was carried out using the "Register of Interested Parties" checklist (ISO 9001 – Organizational Context), activated to examine the structure of interested parties, the identification of their requirements, the traceability of normative sources, and the associated monitoring procedures. A total of five questions from this checklist were used and automatically evaluated by the application.

Table 10 - doc 1 : Mota-Engil

Checklist question	**Verdict**	**Concise explanatory sentence for the application**
Is the register available and dated?	Non-compliant	The document mentions dates of adoption and revision of the code but does not present a formal, dated and structured register of interested parties.
Does the document list the categories of interested parties (customers, employees, suppliers, authorities, etc.)?	Compliant	The code explicitly lists different categories of stakeholders such as customers, employees, suppliers, authorities and other external entities.
For each party listed, does the document indicate the requirements/expectations?	Compliant	The document specifies the expectations and rules applicable to the different parties identified (suppliers, partners, third parties, etc.).
Does the document identify the source of each requirement (contract, law, customer request, etc.)?	Partially compliant	General references to laws and regulations are mentioned, without explicitly linking each requirement to a specific source.

Checklist question	Verdict	Concise explanatory sentence for the application
Does the document specify the frequency/method of monitoring the requirements of interested parties (who monitors, when)?	Non-compliant	No details are provided regarding the method, frequency, or responsibility for monitoring the requirements of interested parties.

B. Document 2 – Social Responsibility Policy

The second document analyzed for Mota-Engil was the Corporate Social Responsibility Policy. The assessment was conducted using the "Register of Interested Parties" checklist (ISO 45001 – Organizational Context), activated to examine the identification of relevant stakeholders, the formalization of their needs and expectations, and the existence of a periodic review mechanism. A total of three questions from this checklist were automatically selected and evaluated by the application.

Table 11 - doc 2 : Mota-Engil

Checklist question	Verdict	Concise explanatory sentence for the application
Are relevant stakeholders identified along with their type, influence, and occupational safety and health expectations?	Partially compliant	The document acknowledges the stakeholders and their involvement, but does not explicitly specify their type, level of influence, or specific expectations regarding occupational health and safety.
Are the needs and expectations of interested parties identified and defined through documented consultations, surveys, meetings, and worker representation mechanisms?	Compliant	The policy describes a materiality analysis and structured consultations to identify and formalize stakeholder expectations.
Are the needs and expectations of other interested parties documented and reviewed periodically?	Partially compliant	The needs of stakeholders are identified and documented, but the existence of a formalized periodic review is not explicitly demonstrated.

C. Document 3 – SHEQ Policy

The third document analyzed for Mota-Engil is the Occupational Health and Safety Policy (SHEQ). The assessment was conducted using the "Occupational Health and Safety

Policy" checklist (ISO 45001 – Leadership). A total of nine questions were activated and automatically evaluated by the application.

Table 12 - doc 3 : Mota-Engil

Checklist question	**Verdict**	**Concise explanatory sentence for the application**
Is the document available as documented information?	Non-compliant	The document describes the approval and implementation of the policy, but does not explicitly state that it constitutes formally available documented information.
Does the document include a commitment to provide safe and healthy working conditions?	Compliant	The policy explicitly sets out commitments regarding safety, health and well-being, including minimum quality standards and risk assessment.
Is the document appropriate for the purpose, size and context of the organization?	Compliant	The policy is aligned with the group's strategic plan and applicable to all units, projects and contracts, demonstrating its suitability to the organizational context.
Does the document provide a framework for the establishment of Occupational Health and Safety objectives?	Compliant	The document provides for the definition and communication of proactive objectives, targets and indicators in the field of safety and health, constituting a structuring framework.
Does the document include a commitment to meet legal and other requirements?	Compliant	The policy explicitly mentions compliance with legal and contractual requirements and stakeholder expectations.
Does the document include a commitment to eliminate hazards and reduce Occupational Health and Safety risks?	Partially compliant	The document commits to reducing risks and promoting prevention, without explicitly mentioning the systematic elimination of hazards.
Does the document include a commitment to continuous improvement of the Occupational Health and Safety Management System?	Compliant	The policy makes explicit reference to the continuous improvement of the integrated management system, including the health and safety component.
Does the document include a commitment to consult and participate workers?	Compliant	The establishment of SHEQ committees is mentioned in order to ensure consultation and participation of workers.

Checklist question	Verdict	Concise explanatory sentence for the application
Is the document made available to interested parties, if applicable?	Partially compliant	Some elements indicate external communication, but others specify restrictions on dissemination, which does not clearly confirm that it was made available to interested parties.

C. Factual Summary – Document 3

In total, the analysis of the three documents involved: • 5 requirements (Document 1) • 3 requirements (Document 2) • 9 requirements (Document 3)

That makes 17 requirements evaluated.

The overall breakdown of verdicts is as follows:

• Compliant: 10

• Partially compliant: 5

• Non-compliant: 2

These results correspond exactly to the verdicts generated by the application for the documents and checklists activated within the framework of this case study.

4.3.3 Case 3 – TDF

A. Document 1 – HSE Charter (2021)

The first document analyzed for TDF was the HSE Charter (2021). The assessment was carried out using the checklist "Organization Chart, Job Descriptions, or Occupational Health and Safety Responsibility Matrix" (ISO 45001 – Leadership). A total of 6 questions were activated and automatically evaluated by the application.

Table 13 – Doc 1 : TDF

Checklist question	**Verdict**	**Concise explanatory sentence for the application**
Is the document kept up to date in the form of documented information?	Non-compliant	The document mentions continuous improvement and periodic evaluations, but does not explicitly state that the charter itself is kept up to date as documented information.
Are the responsibilities related to the Occupational Health and Safety Management System assigned and documented?	Partially compliant	The charter states that each employee is responsible for their role, but does not formally demonstrate structured documentation of responsibilities in the OHS management system.
Are the authorities related to the Occupational Health and Safety Management System assigned and documented?	Non-compliant	The extracts describe general commitments without identifying or formalizing the specific authorities related to the OHS management system.
Does the document assign responsibility for compliance with ISO 45001?	Partially compliant	The document mentions risk management aligned with standards similar to ISO 45001, without explicitly assigning responsibility for compliance to a specific role.
Does the document give management the responsibility for reporting on performance?	Partially compliant	The evaluation of managers' performance is mentioned, but the explicit allocation of reporting responsibility by management is not clearly established.
Does the document demonstrate that roles and responsibilities are communicated at all levels?	Compliant	The charter specifies that each individual, including managers and directors, must be aware of their role and responsibilities, indicating communication at all levels.

B. Document 2 – TDF Group Code of Ethics

The second document analyzed for TDF is the TDF Group Code of Ethics. The assessment was carried out using the "Register of Interested Parties" checklist (ISO 45001 – Context of the organization). A total of three questions were activated and automatically evaluated by the application.

Table 14 - Doc 2 : TDF

Checklist question	Verdict	Concise explanatory sentence for the application
Are relevant stakeholders identified along with their type, influence, and occupational safety and health expectations?	Non-compliant	The document identifies categories of stakeholders (employees, partners, shareholders, civil society), but does not specify their detailed type, level of influence, or specific expectations regarding OHS.
Are the needs and expectations of interested parties identified and defined through documented consultations, surveys, meetings, and worker representation mechanisms?	Non-compliant	The Code of Ethics describes general principles and relationships with stakeholders, without demonstrating the formal identification of their needs through documented mechanisms of consultation or representation.
Are the needs and expectations of other interested parties documented and reviewed periodically?	Non-compliant	Although the document mentions a periodic review of the Code, it does not demonstrate that the needs and expectations of interested parties are formally documented and reviewed in a structured manner.

C. Document 3 – TDF HSE Policy

The third document analyzed for TDF is the TDF HSE Policy (2021). The assessment was carried out using the "Occupational Health and Safety Policy" checklist (ISO 45001 – Leadership). A total of 10 questions were activated and automatically evaluated by the application.

Table 15 - Doc 3 : TDF

Checklist question	Verdict	Concise explanatory sentence for the application
Is the document available as documented information?	Non-compliant	The document mentions annual updates and its content, but does not explicitly confirm that it is formally controlled and available as documented information.

Checklist question	Verdict	Concise explanatory sentence for the application
Does the document include a commitment to provide safe and healthy working conditions?	Compliant	The policy is based in particular on the pillar of "safety and health of persons" and the improvement of working conditions, reflecting an explicit commitment to safe and healthy conditions.
Is the document appropriate for the purpose, size and context of the organization?	Partially compliant	The document describes the objectives and pillars of the HSE policy without explicitly specifying its formal suitability to the context, size and purpose of the organization.
Does the document provide a framework for the establishment of Occupational Health and Safety objectives?	Compliant	The policy includes an annual prevention plan in conjunction with the CSSCT, structuring the definition of actions and objectives in the area of occupational health and safety.
Does the document include a commitment to meet legal and other requirements?	Compliant	The document mentions periodic regulatory checks and compliance with applicable legal obligations.
Does the document include a commitment to eliminate hazards and reduce Occupational Health and Safety risks?	Compliant	The policy sets a target of "zero high-risk events" and provides for the elimination or substitution of hazards.
Does the document include a commitment to continuous improvement of the Occupational Health and Safety Management System?	Compliant	The HSE management system is presented as being based on a continuous improvement process aligned with ISO 45001 and ISO 14001 standards.
Does the document include a commitment to consult and participate workers?	Compliant	The update in connection with the CSSCT and the involvement of employees and their representatives demonstrate a commitment to consultation and participation.
Is the Occupational Health and Safety policy communicated, understood and accessible at all levels of the organization (eg, posting, intranet, reception, training)?	Partially compliant	Dissemination and training activities are mentioned, without explicitly specifying the full modalities of accessibility and understanding at all levels.

Checklist question	Verdict	Concise explanatory sentence for the application
Is the document made available to interested parties, if applicable?	Non-compliant	The document describes the internal dissemination of security information, without explicitly indicating its formal availability to external interested parties.

D. Document 4 – CSR Policy

The fourth document analyzed for TDF is the CSR Policy. The assessment was carried out using the "Environmental context analysis (global report)" checklist (ISO 14001 – Context of the organization). A total of 8 questions were activated and automatically evaluated by the application.

Table 16 - Doc 4 : TDF

Checklist question	Verdict	Concise explanatory sentence for the application
Is the document identified, dated, and versioned?	Partially compliant	The document is clearly identified and refers to a year of publication, without explicitly specifying a formalized version number.
Are the internal environmental issues clearly identified?	Non-compliant	The document presents objectives and environmental governance without explicitly listing internal environmental issues in a structured manner.
Are the external environmental issues clearly identified?	Partially compliant	Issues such as carbon footprint and climate risks are mentioned, without providing a detailed and exhaustive identification of external issues.
Are the environmental conditions affected by the organization described?	Compliant	The document describes the environmental impact of the activities and the actions taken to reduce this impact.
Are the environmental conditions that are likely to affect the organization described?	Partially compliant	The risks associated with climate change are mentioned, without a complete analysis of the environmental conditions likely to affect the organization.
Are the issues analyzed in relation to the expected results of	Partially compliant	An approach to environmental management and periodic evaluation is described, without explicitly demonstrating the formalized link

Checklist question	Verdict	Concise explanatory sentence for the application
the Environmental Management System?		between the identified issues and the expected results of the EMS.
Is the method for context analysis explained?	Non-compliant	The document describes the governance and ESG assessments, without explaining the methodology used for the analysis of the environmental context.
Is a review or periodic update planned?	Compliant	Annual ESG assessments and periodic steering committees are mentioned, indicating the existence of a regular review mechanism.

E. Factual summary – Case 3: TDF

The analysis of the four documents from the company TDF involved successively using different checklists from the ISO 45001 and ISO 14001 standards, depending on the nature of each document.

Summary of the number of requirements assessed

- Document 1 – HSE Charter 6 requirements
- Document 2 – TDF Group Code of Ethics 3 requirements
- Document 3 – TDF HSE Policy 10 requirements
- Document 4 – CSR Policy 8 requirements

Total number of requirements assessed: 27

Overall distribution of verdicts

By strictly aggregating the results generated by the application:

- Compliant: 9
- Partially compliant: 9
- Non-compliant: 9

These results correspond exclusively to the automatic verdicts produced by the application for the checklists activated within the framework of the case study relating to the company TDF, without additional interpretation.

4.3.4 Case 4 – Total Energies

A. Document 1 – Safety, Environment and Quality Charter

The first document analyzed for TotalEnergies was the Safety, Environment, and Quality Charter. The assessment was carried out using the checklist "Organization Chart, Job Descriptions, or Occupational Health and Safety Responsibility Matrix" (ISO 45001 – Leadership). A total of six questions were activated and evaluated by the application.

Table 17 – Doc1 : Total Energies

Checklist question	**Verdict**	**Concise explanatory sentence for the application**
Is the document kept up to date in the form of documented information?	Partially compliant	A document date is mentioned and periodic updates are referred to, without explicit confirmation that the charter is maintained as formalized documented information.
Are the responsibilities related to the Occupational Health and Safety Management System assigned and documented?	Partially compliant	Individual and hierarchical responsibilities regarding safety are indicated, without explicit formalization within the framework of the OHS management system.
Are the authorities related to the Occupational Health and Safety Management System assigned and documented?	Non-compliant	The extracts describe general responsibilities, without specifying the formal allocation of authorities related to the OHS management system.
Does the document assign responsibility for compliance with ISO 45001?	Non-compliant	No explicit mention of the ISO 45001 standard or of a dedicated responsibility for compliance with this standard is identified.
Does the document give management the responsibility for reporting on performance?	Partially compliant	Performance measurement and evaluation are mentioned, without explicit attribution to management of formal reporting responsibility.
Does the document demonstrate that roles and responsibilities are communicated at all levels?	Compliant	The document indicates that each person must be aware of their role and responsibilities, which demonstrates cross-functional communication of responsibilities.

B. Document 2 – Code of Conduct (2025)

The first document analyzed for Total Energies was the 2025 Code of Conduct (totalenergies_code-de-conduite_2025.pdf). The assessment was carried out using the

"Register of Interested Parties" checklist (ISO 45001 – Organizational Context). A total of three questions were activated and evaluated by the application.

Table 18 – Doc 2 : Total Energies

Checklist question	**Verdict**	**Concise explanatory sentence for the application**
Are relevant stakeholders identified along with their type, influence, and occupational safety and health expectations?	Non-compliant	The document mentions stakeholders in general terms, without explicitly identifying their type, level of influence, or specific expectations regarding occupational health and safety.
Are the needs and expectations of interested parties identified and defined through documented consultations, surveys, meetings, and worker representation mechanisms?	Partially compliant	The code mentions dialogue, transparency and commitment to stakeholders, but does not formally demonstrate that their needs and expectations are identified and defined through documented mechanisms for consultation or worker representation.
Are the needs and expectations of other interested parties documented and reviewed periodically?	Non-compliant	Commitments to stakeholders are described, but no explicit formalization of documentation and periodic review of their needs and expectations is identified.

C. Document 3 – Booklet “The 12 Golden Rules” (Safety Rules)

The third document analyzed for Total Energies was the "12 Golden Rules" booklet, an operational document outlining the safety requirements applicable to the Group's sites and activities. The assessment was conducted using the "Occupational Health and Safety Operational Procedures" checklist (ISO 45001 – Operation). A total of five questions were activated and evaluated by the application.

Table 19 - Doc 3 : Total Energies

Checklist question	**Verdict**	**Concise explanatory sentence for the application**
Does the document establish clear criteria for carrying out the process concerned?	Compliant	The document sets out precise and operational rules (tool verification, equipment inspection, work permit, wearing of PPE) which define clear criteria for performing tasks safely.

