computers

Article

A Methodological Framework for Designing Personalised Training Programs to Support Personnel Upskilling in Industry 5.0

Francisco Fraile [1], Foivos Psarommatis [1,2,*], Faustino Alarcón [1] and Jordi Joan [3]

[1] Research Centre on Production Management and Engineering (CIGIP), Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia, Spain; ffraile@cigip.upv.es (F.F.); faualva@omp.upv.es (F.A.)
[2] SIRIUS, Department of Informatics, University of Oslo, Problemveien 11, 0313 Oslo, Norway
[3] VRAIN—Valencian Research Institute for Artificial Intelligence, Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia, Spain; jlinares@dsic.upv.es
* Correspondence: foivosp@ifi.uio.no

Abstract: Industry 5.0 emphasises social sustainability and highlights the critical need for personnel upskilling and reskilling to achieve the seamless integration of human expertise and advanced technology. This paper presents a methodological framework for designing personalised training programs that support personnel upskilling, with the goal of fostering flexibility and resilience amid rapid changes in the industrial landscape. The proposed framework encompasses seven stages: (1) Integration with Existing Systems, (2) Data Collection, (3) Data Preparation, (4) Skills-Models Extraction, (5) Assessment of Skills and Qualifications, (6) Recommendations for Training Program, (7) Evaluation and Continuous Improvement. By leveraging Large Language Models (LLMs) and human-centric principles, our methodology enables the creation of tailored training programs to help organisations promote a culture of proactive learning. This work thus contributes to the sustainable development of the human workforce, facilitating access to high-quality training and fostering personnel well-being and satisfaction. Through a food-processing use case, this paper demonstrates how this methodology can help organisations identify skill gaps and upskilling opportunities and use these insights to drive personnel upskilling in Industry 5.0.

Keywords: NLP; Large Language Models; skills extraction; workers upskilling; Zero Defect Manufacturing; Industry 5.0

check for updates

Citation: Fraile, F.; Psarommatis, F.; Alarcón, F.; Joan, J. A Methodological Framework for Designing Personalised Training Programs to Support Personnel Upskilling in Industry 5.0. *Computers* **2023**, *12*, 224. https://doi.org/10.3390/computers12110224

Academic Editors: Paolo Bellavista, Katia Lida Kermanidis, Manolis Maragoudakis and Phivos Mylonas

Received: 7 September 2023
Revised: 26 October 2023
Accepted: 31 October 2023
Published: 2 November 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Contemporary manufacturing is undergoing radical changes with the adoption of Industry 4.0 and the need for higher manufacturing sustainability. Therefore, manufacturers are constantly seeking to adopt new technologies that will allow them to be more efficient and sustainable [1]. To that end, it is necessary to educate employees regarding the new digital, data-based, knowledge-based, and interoperable technologies. The rapid technological advancements in the manufacturing and Information and Communication Technologies (ICT) domains make manufacturing education critical [2,3]. As they combine digital technologies, manufacturing systems are becoming increasingly complex [4–6]. Therefore, new technologies require new educational methods to update or upgrade the skills of engineers and blue-collar workers [7–9]. Especially in the manufacturing domain, the education of personnel is a necessary and continuous process for keeping up with the pace of changes, increasing the safety and sustainability of manufacturing systems, achieving high-quality production with the least amount of resources, and staying competitive [10]. Furthermore, careful attention to the human aspect in manufacturing systems is mandatory not only from the business or economic perspective, but also with regard to the social aspect of manufacturing [11,12]. This approach is highly aligned with the concept of Industry 5.0, which was newly introduced by the European Commission. Industry 5.0 is not completely new; it has deep roots in Industry 4.0, but the difference lies in the fact that the goal of Industry 5.0 is not only to produce goods and services for profit, but to give great focus to environmental and social aspects of manufacturing [13–15].

In this Industry 5.0 context, quality control and assurance is one of the most crucial domains in manufacturing, and labour training is essential [16–19]. This domain is facing radical changes, and many new technologies and approaches are being introduced, rendering traditional methods and technologies obsolete. Quality control and assurance is a vital part of any manufacturing system. Achieving high-quality production is something that manufacturers are aiming for and spend significant resources to accomplish, but they are not always able to achieve this goal. Here, "quality" refers not only to the product quality, but also to process quality [10,20,21]. Production quality is related directly to manufacturing sustainability, which is a key business target for manufacturing companies. Poor-quality operations have multi-level negative implications, from direct economic losses to indirect losses, such as the negative impact on the reputation of the company [22,23]. To achieve high product and process quality, manufacturing companies are using various quality-management tools to improve their operational performance [24]. The latest approach for quality assurance is Zero Defect Manufacturing (ZDM) [10,21]. ZDM imposes some new rules; therefore, the training of engineers and blue-collar workers in ZDM technologies is imperative.

The literature has identified numerous factors that discourage companies from offering training courses to their employees. The most important such factors are the cost and the working time required for the courses [25]. Due to the investment that companies should make in the training of their employees, it is of vital importance to identify the skills of each employee and to adapt their training sessions to the acquisition of new competences. Therefore, in this paper, we propose the use of sentence embeddings using transformers and Natural Language Processing (NLP) to identify the skills of each worker and at the same time identify areas in which each worker may require training [26,27]. This approach will give manufacturers the opportunity to offer personalised training courses to each worker and thus save time and money, while giving their employees the exact knowledge that is required to complete their assigned tasks.

Our proposal relies on analysis of the internal communications issued by personnel during operations. Many companies use information systems in which personnel describe events that occurred during the manufacturing process using natural language. One prominent example is maintenance incident requests, in which operators report failures that require intervention by the maintenance department. These communications capture the tacit knowledge, skills, and competences of personnel, as each communication uses language and terminology specific to the skill domain (for instance, electric work). State-of-the-art Large Language Models (LLMs) allow to companies to obtain insight into these internal communications. This research work focuses on the improvement of companies' training and upskilling programs through the extraction of this tacit knowledge of personnel skills from unstructured data and comparison of the extracted model with an explicit model of the skills provided by qualification results. For instance, this comparison allows companies to identify the interests of personnel in specific competence areas for which they have not yet acquired qualifications and to identify areas in which training needs to be reinforced. From this analysis, it is possible to design personalised training programs to support personnel training and upskilling.

The rest of the paper is structured as follows: Section 2 contains a description of the state of the art in worker training in Industry 5.0, ZDM, and NLP. Section 3 describes the methodological framework of the proposal. Section 4 provides details about the implementation use case. Section 5 shows the results, and Section 6 provides some final conclusions and remarks.

2. State of the Art

In Industry 4.0 and Industry 5.0, digitalization is one of the most important drivers of innovation in reducing waste and improving overall quality, hence saving cost and time and optimising production systems. Digital technologies can be applied to all manufacturing stages, including product design, the overall operation of company, and personnel training [28,29]. Training is key to allowing companies to stay competitive; therefore, operators and employees should receive regular up-to-date training to improve not only process performance, but also safety and sustainability [30–33]. Furthermore, training could increase responsiveness to unexpected events and foster the knowledge required to prevent them, increasing the companies' resilience [33,34].

However, despite the importance of training to companies' competitiveness, few frameworks exist in the literature for its empowerment and management. One such framework is the Conceptual Learning Framework proposed by [35], the goal of which is to understand the essential future skills that are expected from the digital workforce. That exploratory study identified nine critical skills in three categories: cognitive and metacognitive skills, social-emotional skills, and practical skills. That study further identified community management, data-analytic skills, and web-development skills as critical. One of the important conclusions of these authors was that organizations should monitor the abilities of their employees and offer opportunities for individuals to develop their skills, which is precisely the basis of our work.