Checklist question	Verdict	Concise explanatory sentence for the application
Does the document describe the implementation of mastery in accordance with these criteria?	Non-compliant	The extracts present safety rules and requirements, but do not explicitly describe a structured control or monitoring system demonstrating the formalized implementation of the criteria.
Does the document mention the adaptation of work to workers (ergonomics)?	Compliant	The booklet explicitly mentions postures, gestures, body positioning and adaptation of tools to the task, which falls under ergonomics and adapting work to workers.
Does the document describe the coordination of Occupational Health and Safety processes for multiple employer workplaces?	Non-compliant	No explicit mention of the coordination of OHS processes between several employers or companies operating simultaneously on the same site is identified.
Is the document kept as documented information proving that the process was carried out as planned?	Non-compliant	The document provides a framework for safety rules, but does not demonstrate that it is kept as documented information serving as evidence of the actual implementation of planned processes.

D. Document 4 – Sustainability & Climate 2024 Progress Report

The fourth document analyzed for Total Energies is the Sustainability & Climate 2024 Progress Report. The assessment was carried out using the "Environmental context analysis (global report)" checklist (ISO 14001 – Organizational context). A total of eight questions were activated and evaluated by the application.

Table 20 - Doc 4 : Total Energies

Checklist question	Verdict	Concise explanatory sentence for the application
Is the document identified, dated, and versioned?	Partially compliant	The report clearly mentions its title and the year 2024, without explicit indication of a version number or a formal versioning system.
Are the internal environmental issues clearly identified?	Compliant	The document explicitly identifies internal issues such as methane emissions from operated sites, CO_2 emissions and associated reduction targets.
Are the external environmental issues clearly identified?	Compliant	External issues are explicitly mentioned (energy transition, decarbonization, emissions related to uses, environmental risks).

Checklist question	Verdict	Concise explanatory sentence for the application
Are the environmental conditions affected by the organization described?	Compliant	The report describes the impact of the Company's activities on greenhouse gas emissions and the actions taken to reduce these impacts.
Are the environmental conditions that are likely to affect the organization described?	Non-compliant	The analyzed extracts do not explicitly identify external environmental conditions (climate, natural hazards, ecological changes) likely to affect the organization.
Are the issues analyzed in relation to the expected results of the Environmental Management System?	Non-compliant	Although climate and environmental risks are mentioned, the explicit link with the expected results of an environmental management system (EMS) is not established.
Is the method for context analysis explained?	Non-compliant	No formal methodology for context analysis (such as SWOT, PESTEL, analysis matrix) is explicitly described in the extracts.
Is a review or periodic update planned?	Partially compliant	A follow-up period (2023–2025) and performance indicators are mentioned, without a formal description of a structured mechanism for periodic review of the context.

E. Factual Summary – Case 4 TotalEnergies

The analysis of the four documents from the company Total Energies successively mobilized different checklists from the ISO 45001 and ISO 14001 standards, depending on the nature of each document.

Summary of the number of requirements assessed

- Document 1 – Safety, Environment and Quality Charter 6 requirements
- Document 2 – Code of Conduct 2025 3 requirements
- Document 3 – Golden Rules for Safety (2022 booklet) 5 requirements
- Document 4 – Sustainability & Climate 2024 Progress Report 8 requirements

Total number of requirements assessed: 22

Overall distribution of verdicts By strictly aggregating the results generated by the application:

- Compliant: 7
- Partially compliant: 5
- Non-compliant: 10

These results correspond exclusively to the automatic verdicts produced by the application for the checklists activated within the framework of the case study relating to the company Total Energies, without additional interpretation or analytical adjustment.

5 Discussion

5.1 General discussion of experimental tests

5.1.1 Discussion of results – Test 1 : Reproducibility of verdicts

The reproducibility test aimed to evaluate the temporal stability of the verdicts generated by the application using the same document corpus analyzed twice under identical experimental conditions. The analysis focused on 88 activated questions, covering all 12 documents studied. The results obtained highlight a particularly high level of agreement between the two successive runs, which constitutes a key indicator of the system's internal reliability.

The observed crude agreement rate of 87.5% reflects a high overall consistency in the classifications produced. Of the 88 verdicts compared, 77 are strictly identical, while 11 show a divergence. This level of agreement far exceeds the threshold generally considered satisfactory in methodological reproducibility studies. However, crude agreement does not account for the probability of agreement due to chance, which justifies the use of a more robust indicator incorporating the ordinal structure of the scale used.

Examination of the marginal distributions confirms the overall stability of the model's classification behavior. The respective proportions of the categories "Non-compliant," "Partially compliant," and "Compliant" remain close between Run 1 and Run 2. The slight variations observed—notably a marginal increase in "Non-compliant" and "Compliant" verdicts in the second run—do not reflect a structural shift in the decisions, but rather ad hoc adjustments. This homogeneity of the distributions suggests the absence of systematic bias from one run to the next.

Cohen's quadratic weighted coefficient κ, equal to 0.911, is the central element of the statistical interpretation. By incorporating the ordinal nature of the scale (0/1/2) and weighting disagreements differently according to their magnitude, this indicator provides a more precise measure of actual agreement. The value obtained corresponds to a level of agreement described as "near-perfect" according to standard interpretation frameworks. This result means that the vast majority of observed discrepancies correspond to differences of only one level on the scale and not to major contradictions.

The qualitative analysis of the 11 discrepancies confirms this interpretation. No direct shifts between the extreme categories ("Non-compliant" to "Compliant" or vice versa) were observed. The discrepancies are limited to adjacent transitions (NC ↔ PC or PC ↔ C), indicating that the variations concern the intensity of the assessment rather than the fundamental nature of the judgment. Furthermore, no concentration of discrepancies on a specific document was identified, ruling out the hypothesis of localized instability linked to a particular type of documentary content.

From a methodological standpoint, these results confirm that the application exhibits high decision stability in an identical reanalysis context. Reproducibility is a fundamental criterion for any document auditing tool, particularly when it employs artificial intelligence models that may introduce stochastic variability. The observed level of near-perfect agreement demonstrates that, under the chosen parametric conditions, the model's intrinsic variability remains contained and does not significantly affect the overall consistency of the verdicts.

In summary, Test 1 demonstrates substantial temporal robustness of the automated analysis system. The high agreement rate, the stability of the marginal distributions, and the value of the weighted κ coefficient all point to a consistent conclusion: the application produces highly reproducible verdicts. This property is an essential prerequisite before any further evaluation of the validity, sensitivity to changes in documentation, or normative relevance of the classifications produced.

5.1.2 Discussion of results – Test 2 : Robustness to variations in document form

Test 2 aimed to examine the system's sensitivity to purely formal variations in a document, in the absence of any modification to the underlying normative content. This step is methodologically crucial, as it allows us to distinguish a true capacity for semantic analysis from simple lexical or structural identification. By submitting each question to three versions of the same document—original, reformulated, and enriched with a neutral text—the experiment aimed to assess decision stability in the face of controlled editorial disruptions.

The descriptive analysis shows that a clear majority of questions (61.5%) exhibit perfect stability of verdicts across the three versions. This result indicates that, in more than half

of the cases, reformulating or adding a paragraph without normative scope does not affect the conformity assessment. This stability constitutes a first positive indicator of semantic robustness. It suggests that the model is not limited to a strict dependence on the initial wording, but maintains decisional consistency when the normative meaning remains unchanged.

In addition, 30.8% of the questions show variations of only one level on the ordinal scale. These discrepancies correspond to adjacent transitions (for example, from "Partially compliant" to "Compliant"), without any extreme shifts. Thus, 92.3% of the questions show a maximum variation of only one level, which confirms that the observed differences are generally moderate and proportionate. The cases of extreme variation, limited to 7.7%, remain marginal compared to the overall observations.

Statistical evaluation using Krippendorff's alpha, adapted for ordinal data and multiple comparisons, provides further insight. The value obtained ($\alpha \approx 0.72$) places the reliability within the so-called "acceptable" range. This result reflects substantial consistency among the three versions evaluated, while also revealing significant variability. Unlike Test 1, which measured strict reproducibility under identical conditions, Test 2 intentionally introduces a formal perturbation; therefore, a lower level of agreement is expected. The relative decrease in the indicator (compared to the κ of Test 1) precisely reflects this controlled sensitivity.

Analysis of the observed disagreement compared to the expected disagreement confirms this interpretation. The majority of the compared pairs exhibit a zero distance, and maximum distances remain rare. The average observed disagreement remains significantly lower than the expected disagreement by chance, which attests that the variations do not stem from random instability of the model, but from a measured responsiveness to editorial changes.

From an interpretative standpoint, these results highlight a dual characteristic of the system. On the one hand, the application demonstrates overall robustness in the face of reformulations and additions of non-normative text, which is essential in a document audit context where companies exhibit a high degree of stylistic diversity. On the other hand, the existence of a measurable sensitivity indicates that certain reformulations can alter the perception of precision, explicitness, or normative completeness. This sensitivity

is not necessarily a flaw; it may reflect the model's ability to detect editorial nuances that could influence the interpretation of conformity.

Ultimately, Test 2 confirms that the application does not rely solely on a superficial recognition of fixed text structures, but maintains a majority consistency in decision-making when the normative background is stable. The level of reliability achieved, deemed acceptable, validates the overall robustness of the system while highlighting the moderate influence of editorial choices on the verdicts. This observation opens up avenues for improvement, particularly regarding the normalization of prompts or the adjustment of model parameters, in order to further reduce the variability induced by formal variations.

5.1.3 Discussion of results – Test 3: Sensitivity to the detection of expected non-conformities

Test 3 aimed to evaluate the system's performance in a targeted situation involving the detection of obvious non-conformities defined a priori. Unlike the first two tests, which focused on decision stability, this test examined the application's discriminant validity, that is, its ability to correctly identify explicit normative absences and avoid misclassifications that would lead to unjustified conformity. The selected positive class corresponded to "Non-compliant" verdicts (0), thus placing the analysis within a logic of deviation detection.

The confusion matrix highlights two true positives out of three expected nonconformities, as well as one false negative. The recall rate of 66.7% means that the system detects approximately two-thirds of the nonconformities actually present. This level of sensitivity can be described as moderate. It indicates that the application is capable of identifying a majority of obvious deviations from the standards, but that it remains susceptible to under-detecting certain documentation deficiencies. The observed false negative reveals that a nonconformity explicitly defined as such in the experimental protocol was not classified as "Non-compliant," which is a critical point in an audit context.

The accuracy (precision), at 50%, reflects another dimension of performance: among the "Non-compliant" verdicts issued by the system, one out of two actually corresponds to an expected non-compliance. The two false positives observed out of 22 negative cases

indicate that the model can interpret certain ambiguous or incomplete formulations as non-compliant, even though they were not identified as such in the experimental reference. This tendency suggests a certain propensity for classificatory caution, potentially linked to a strict interpretation of the normative requirements.

The F1 score of 0.57 summarizes the trade-off between sensitivity and precision. This value reflects intermediate performance, consistent with the observed balance between false negatives and false positives. It should be noted that the number of positive cases is small (n = 3), which limits the statistical stability of the indicators. In a small sample size, the variation of a single case strongly influences the metrics, calling for a cautious interpretation of the results.

From a methodological standpoint, these results reflect a real but incomplete capacity to detect obvious non-conformities. The system is neither excessively permissive—the number of false positives remains limited—nor entirely conservative—it effectively identifies the majority of deviations defined a priori. However, the presence of a false negative highlights a significant operational risk: in a QHSE audit context, the failure to detect a non-conformity can have significant implications for regulatory compliance and risk management.

In summary, Test 3 reveals partial sensitivity in detecting expected non-conformities, coupled with moderate accuracy. The results confirm that the application possesses a functional, but improvable, discriminatory capacity, particularly regarding the reduction of false negatives. These observations suggest potential improvements could involve adjusting the decision parameters or enriching the analysis instructions to optimize the balance between regulatory vigilance and classification accuracy.

5.1.4 Discussion of results – Test 4: Logical coherence before/after document correction

Test 4 aimed to assess the system's directional consistency in response to an explicit normative intervention. Unlike previous tests, which measured stability and detection capabilities respectively, this test examines the model's logical sensitivity to a targeted improvement in the document content. The underlying assumption is that a structuring addition—explicitly incorporating a responsible party, a defined action, a deadline, and

an element of evidence—should lead to a consistent improvement in the verdicts for the relevant requirements.

The analysis focused exclusively on eight questions directly related to initially identified nonconformities. In each case, the verdict changed from "Non-compliant" (0) to "Compliant" (2), representing a uniform improvement of two levels on the ordinal scale. This consistency is a strong indicator of internal coherence: no stagnation or deterioration was observed. The system systematically integrated the added normative information and adjusted its assessment accordingly.

The use of the Wilcoxon signed-rank test is justified by the ordinal nature of the data and the paired before/after comparison on the same units of analysis. Since all differences are strictly positive, the sum of negative ranks is zero, and all ranks are represented by positive differences. The exact probability associated with such a configuration under the null hypothesis is low ($p \approx 0.0039$), which leads to the rejection of the hypothesis of no effect. Statistically, the observed improvement cannot therefore be attributed to chance.

Beyond their significance, the scope and uniformity of the changes deserve particular attention. The direct shift from 0 to 2 indicates that the implemented corrections fully met the normative requirements assessed by the system. This is not simply a marginal adjustment (e.g., from "Non-compliant" to "Partially compliant"), but a complete reclassification to compliance. This responsiveness suggests that the application is sensitive to the structuring markers of document formalization, especially when they explicitly correspond to the expected normative criteria.

From a methodological standpoint, this result reinforces the hypothesis that the model does not merely perform a superficial analysis of the text, but rather incorporates elements constituting the logic of conformity. The explicit addition of the components "responsible party – action – deadline – evidence" appears to act as a sufficiently discriminating structuring signal to alter the verdict. The perfect directional consistency observed in all eight cases constitutes an indicator of internal validity within the chosen experimental framework.

It should be noted, however, that the sample remains small and targeted. The corrections were introduced intentionally explicitly and aligned with the evaluation criteria, which

may enhance the detectability of the improvement. Nevertheless, from an operational perspective, the system's ability to recognize a relevant normative correction is an essential property for a document auditing tool.

In summary, Test 4 highlights a systematic and statistically significant improvement in verdicts after targeted document correction. The complete absence of degradation, the uniformity of the adjustments, and the significance of the Wilcoxon test confirm the system's strong logical consistency in the face of normatively relevant improvements. This result complements previous tests by demonstrating not only the model's stability and robustness but also its ability to respond appropriately to actual improvements in document content.

5.1.5 General discussion of the tests

The four tests carried out made it possible to evaluate in a structured way the overall performance of the application according to four complementary dimensions: reproducibility, formal robustness, the ability to detect non-conformities and logical consistency in the face of improved documentation.

Test 1 demonstrated high temporal stability of the verdicts, attesting to the system's strong internal reliability. Test 2 confirmed generally satisfactory robustness to variations in wording, although a moderate sensitivity to reformulations remained. Test 3 highlighted a real but improvable ability to detect expected non-conformities, with an intermediate compromise between sensitivity and accuracy. Finally, Test 4 showed a systematic and statistically significant improvement in the verdicts after targeted normative correction, reflecting solid directional consistency.

Taken together, these results indicate that the application has a generally consistent and methodologically satisfactory level of performance, while revealing areas for optimization, particularly in terms of exhaustive detection of non-conformities.

5.2 Discussion of case studies

5.2.1 Case 1 – Jan De Nul

The case study of Jan De Nul falls within the framework of the standardized automated analysis protocol described earlier. The four selected documents were associated with specific normative checklists, activated within the application according to the predefined document correspondence. The results obtained stem exclusively from the automated processing of the textual content accessible in the public versions of the analyzed documents. It is important to reiterate that the objective of this section is not to make an overall judgment on the company's level of compliance, but rather to examine the application's behavior when faced with real organizational documents.

Analysis of the Code of Conduct, compared with the "Register of Interested Parties" checklist (ISO 9001 – Organizational Context), reveals a disconnect between the ethical dimension and the expected formalization of a structured register of interested parties. The results show that, although certain categories of stakeholders and their expectations are mentioned, the absence of a formalized, dated, and structured register leads primarily to verdicts of non-compliance or partial compliance. This result illustrates the tool's ability to distinguish between a general mention of stakeholders and a formalization that explicitly meets the normative requirements for traceability and monitoring.

The QHSE Policy Statement, analyzed using the "Environmental Policy" checklist (ISO 14001 – Leadership), presents a significantly more favorable profile. The majority of requirements are rated as compliant, particularly regarding management commitment, regulatory compliance, and continuous improvement. The reservations expressed mainly concern the level of detail in contextualizing environmental impacts and the document's usable structure. This internal contrast demonstrates that the application assigns different verdicts based on the level of explicit formalization detected, rather than on an overall assessment of the document.

The Energy Management Action Plan, assessed using the "Environmental Action Plan" checklist (ISO 14001 – Planning), reveals majority compliance with several partial assessments. Elements relating to action planning, monitoring, and integration into the management system are generally recognized. However, the lack of explicit links to a

formal register of risks and opportunities, as well as the absence of references financial or operational constraints, leads to less favorable ratings. This result highlights the system's sensitivity to the presence of structured formulations establishing explicit connections between normative requirements and organizational structures.

The analysis of the 2017 Annual Report, compared to the "Environmental Context Analysis (Global Report)" checklist (ISO 14001 – Organizational Context), reveals a more heterogeneous profile. While some dimensions, notably the description of external environmental issues and impactful activities, are deemed compliant, the majority of requirements relating to the formal structuring of the context analysis are considered non-compliant. The absence of an explicitly described methodology, a formal link to the expected outcomes of the environmental management system, and a periodic review plan explains these assessments. This case illustrates the distinction made by the application between a descriptive institutional report and a document structured according to a management system logic.

Across all 30 requirements assessed, the distribution of verdicts shows a majority of compliant results, accompanied by a significant number of non-compliant and partially compliant results. This heterogeneous distribution reflects the variability in levels of document formalization among the different types of public documents analyzed. It also reflects the methodological limitations inherent in the study, insofar as the accessible documents do not necessarily correspond to the internal documents formalized within a certified system.

Ultimately, Case 1 highlights the application's ability to produce differentiated verdicts based on the degree of normative structuring detected in each document. The results obtained do not constitute a comprehensive assessment of organizational compliance, but rather an analytical rendering of the automated processing applied to public documents. This distinction is essential to situate the interpretation within the defined methodological framework and to avoid any normative extrapolation beyond the chosen experimental scope.

5.2.2 Case 2 – Mota-Engil

The case study of Mota-Engil was conducted according to the standardized automated analysis protocol described in the methodology section. Three distinct organizational documents were examined against normative checklists associated with the ISO 9001 and ISO 45001 standards. The results presented correspond exclusively to the verdicts generated by the application based on the textual content accessible in the analyzed public documents. They do not constitute a comprehensive evaluation of the company's management system, but rather an analytical report of the tool's performance within this specific document context.

Analysis of the Code of Ethics and Professional Conduct reveals a relatively explicit structuring of stakeholder categories and their expectations. The "Compliant" verdicts assigned to these dimensions reflect the application's ability to identify structured lists and behavioral rules clearly associated with defined groups. Conversely, the absence of a formal, dated, and structured register, as well as the lack of explicit monitoring mechanisms, leads to non-compliance verdicts. This distinction underscores that the tool differentiates between descriptive stakeholder recognition and procedural formalization consistent with the logic of ISO management systems.

The Social Responsibility Policy presents a more nuanced profile. The application identifies the existence of structured consultation and materiality analysis mechanisms, thus justifying a favorable verdict regarding the identification and formalization of expectations. However, the lack of explicit indications concerning the level of stakeholder influence or the existence of a formalized periodic review leads to partial assessments. This configuration highlights the system's sensitivity to the presence of explicitly demonstrable documentary evidence, particularly when dealing with recurring or institutionalized processes.

The SHEQ Policy, analyzed against the "Occupational Health and Safety Policy" checklist (ISO 45001 – Leadership), received mostly compliant verdicts. Commitments related to regulatory compliance, continuous improvement, worker consultation, and the structuring of health and safety objectives are clearly recognized by the application. Partial verdicts mainly concern the degree of precision of the commitments, particularly regarding the explicit elimination of hazards and the detailed methods for communicating

and making the policy accessible. The single non-conformity related to formalization as documented information indicates that the tool requires an explicit mention of document availability to fully meet the normative requirement.

The overall distribution of verdicts—mostly compliant, with a limited number of non-compliant—reflects internal consistency among the various documents analyzed. The identified discrepancies primarily concern aspects of explicit formalization, traceability, and procedural structuring, rather than the absence of commitments in principle. This observation is consistent with the nature of the public documents studied, which often aim for external institutional communication rather than a comprehensive technical formalization of ISO requirements.

From a methodological standpoint, this case confirms the application's ability to produce differentiated and proportionate verdicts based on the degree of normative explicitness detected in each document. The results also illustrate the inherent limitations of analyzing public documents, which do not necessarily reflect all the internal tools formalized within a certified system. In this sense, the case study does not aim to determine overall organizational compliance, but rather to observe how the tool interprets and classifies diverse documentary content with regard to specific normative requirements.

5.2.3 Case 3 – TDF

The case study of TDF was conducted according to a standardized automated analysis protocol. Four organizational documents were examined using checklists from ISO 45001 and ISO 14001, activated based on the identified document type. The results analyzed in this section correspond exclusively to the verdicts generated by the application from the text extracts available in the public documents studied. They do not constitute a comprehensive assessment of the organization's management system, but rather an observation of the tool's performance in this specific context.

Analysis of the HSE Charter (2021) reveals a predominance of partial or negative assessments regarding the formalization of responsibilities and authorities related to the occupational health and safety management system. While the document demonstrates widespread communication of roles at all hierarchical levels, the application identifies a lack of explicit structuring in the form of a responsibility matrix or formal documentation of authorities. This finding illustrates the system's ability to distinguish between a

statement of commitment or principle and a structured organizational formalization that complies with ISO 45001 requirements.

The TDF Group Code of Ethics receives only non-compliance verdicts on the checklist for the stakeholder register (ISO 45001 – Context). The application acknowledges the mention of general stakeholder categories but considers that the required elements—identification of the level of influence, formalization of health and safety expectations, documented consultation mechanisms, and structured periodic review—are not explicitly demonstrated. This result highlights a clear distinction made by the tool between an ethical document intended for institutional use and a structured register that conforms to the logic of management systems.

The TDF HSE Policy (2021) presents a more balanced profile. Commitments related to risk prevention, regulatory compliance, continuous improvement, and worker consultation are clearly recognized as compliant. Partial and negative assessments mainly concern explicit documentation and the precise methods of external dissemination. The lack of a clear indication of formal availability as documented information, as well as the absence of an explicit mention of making it available to external stakeholders, influences the assigned rating. This result confirms the system's sensitivity to formal traceability and explicitly normative wording.

The CSR Policy, assessed against the "Environmental Context Analysis (Global Report)" checklist (ISO 14001 – Context), reveals significant heterogeneity. While the application acknowledges the description of environmental impacts and the existence of a periodic review mechanism, it identifies shortcomings in the formal structuring of the context analysis. The absence of an explicitly described methodology and a structured list of internal issues leads to partial or negative assessments. This result further illustrates the distinction between strategic ESG communication and a formalized analysis that meets the methodological requirements of the ISO 14001 standard.

The overall case study reveals a perfectly balanced distribution between compliance, partial compliance, and non-compliance. This three-part distribution reflects a contrasting documentary profile. The documents contain substantial commitments and descriptions of operational actions, but exhibit shortcomings in terms of explicit procedural

formalization, detailed normative structuring, and document traceability. This pattern is consistent with the public and communicative nature of the analyzed documents.

From a methodological standpoint, Case 3 confirms the application's ability to discriminate between different levels of normative structuring within the same organizational corpus. The assigned verdicts appear proportionate to the degree of explicitness detected in the documents, without a uniform tendency toward positive or negative over-qualification. As with the other case studies, these results must be interpreted in light of the inherent limitations of analyzing public documents and should not be considered a comprehensive organizational compliance assessment.

5.2.4 Case 4 – Total Energies

The Total Energies case study was conducted using a standardized automated analysis protocol, employing checklists from ISO 45001 and ISO 14001 standards based on the document type. The analyzed results correspond precisely to those generated by the application from extracts accessible in the selected public documents. This discussion aims to examine the tool's classification behavior with these documents, without offering an overall assessment of the organization's level of compliance.

Analysis of the Safety, Environment, and Quality Charter reveals a predominance of partial and negative assessments regarding the formalization of responsibilities and authorities related to the occupational health and safety management system. While the document acknowledges cross-functional communication of responsibilities and mentions performance evaluation mechanisms, its implementation highlights the lack of formal authority assignments, as well as the absence of explicit references to ISO 45001 and a dedicated compliance responsibility. This finding illustrates the system's sensitivity to the presence of formalized elements explicitly linked to the regulatory framework, beyond general commitments.

The 2025 Code of Conduct, assessed against the checklist for stakeholder registers (ISO 45001 – Context), yielded mostly negative results. The application acknowledges the existence of a structured discourse on stakeholders and dialogue, but does not detect any explicit formalization of specific health and safety expectations, nor any documented mechanisms for consultation and periodic review. This configuration confirms the

distinction made by the tool between a document with an ethical and communicative purpose and a structured register that meets the methodological requirements of a management system.

The "12 Golden Rules" booklet yielded mixed results. Operational safety criteria were clearly identified as compliant, particularly regarding the explicit definition of task execution rules and the integration of ergonomic elements. However, the lack of a description of a structured, formalized control mechanism, inter-company coordination, or the preservation of information as verifiable documentation led to several non-conformities. This result illustrates the system's ability to recognize precise operational requirements while demanding demonstrable procedural formalization.

The Sustainability & Climate 2024 Progress Report presents a heterogeneous profile. The application clearly identifies internal and external environmental issues and describes the impacts of activities, leading to several positive assessments. However, the lack of methodological explanation for the context analysis, the absence of a formal link to the expected outcomes of an environmental management system, and the insufficient formalization of the periodic review process result in partial or negative assessments. This outcome confirms that the tool distinguishes between strategic communication on sustainability and a context analysis formalized according to ISO 14001 requirements.

The overall case study highlights a significant proportion of non-conformities, coupled with a smaller number of conformities and partial conformities. The identified discrepancies primarily concern explicit document formalization, normative traceability, and the methodological structuring of the described systems. The analyzed documents present substantial commitments and descriptions of concrete actions, but do not systematically demonstrate formalization aligned with the documentary logic of the activated ISO standards.

From a methodological standpoint, this case confirms that the application is not limited to recognizing the presence of general commitments or institutional discourse, but rather operates a discrimination based on the formal explication of normative requirements. As with the other case studies, the results must be interpreted in light of the inherent limitations of analyzing public documents, which do not necessarily reflect all the internal tools formalized within the organization's management system.

5.3 Discussion of Results in Light of the Literature

This section relates the experimental results to the findings of the literature review, identifying both confirmations and observed divergences.

5.3.1 Empirical Confirmation of the Literature

5.3.1.1 Reproducibility and Reliability in Regulated Contexts

The literature emphasizes that AI systems deployed in compliance and auditing environments must demonstrate reproducibility and traceability to be considered operationally reliable (Popara et al., 2023; Winkler et al., 2025). In regulated domains, decision stability is directly linked to accountability and defensibility.The results of Test 1 confirm this requirement. The high agreement between successive analyses demonstrates that, under controlled experimental conditions, the system produces stable and consistent verdicts. This finding corroborates the governance perspective outlined by Raji et al. (2020), according to which automated systems must ensure repeatability in order to be integrated into compliance-sensitive processes. The experimental evidence therefore confirms that structured AI-assisted documentary evaluation can achieve reliability levels compatible with regulatory audit expectations.

5.3.1.2 Robustness and Sensitivity to Editorial Variations

The literature on AI robustness indicates that language models may exhibit sensitivity to controlled textual perturbations (Ribeiro et al., 2020; Jia & Liang, 2017). Such sensitivity does not necessarily invalidate the model but highlights structural limitations in semantic invariance. Test 2 partially confirms these findings. While most verdicts remained stable when documents were reformulated or enriched with neutral content, some moderate variations were observed. This outcome aligns with the robustness concerns raised in the literature and confirms that AI-based documentary analysis is generally resilient but not entirely immune to editorial influence. The results therefore support the view that semantic robustness in compliance auditing remains substantial yet imperfect.

5.3.1.3 Operationalization of Normative Clauses

Several studies in your literature review stress that AI performs more effectively when regulatory requirements are translated into structured, machine-interpretable elements (Arora et al., 2015; Adebayo et al., 2022). The mapping of standards to internal

documents is identified as a central condition for effective compliance automation. Test 4 provides strong empirical confirmation of this principle. The explicit addition of structured compliance elements (responsible party, defined action, deadline, and evidence) led to systematic improvements in verdicts. This finding directly supports the hybrid intelligence model described by Dellermann et al. (2021), in which AI operates most effectively when normative expectations are formalized and embedded in structured decision frameworks rather than inferred from implicit statements.

5.3.2 Nuanced and Divergent Findings

5.3.2.1 Limits of Automated Detection

While Noordin et al. (2022) and Xu Yao (2022) report significant improvements in anomaly detection and audit efficiency through AI technologies, the results of Test 3 reveal only moderate performance in detecting predefined non-conformities. Although the majority of expected deviations were identified, at least one remained undetected. This result diverges from the most optimistic performance claims and instead aligns with the caution expressed by Winkler et al. (2025) and Raji et al. (2020), who emphasize that automated systems in regulated environments cannot fully eliminate classification risks. The divergence observed may be explained by the complexity of normative language highlighted by Guha et al. (2023), as well as by the dependency of AI evaluation on explicit documentary formulation. When requirements are implicit or ambiguously expressed, detection performance decreases.

5.3.2.2 Documentary Nature and ISO Logic

The case studies reveal a recurring pattern: public corporate documents frequently receive partial or non-compliant verdicts due to insufficient formal structuring, absence of explicit registers, or lack of documented traceability mechanisms. This outcome is consistent with the ISO documentary requirements described in ISO 9001, ISO 14001, ISO 45001, and ISO 19011, which emphasize structured documented information and objective evidence. However, the results also highlight a structural nuance not extensively discussed in prior literature: documents designed for institutional communication (codes of conduct, sustainability reports, charters) are not necessarily intended to function as formal management-system evidence. This explains why AI evaluation may identify formal gaps even when an organization operates a certified system internally. The

divergence is therefore attributable to the nature of the analyzed corpus rather than to a systemic deficiency of the automated system.

5.3.3 Integrated Answer to the Research Question

The research question sought to determine how artificial intelligence can automate QHSE document audits while improving efficiency and reliability. The findings allow for a structured conclusion aligned with the literature:

- AI enhances standardization and traceability of documentary assessments (Popara et al., 2023).
- Its performance improves when normative clauses are explicitly structured (Adebayo et al., 2022; Arora et al., 2015).
- Detection capacity is significant but not exhaustive (Noordin et al., 2022; Winkler et al., 2025).
- Governance and human oversight remain essential (Dellermann et al., 2021; Raji et al., 2020).

Overall, the results largely confirm the literature while refining it through empirical validation in a QHSE multi-standard environment.