In this line, ref. [36] highlights current efforts to implement and research personalised learning within K-12 and Higher Education and suggests that personalized learning may be a promising strategy to solve many of the problems of workforce training and development programs. In response, they propose a Personalized Learning Interaction Framework based on a three-level interaction model and five types of interactions (among learners, coaches, Artificial Intelligence Assistants, small groups, networks, etc.). The findings suggest that personalized learning may be an effective strategy to increase engagement in workforce training and development programs.

On the other hand, ref. [37] propose a six-step methodological framework with the objective of determining whether it is possible to create a skill-focused climate in a software organization through a software system proposed by the researcher and whether such a system will help in utilizing in-house employees for new opportunities, rather than hiring new employees. The steps of the methodological framework are: (1) identify minimum viable features; (2) identify the necessary technologies and approaches; (3) design, implement and test; (4) prepare a questionnaire for HR and software professionals; (5) obtain feedback on the skill-based system; and (6) evaluate feedback to generate insights. Additionally, the authors model the workflow among the actors in the system (employee, manager, skills expert).

In addition to these frameworks, other frameworks and literature on personalised learning and education can be found [38,39], but no NLP-based frameworks like the one proposed in this work have been found.

The rest of the section is organised into operators training, ZDM, and NLP to identify the main functions and concepts and thus to allow the design of a methodology for skills extraction.

2.1. Personnel Training

The most common approaches for personnel training are classroom or online lectures, online learning packages, pilot-plant exercises for hands-on training, and computer simulations [28,40]. Studies have shown that, very frequently, lectures do not have the desired outcome, as they do not provide an engaging and challenging experience [41–43]. In general, lectures and massive-passive methods no longer seem to adapt well to the training needs of current workers, especially those from younger generations, who are accustomed to consuming on-demand audiovisual content in their private lives.

On the other hand, the rise of Industry 4.0 and the growing incorporation of all the technologies associated with this concept require a renewal of the skills and knowledge of workers [44–47]. Some researchers try to find solutions to address this new challenge and analyse the advantages and possibilities of new training methodologies. Another study introduced an approach for familiarising both low- and high-experience workers with Industry 4.0 hybrid workstations and assembly procedures using an intuitive training methodology [48]. Furthermore, they propose that the training should be performed on site, at the physical workstation during production, to accelerate learning. However, this practice could impact normal production performance and might not always be advisable.

Immersive experiences may help companies respond to the industry's requirement for more engaging training methodologies. Overall, the data support the idea that a growing number of immersive technologies are being used for operator training [48,49]. It is projected that the number of immersive training apps will continue to rise in the coming years. For example, immersive experiences have been used in the chemical, nuclear, manufacturing, and industrial environments [28]. Even though each industry's ultimate output differs, they have components in common, such as hazardous working conditions and technical or functional complexity. Immersive experiences have gained popularity as employee training due to the ability to train and to perform tasks safely (in a virtual environment when the tasks would be too dangerous, very expensive or even impossible to perform in the real world) [50,51]. The research by García et al. emphasises the essential issue of evaluating the training experience [28]. Despite its importance, these authors highlight that only a tiny percentage of researchers carry out and report a comparison of immersive training with standard approaches along with the performance metrics that were measured. The findings show that more work is needed to ensure a thorough examination of the usefulness and efficiency of immersive experiences in the process industry. To this end, they construct a model from the learner's perspective to analyse immersive experiences in the process industry, aided by the analysis of performance indicators. This model constitutes a starting point for determining which parameters could be used to evaluate the effectiveness and efficiency of these experiences.

The state of the art now shows a clear picture of the research and application of recent immersive technologies and developments. According to the findings, the three key immersive technologies used for operator training that are most widely mentioned in the literature are, in this order [28], 3D immersive training, Virtual Reality (VR), and Augmented Reality (AR). This categorization of the types of training was helpful in separately clarifying the use of each one and made it possible to ascertain that some beneficial and promising objectives are not achieved because the training was not implemented in emergencies and dangerous situations.

3D immersive training includes all developments that create a 3D representation of a specific environment or situation within which certain tasks can be executed [28,52,53]. This 3D representation can be visualised on a computer screen, on a powerwall, or in a cave automatic virtual environment. Some modern experiences in 3D immersive training that have been applied to an industrial context include simulations of industrial plants, where operators can move around the plant to learn about the different plant sections [54,55] and about some of the elements or devices inside. This application represents an essential advantage of virtual environments in training workers for extreme and dangerous situations that, with these technologies, can be reproduced in safe conditions.

Virtual reality (VR) refers to virtual environments reproduced by using a head-mounted display (HMD) or VR headset [28]. This technology is also a leader in training systems because it can offer interactive training environments, which can simulate testing and execution in the virtual environment prior to the execution of the actual processes in the production line [56,57]. With such training systems, production productivity and safety are increased significantly. Another study proposed a structured approach to assess operator performance and cognitive conditions during assembly training using virtual reality applications [58–60]. The findings of these experiments will enable the human-centred optimization of the VR training app, resulting in an improvement in the operators' knowledge and skills. However, although VR technology is very promising, there is still a need for research to make VR a viable training method [53].

Another contemporary technology, AR, has been used for personnel training. AR overlays virtual information with the real world and shows it simultaneously as a real-time interaction [48,52]. In most literature applications, AR has been used to familiarise operators with assembly operations and digital technologies. AR can help operators understand a problem and even show the next steps to solve it, providing additional relevant information. This information can be, for example, photos, videos, flow diagrams, instructions, technical documentation, voice assistance, and 3D models or representations. The information can include warnings, alerts or anything that can simplify or complement the operator's understanding of the situation. For these reasons, AR training is primarily used to provide virtual guidance in the industry [41]. Among the main advantages of using this technology are the reduction of errors in problem solving, which is directly related to reducing the risk of accidents, and the reduction in the time needed to complete tasks.

Operator 4.0 is another important concept within the Industry 4.0 paradigm and describes the operators that have become familiar and use Industry 4.0 technologies [61–64]. In modern manufacturing systems, Industry 4.0 technologies are being increasingly used, which this creates the need to adopt the Operator 4.0 approach and to shift from traditional training approaches to training models that utilize more engaging approaches and digital technologies. In this setting, a human-centred approach is required for the development of effective training pathways [65]. The usefulness of modern training methods in this new context seems obvious, and research about this topic has been growing in recent years; however, to the authors' best knowledge, there are no studies in the literature discussing immersive training experiences tailored to each individual.

2.2. Zero Defect Manufacturing (ZDM)

Manufacturing companies have traditionally used at least one Quality Improvement (QI) strategy to maintain and improve quality during production while minimising performance loss [20,66]. Six sigma (SS), lean manufacturing (LM), theory of constraints (TOC), and total quality management (TQM) are all traditional QI approaches, but they all focus on product quality and use statistical tools to improve it. The modern manufacturing landscape has shifted away from statistical methodologies and toward data-driven technology as a result of the Industry 4.0 paradigm. This shift results from the fact that classical QI approaches were created utilising the technologies available at the time, without recent technical breakthroughs, in the context of Industry 4.0 [66]. Artificial intelligence, machine learning, and semantic modelling, for example, open up an entirely new horizon for manufacturers, allowing them to achieve goals that were previously unattainable [67]. The significant rise in computational power and reductions in the price of sensors have further expanded the usage of data-driven technologies [68–71].