5.3.4 Scientific Contribution

This study makes three contributions grounded in the existing body of research. First, it provides experimentally controlled validation of AI reproducibility and robustness in a QHSE documentary auditing context, addressing the scarcity of empirically tested compliance automation frameworks noted in the literature (Kokina et al., 2025).

Second, it operationalizes ISO documentary requirements through structured checklist architectures combined with AI-based textual evaluation, thereby concretizing the computational alignment approach described by Adebayo et al. (2022). Third, it empirically demonstrates that AI performance in compliance auditing is strongly mediated by the degree of documentary formalization. This finding refines the hybrid intelligence perspective proposed by Dellermann et al. (2021), by identifying documentary explicitness as a key determinant of automated audit reliability. Together,

these contributions strengthen the theoretical and operational foundations of AI-assisted QHSE auditing within regulated environments.

6 General conclusion

This research set out to examine how artificial intelligence can be used to automate QHSE document audits and enhance both efficiency and accuracy in compliance verification. By combining systematic literature analysis with experimental validation and applied case studies, the study provides a structured and empirically grounded response to this question.

The findings demonstrate that AI can meaningfully support the automation of documentary audit tasks, particularly in areas requiring systematic comparison of textual content against structured normative requirements. The developed platform exhibited high decision reproducibility under controlled conditions, indicating strong internal reliability. Robustness testing further showed that, while moderate sensitivity to textual reformulation remains, the system maintains consistent classification patterns when normative meaning is preserved.

In terms of compliance detection, the system demonstrated functional but not exhaustive sensitivity to predefined non-conformities. The presence of false negatives highlights a critical operational insight: AI-assisted auditing enhances efficiency but does not eliminate the need for expert supervision. Compliance verification remains a high-stakes domain where interpretative judgment and contextual understanding are indispensable.

Notably, the system showed coherent directional responsiveness when documents were normatively improved. The introduction of explicit compliance markers—such as assigned responsibility, defined actions, timelines and documented evidence—resulted in systematic and statistically significant improvements in compliance verdicts. This finding underscores a key structural principle: AI-assisted compliance evaluation performs best when documentation is explicit, structured and demonstrably aligned with ISO logic.

The applied case studies reinforce an important distinction between institutional communication documents and formally structured management system documentation. While public-facing policies and sustainability reports often express commitments and intentions, they do not consistently provide the traceable, methodologically structured evidence required by ISO standards. The system's evaluations reflect this normative distinction rather than organizational performance itself.

From a strategic perspective, this research supports a hybrid intelligence model. Artificial intelligence can enhance audit efficiency by reducing manual workload, standardizing analytical criteria and systematically generating traceable justifications. It can improve accuracy by applying consistent evaluative logic across documents. However, in alignment with ISO 19011:2018 principles of auditor competence, independence and professional judgment, AI must function as an assistive instrument rather than a substitute decision-maker. The contribution of this study lies in demonstrating that AI-driven QHSE document audit automation is feasible, measurable and conditionally reliable when embedded within a structured methodological framework. The research clarifies three critical conditions for effective integration:

1. Documentation must be normatively explicit and structured.
2. AI systems must operate within transparent and reproducible governance parameters.
3. Human oversight must remain central to final compliance validation.

Several limitations should be acknowledged. The experimental corpus was limited in size and primarily based on publicly available documents. Additionally, system performance remains influenced by language model configuration and evolving AI architectures. Future research may extend the dataset to internal certified documentation, compare alternative AI models, and integrate advanced explainability mechanisms to further strengthen regulatory trust. Beyond the empirical results obtained, an important methodological limitation concerns the level of expertise required to replicate the evaluation protocol. The statistical assessment of the system's performance relies on applied inferential statistics specific to automated classification systems, including the interpretation of weighted agreement coefficients (Cohen's κ), multi-rater reliability indices (Krippendorff's α), and standard classification metrics derived from confusion matrices. Replicating such analyses therefore presupposes a solid understanding of ordinal reliability modelling, performance evaluation frameworks, and the use of computational tools such as Python or R for statistical implementation. It is essential, however, to distinguish between operational use and scientific evaluation. The practical use of the application by a QHSE auditor does not require advanced knowledge in artificial intelligence or statistical modelling. The interface is designed for applied decision support, and its routine deployment remains accessible to non-specialist users.

In contrast, the rigorous scientific assessment of its performance necessarily requires an intermediate to advanced methodological competence in quantitative analysis. Consequently, while the reproducibility of the present study is technically achievable, it is conditional upon the methodological qualification of the replicating researcher. This constraint constitutes an external limitation related to expertise requirements, rather than an intrinsic weakness of the system itself.

In conclusion, artificial intelligence can meaningfully automate key components of QHSE document audits and improve efficiency while maintaining acceptable levels of accuracy—provided that it operates within a hybrid governance framework that preserves professional accountability. Rather than displacing auditors, AI has the potential to transform documentary compliance verification into a more structured, transparent and analytically consistent process.

7 Bibliographic references

Adebayo, A., Sow, D., & Bulut, M. F. (2022). Automated compliance blueprint optimization with artificial intelligence. *Journal of Regulatory Compliance, 7*(1), 45–62.

Arora, C., Sabetzadeh, M., Briand, L., & Zimmer, F. (2015). Automated checking of conformance to requirements templates using natural language processing. *IEEE Transactions on Software Engineering, 41*(10), 944–968.

Balabanov, P., & Davletshin, M. (2018). Optimizing document flow in QHSE management systems. *Journal of Quality Management, 22*(3), 112–125.

Binh, D. T. (2025). Artificial intelligence in auditing : A thematic review based on latent Dirichlet allocation. *International Journal of Digital Accounting Research, 25*, 1–24.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46.

Constant, M., Grosclaude, R., Denecker, M., & van Laere, O. (2025). AnnoCaseLaw : A richly-annotated dataset for benchmarking legal reasoning in AI systems. *Artificial Intelligence and Law, 33*(1), 88–110.

Cutting-Decelle, A. F., Digeon, A., Young, R. I., Barraud, J. L., & Lamboley, P. (2018). Extraction of technical information from normative documents using automated methods based on ontologies. *Advanced Engineering Informatics, 38*, 502–515.

Darji, H., Mitrovic, J., & Granitzer, M. (2023). Challenges and considerations in annotating legal data : A comprehensive overview. *Legal Informatics Review, 5*(2), 14–31.

Dellermann, D., Ebel, P., Söllner, M., & Leimeister, J. M. (2021). Hybrid intelligence. *Business & Information Systems Engineering, 63*(5), 637–643.

Dellermann, D., Lipusch, N., Ebel, P., Popp, K. M., & Leimeister, J. M. (2021). Design principles for hybrid intelligence systems. *Journal of Business Economics, 91*(3), 357–392.

Guha, N., Bhattacharya, P., & Ghosh, S. (2023). *How disagreement is handled in annotation of legal texts*. Springer.

Haataja, M., Hyrynsalmi, S., & Leppänen, V. (2021). A systematic review of artificial intelligence applications in the context of software engineering. *Journal of Systems and Software, 174*, Article 110886.

Imperial, J. M., Jones, M. D., & Madabushi, H. T. (2025). Standardizing intelligence : Aligning generative AI for regulatory and operational compliance. *Journal of AI Standards, 12*(1), 45–67.

International Organization for Standardization. (2015). *Quality management systems — Requirements* (ISO Standard No. 9001:2015).

International Organization for Standardization. (2015). *Environmental management systems — Requirements with guidance for use* (ISO Standard No. 14001:2015).

International Organization for Standardization. (2018). *Guidelines for auditing management systems* (ISO Standard No. 19011:2018).

International Organization for Standardization. (2018). *Occupational health and safety management systems — Requirements with guidance for use* (ISO Standard No. 45001:2018).

International Organization for Standardization/International Electrotechnical Commission. (2023). *Artificial intelligence management system — Requirements* (ISO/IEC Standard No. 42001:2023).

Iwuanyanwu, U., Okafor, F., & Nwogu, C. (2023). Challenges in AI adoption for audit processes : Data quality, explainability, and auditor skills. *International Journal of Auditing Technology, 12*(1), 89–105.

Jia, R., & Liang, P. (2017). Adversarial examples for evaluating reading comprehension systems. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* (pp. 2021–2031). Association for Computational Linguistics.

Kitchenham, B., & Charters, S. (2007). *Guidelines for performing systematic literature reviews in software engineering* (Technical Report EBSE-2007-01). Keele University.

Kokina, J., Jain, S., & O'Leary, D. E. (2025). Adoption challenges and systemic barriers of artificial intelligence in audit. *Journal of Emerging Technologies in Accounting, 22*(1), 55–78.

KPMG. (2024). *ISO/IEC 42001 : A new standard for AI governance* [Insight report].

Krippendorff, K. (2011). *Computing Krippendorff's alpha-reliability* (Departmental Paper No. 43). Annenberg School for Communication, University of Pennsylvania.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174.

Leocádio, J., Malheiro, A., & Reis, E. (2024). Artificial intelligence in auditing : A conceptual framework for auditing practices. *Journal of Information Systems and Technology Management, 21*, Article e202421005.

MDPI. (2023). The use of AI and audit quality : A critical assessment. *Journal of Risk and Financial Management, 9*(4), 233–251.

Nannini, L., Balayn, A., & Smith, A. L. (2023). Explainability in AI policies : A critical review of communications. *Ethics and Information Technology, 25*(3), 42–59.

Noordin, S., Rahman, N. A., & Abdullah, N. H. (2022). Enhancing audit document analysis using natural language processing techniques. *Journal of Financial Compliance, 7*(3), 210–225.

Ojewale, V., Steed, R., Vecchione, B., et al. (2024). Towards AI accountability infrastructure : Gaps and opportunities in AI audit tooling. *Journal of Responsible Technology, 18*, 100076.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E.,

McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement : An updated guideline for reporting systematic reviews. *BMJ, 372*(n71).

Popara, J., Savkovic, M., Ciric Lalic, D., & Lalic, B. (2023). Application of digital tools, data analytics and machine learning in internal audit. *Procedia Computer Science, 217*, 123–132.

Popara, O., Rusu, L., & Marinescu, D. (2022). Optical character recognition in digital auditing : Improving data extraction from scanned documents. *Computers in Industry, 138*, Article 103628.

Powers, D. M. W. (2011). Evaluation : From precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies, 2*(1), 37–63.

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

Ribeiro, M. T., Wu, T., Guestrin, C., & Singh, S. (2020). Beyond accuracy : Behavioral testing of NLP models with CheckList. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 4902–4912).

Shabani, H., et al. (2024). Challenges in document management for safety audits : A case study from Zimbabwe. *Safety Science, 170*, 106341.

SpringerLink. (2023). *Explainable AI and law : An evidential survey*. Springer Nature.

Wang, C., Pastore, F., Göknil, A., Briand, L., & Iqbal, Z. (2019). Automatic generation of acceptance test cases from use case specifications. *IEEE Transactions on Software Engineering, 46*(12), 1312–1334.

Wang, F., et al. (2020). Development and validation of a natural language processing tool to generate the CONSORT reporting checklist. *JAMA Network Open, 3*(7), e209179.

Wang, J., Ye, X., & Pan, R. (2025). LLM-based HSE compliance assessment : Benchmark, performance, and advancements. *Safety and Health at Work*, *16*(1), 12–28.

Wijaya, A., et al. (2025). Current trends and challenges in AI adoption for audit professionals : A global perspective. *International Journal of Auditing*, *29*(1), 102–119.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, *1*(6), 80–83.

Winkler, T. J., Bennek, L., & Walke, F. (2025). Artificial intelligence explainability requirements of the AI Act and metrics for measuring compliance. *Computer Law & Security Review*, *56*, 105942.

Xu, Y. (2022). Machine learning applications in audit anomaly detection : Current state and future directions. *Auditing Journal*, *29*(2), 143–161.

Yew, L. L., & Goh, C. F. (2015). Adoption of cloud-based integrated management systems in SMEs : Compliance with ISO 9001 and ISO 14001. *International Journal of Production Research*, *53*(19), 5678–5692.

Zhang, Y., Cho, M., & Vasarhelyi, M. (2022). Explainable AI in auditing. *Journal of Emerging Technologies in Accounting*, *19*(2), 1–15.

Zhang, Z. (2021). Integrating AI and OCR for automated detection of irregularities in financial audits. *Journal of Digital Accounting*, *15*(1), 32–47.

Zin, M. M., Nguyen, H. T., Satoh, K., & Nishino, F. (2024). Addressing annotated data scarcity in legal information extraction. *Information Processing & Management*, *61*(4), 103721.

Appendix A - checklist Iso 9001 – Iso 14001 – Iso 45001

A. Context of the organization (ISO 9001)	
SWOT Analysis	Is the document dated and versioned?
	Are internal strengths clearly identified and documented?
	Are internal weaknesses clearly identified and documented?
	Are external opportunities explicitly listed and justified?
	Are external threats explicitly listed and justified?
	Is the methodology used to identify strengths, weaknesses, opportunities, and threats described (sources, consultations, data)?
	Is the impact of each SWOT element on the QMS explained?
	Are actions or follow-up plans proposed for the identified elements (responsible party, deadline)?
	Is the update frequency or the triggers for revising the SWOT analysis specified?
PESTEL Study	Is the document dated and versioned?
	Are Political factors identified and analyzed?
	Are Economic factors identified and analyzed?
	Are Social factors identified and analyzed?
	Are Technological factors identified and analyzed?
	Are Environmental factors identified and analyzed?
	Are Legal factors identified and analyzed?
	Is the methodology used to collect and analyze PESTEL factors described (sources, studies, consultations)?
	Is the potential impact of these factors on the QMS explained?
	Are actions or follow-ups proposed in response to the identified factors (responsible party, deadline)?
	Is the update frequency or the triggers for revising the PESTEL study specified?
Document "Scope of application" (QMS scope)	Is the document dated and available?
	Does the document explicitly list the products and services covered by the QMS?
	Does the document clearly describe the limits and scope (sites, activities, interfaces)?
	Does the document mention any exclusions from the standard?
	For each exclusion, does the document justify why it is not applicable and explain the absence of impact on the conformity of products/services?
	Does the document indicate where the scope of application is published/accessible?
Register of interested parties	Is the register available and dated?
	Does the document list the categories of interested parties (customers, employees, suppliers, authorities, etc.)?
	For each party listed, does the document indicate the requirements/expectations?
	Does the document identify the source of each requirement (contract, law, customer request, etc.)?
	Does the document specify the frequency/method of monitoring the requirements of interested parties (who monitors, when)?