These circumstances resulted in the development of a new quality management technique known as Zero-Defect Manufacturing (ZDM) [21], which originated in the discrete-production domain but is now also relevant to the continuous-manufacturing domain. The difference between traditional QI and ZDM is that the latter is a holistic approach that looks at all stages of production, from product design to production and finally to the actual operation of the system [20]. ZDM aims to eliminate waste and improve the long-term viability of industrial systems [10]. This aim is accomplished using four ZDM strategies: detection, prediction, prevention, and repair. Those ZDM techniques are used in pairs according to Psarommatis et al.: detecting a problem, repairing it, attempting to prevent future defects, and using data from identified anomalies to forecast when defects will arise in the near future and prevent those defects [21]. The term "defect" refers to either a product defect or a defect in the manufacturing process. Maintenance is a component of ZDM; when a quality issue is discovered at the process level, maintenance is the solution. For example, when an equipment failure is predicted, predictive maintenance is necessary to keep KPIs at target levels [72]. The design of both the product and the production system plays a vital role in achieving ZDM [73–75]. Psarommatis et al. presented a methodology utilising a digital twin to appropriately design a manufacturing system leveraging the four ZDM strategies (detect, correct, predict, and prevent), with the goal of achieving zero defects in manufacturing [20]. There is very little research on how to properly design a manufacturing system to reach ZDM. The semantic modelling of data and information models in general play a major role in the efficiency and capacities of data-driven technologies [76,77].

2.3. Natural Language Processing (NLP) and Large Language Models (LLM)

Natural Language Processing (NLP) is a field of computer science and artificial intelligence that involves the development of algorithms and techniques that enable computers to understand, interpret, and generate human language. Although its roots in the world of computational linguistics date to the 1950s, it has been the emergence of deep learning that has made it possible for NLP to achieve its most remarkable results. The use of word-embedding techniques allowed the semantic representation of words in multidimensional spaces and powerful semantic-similarity solutions between words or sentences. Recurrent Neural Networks (such as LSTM and GRU) were soon being used for text-generation tasks. The emergence of the transformer architecture [78], with its numerous variations, has allowed the implementation of the so-called Large Language Models (LLMs). Although there are differences depending on the transformer architecture used, the training of these models is initially carried out in an unsupervised way, with large amounts of text, so that the models are finally able to complete subsequent text from a sequence, resolve instructions, identify semantic similarities between texts, and perform a wide variety of tasks typical of high-level natural language management.

Regarding specific applications in the manufacturing domain, Ayadi et al. presented a methodology for utilising NLP methods to perform Named Entity Recognition (NER), extract details about environmental exposure to engineered nanomaterials (ENMs) from text sources, classify them in line with ontological ideas, and automatically improve knowledge [79–83]. They did this by using NLP methods combined with a domain ontology. On the one hand, NER is used to extract relevant information. By contrast, the function of the domain ontology is to semantically classify and categorise the extracted data in accordance with its concepts, thereby connecting the data with its corresponding attributes [47]. Müller and Metternich [27] suggested that using NLP in digital shop-floor management could allow the development of help features based on information generated during order-fulfilment procedures [84]. The valuable domain-specific knowledge of the employees is found in the free-text data, even though the performance-indicator data may be leveraged to improve anomaly detection [27]. Sala et al. utilised NLP methods for text-mining to produce new knowledge in the maintenance domain; the objective was to demonstrate how businesses may use maintenance-report analysis to extract data that might influence asset design and maintenance-service delivery, resulting in an improvement from a dual perspective [85]. Mourtzis et al. suggested a methodology for assessing the level of expertise in human resources based on their abilities and competencies. That method makes use of unprocessed text data that was obtained from an operator's interactions with an industrial social network (ISN), and with the help of NLP methods, they managed to analyse and classify the level of expertise of each operator [85].

As stated previously, in recent years, large language models (LLMs) have made remarkable progress in various NLP tasks, such as language generation, text classification, and question answering. Relying heavily on data and information processing, ZDM has explored the potential of LLMs to improve manufacturing quality. LLMs can analyse data from various sources, such as images, sensor readings, and textual descriptions, to detect anomalies (defect detection), identify deviations from quality standards (quality control), and predict the need for maintenance operations (predictive maintenance). LLMs have been also applied in operator training, generating easy-to-comprehend summaries of key training concepts, providing natural-language feedback to guide the operator during the execution of tasks, and providing personalized feedback based on the strengths and weaknesses detected during execution. LLMs can be combined with other multimodal interactions on industrial shop floors; for example, Augmented Reality can be used to connect natural-language interactions with visual clues anchored in the real environment of the operator [34].

3. Methodological Framework

The proposed methodological framework is divided into several structured steps aimed at designing personalized programs. The steps are:

3.1. Integration with Existing Systems

3.1.1. Digital Transformation

Leverage digital technologies to achieve effective data collection, transitioning from face-to-face or paper-based interactions to digital platforms for reporting and documenting internal communications and establishing workplace requirements (roles, tasks, and skills).

3.1.2. Operational Systems Compatibility

Ensure compatibility with operational support systems like Enterprise Resource Planning (ERP), Manufacturing Execution Systems (MES), Manufacturing Operations Management (MOM), Computer Maintenance Management System (CMMS), and Content Management Systems (CMS), utilizing ISA-95 [86] and SCOR Information [87] model standards for seamless integration and data modelling.

3.2. Data Collection

3.2.1. Identification of Workplace Requirements (Roles, Task, and Skills), Data Collection, and Dataset Preparation

Each person involved in the manufacturing process plays a different role, handling a specific set of tasks that require specific competences. For instance, specialists (role) handle the configuration of a machine during start-up (task), which requires specific technical knowledge and skills (competences). The first step is to identify the different roles, tasks, and skills, then to collect data describing this information.

3.2.2. Identification of Training Materials, Data Collection, and Dataset Preparation

The second step is to collect and possibly digitize existing training materials linked to each competence.

3.2.3. Collection of Internal Communications Data and Dataset Preparation

In-line personnel are requested to report and document situations and events occurring in their workplace (for instance, security incidents). The third step is to obtain digital records of internal communications from personnel detailing different situations occurring in manufacturing settings.

3.2.4. Collection of Qualifications Data and Dataset Preparation

The organisation may require specific qualification tests (an examination or formal completion of specific training tasks) to formally acknowledge skills as competences. The methodological framework requires the collection of these formal qualifications.

3.3. Dataset Preparation

This step involves cleaning, organizing, and transforming data into a usable format for subsequent analysis. This step involves, for instance, extracting, labelling, or further processing syntactic units from internal communication records and training materials that are stored in digital formats.

3.4. Skills-Models Extraction

3.4.1. Generation of Embeddings Using Transformers

Employ Transformers to generate embeddings representing the semantics of training materials and internal communications.

3.4.2. Semantic Correspondence

Extract skills models by comparing internal communication to training materials, applying semantic correspondence via similarity functions applied to extracted embeddings.

3.5. Assessment of Skills and Qualifications

3.5.1. Skills-Gap Analysis

Contrast the extracted skills model to the required qualifications for the respective roles and workplaces, identifying missing or underrepresented skills.

3.5.2. Identification of Advanced Skills

Detect skills in the model that surpass the required qualifications, providing insights into the potential for personnel upskilling.

3.6. Recommendations for Training Programs

3.6.1. Competence Reinforcement

Use the insights from the skills-gap analysis to deliver personalized training recommendations to reinforce essential competences.

3.6.2. Personalized Upskilling

Use the insights from the skills assessment to deliver personalized training and upskilling recommendations, facilitating personnels' transition to higher-value roles within the organization.

3.7. Evaluation and Continuous Improvement

3.7.1. Personalized Upskilling Assessment

Implement control mechanisms and mitigation plans to ensure that the recommendations are adequate; implement mitigation plans to continuously improve performance and detect and reduce bias.

3.7.2. Qualification Testing

Implement regular qualification tests to formally acknowledge personnel skills, fostering a culture of continual learning and improvement.