Forms / table of requirements of interested parties (if distinct)	Is each requirement associated with an owner responsible for taking it into account?
	Does each requirement have a status (met / in progress / to be processed)?
	Does the document provide for actions or measures to comply with the identified requirements?
B. Leadership (ISO 9001)	
Minutes / Report of management review	Is the minutes dated and signed/validated by management?
	Do the minutes explicitly mention the review of the context (internal/external issues)?
	Does the minutes or report show that management assumes responsibility for the effectiveness of the QMS?
	Does the document prove that the quality policy and quality objectives are established and compatible with the context and strategy?
	Do the minutes show that the QMS requirements are integrated into business processes?
	Does the document mention the promotion of the process approach and the risk approach?
	Is there evidence that management ensures that the resources necessary for the QMS are available?
	Does the document contain explicit communication about the importance of an effective QMS and compliance with requirements?
	Do the minutes show that management verifies that the QMS achieves its expected results?
	Does the document show that management encourages and supports staff to contribute to the QMS?
	Is there evidence that management promotes continuous improvement?
	Does the document show that management supports other management representatives in their quality roles?
	Do the minutes contain decisions or actions taken resulting from the context review?
	Do the decided actions have a clearly indicated responsible party and deadline?
	Is information on the performance and effectiveness of the QMS documented (trends, satisfaction, objectives, process performance, non-conformities, audits, providers)?
	Are changes in internal and external issues taken into account?
	Is the effectiveness of actions in the face of risks and opportunities evaluated?
	Are opportunities for improvement identified and documented?
	Are the needs for changes to be made to the QMS documented?
	Are resource needs documented?
	Is documented information relating to the management review retained?
Quality manual	Does the quality manual (or equivalent document) clearly present the structure and processes of the QMS?
	Are the interactions between processes identified and described?
	Is the scope of the QMS defined and justified?
	Are exclusions to the standard mentioned and justified in accordance with article 4.3?
	Is this document kept up to date and approved according to the document control procedure?
C. Planning (ISO 9001)	

Quality policy	Is the policy available, dated and kept up to date?
	Is the policy appropriate to the context and purpose of the organization?
	Does the policy support the strategic direction?
	Does the policy provide a framework for establishing quality objectives?
	Does the policy include a commitment to satisfy applicable requirements?
	Does the policy include a commitment to continuous improvement?
	Is the policy disseminated internally (posters, intranet, notes) and understood by staff?
	Is the policy accessible to external interested parties?
Register / plan for risk and opportunity analysis and treatment	Are the risks related to the context and interested parties identified?
	Are the opportunities relevant to the organization identified?
	Is each risk or opportunity linked to an expected impact on the QMS?
	Are the planned actions to address each risk/opportunity specified and integrated into the processes?
	Is a mechanism for evaluating the effectiveness of the actions planned?
	Are the actions proportional to the potential impact?
Quality Objectives Register / Implementation Plan	Are the quality objectives documented, dated, and versioned?
	Is each objective consistent with the quality policy?
	Is each objective measurable, relevant, and regularly monitored?
	Do the objectives take applicable requirements into account?
	Is each objective associated with an action plan, responsible person, deadline, and resources?
	Are the results evaluated and the objectives updated if necessary?
D. Support (ISO 9001)	
Official Organization Chart / Job Descriptions	Does the organization chart designate quality managers and other key roles related to the QMS?
	Do the job descriptions clearly indicate the quality responsibilities and authorities for each role?
	Does each job description assign responsibility for QMS compliance, process performance, customer focus, and promotion of continuous improvement?
Register / inventory of documented information	Does the register list the documents/records necessary for each process (title, type)?
	For each document listed, does the register indicate the owner, location and retention period?
	Does the register indicate the control status (version, approval, access) for each document?
	Does the register show that the necessary documents are accessible and protected according to internal rules?
Plan / register of resources	Does the document identify the internal resources available and their constraints?
	Are the resources to be procured from external providers mentioned?
	Are the responsibilities for resource management clearly defined?
	Is the document dated and versioned?

Procedure / infrastructure register	Does the document list all the necessary infrastructures (buildings, equipment, software, means of transport, ICT)?
	Are the infrastructures maintained (maintenance plans, contracts, budgets)?
	Are the responsibilities related to the infrastructure specified?
	Is the document dated and versioned?
Report on the working environment	Are the necessary social conditions (non-discrimination, working climate, absence of conflicts) described?
	Are the psychological conditions (stress prevention, psychological safety) described?
	Are the physical conditions (temperature, lighting, hygiene, noise, ventilation) described?
	Does the document show that these conditions are regularly monitored and maintained?
	Is the document dated and versioned?
Procedure / register of measuring equipment	Are the measuring equipment necessary to ensure reliable results identified?
	Are the methods for maintaining and verifying the suitability of the equipment described?
	Are documented proofs of suitability (certificates, registers) kept?
	Are the responsibilities for the management of the equipment specified?
	Is the document dated and versioned?
Metrology register / measuring equipment	Is each piece of equipment calibrated or verified at specified intervals?
	Are calibration certificates or records kept?
	Is each piece of equipment identified (number, label) to ensure calibration traceability?
	Is the equipment protected against adjustments or damage that could invalidate the results?
	Is an analysis planned if equipment is deemed unsuitable (impact on past results, corrective actions)?
	Is the document dated and versioned?
Organizational Knowledge Register	Is the knowledge necessary to implement the processes and ensure product/service compliance identified?
	Is the knowledge kept up to date?
	Are methods for acquiring or accessing additional knowledge defined (training, partnerships, monitoring)?
	Are the sources of knowledge identified (internal and external)?
	Is the document dated and versioned?
Skills Matrix / Training Records	Are the skills required for each position related to the QMS identified?
	Are the people in these positions competent (evidence: training, diploma, experience)?
	Are actions planned to address skill gaps (training, coaching, recruitment)?
	Are the proofs of competence retained (certificates, attestations, CVs, records)?
	Is the document dated and versioned?
Internal Awareness Plan	Is staff aware of the quality policy?

	Is staff aware of the relevant quality objectives?
	Does staff understand the importance of their contribution to the effectiveness of the QMS?
	Are the consequences of non-compliance with QMS requirements communicated?
	Is the document dated and versioned?
Internal and External Communication Plan / Procedure	Are the topics to be communicated (internal/external) defined?
	Is the schedule or frequency of communication specified?
	Are the recipients or interested parties identified?
	Are the means and media of communication specified?
	Are responsibilities for communication assigned?
	Is the document dated and versioned?
Document Control Procedure / Register	Is all documented information required by the standard included?
	Is the information necessary for the effective operation of the QMS included?
	Are the documents correctly identified and described (title, author, date, reference)?
	Are the format and medium specified (paper, digital, language, software version)?
	Have the documents been reviewed and approved before distribution?
	Is the document dated and versioned?
Document control procedure	Are the documents available where needed?
	Are the documents protected against loss, inappropriate use or alteration?
	Are the rules for distribution, access, retrieval and use defined?
	Are the rules for storage and protection (readability, backup) specified?
	Is the control of changes and versions defined?
	Are the rules for document retention and disposal defined?
	Are external documents identified and controlled?
	Are documents used as evidence of conformity protected against unintentional alteration?
	Is the document dated and versioned?
Operational change management form / procedure	Are changes (equipment, processes, resources, documents) recorded in a form or register?
	Is a risk assessment related to the change carried out before approval?
	Are the persons in charge and implementation dates identified?
	Are changes communicated to the stakeholders concerned?
	Are the effects of the change reviewed to confirm its effectiveness or detect unforeseen impacts?
E. Operation (ISO 9001)	
Process map	Is the map dated/versioned?
	Does the map show the macro-processes (management / operational / support)?
	Does the map represent the sequence and interactions between processes (flows, links)?
	Does the map identify key interfaces (suppliers, customers, upstream/downstream processes)?

	Does the map designate a responsible person/owner per process?
Process sheet (for each process)	Does the sheet indicate the title of the process and the owner?
	Does the sheet clearly describe the inputs and outputs of the process?
	Does the sheet state the criteria and methods to ensure process control?
	Does the sheet specify the necessary resources?
	Does the sheet define the responsibilities and authorities related to the activities of the process?
	Does the sheet identify the performance indicators (KPIs) and the frequency of measurement?
	Does the sheet document the records to be kept?
	Does the sheet include the risks and opportunities specific to the process and the associated actions?
	Does the sheet describe the sequence/interactions with other processes?
Procedure / file for QMS modification	Is the document available and dated/versioned?
	Are the possible consequences of the changes on the processes and the QMS evaluated?
	Is the objective of the planned changes clearly identified?
	Is the availability of the resources needed for the modification verified?
	Are the responsibilities and authorities related to the modification specified or reassigned if necessary?
	Are the changes planned and not treated in an ad hoc manner?
Operational Control Plan	Are all processes necessary for the provision of products/services listed?
	Are the criteria for controlling the processes defined?
	Are the resources needed for each process identified?
	Is the documented information updated and retained?
	Are planned or unplanned changes analyzed and controlled?
	Are outsourced processes identified and controlled?
	Is the document dated and versioned?
Change management procedure	Is there a procedure for managing changes affecting the QMS?
	Are the impacts of changes (human, technical, organizational) assessed before implementation?
	Are the changes validated and approved before application?
	Are change records kept and accessible?
	Have the changes been the subject of adequate internal communication?
Work instructions / operating procedures	Are work instructions available for critical activities?
	Are they up-to-date, dated, approved, and accessible to staff?
	Do operators know and apply the instructions?
	Are instruction changes controlled?
Records of verification and validation of products/services	Are product or service controls and validations recorded?
	Are the acceptance criteria clearly defined?
	Do the records indicate who performed the verification and the date?

	Are non-conforming products identified and isolated?
Register for monitoring and measuring processes	Does each process have performance or measurement indicators?
	Are the measurement results monitored and analyzed?
	Are improvement actions initiated when the results are unsatisfactory?
	Is the register kept up to date and verified by management?
Product traceability (if applicable)	Does the traceability system make it possible to link a product to its batch, its raw materials, or its service provider?
	Is traceability ensured throughout the production cycle?
	Are records kept for the required duration?
Non-conforming product management procedure	Is there a specific procedure for managing non-conforming products?
	Are non-conforming products identified, isolated and handled according to the procedure?
	Are decisions (repair, scrap, reuse) authorized by a competent person?
	Are processing records kept?
Preventive maintenance plan	Is a preventive maintenance plan established for critical equipment?
	Does the plan specify the frequencies, responsibilities and necessary resources?
	Are interventions tracked and archived?
	Are defective equipment identified and removed from service?
Procedure for evaluating and selecting suppliers	Are the criteria for evaluating and selecting suppliers defined?
	Is a list of approved suppliers kept up to date?
	Are supplier performances reviewed periodically?
	Are corrective actions applied to non-conforming suppliers?
List of approved suppliers / annual evaluation	Is the list of approved suppliers dated and validated?
	Are the annual evaluation results documented?
	Are decisions (maintenance, suspension, withdrawal) tracked?
Procedure for welcoming and integrating staff	Is there a welcoming and integration procedure for new staff?
	Does the integration include training on quality, safety and the environment?
	Are the integration sheets signed and archived?
Analysis of customer needs and expectations	Are customer needs and expectations identified and updated regularly?
	Are these needs taken into account in the quality objectives?
	Are satisfaction surveys carried out?
	Are the results analyzed and communicated to management?
Communication plan with interested parties	Is the internal and external communication plan formalized?
	Are communication methods and frequencies defined?
	Are relevant interested parties identified?
	Is the effectiveness of communication evaluated?

F. Performance evaluation (ISO 9001)	
Internal audit reports	Are internal audits conducted according to an annual plan?
	Do audit reports mention findings, non-conformities, and recommendations?
	Are audit results communicated to management and taken into account for improvement?
	Are corrective actions resulting from audits monitored?
Internal audit program	Does an internal audit program exist and does it cover all QMS processes?
	Is the frequency of audits adapted to the importance and results of the processes?
	Are the auditors competent and independent of the audited activities?
	Does the program take into account previous results and organizational changes?
Performance indicators / quality dashboard	Are quality indicators defined for each process?
	Are the objectives measurable and consistent with the quality policy?
	Are the results regularly monitored and analyzed?
	Are actions implemented in the event of deviations from the objectives?
	Are the indicators communicated to management and relevant personnel?
Customer complaints register	Is there a register centralizing all customer complaints received?
	Does each complaint include a description, date, responsible person and associated action?
	Are the causes of complaints analyzed to avoid their recurrence?
	Are processing times monitored and controlled?
	Are complaint trends analyzed to identify areas for improvement?
Register of external service providers (excluding material suppliers)	Does the register identify external service providers (maintenance, cleaning, security, etc.)?
	Are the criteria for selecting and evaluating these service providers defined?
	Are the evaluation results recorded and updated regularly?
	Do contracts specify applicable quality requirements?
	Are corrective actions implemented in the event of non-compliance by a service provider?
Customer satisfaction report or analysis	Are formal methods for measuring customer satisfaction defined?
	Are the results consolidated in a report or analysis table?
	Are satisfaction indicators defined, monitored and compared over time?
	Are customer satisfaction results communicated to management?
	Are improvement actions decided based on the results of this analysis?
G. Improvement (ISO 9001)	
Continuous Improvement Plan (corrective actions, QMS improvement)	Is the continuous improvement plan formalized, dated and versioned?
	Are the identified improvement actions resulting from the analysis of non-conformities, audits, or customer feedback?

	Are the actions monitored until their closure and evaluated for their effectiveness?
	Is the plan updated periodically?
	Are responsibilities and deadlines clearly defined for each action?
Procedure for managing non-conformities and corrective actions	Is there a documented procedure for managing non-conformities?
	Are non-conformities recorded with the cause, immediate actions, and corrective actions?
	Does the process include an evaluation of the effectiveness of corrective actions?
	Are processing times respected?
	Are non-conformity trends analyzed to prevent their recurrence?
A. Context of the organization (ISO 14001)	
Environmental SWOT analysis	Is the document clearly identified, dated, and versioned?
	Does the analysis explicitly cover the environmental context of the organization?
	Does the analysis identify at least one internal environmental weakness?
	Does the analysis identify at least one internal environmental strength?
	Does the analysis identify at least one external environmental threat?
	Does the analysis identify at least one external environmental opportunity?
	Are the elements identified related to environmental conditions affected by the organization?
	Are the elements identified related to environmental conditions that may affect the organization?
	Does the analysis show the impact of the identified issues on the organization's ability to achieve its objectives?
	Is a periodic review or update of the analysis planned or mentioned?
PESTEL environmental analysis	Is the document identified, dated, and versioned?
	Are environmental policy factors analyzed?
	Are environmental economic factors analyzed?
	Are social factors related to the environment analyzed?
	Are environmental technological factors analyzed?
Environmental context analysis (global report)	Is the document identified, dated, and versioned?
	Are the internal environmental issues clearly identified?
	Are the external environmental issues clearly identified?
	Are the environmental conditions affected by the organization described?
	Are the environmental conditions that are likely to affect the organization described?
	Are the issues analyzed in relation to the expected results of the Environmental Management System?
	Is the method for context analysis explained?
	Is a review or periodic update planned?
Register of interested parties – Environment	Is the document identified, dated, and versioned?
	Are the environmentally relevant stakeholders listed?
	Are the environmental needs and expectations of each interested party described?

	Are the applicable requirements clearly distinguished from expectations?
	Does the document identify which requirements become compliance obligations?
	Is a periodic update of the register planned?
Register of Environmental Compliance Obligations	Is the environmental management system register identified, dated, and versioned?
	Are the applicable environmental legal obligations identified?
	Are customer environmental requirements included?
	Are voluntary commitments, such as charters, labels, or agreements, included?
	Is each environmental obligation related to an activity, product, or service?
	Does the document provide for regulatory monitoring and updates?
Field of application of the Environmental Management System	Is the field of application formalized in the form of documented information?
	Are organizational and physical boundaries clearly defined?
	Are the activities, products, and services covered clearly listed?
	Are environmental compliance obligations taken into account?
	Are the authority and capacity for control or influence specified?
	Is the field of application consistent with the content of the document?
Environmental manual	Does the document describe the processes of the environmental management system?
	Are interactions between processes identified?
	Is the environmental management system built on the basis of the results of the organizational context and the expectations of interested parties?
	Does the document explicitly aim to improve environmental performance?
	Is a logic of continuous improvement described?
B. Leadership (ISO 14001)	
Management Review Minutes / Executive Committee Minutes	Is the document identified, dated and validated by management?
	Does management explicitly assume responsibility for the effectiveness of the Environmental Management System?
	Is the Environmental Management System addressed as a strategic topic in the document?
	Does the document mention the follow-up of the expected results of the Environmental Management System?
	Are decisions or guidelines related to the continuous improvement of the Environmental Management System formalized?
	Does management show support for environmental managerial roles?
Environmental policy	Is the document identified as an official environmental policy, dated and versioned?
	Is the policy appropriate to the purpose and context of the organization?
	Does the policy take into account the nature, size and environmental impacts of activities, products and services?
	Does the policy provide a framework for setting environmental objectives?
	Does it contain a commitment to protect the environment (e.g., pollution prevention)?
	Does it contain a commitment to comply with compliance obligations?
	Does it contain a commitment to continuous improvement of the organization and environmental performance?