The concept diagram in Figure 1 illustrates the different steps in the methodology.

The next sub-sections describe in greater detail the core steps of the methodology (namely, Dataset Preparation, Skills-Models Extraction, Skills and Qualification Assessment, Training Program Recommendations, and Evaluation and Continuous Improvement). The next section, Implementation Use Case, uses concrete use-case examples to further describe the different steps in a practical implementation.

3.8. Dataset Preparation and Skills-Model Extraction

Figure 2 provides a high-level overview of the data-processing pipeline used to extract a model of personnel skills from internal communications. It identifies the three different data sources identified in the data-collection stage: Workplace Requirements, Qualification and Training Materials, and Internal Communications. Note that this separation into three data sources is mainly for the sake of clarity in explaining the data processing and NLP involved. In practical implementations, data might be organised in a different way. The Internal Communications database contains internal-communications text messages, as well as related information like the issuer, date, role, and workplace. The Workplace Requirements database contains information related to the skills required for every role and workplace. Finally, the training-materials data consist of training materials stored in separate documents and metadata specifying the skill to which each document is linked.

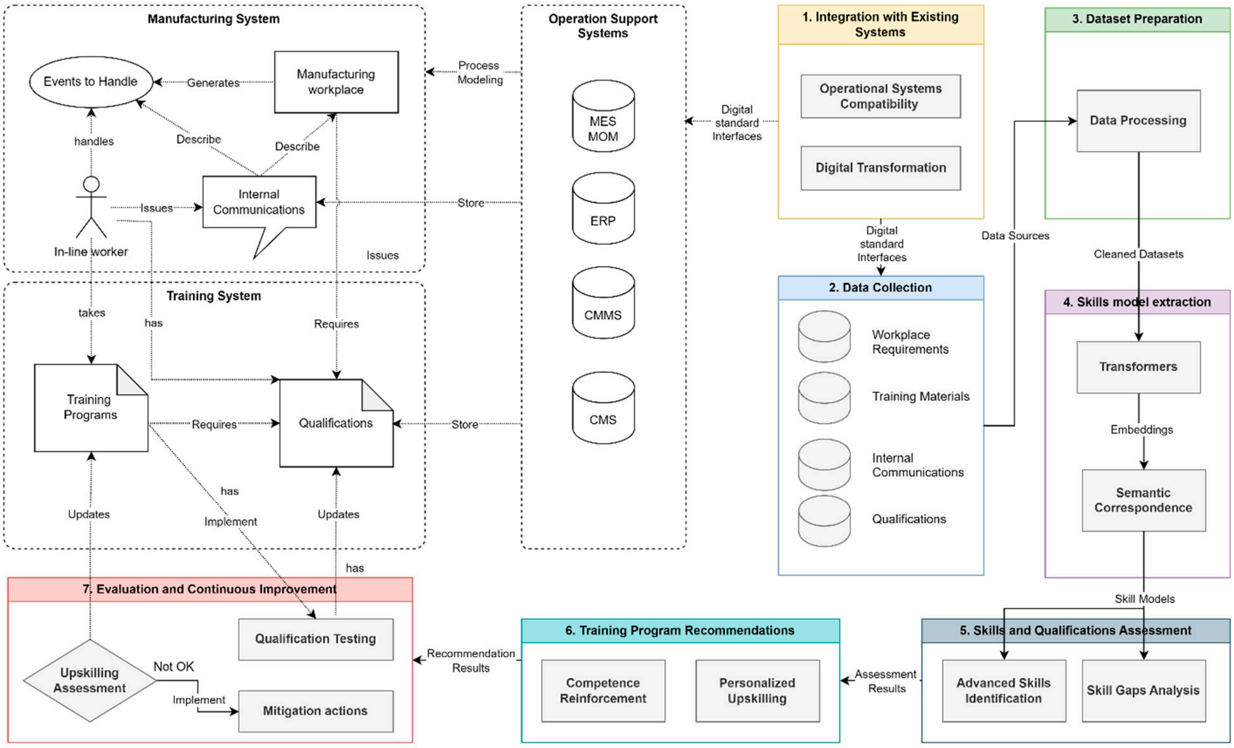

Figure 1. Methodological framework concept diagram.

3.8.1. Dataset Preparation

In the internal communications-data pipeline, communication are filtered and labelled according to different criteria (workplace, issuer, date, etc.). Then, the text is split into sentences so that the length is more homogeneous and the data are cleaned to eliminate redundant parts and noise. The text-cleaning process consists of Part of Speech (PoS) tagging to learn the part of speech of each word in each sentence, stop-word removal to remove irrelevant parts of the sentence (taking into account the PoS so that the word removal does not change the meaning of sentences), lemmatization, and normalisation, including the normalisation of specific terms used in the manufacturing process (for instance, to correct spelling errors in reference to specific products). In the implementation described below, the cleaning process is performed with available open-source models and pipelines for Spanish-language text [88]. Thus, the result is a dataset in which each row contains a normalised text field that summarises a sentence of an internal-communication message, labelled with the unique identifier of the issuer, the workplace of origin, the date and time it was issued, and a unique sentence identifier. The raw text of the communication messages is kept in a different column to facilitate analysis. It is important to note that S-BERT models and LLMs in general do not require this processing and text normalisation [89]. State-of-the-art models allow for the comparison of long texts and for asymmetric comparisons (comparing long texts and short terms; it is thus not required to split texts into sentences). However, these features are model-dependent, and even if they are implementation-dependent, these processing functions are included in the methodological problem. Moreover, for the model used in the use-case implementation, the text-normalisation pipeline reduces the cosine distance between sentences that are semantically related and increases the distance between sentences that are not related, thereby improving the overall results. Similarly, this model performs better with symmetric searches. For that reason, it is important that the encoded texts that are compared are of similar length, so in this particular case, the comparison is made sentence-by-sentence to ensure symmetry of semantic searches and improve the overall accuracy of the setup.

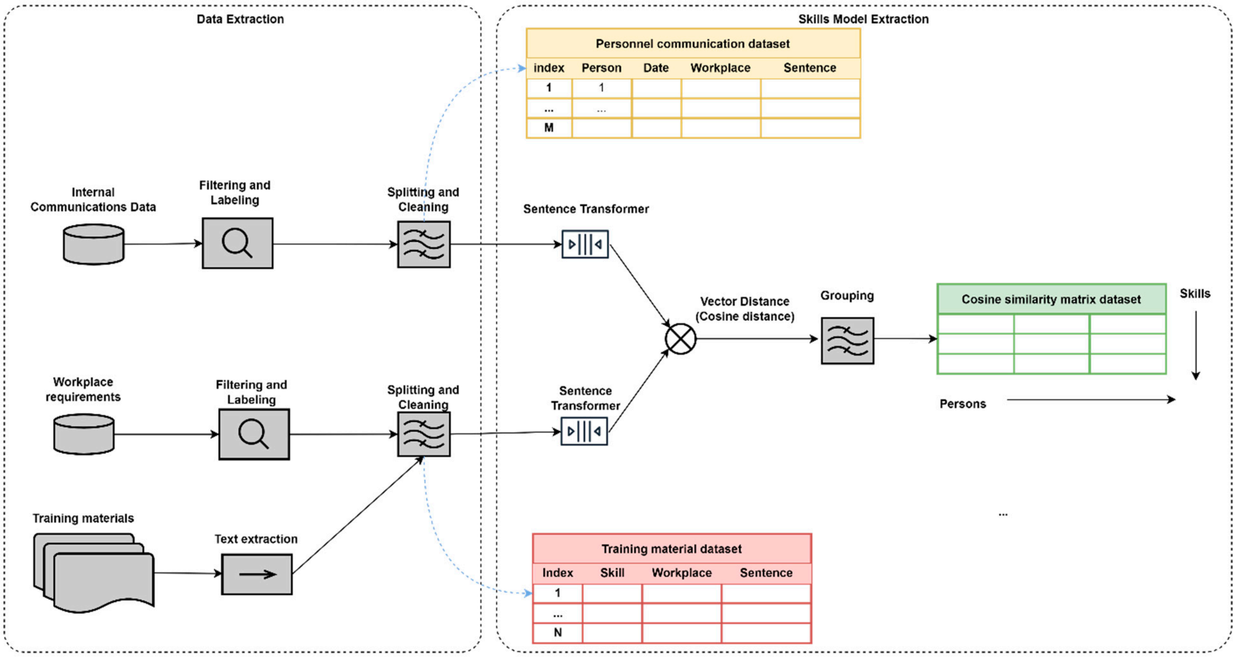

Figure 2. Overview of dataset preparation and skills-model extraction.

Additionally, the content of the training materials is extracted, divided into sentences, cleaned and denoised in a way similar to that described above for communications, and finally labelled with the unique identifier of each training material, the training module the material belongs to, the qualification associated to the training material, and a unique sentence identifier. The text extraction for the training material in the implementation uses the Apache Tika toolkit to extract the text from different document file formats (e.g., PDF, Microsoft Word, Microsoft PowerPoint, etc.). Thus, these processing steps produce two datasets of sentences: one corresponding to communication messages and the other corresponding to training material.

3.8.2. Skills Extraction

The central AI task of the proposed framework is the extraction of skills models for personnel based on internal communications. To better illustrate the methodology and for the sake of repeatability, the methodological framework divides this task into two steps: embeddings generation using transformers and semantic correspondence.

Embeddings are compact numeric vector representations of data that encapsulate essential information about the data they represent. In this methodology, embeddings are generated from internal communications and sentences from training materials. These vector representations enable the application of mathematical operations to analyse and compare data downstream. Therefore, the result after embeddings generation consists of multi-dimensional vectors representing the semantics of internal communications sent by personnel and vectors representing sentences from training materials.

Transformers are a class of model in NLP that have shown remarkable success in generating rich text embeddings. In the context of this work, from the possible transformer architectures currently available for NLP, the BERT [90–94] family was chosen. The BERT transformer family has been proven to be especially well suited to text classification, with NER (Name Entity Recognition) being the type of task that is best aligned with the work proposed here. The pre-trained BERT model used in this implementation is a model specifically built for Spanish (the language used in the use case described below) [95]. To look for semantic similarities between sentences, Sentence-BERT networks [95], a Siamese BERT extension for sentence-level semantic similarity, is used to generate sentence embeddings of the normalised text.

The generated embeddings can be used downstream to perform a semantic-similarity comparison. As embeddings are essentially numeric vectors, they can be compared with each other using vector distance functions like heuristic distance or cosine distance. Furthermore, as each embedding captures the semantics of a sentence, this comparison quantifies the semantic similarity between two sentences.

In this way, each internal-communication and training-material sentence pair is compared, resulting in a $M \times N$ matrix C in which each element c_{lk} represents the cosine distance between the internal communication l ($i \in [1, \ldots, M]$) and the training-material sentence k ($j \in [1, \ldots, N]$).

Finally, as all sentences in the datasets have unique identifiers, it is possible to leverage the dataset labels to group these distance coefficients according to different criteria. Applying a grouping function (like minimum distance or mean distance) to grouped coefficients allows the extraction of information related to the semantic relation between higher-level semantic entities composed of groups of sentences. For instance, to infer relationships between personnel and skills, the cosine-distance elements are grouped by the unique identifier of each individual person and skill, and the minimum grouping function returns a C^* matrix in which each element c^*_{ij} represents the minimum semantic distance between any sentence in the training materials associated with skill i and any sentence issued by person j.

3.9. Assessment of Skills and Qualifications and Training Recommendations

The Training Recommendations system is based on the comparison of the actual skills of personnel, as acknowledged by the company through qualifications tests, with the skills models extracted from internal communications (obtained through the procedure described in the previous section) (Figure 3). The qualifications, which are stored in a Qualifications database, are filtered and labelled with the unique identifiers of the qualifications and operators. Then, the table is pivoted to obtain a matrix A in which each element a_{ij} is 1 if operator j has skill i and 0 otherwise. This matrix A is used to split the cosine similarity matrix dataset C^* described above, where element c^*_{ij} represents the minimum semantic distance between sentences in training materials related to skill i and internal communications sent by worker j.

The indices where A is equal to 1, that is (i, j): $a_{ij} = 1$, are used to obtain the values of C^* corresponding to all the acknowledged qualifications of every worker. These values should be close to 0, representing high similarity, as it is expected that the persons who have a skill will use expressions similar to those found in the training materials related to that skill. The values form three different clusters: a high-similarity cluster (low semantic distance), a medium-similarity cluster (moderate semantic distance), and a low-similarity cluster (high semantic distance). If c^*_{ij} is in the high-similarity cluster, the output of the recommender system is to not recommend any action for skill i to worker j. If c^*_{ij} is in the medium-similarity cluster, the output of the recommender system is to recommend re-training in skill i to worker j. Finally, if c^*_{ij} is in the low-similarity cluster, then the output of the recommender system is to highly recommend re-training in skill i to worker j.

Similarly, the indices where A is equal to 0, that is (l, k): $a_{lk} = 0$, are used to obtain the values of C^* corresponding to the missing skills of every worker missing a qualification test. These values should represent low similarity, as it is expected that the persons who have not passed a qualification test for a skill will not express themselves in internal communications using the same language that is used in training materials for missing skills. The values again form three clusters: a high-similarity cluster, a medium-similarity cluster and a low-similarity cluster. If c^*_{lk} is in the high-similarity cluster, the output of the recommender system is to highly recommend that worker k perform the competence test for skill l. If c^*_{lk} is in the medium-similarity cluster, the output of the recommender system is to recommend that worker k perform the competence test for the corresponding skill, and if c^*_{lk} is in the low-similarity cluster, there is no recommendation.

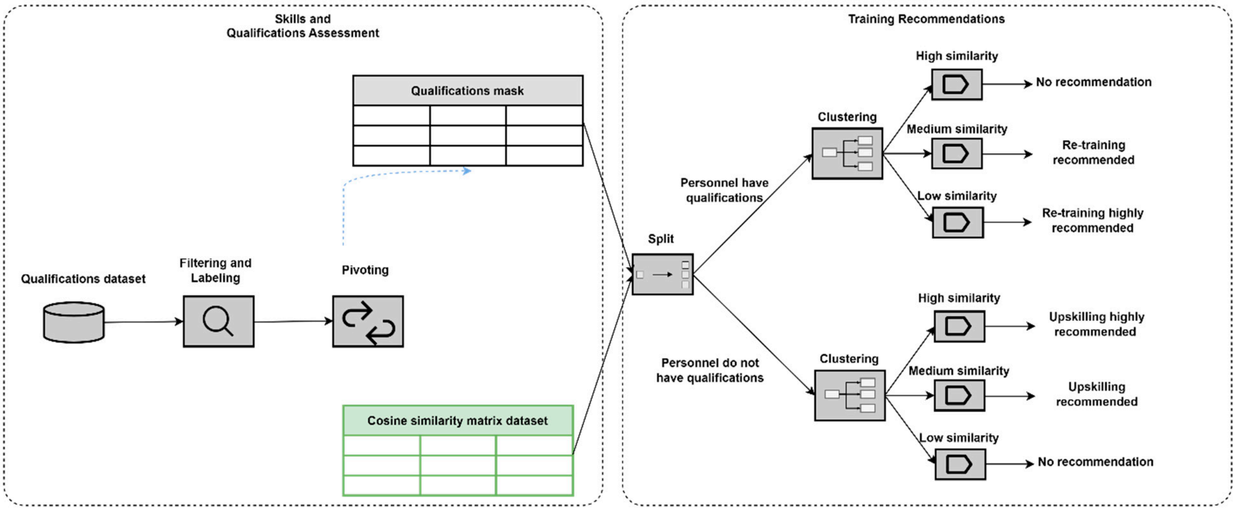

Figure 3. Skills and qualifications assessment and training recommendations.