	Is the policy formalized in the form of documented information?
	Is the policy written in an understandable and exploitable way?
Evidence of environmental policy communication	Does the document demonstrate that the environmental policy is communicated internally?
	Does management ensure that the policy is communicated to all relevant personnel?
	Is the policy presented in a clear and accessible way?
	Does the document show that the policy is made available to interested parties (website, external display, official communication)?
	Is the evidence current and consistent with the current version of the policy?
Environmental objectives	Is the document identified, dated and validated?
	Are the environmental objectives consistent with the environmental policy?
	Are the objectives compatible with the context and strategic orientation of the organization?
	Are the objectives formulated in a clear and understandable way?
	Are environmental objectives measurable or assessable?
	Does the document show the involvement of management in the definition or validation of objectives?
Organization chart or organizational sheet	Is the document identified, dated and up to date?
	Are the relevant roles for the Environmental Management System clearly identified?
	Are environmental responsibilities explicitly attributed?
	Are the authorities associated with environmental roles defined?
	Does the document make it possible to understand who is responsible for what in the Environmental Management System?
Function sheets / mission letters (environmental roles)	Does the document specify the responsibilities related to the Environmental Management System for the position concerned?
	Does the document mention the Environmental Management System's responsibility for compliance with ISO 14001?
	Does the document include responsibility for monitoring environmental performance?
	Are the responsibilities consistent with the organizational chart of the organization?
	Does the document specify the authority to report environmental non-conformities?
	Is the document formalized, dated and communicated?
Environmental performance reporting to management	Is the document explicitly intended for management?
	Is the performance of the organization analyzed?
	Is environmental performance measured using indicators?
	Are the results presented in an understandable and exploitable way?
	Does the reporting include an analysis of trends (improvement or degradation)?
	Does the document allow management to make decisions concerning the Environmental Management System?
C. Planning (ISO 14001)	

Register of Environmental Risks and Opportunities	Is the document identified, dated, and versioned?
	Are the issues from the context taken into account in the analysis?
	Are the requirements and obligations arising from the interested parties taken into account?
	Is the scope of the Environmental Management System clearly considered in the analysis?
	Have the risks and opportunities related to environmental aspects been identified?
	Are the risks and opportunities associated with compliance obligations identified?
	Are the risks and opportunities related to other identified issues taken into account?
	Does the document show that risks and opportunities are analyzed to achieve the expected results of the Environmental Management System?
	Does the document show that risks and opportunities are analyzed to prevent or reduce negative environmental impacts?
	Does the document show that risks and opportunities are analyzed to promote continuous improvement?
	Are potential emergencies with environmental impacts identified?
	Is the document kept up to date and usable for the planning of the Environmental Management System?
Environmental Analysis	Is the document identified, dated, and versioned?
	Are the activities, products, and services included in the scope analyzed?
	Are the environmental aspects that the organization controls identified?
	Are the environmental aspects on which the organization can have an influence identified?
	Are the environmental impacts associated with each aspect described?
	Is the analysis carried out according to a life cycle perspective?
	Are planned or new changes (activities, products, services) taken into account?
	Are abnormal conditions considered?
	Are reasonably foreseeable emergency situations taken into account?
Criteria for determining significant environmental aspects	Does the document describe criteria to assess the significance of the aspects?
	Are the criteria clear, consistent, and understandable?
	Do the criteria make it possible to distinguish significant from non-significant aspects?
	Do the criteria take into account the severity of environmental impacts?
	Is the document identified, dated, and validated?
List of significant environmental aspects	Is the document identified, dated, and versioned?
	Are the significant environmental aspects clearly listed?
	Is each significant aspect related to one or several environmental impacts?
	Are the significant aspects consistent with the defined criteria?
	Is the document exploitable for risk and opportunity management?
	Does the document indicate that these aspects are communicated to the levels and functions concerned?
Register of Environmental Compliance Obligations	Is the document identified, dated, and versioned?
	Are the applicable environmental legal obligations identified?
	Are the obligations related to the environmental aspects concerned?

	Does the document specify how each obligation applies to the organization?
	Are compliance obligations taken into account in the Environmental Management System (planning, implementation)?
Environmental Action Plan	Is the document identified, dated, and versioned?
	Are actions defined to address significant environmental aspects?
	Are actions defined to deal with compliance obligations?
	Are actions defined to address the identified risks and opportunities?
	Are the actions integrated into the Environmental Management System processes or into the business processes?
	Does the document specify how the effectiveness of the actions will be evaluated?
	Are technological options taken into account?
	Are financial, operational, and commercial constraints considered?
Environmental objectives	Is the document identified, dated, and versioned?
	Do the objectives take into account significant environmental aspects?
	Do the objectives take into account the associated compliance obligations?
	Are the objectives consistent with the environmental policy?
	Are the objectives measurable (when this is achievable)?
	Are the objectives defined for the functions and levels concerned?
	Are the objectives communicated?
Planning actions to achieve environmental objectives	Does the document specify what will be done to achieve each objective?
	Are the necessary resources identified?
	Are the responsibilities clearly assigned?
	Are the deadlines set?
	Are the monitoring and evaluation indicators defined?
	Does the document show how actions are integrated into business processes?
	Is the document consistent with the defined environmental objectives?
D. Support (ISO 14001)	
EMS Resource Plan / Environmental Resources Plan	Is the document identified, dated and versioned?
	Are the resources necessary for the Small and Medium-sized Enterprise identified (human, technical, financial)?
	Do the identified resources cover the establishment of the Environmental Management System?
	Do the identified resources cover the implementation of the Environmental Management System?
	Do the identified resources cover up-to-date maintenance of the Environmental Management System?
	Do the identified resources cover continuous improvement of the Environmental Management System?
	Does the document show that resources are adapted to environmental issues?
	Are the planned resources consistent with the environmental objectives and actions?
	Is the document identified, dated and versioned?

Matrix of environmental skills	Are positions or functions with an environmental impact identified?
	Are the environmental skills required for each function defined?
	Do the skills take into account significant environmental aspects?
	Do the skills take compliance obligations into account?
	Does the document make it possible to identify skills gaps?
Skills files / proofs of competence	Is the document a formal proof of competence?
	Are the skills demonstrated consistent with the requirements of the position?
	Do the skills cover the environmental requirements related to the position?
	Is the evidence up to date and exploitable?
	Is the document kept as documented information from the Environmental Management System?
Environmental Training Plan	Is the document identified, dated and validated?
	Are environmental training needs determined?
	Is the training related to environmental aspects?
	Are the trainings related to the Small and Medium-sized Enterprise?
	Is the training related to compliance obligations?
	Are the planned training actions clearly defined?
	Does the document provide for an evaluation of the effectiveness of the training?
Environmental awareness materials	Is the document identified and up to date?
	Does the document deal with environmental policy?
	Are the significant environmental aspects presented?
	Are real or potential work-related environmental impacts explained?
	Does the document explain the importance of the individual contribution to the Environmental Management System?
	Are the consequences of non-compliance with environmental requirements mentioned?
Environmental Communication Procedure	Is the document identified, dated and versioned?
	Does the procedure define the topics of communication?
	Does the procedure define the moments of communication?
	Does the procedure define the recipients of communication?
	Does the procedure define the means of communication?
	Have compliance obligations been taken into account in the communication?
	Does the procedure guarantee the reliability and consistency of environmental information?
	Does the document provide for management of communications received?
Evidence of internal environmental communication	Does the document demonstrate internal communication on the Environmental Management System?
	Does the information communicated concern the functioning of the Environmental Management System?
	Does the information communicated relate to the changes of the Environmental Management System?
	Does communication allow employees to contribute to continuous improvement?
	Is the evidence consistent with the communication procedure?

	Does the document constitute an external environmental communication?
	Is the information communicated consistent with that of the Environmental Management System?
	Do communications meet compliance obligations?
	Is the information reliable and not misleading?
	Is the document kept as proof of communication?
Procedure for controlling documented information	Is the document identified, dated and versioned?
	Does the procedure define the distribution and access rules?
	Does the procedure define the storage and protection rules?
	Does the procedure define the rules for mastering versions?
	Does the procedure define the rules for conservation and disposal?
	Is the documented information of the Environmental Management System protected against loss or alteration?
	Are documents of origin identified and controlled externally?
Master list of environmental documents	Does the document list the documents required by ISO 14001?
	Are the documents necessary for the effectiveness of the Environmental Management System identified?
	Are the current versions clearly indicated?
	Does the document make it possible to ensure the availability of documented information?
E. Operation (ISO 14001)	
Planning and operational control procedure	Is the document identified, dated and versioned?
	Does the procedure describe the operational processes necessary for the Environmental Management System?
	Are actions arising from clauses 6.1 (risks and opportunities) and 6.2 (objectives) integrated?
	Are operational criteria defined for the processes concerned?
	Are the processes clearly described and their methods of mastery specified?
	Are the means of mastery (technical, procedural, organizational) specified?
	Does the procedure provide for up-to-date processes?
	Does the document provide assurance that the processes are carried out as planned?
Environmental Operational Procedures	Is the document identified, dated and versioned?
	Does the procedure cover an activity with a significant environmental aspect?
	Are the environmental operating criteria clearly defined?
	Are the environmental impacts associated with the activity taken into account?
	Are the applicable environmental requirements integrated?
	Is the procedure consistent with environmental objectives and actions?
	Are operational responsibilities defined?
Change Management Procedure	Is the document identified, dated and versioned?
	Does the procedure provide for the control of planned changes?
	Are the environmental consequences of unforeseen changes analyzed?

	Are actions planned to limit negative environmental effects?
	Is the procedure integrated into the Environmental Management System?
	Are responsibilities related to change management defined?
Procedure for controlling outsourced processes	Is the document identified, dated and versioned?
	Are external processes with an environmental impact identified?
	Are the type and degree of control or influence exerted on these external processes defined?
	Are the applicable environmental requirements communicated to service providers?
	Does the procedure specify how the environmental compliance of service providers is monitored?
	Is the document consistent with the Environmental Management System?
Design and development procedure (environmental integration)	Is the document identified, dated and versioned?
	Are environmental requirements integrated into design and development?
	Are the different phases of the life cycle of the product or service taken into account?
	Are potential environmental impacts analyzed?
	Do environmental criteria influence design decisions?
	Is the document consistent with environmental policy and objectives?
Responsible purchasing procedure / acquisition of products and services	Is the document identified, dated and versioned?
	Are the environmental requirements for purchasing defined?
	Are the environmental criteria integrated in the selection of suppliers?
	Are environmental requirements communicated to suppliers?
	Does the document provide for monitoring of compliance with environmental requirements by suppliers?
Supplier Environmental Requirements / Contractual Clauses	Does the document specify the environmental requirements applicable to suppliers?
	Are the requirements adapted to the subcontracted activities?
	Is the document consistent with purchasing and subcontracting procedures?
	Are the requirements clearly formulated and exploitable?
Environmental information products and services	Does the document provide information on potential environmental impacts?
	Are the phases concerned specified (transport, use, end of life, disposal)?
	Is the information tailored to the interested parties concerned?
	Is the document consistent with significant environmental aspects identified?
	Is the information kept up to date?
Environmental Emergency Plan	Is the document identified, dated and versioned?
	Are potential emergencies with environmental impact identified?
	Are there actions planned to prevent or mitigate environmental impacts?
	Are the roles and responsibilities in an emergency situation defined?
	Is the plan consistent with the risks identified in clause 6.1.1?
	Does the document allow a response adapted to the magnitude of the emergency?

Exercise reports or tests of emergency situations	Does the document provide evidence that periodic tests have been carried out?
	Are the results of the exercises documented?
	Are deviations or areas for improvement identified?
	Are corrective actions proposed or implemented?
	Is the document usable for improving the emergency preparedness and response system?
Real Environmental Incident or Emergency Report	Does the document describe a real environmental emergency?
	Are the response actions implemented documented?
	Are environmental impacts being analyzed?
	Are actions to prevent recurrence identified?
	Does the document show a review of the emergency processes following the event?
Emergency Management Training Materials	Is the document identified and up to date?
	Does the training cover the preparation and response to emergency situations?
	Are the people concerned clearly targeted?
	Is the content consistent with the environmental emergency plan?
	Does the document demonstrate effective awareness of environmental emergency procedures?
F. Performance evaluation (ISO 14001)	
Procedure for monitoring, measurement, analysis and evaluation	Is the document identified, dated and versioned?
	Does the procedure define what should be monitored and measured?
	Are the monitoring and measurement methods described?
	Does the procedure specify the provisions guaranteeing the validity of the measurement results?
	Are the criteria for evaluating environmental performance defined?
	Are environmental performance indicators identified?
	Are the monitoring and measurement frequencies specified?
	Are the methods of analysis and evaluation of the results described?
	Does the procedure provide for the evaluation of the effectiveness of the Environmental Management System?
	Are the requirements for the retention of documented information defined?
Dashboard / iEnvironmental performance indicators	Does the document present defined environmental indicators?
	Are the indicators consistent with the significant environmental aspects?
	Are the evaluation criteria associated with the indicators specified?
	Are the results measured and presented in a usable way?
	Are performance trends visible or analyzable?
	Is the document up to date and traceable over time?
Monitoring and Measurement Recordings	Is the document proof of supervision or measurement?
	Are the measured parameters clearly identified?
	Are measurement dates and frequencies visible?
	Are the results usable for performance analysis?

	Is the document consistent with the supervision procedure?
	Does the data allow an assessment of environmental performance?
Procedure for the management of measuring equipment	Is the document identified, dated and versioned?
	Are the monitoring and measurement equipment identified?
	Are the calibration or verification requirements defined?
	Are the equipment maintenance procedures specified?
	Does the procedure ensure the reliability of the results?
	Is the evidence associated with calibration planned?
Environmental Performance Communication Procedure	Does the document specify what performance information is communicated?
	Are the modalities of internal communication defined?
	Are the modalities of external communication defined?
	Are the compliance obligations related to communication taken into account?
	Are the training courses communicated consistent with the measured data?
	Does the document provide for the retention of evidence of communication?
Regulatory Compliance Assessment Procedure	Is the document identified, dated and versioned?
	Are environmental compliance obligations taken into account?
	Does the procedure define the frequency of conformity assessment?
	Are the methods of conformity assessment described?
	Does the procedure provide for actions in the event of non-compliance?
	Is the maintenance of knowledge of the state of conformity ensured?
	Are the requirements for the retention of evidence defined?
Compliance Assessment Reports	Is the document a proof of conformity assessment?
	Are the items assessed clearly identified?
	Are the results of conformity or non-compliance explained?
	Are actions proposed or followed in the event of a gap?
	Does the document make it possible to know the state of conformity of the organization?
Internal audit procedure	Is the document identified, dated and versioned?
	Does the procedure define the objectives of the internal audit?
	Are the audit criteria defined?
	Is the scope of the audits specified?
	Are the ISO 14001 requirements taken into account?
	Are the roles and responsibilities of auditors defined?
	Are the rules of objectivity and impartiality specified?
	Are the methods for reporting the results described?
Internal Audit Program	Does the audit program cover a defined period?
	Is the frequency of the audits specified?
	Are the audited processes identified?
	Is the environmental importance of processes taken into account?