3.10. Evaluation and Continuous Improvement

Once results have been obtained, the framework includes a systematic approach for evaluation and continuous improvement. The primary objective is to ensure the accuracy and effectiveness of the training and upskilling recommendations. Another important objective is to ensure that the persons in charge of in-line personnel training can assess the quality of the results and pinpoint and address any discrepancies or issues. For instance, the skills-gap analysis may identify an underrepresented skill and the system may therefore issue a competence-reinforcement recommendation, but this result could be due to an error if the dataset was not updated when the individual obtained the qualification. The evaluation and continuous-improvement processes consist of the following steps and mitigation actions:

1. Quality Assessment of Recommendation Results: Review and assess the alignment between the inferred skill models and formal qualifications through feedback loops involving individuals and other stakeholders, like trainers. Identify potential mismatches and inconsistencies that might suggest inaccuracies in the dataset or the system.

2. Check for Ethical Compliance: Evaluate the fairness and ethical implications of the training recommendations to ensure they are unbiased. Establish mechanisms to monitor imbalance or demographic bias in training materials or communications and implement mitigation plans to correct detected problems. Establish mechanisms to ensure compliance with privacy and ethical regulations in data collection, analysis, and dissemination of the training recommendations.

3. Identification and Mitigation of Dataset Errors: When discrepancies are found, identify the source of error (for instance, insufficient communication data, insufficient training-material content, dataset inaccuracies), formulate a mitigation plan to correct the dataset, and re-evaluate the skills-gap analysis. In cases in which training materials are insufficient, the mitigation plan involves enriching and updating the content to provide a robust context for both personnel training and accurate analysis. In cases in which there are insufficient communication data, mitigation plans may contain specific actions to foster internal communications and overall involvement in the continuous improvement of all personnel.

4. Performance Evaluation and System Updates: Implement test cases to evaluate skills assessment under different scenarios. If recommendations are not aligned with actual training needs, refine the algorithm, e.g., update the model, adjust thresholds, or incorporate other techniques. Update test cases to facilitate the detection of errors. Generate internal documentation to inform future iterations of the continuous-improvement framework and enrich the knowledge base.

4. Implementation Use Case

4.1. Description

The main objective of this implementation is to present a specific use case, explain how the data are collected and prepared in a practical context, present some experimental results to help other researchers apply the methodology, and present the results of a specific case study.

The implementation use case is a food-processing factory, and the analysis is based on different Manufacturing Operations Management (MOM) application modules. These modules, namely, Operators Training and Internal Communications, provide support to the Total Productivity Management (TPM) methodology used for continuous improvement.

The Human Resources department uses the Operators Training module to manage internal and external training programs for line personnel. It is worth mentioning that one of the main pillars of Total Productivity Management (TPM)—the continuous-improvement methodology used in the factory—is autonomous maintenance, which consists of providing tools to operators so that they can perform maintenance tasks on their own equipment with no support from the maintenance department. In this sense, operator training is crucial to obtaining this objective. Through the management interfaces, the training and upskilling managers can create courses, manage the course calendar, add training materials to the course, and assign instructors (either internal or external personnel) and trainees (internal personnel). The courses are composed of training tasks, and each task can be assigned to different calendar events in the course. Trainers can use the application user interfaces to manage their courses: update the course materials, upload new materials, and manage the attendance list for each training event. Similarly, trainees receive calendar notifications and reminders to access the course program and attend the courses. Figure 4 highlights how these courses are linked to the personnel model used in MOM application modules. The completion of tasks is linked to a *qualification test specification* to indicate that the training tasks are part of the procedure required to obtain a specific qualification: a demonstrated competency that ensures that the person has the required training and experience to perform specific operations. The operations definitions have specific personnel specifications, which correspond to the personnel and personnel qualifications requirements for the operation. In this way, it is possible to consider these requirements in production personnel scheduling, during the execution of production orders, and during performance evaluations. All these definitions are consistent with the ISA-95 standard concepts and models.

The internal communications application is used primarily by the continuous-improvement and maintenance departments to collect additional information about the execution of production orders. The objective is to leverage operators' in-depth knowledge of the execution process. Through this application, in-line personnel report events that occurred during the execution of the production order using natural language. The events they report vary in content: for the most part, the events reported are incidents that negatively affected safety, productivity, process performance, or product quality. Examples include maintenance incidents, whether they were solved by the personnel or triggered a maintenance request; incidents that resulted in a loss of performance or product quality; and safety incidents. Personnel can also report observations regarding re phenomena that did not negatively affect performance yet but may do so in the future (for instance, presence of rust in a machine), and proposals for improvement to safety, productivity, process performance or product quality. Maintenance and production managers review these internal communications and design an action plan as part of the continuous-improvement practise of the factory.

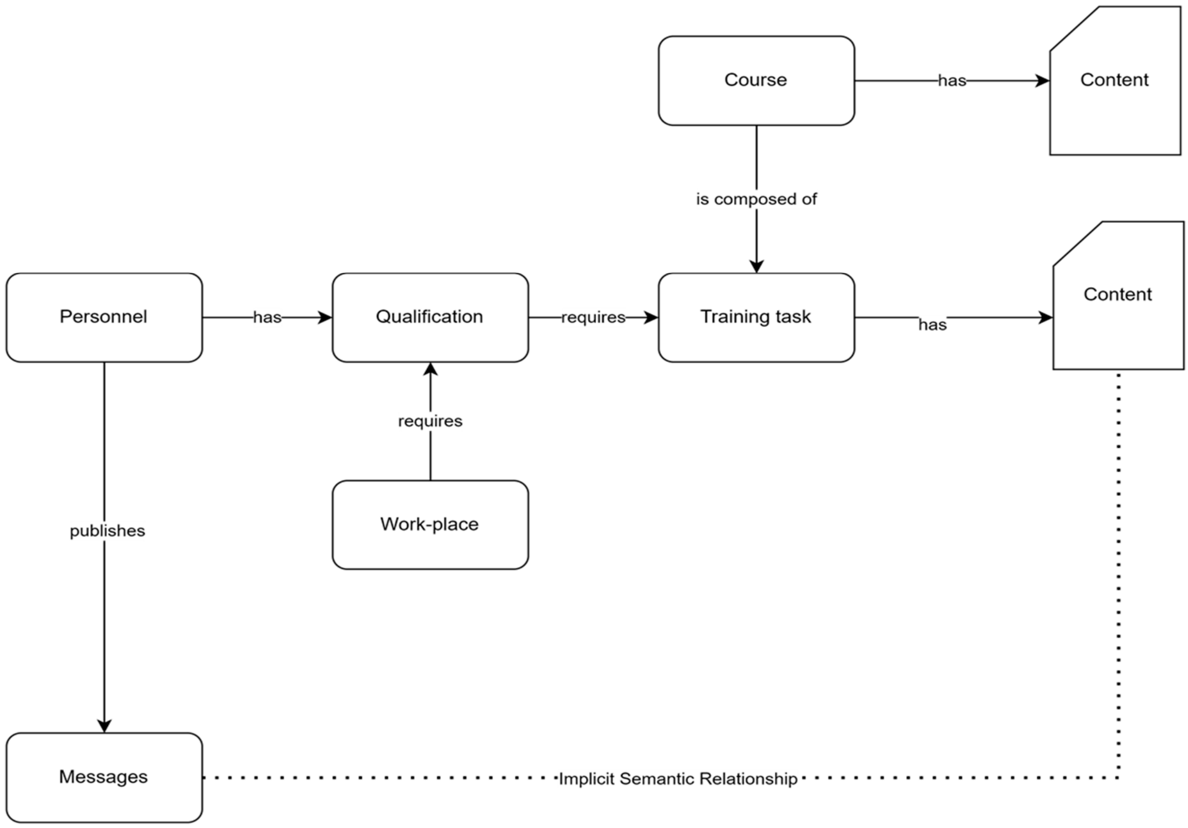

Figure 4. Conceptual model of personnel qualifications and internal communications.

Both applications are linked by the personnel model, as shown in Figure 4: persons have qualifications and publish communications, so there is an explicit relationship between the communications and the entities that model skills and training. All of the data are stored in proprietary Structured Query Language (SQL) databases, so the datasets can be extracted using SQL queries based on the model described above.

4.2. Preliminary Data Analysis

The study focuses narrowly on a specific section of the food-processing factory, the packing section, and on a specific set of qualifications related to packing-process operations, security, and quality. Table 1 shows the qualifications considered and classified in training modules, as well as the different roles in which the qualification is required: Team Leader (1), First Officer (2), and Specialist (5). The number in parentheses is the degree, or level in the hierarchical organization of the line, with degree 1 representing the highest degree in the hierarchy. Some qualifications in packing operations are specific to each machine (there are 19 different machines), but only one row is included in the table text (one per machine).

The study group was composed of 60 persons who have experience in the packing section and who had sent communication messages in the last year. These messages were processed to obtain a model of their skills, and this model was compared to the qualification tests the operators had passed to obtain personalized training and upskilling recommendations. Figure 5 shows a heatmap that represents the qualifications of all personnel with experience in the packing section. The X axis represents the operator order, sorted by the number of qualification tests passed, and the Y axis represents the index of the qualification test, ordered by module. The Y-axis labels show the module identifier. The heatmap uses a binary colour map to represent whether a specific person has acquired a qualification through a qualification test (yellow means yes; purple means no). There is a high number of workers with the minimum required qualifications (at the right side of the figure).

Table 1. Description of qualifications.

Modules	Qualifications	Roles
10—Packing Operations	Routing and changeover (1 per machine)	Team Leader (1), First Officer (2)
10—Packing Operations	Initial conditions, alarms, and failure states (1 per machine)	Team Leader (1), First Officer (2)
10—Packing Operations	Routing and changeover (1 per machine)	Team Leader (1), First Officer (2)
10—Packing Operations	Initial conditions, alarms, and failure states (1 per machine)	Team Leader (1), First Officer (2)
10—Packing Operations	Cleaning procedures	First Officer (2)
10—Packing Operations	General principles: packing section	First Officer (2), Specialist (5)
15—Quality	Basic hygiene norms and food handling	Team Leader (1), First Officer (2), Specialist (5)
15—Quality	Food-handling certification	Team Leader (1), First Officer (2), Specialist (5)
15—Quality	Allergens management	Team Leader (1), First Officer (2), Specialist (5)
15—Quality	Hazard analysis and Critical control point	Team Leader (1), First Officer (2), Specialist (5)
15—Quality	Potential risks and Individual protective equipment	Team Leader (1), First Officer (2)
15—Quality	Quality procedures	First Officer (2)
15—Quality	Process control	First Officer (2)
16—Packing Cleaning	General principles: cleaning	First Officer (2), Specialist (5)
16—Packing Cleaning	Specific cleaning procedures	First Officer (2), Specialist (5)
17—Specialist Training	Line feeding: operations, security, and process control	Specialist (5)
22—Security	Emergency management and evacuation plan	Team Leader (1), First Officer (2), Specialist (5)
22—Security	General security risks	First Officer (2)
22—Security	Fall-prevention plan	Team Leader (1), First Officer (2), Specialist (5)
22—Security	General risks: hygiene	Team Leader (1), First Officer (2), Specialist (5)
22—Security	General risks: ergonomics	Team Leader (1), First Officer (2), Specialist (5)
22—Security	Action in the event of an accident	Team Leader (1)

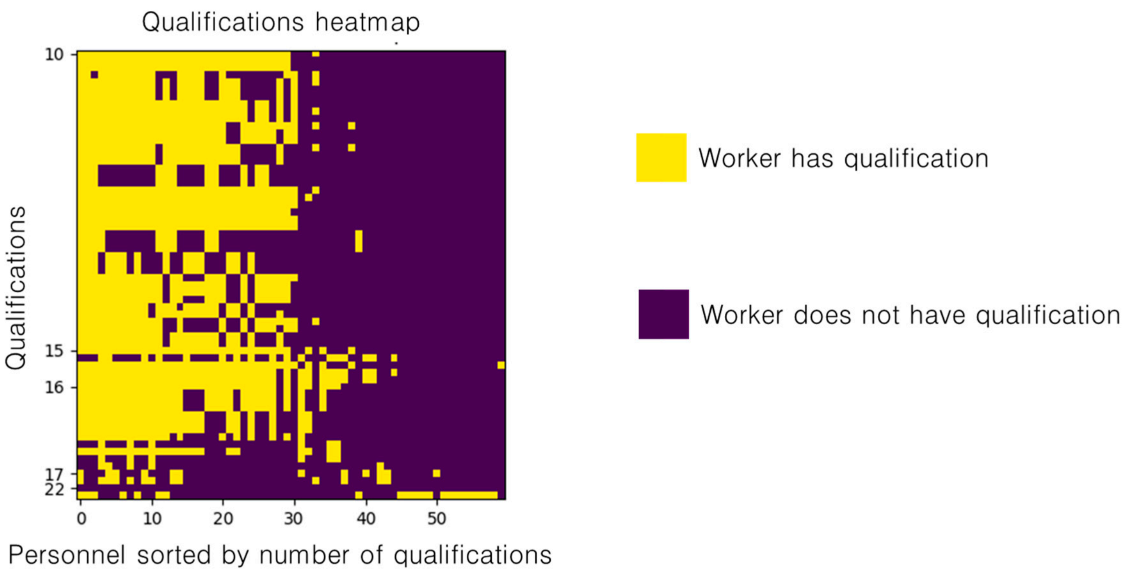

Figure 5. Qualifications heatmap (study group).

4.3. Experimentation Example

This section contains a practical example utilizing use-case data to illustrate the workings of the proposed methodological framework. Figure 6 shows a literal translation of one of the messages used in internal communications, together with a fictional unique identifier and the role of the person who issued the communication.

Personnel Id: 555

Role: Team Leader

Communication text: The hose in the cabin is broken. When you turn the water on, the pressure is too high. If we are using soap, it could be dangerous.

Figure 6. Example communication.

The message highlights a maintenance issue describing a malfunctioning hose installed in a cabin that may lead to a safety incident. This communication has high semantic similarity with sentences found in the training materials that describe the end-of-shift inspection procedures, as shown in Figure 7.

Training module: Quality

Qualification: Packing Process Control

Training material: End-of-Shift Inspection

Sentence: Using a water hose that does not have too much pressure, we will clean up the remainder of the product.

Figure 7. Example sentence from training material.