	Do the results of previous audits influence planning?
	Are the responsibilities clearly assigned?
Internal Audit Reports	Does the document constitute an internal audit report?
	Are the criteria and scope of the audit specified?
	Are the findings (conformities / non-conformities) clearly formulated?
	Is the evidence associated with the findings mentioned?
	Are the results communicated to the management concerned?
	Is the document usable for the improvement of the Environmental Management System?
Management Review Procedure	Is the document identified, dated and versioned?
	Is the frequency of management reviews defined?
	Are the input items required by ISO 14001 listed?
	Are the responsibilities related to the management review specified?
	Does the procedure provide for documented output elements?
	Is the link with continuous improvement explained?
Management Review Report	Is the document proof of a management review?
	Are the required input elements addressed?
	Are the conclusions on the effectiveness of the Environmental Management System formulated?
	Are the decisions on continuous improvement present?
	Are the actions to be taken identified?
	Are the resources and changes needed mentioned?
	Is the document usable for the strategic management of the Environmental Management System?
G. Improvement (ISO 14001)	
Environmental Management System Improvement Procedure	Is the document identified, dated and versioned?
	Does the procedure define how opportunities for improvement are identified?
	Are the sources of improvement resulting from surveillance (clause 9.1) taken into account?
	Are the results of internal audits (clause 9.2) used as a source of improvement?
	Are the conclusions of the management reviews (clause 9.3) integrated?
	Does the document provide for the implementation of actions to achieve the expected results of the Environmental Management System?
	Is the link between improvement and environmental performance explained?
Register / improvement action plan	Does the document identify the opportunities for improvement?
	Are the improvement actions clearly described?
	Is the origin of each action (audit, review, indicator, non-compliance, etc.) identifiable?
	Are the responsibilities associated with the actions defined?
	Are the deadlines specified?
	Does the document allow tracking the progress of actions?

	Are the actions aimed at improving the Environmental Management System or environmental performance?
Management procedureNon-conformities and corrective actions	Is the document identified, dated and versioned?
	Does the procedure define what environmental non-compliance is?
	Are the modalities of immediate reaction to non-compliance described?
	Are the control and correction of non-compliance planned?
	Are the consequences of non-compliance taken into account?
	Does the procedure provide for the analysis of the causes of non-compliance?
	Is the search for similar non-compliances integrated?
	Is the implementation of corrective actions supervised?
	Is the evaluation of the effectiveness of corrective actions planned?
	Is the possibility of modifying the Environmental Management System mentioned?
	Are the requirements for the retention of documented information defined?
Non-conformity sheets / registers	Does the document constitute evidence of identified non-compliance?
	Is the nature of non-compliance clearly described?
	Are the date and context of detection specified?
	Are real or potential environmental impacts mentioned?
	Are the immediate corrective actions described?
	Is the document usable for cause analysis?
Cause analysis (5 whys, Ishikawa, etc.)	Is the document related to an identified non-compliance?
	Is a structured analysis of the causes carried out?
	Are the root causes clearly identified?
	Is the used method formalized and understandable?
	Does the analysis prevent the reappearance of non-compliance?
	Is the document consistent with the non-compliance procedure?
Corrective Action Plan	Does the document list the corrective actions decided?
	Is every action related to an identified cause?
	Are the responsibilities assigned?
	Are the deadlines set?
	Are the actions proportional to the seriousness of the non-compliance?
	Does the document allow clear monitoring of implementation?
Evaluation of the effectiveness of corrective actions	Is the document proof of the evaluation of effectiveness?
	Are the criteria for effectiveness defined?
	Do the results show that non-compliance does not happen again?
	Have the environmental impacts been reduced or eliminated?
	Are additional actions considered if necessary?
	Is the document usable for continuous improvement?
	Does the document track the changes made to the Environmental Management System?

History of Changes to the Environmental Management System	Are the changes related to non-compliances or improvements?
	Are the reasons for the changes justified?
	Is the impact of the changes on the Environmental Management System assessed?
	Is the document consistent with other documented information?
Continuous improvement policy or approach	Is the document identified, dated and versioned?
	Is the commitment to continuous improvement explicit?
	Does the improvement deal with the relevance of the Environmental Management System?
	Does the improvement relate to its adequacy?
	Does the improvement relate to its effectiveness?
	Is the link with improving environmental performance clear?
	Is the document consistent with environmental policy?
Examples of improvements implemented	Does the document demonstrate a real improvement in the Environmental Management System?
	Are the actions implemented clearly described?
	Are the results obtained measurable or observable?
	Are positive environmental impacts identified?
	Does the document show a dynamic of continuous improvement?
A. Context of the organization (ISO 45001)	
Context Analysis (or Issue Register)	Is the document dated and versioned?
	Are relevant external issues identified and documented?
	Are the relevant internal issues identified and documented?
	Is each issue evaluated in terms of its impact on occupational health, worker safety, and occupational safety and health management system performance, including legal compliance, occupational risks, and occupational safety and health objectives?
	Is the methodology for identifying issues documented?
	Are the frequency or conditions for updating the issue analysis specified?
Register of Interested Parties	Are relevant stakeholders identified along with their type, influence, and occupational safety and health expectations?
	Are the needs and expectations of interested parties identified and defined through documented consultations, surveys, meetings, and worker representation mechanisms?
	Are the needs and expectations of other interested parties documented and reviewed periodically?
Legal Requirements Matrix	Are the needs or expectations constituting legal requirements identified?
	Are the needs or expectations of other requirements (contractual, voluntary, etc.) identified?
	Is the document kept as documented information and kept up to date?
Occupational Health and Safety Management	Is the scope of application available as documented information?
	Are the physical boundaries of the occupational safety, health, and environment management system defined?

System Application Perimeter	Are the organizational boundaries of the occupational safety, health, and environment management system defined?
	Is the scope of the SMSST consistent with the internal and external issues identified in the context analysis?
	Does the scope take into account the relevant requirements of the identified stakeholders?
	Are work-related activities, exercised or planned, taken into account?
	Do the activities, products and servicesUnder the control of the organization are included?
	Are the activities, products and services under the influence of the organization included?
B. Leadership (ISO 45001)	
Occupational Health and Safety policy	Is the document available as documented information?
	Does the document include a commitment to provide safe and healthy working conditions?
	Is the document appropriate for the purpose, size and context of the organization?
	Does the document provide a framework for the establishment of Occupational Health and Safety objectives?
	Does the document include a commitment to meet legal and other requirements?
	Does the document include a commitment to eliminate hazards and reduce Occupational Health and Safety risks?
	Does the document include a commitment to continuous improvement of the Occupational Health and Safety Management System?
	Does the document include a commitment to consult and participate workers?
	EIs the Occupational Health and Safety policy communicated, understood and accessible at all levels of the organization (e.g. posting, intranet, reception, training)?
	Is the document made available to interested parties, if applicable?
Procedure/Resource Management System	Does the document describe the process of allocating human, financial and material resources to the Occupational Health and Safety Management System?
	Does the document state that management ensures the availability of the necessary resources?
Organization Chart, Job Descriptions, or Occupational Health and Safety Responsibility Matrix	EIs the document kept up to date in the form of documented information?
	Are the responsibilities related to the Occupational Health and Safety Management System assigned and documented?
	Are the authorities related to the Occupational Health and Safety Management System assigned and documented?
	Does the document assign responsibility for compliance with ISO 45001?
	Does the document give management the responsibility for reporting on performance?
	Does the document demonstrate that roles and responsibilities are communicated at all levels?
Consultation and Worker Participation Procedure	Does the document describe the consultation processes and worker participation?
	Does the document provide for the modalities and time necessary for consultation and participation?
	Does the document provide for the necessary training for consultation and participation?

	Does the document provide workers with access to clear and relevant information on the Occupational Health and Safety Management System?
	Does the document describe the process of identifying and removing barriers to participation?
	Does the document mention support for the establishment of Occupational Health and Safety committees (if applicable)?
	Does the document include provisions to protect workers from reprisals?
Minutes of the Occupational Health and Safety Committee Meetings	Do the meeting minutes demonstrate the consultation of workers on Occupational Health and Safety policy?
	Do the meeting minutes demonstrate the participation of workers in the identification of hazards?
	Do the meeting minutes demonstrate the participation of workers in the determination of risk reduction actions?
	Do the meeting minutes show that the dirEction reports on the effectiveness of the Occupational Health and Safety Management System?
C. Planning (ISO 45001)	
Procedure for Hazard Identification, Risk and Opportunity Evaluation	Does the document describe an identification process?
	Do we continue and proactively of the dangers?
	Does the document describe an Occupational Health and Safety risk assessment process?
Procedure for Hazard Identification, Risk and Opportunity Evaluation	Does the document describe a process for assessing other risks related to Occupational Health and Safety Management System.?
	Does the document describe a process of evaluating Occupational Health and Safety opportunities?
	Does the document describe a process of evaluation ofOpportunities for improving the Occupational Health and Safety Management System.?
	Does the document specify that the organization of work is taken into account in the identification of hazards?
	Does the document specify that past adverse events are taken into account?
	Is there evidence that risks are assessed before any planned, permanent or temporary change (e.g. MOC, projects, reorganizations)?
Occupational Health and Safety Risk Assessment Matrix/Grid	Does the document define the Occupational Health and Safety risk assessment methods?
	Does the document define the Occupational Health and Safety risk assessment criteria?
	Does the method make it possible to identify and treat risks before incidents occur (e.g. task analysis, non-routine situations, changes)?
	Is the document kept as documented information?
Register of Hazards, Risks and Opportunities (or Single Document)	Does the document identify all hazards, including those related to people?
	Does the document assess the Occupational Health and Safety risks arising from the identified hazards?
	WhatDoes the document evaluate the effectiveness of existing prevention measures?
	Does the document evaluate Occupational Health and Safety opportunities, including work adaptation to workers?

	Does the document take into account the issues, requirements and scope?
	EIs the document kept up to date in the form of documented information?
Register of Legal and Other Requirements	Are hazards identified for all activities, normal, abnormal and emergency situations, including fHuman and psychosocial actors?
	Are the associated Occupational Health and Safety risks evaluated according to the defined method?
	The effectiveness of pIs existing revention evaluated and documented?
	Do Occupational Health and Safety opportunities include adapting work to workers' abilities?
	Is the register consistent with the scope, issues and applicable requirements?
Occupational Health and Safety Action Plan (6.1.2)	Does the plan define the actions to deal with risks and opportunities? EQUAL FOR OTHERS
	Does the plan define the actions to meet legal and other requirements (6.1.3)?
	Does the plan define actions related to emergency situations (8.2)?
	Does the plan describe how actions will be integrated into business processes?
	Does the plan describe how the effectiveness of the actions will be evaluated?
	Is itE that the plan demonstrates the consideration of the hierarchy of prevention measures (8.1.2)?
Action Plan for the Achieving of Occupational Health and Safety Objectives	Does the document define the Occupational Health and Safety objectives (6.2.1)?
	Are the objectives consistent with the Occupational Health and Safety policy?
	Are the objectives measurable when possible?
	Does the plan specify what will be done to achieve each objective?
	Does the plan specify the resources needed for each objective?
	Does the plan specify those responsible for each action?
	Does the plan define the deadlines for each action?
	Does the plan define the indicators necessary to evaluate the results?
	Is the document kept up to date and conServed as documented information?
D. Support (ISO 45001)	
Occupational Health and Safety Resource Allocation Plan (or Manual/Budget chapter)	Does the document identify the human, technical and financial resources necessary for the Occupational Health and Safety Management System.?
	Does the document provide evidence that the necessary resources have been provided or are available?
Occupational Health and Safety Skills Matrix / Job Sheets	Does the document determine the skills required for positions that have an impact on Occupational Health and Safety performance?
	Does the document specify the skill required to identify hazards?
Individual Personnel Files (CV, Diplomas, Training Certificates, Certificates)	Does the file contain evidence of relevant initial or vocational training?
	Does the file contain evidence of the worker's proper experience?
	Does the file contain documented information demonstrating the skills of the worker?
	Does the file contain an evaluationOn the effectiveness of the training followed?

Occupational Health and Safety Awareness/Consciousness Program	Does the program include awareness of Occupational Health and Safety policy?
	Does the program include awareness of Occupational Health and Safety objectives?
	Does the program inform about the consequences of non-compliance with Occupational Health and Safety Management System. requirements?
	Does the program inform about the dangers, risks and actions put in place?
	Does the program informWorkers of their right of withdrawal in the event of serious and imminent danger?
Internal and External Communication Procedure	Does the procedure clearly identify the topics to be communicated?
	Does the procedure clearly identify the times when communication should take place?
	Does the procedure clearly identify the internal and external parties to whom to communicate?
	Does the procedure clearly identify the communication methods to be used?
	Does the procedure take into coMpte diversity (languages, disabilities, access to information...)?
	Does the procedure describe how the organization reacts to relevant observations related to the Occupational Health and Safety Management System.?
	Does the procedure guarantee that the information communicated is reliable and consistent?
	Does the procedure ensure that workers can contribute to the continuous improvement of the Occupational Health and Safety Management System?
Evidence of Communication (E.g. Memos, Emails, Meeting Minutes, Dated Displays)	Do the documents prove the internal communication of relevant information from the Occupational Health and Safety Management System.?
	Do the documents prove the internal communication of the changes made to the Occupational Health and Safety Management System.?
	Do the documents prove that the feedback of external interested parties has been taken into account?
	Do the documents prove the communicationExternal ation of the relevant information of the Occupational Health and Safety Management System.?
Procedure for Controlling Documented Information	Does the document detail the identification and description of the information (title, date, reference)?
	Does the document specify the management of the format and support of the documented information?
	Does the document describe the review and approval of documents before dissemination?
	Does the document detail the control of the distribution, access and retrieval of documents?
	Does the document detail the control of the storage and protection of documents (confidentiality, integrity)?
	Does theDocument details the control of changes and version management?
	Does the document detail the control of the preservation and disposal of documents?
	Does the document describe the mastery of documented information of external origin?
E. Operation (ISO 45001)	
	Does the document establish clear criteria for carrying out the process concerned?

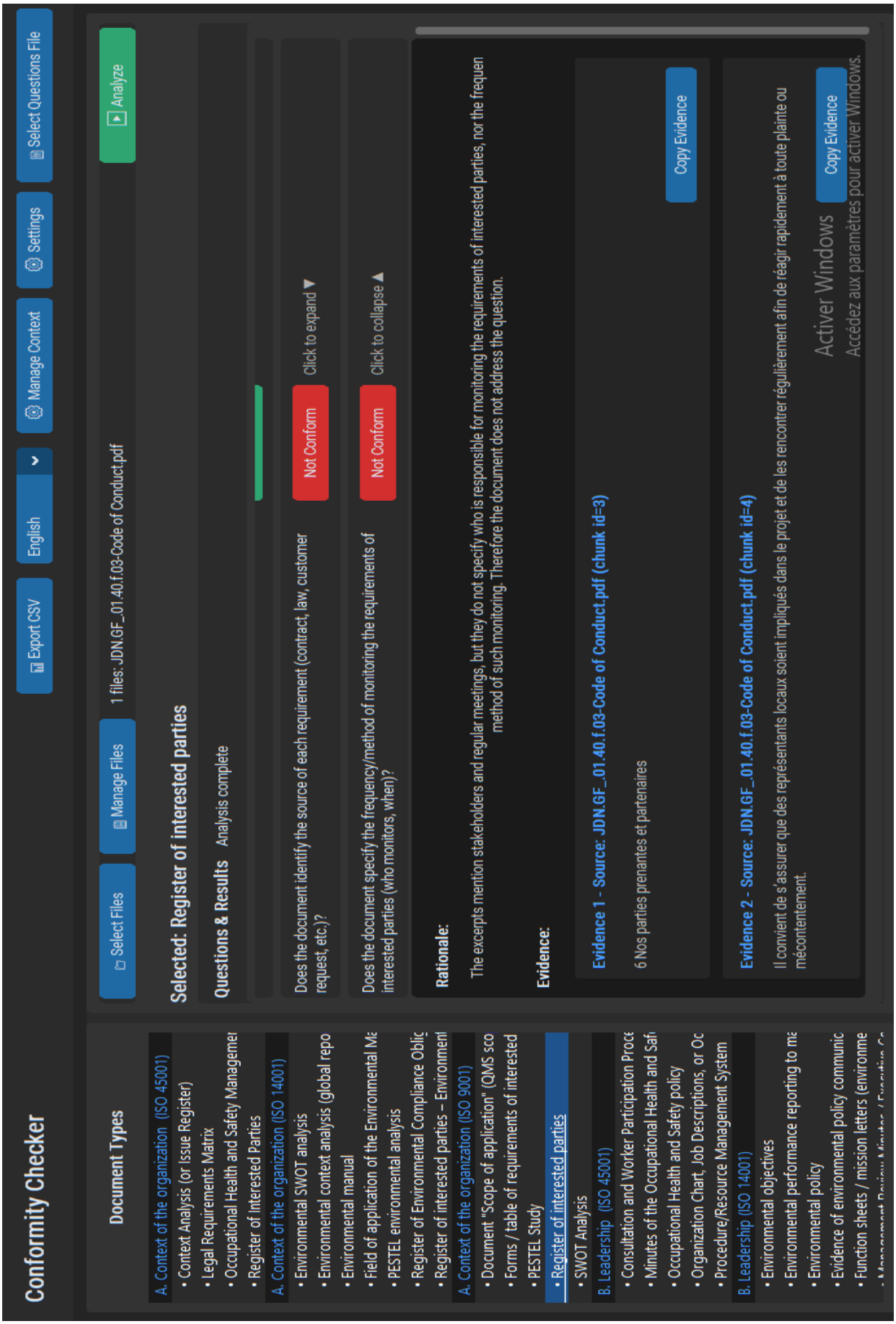
Occupational Health and Safety Operational Procedures	Does the document describe the implementation of mastery in accordance with these criteria?
	Does theDocument mentions the adaptation of work to workers (ergonomics)?
	Does the document describe the coordination of Occupational Health and Safety processes for multiple employer workplaces?
	Is the document kept as documented information proving that the process was carried out as planned?
Risk Management Procedure (Hierarchy of Prevention Measures)	Does the document describe a process based on the hierarchy of prevention measures?
	Does the document prioritize the elimination of danger as the first level of control?
	Does the document mention substitution with less hazardous elements as the second level of control?
	Does the document mention collective protection measures and work organization as the third level of control?
	Does the document mention administrative measures, including training, as the fourth level of control?
	Does the document mention the use of personal protective equipment as the fifth level of control?
Change Management Procedure (Management of Change)	Does the document establish a process of controlling temporary changes?
	What Does the document establish a process of controlling permanent change?
	Does the process include changes related to new products, services or processes?
	Does the process include changes resulting from changing legal requirements?
	Does theThe process requires the analysis of the consequences of unforeseen changes?
Procedure for the Acquisition of Goods and Services (Purchases)	Does the document describe a process to control the acquisition of products and services?
	Does the document guarantee that the products and services acquired comply with the Occupational Health and Safety Management System?
Management Procedure for External Stakeholders and/or Contracts	East-That the document describes coordination with external stakeholders to identify dangers?
	Does the document define and apply Occupational Health and Safety criteria for The selection of external stakeholders?
	Does the document guarantee that external stakeholders meet the Occupational Health and Safety Management System requirements?
	Does the document cover the Occupational Health and Safety risks generated by the organization's activities that may affect external stakeholders?
	Does the document cover Occupational Health and Safety risks generated by external stakeholders that may affect the organization?
Outsourcing Management Procedure and/or Outsourcing Contracts	Does the document define the Type and degree of control to be applied to outsourced functions and processes?
	Does the document guarantee that the outsourced functions are compliantS to the applicable legal requirements?
	Does the document guarantee that the outsourced functions make it possible to achieve the expected results of the Occupational Health and Safety Management System?
	Does the document exist as documented information?

Procedure and Plans for Preparedness and Response to Emergency Situations	Does the planEstablishes a planned response to emergency situations, including first aid?
	Does the document provide for the training necessary to implement this response?
	Does the document provide for the periodic performance of tests and exercises?
	Does the document provide for the communication of obligations and responsibilities to workers?
	Does the document provide for the communication of informationRelevant to emergency services and public authorities?
	Does the document take into account the needs and capacities of all relevant stakeholders?
	Does the document specify the performance evaluation and the review of the planned response?
F. Performance of evaluation (ISO 45001)	
Performance Monitoring, Measurement, Analysis and Evaluation Procedure	Does the document clearly determine what needs to be monitored and measured?
	Does the document define the methods of monitoring, measurement, analysis and evaluation to be applied?
	Does the document define the criteria Used to evaluate performance in Occupational Health and Safety?
	Does the document determine when the monitoring and measurement must be carried out?
	Does the document describe the evaluation of the effectiveness/effectiveness of operational prevention measures?
	Does the document include the assessment of progress towards achieving Occupational Health and Safety objectives?
Monitoring and Measurement Reports / Dashboards (Key Performance Indicator Occupational Health and Safety)	Does the document keep the evidence of the results Monitoring, measurement, analysis and evaluation?
	Does the report demonstrate monitoring and measurement of the degree of satisfaction with legal requirements?
	Does the report demonstrate the performance evaluation in Occupational Health and Safety?
	Does the rappOrt demonstrates the determination of the effectiveness/effectiveness of the Occupational Health and Safety Management System ?
Measurement Equipment Maintenance/ Calibration Files	Does the file contain evidence of calibration or verification of the measurement equipment?
	Does the file demonstrate the proper maintenance of monitoring and measurement equipment?
Legal Compliance Assessment Procedure	Does theDocument defines the frequency of evaluation of legal compliance?
	Does the document define the method used for the assessment of legal compliance?
	Does the document provide for corrective action when a non-compliance is identified?
	Does the document describe how the organization maintains its knowledgeAnd his understanding of its compliance?
Legal Compliance Assessment Reports	Does the document keep the documented information on the results of the conformity assessments?
	Does the report prove that the conformity assessment was carried out at the defined frequency?

Occupational Health and Safety Internal Audit Program	Does the document plan the audits taking into account the importance of the processes concerned?
	Does the document define the audit criteria for each planned audit?
	Does the document define the perimeter of each planned audit?
	Does the audit program take into account the results of previous audits?
	Is the document kept as proof of the implementation of the audit program?
	Does the document demonstrate that the audit has verified compliance with internal requirements (Policy, Objectives)?
	Does the document show that the audit verified compliance with the requirements of ISO 45001?
	Does the document demonstrate that the audit was conducted objectively and impartially?
	Does the document demonstrate that the audit results have been communicated to the relevant management personnel?
	Does the document demonstrate that the audit results have been communicated to workers and their representatives?
	Does the document contain the identification of non-conformities and/or opportunities for improvement?
	Is the document kept as proof of the audit results?
Minutes (meeting minutes) of Occupational Health and Safety Management Review	Does the meeting minutes prove that the management review took place at scheduled intervals?
	Does the meeting minutes take into consideration the follow-up of the actions decided during the previous reviews?
	Does the meeting minutes take into account changes in internal and external issues?
	Does the meeting minutes analyze the level of achievement of the Occupational Health and Safety Policy and Objectives?
	Does the meeting minutes analyze adverse event trends and corrective actions?
	Does the meeting minutes analyze the results of the monitoring and measurement?
	Does the meeting minutes analyze the results of the evaLuation of legal compliance?
	Does the meeting minutes analyze the internal audit results?
	Does the meeting minutes include decisIons regarding the adequacy, relevance and effectiveness of the Occupational Health and Safety Management System?
	Does the minutes include decisions on the resources needed?
	Does the meeting minutes include decisions regarding opportunities for continuous improvement?
	Does the minutes show that the relevant exit elements have been communicated to the workers?
G. Improvement (ISO 45001)	
Procedure for Managing Adverse Events, Non-Conformities and Corrective Actions (NC/AC)	Does the document describe a process of reporting and analyzing adverse events?
	Does the document describe a process of reporting and analyzing non-conformities?
	Does the document require quick action to control and correct the event or the non-conformity?
	Does the document require the participation of workers in the assessment of the need for corrective action?
	Does the doCument requires research and analysis of fundamental causes?
	Does the document require the identification of adverse events or similar nonconformities?
	Does the document require the revision of existing risk assessments where relevant?

	Does the document require the determination of corrective actions in accordance with the hierarchy of prevention measures?
	Does the document require an assessment of the risks related to the new hazards before taking corrective action?
	Does the document require the review of the effectiveness/effectiveness of the corrective actions implemented?
RappOrts of Adverse Events / Non-Compliance and Corrective Action Sheets	Does the document retain the Proof of the nature of the adverse event or non-compliance?
	Does the document retain proof of the initial mastery and correction actions?
	Does the document retain evidence of the analysis of the root causes?
	Does the document retain evidence of the corrective actions implemented?
	Does the document retain proof of effectiveness/Effectiveness of corrective action?
	Does the document prove that the corrective actions were proportionate to the consequences?
	Does the document prove the communication of the results of corrective action to the workers concerned?
Occupational Health and Safety Performance Reports / Continuous Improvement Plan (PAC)	Does the document show that the organization has identified opportunities for improvement?
	Does the document demonstrate the implementation of actions aimed at to improve performance in Occupational Health and Safety?
	Does the document demonstrate actions promoting a positive culture of the Occupational Health and Safety Management System ?
	Does the document prove the participation of workers in continuous improvement actions?
	Does the document prove the communication of the results of continuous improvement to the workers concerned?
	Is the document kept as proof of continuous improvement?

Appendix B – Documentary Analysis Platform Interface



Appendix C – Test 1: Results Reproducibility Assessment

Test 1	0 = NC	1 = PC	2 = C
documents	QUESTION	RUN 1	RUN2
D1 Continuous Improvement Plan	Q1	1	1
	Q2	1	1
	Q3	1	1
	Q4	2	2
	Q5	1	1
D2 Environmental Emergency Plan	Q1	1	1
	Q2	2	2
	Q3	2	2
	Q4	1	2
	Q5	1	1
	Q6	1	0
D3 Environmental manual	Q1	2	2
	Q2	2	2
	Q3	2	2
	Q4	2	2
	Q5	2	2
D4 Internal audit reports 9001	Q1	1	1
	Q2	2	2
	Q3	2	2
	Q4	2	2
D5 Management procedureNon-conformities and corrective actions	Q1	2	2
	Q2	0	0
	Q3	2	2
	Q4	2	2
	Q5	2	2
	Q6	2	2
	Q7	0	0
	Q8	2	2
	Q9	2	2
	Q10	0	0
	Q11	2	2
D6 Minutes (meeting minutes) of Occupational Health and Safety Management Review	Q1	1	1
	Q2	2	2
	Q3	1	1
	Q4	0	0
	Q5	2	1

	Q6	1	1
	Q7	0	0
	Q8	0	0
	Q9	0	0
	Q10	2	2
	Q11	1	1
	Q12	0	0
D7 Occupational Health and Safety Operational Procedures	Q1	2	1
	Q2	0	0
	Q3	0	0
	Q4	1	1
	Q5	0	0
D8 Occupational Health and Safety Policy	Q1	2	2
	Q2	2	2
	Q3	2	1
	Q4	1	2
	Q5	2	2
	Q6	2	2
	Q7	2	2
	Q8	2	2
	Q9	2	2
	Q10	2	2
D9 Organization chart or organizational sheet	Q1	0	0
	Q2	0	0
	Q3	0	0
	Q4	0	0
	Q5	1	1
D10 Planning and operational control procedure	Q1	2	2
	Q2	2	2
	Q3	1	0
	Q4	2	2
	Q5	2	2
	Q6	0	0
	Q7	2	2
	Q8	2	2
D11 Quality manual	Q1	2	2
	Q2	1	0
	Q3	2	2
	Q4	1	2
	Q5	2	2
D12 Register of Environmental Risks and Opportunities	Q1	0	0

	Q2	0	0
	Q3	2	2
	Q4	0	0
	Q5	0	1
	Q6	2	2
	Q7	2	2
	Q8	1	2
	Q9	0	0
	Q10	2	2
	Q11	2	2
	Q12	1	1

Appendix D – Test 2 : Results – Robustness to Variations in Documentary Form

Test 2 :	0 = NC	1 = PC	2 = C	
documents	QUESTION	ORIGINAL	REFORMULER	NEUTRE
D3 Environmental manual	Q1	2	2	2
	Q2	2	2	2
	Q3	2	2	2
	Q4	2	2	2
	Q5	2	2	2
D6 Minutes (meeting minutes) of Occupational Health and Safety Management Review	Q1	1	1	1
	Q2	2	2	2
	Q3	1	2	2
	Q4	0	0	0
	Q5	2	2	1
	Q6	1	1	1
	Q7	0	0	0
	Q8	0	0	0
	Q9	0	0	0
	Q10	2	2	1
	Q11	1	0	2
	Q12	0	1	0
D7 Occupational Health and Safety Operational Procedures	Q1	2	1	1
	Q2	0	0	0
	Q3	0	0	0
	Q4	1	0	1
	Q5	0	0	0
D11 Quality manual	Q1	2	2	2
	Q2	1	2	0
	Q3	2	2	2
	Q4	1	2	1
	Q5	2	2	2
D12 Register of Environmental Risks and Opportunities	Q1	0	1	1
	Q2	0	0	0
	Q3	2	2	2
	Q4	0	0	0

	Q5	0	1	0
	Q6	2	2	2
	Q7	2	1	1
	Q8	1	2	2
	Q9	0	0	0
	Q10	2	1	1
	Q11	2	0	0
	Q12	1	1	1

Appendix E – Test 3: Results of the Sensitivity to the Detection of Expected Non-Conformities

Test 3	0 = NC	1 = PC	2 = C
documents	QUESTION	Expected Number of NC	VERDICT
D2 Environmental Emergency Plan	Q1	_	1
	Q2	_	2
	Q3	_	2
	Q4	_	1
	Q5	x	1
	Q6	_	1
D5 Management procedureNon-conformities and corrective actions	Q1	_	2
	Q2	x	0
	Q3	_	2
	Q4	_	2
	Q5	_	2
	Q6	_	2
	Q7	x	0
	Q8	_	2
	Q9	_	2
	Q10	_	0
	Q11	_	2
D10 Planning and operational control procedure	Q1	_	2
	Q2	_	2
	Q3	_	1
	Q4	_	2
	Q5	_	2
	Q6	_	0
	Q7	_	2
	Q8	_	2

Appendix F – Test 4: Logical Consistency Before and After Documentary Correction

Test 4	0 = NC	1 = PC	2 = C	
documents	QUESTION	Association with the NC	VERDICT 1	VERDICT 2
D5 Management procedureNon-conformities and corrective actions	Q1	no	2	2
	Q2	yes	0	2
	Q3	no	2	2
	Q4	no	2	2
	Q5	no	2	2
	Q6	no	2	2
	Q7	yes	0	2
	Q8	no	2	2
	Q9	no	2	2
	Q10	yes	0	2
	Q11	no	2	2
D12 Register of Environmental Risks and Opportunities	Q1	yes	0	2
	Q2	yes	0	2
	Q3	no	2	2
	Q4	yes	0	2
	Q5	yes	0	2
	Q6	no	2	2
	Q7	no	2	2
	Q8	no	1	2
	Q9	yes	0	2
	Q10	no	2	2
	Q11	no	2	1
	Q12	no	1	2