Within the framework, this semantic similarity plays a crucial role in deducing whether individuals are familiar with the procedures outlined in the training materials, based on their communications. This inference mirrors a human-like approach of deducing someone's knowledge by analysing how they articulate domain-specific matters. To illustrate this approach, consider the sentence in Figure 7. It is semantically very similar to the example internal communication, but it was not selected manually. Instead, we applied a distance function—specifically, cosine distance– to compute the semantic distance between the embeddings of the example communication sentence in Figure 6 and the embeddings of all sentences across all training materials. The embeddings are generated using the pre-trained BERT model selected for the implementation of the use case [88].

The cosine distance function yields a value within the range [0, 1], where a lower value indicates a higher similarity. Consequently, for every sentence in the training materials, a cosine distance value is computed, resulting in a vector v of length 3563 (number of sentences in training materials). The index of the sentence in Figure 7 corresponds to the index where the minimum of v is observed (argmin(v)), indicating the index with the greatest similarity. This vector is represented in Figure 8, providing a clear depiction of how the framework identifies the sentence from the training materials that is most semantically similar to the communication sentence.

Moreover, a closer examination of Figure 8 reveals other indices of sentences from training materials for which the cosine similarity is low, indicating high semantic similarity to the communication sentence. Figure 9 shows the values of vector v sorted in ascending order, positioning sentences with high semantic similarity (lower cosine-distance values) towards the left of the figure. Looking at this figure, we can identify three different regions. First, there is a rapid increase in the value, but this increase starts decaying around 0.4. Beyond this point, the rate of increase in the cosine value slows gradually until it reaches a value of about 0.8. Past this threshold, the cosine distance increases sharply once again.

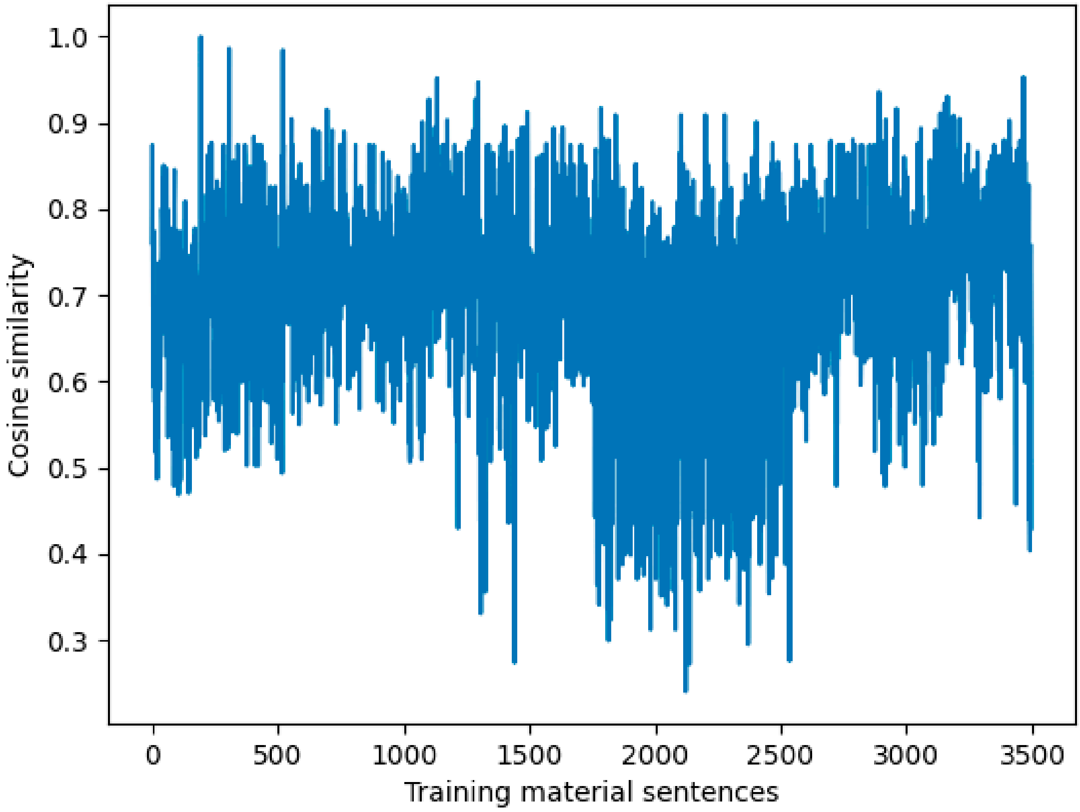

Figure 8. Cosine similarity of sentences in the training materials.

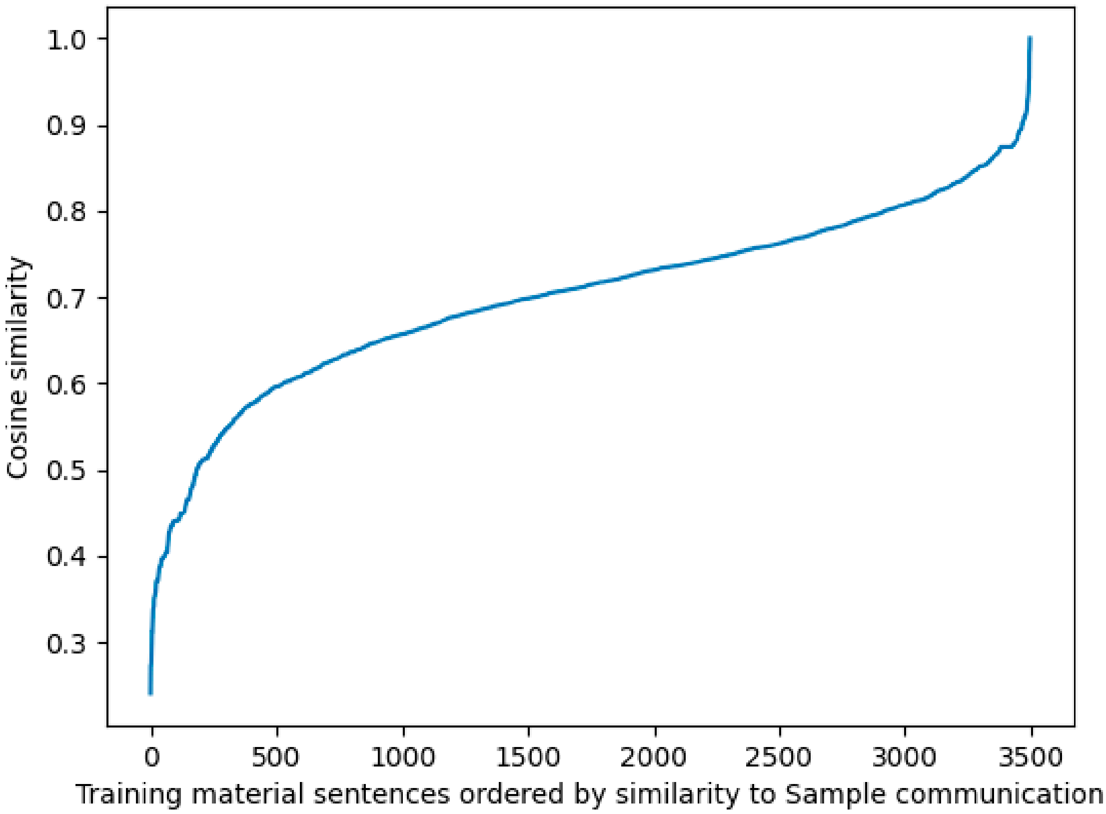

Figure 9. Sorted cosine similarity of training-material sentences.

If we take a conservative approach, one might infer that the training materials in the first region bear high similarity to the communication sentence. If the communicating person does not have these qualifications, we should provide personalised training for them because the person is expressing interest in related activities and procedures. Acquiring related qualifications will yield higher levels of workplace satisfaction and in turn higher added value for the organisation. However, it is important to note that the shape and the specific boundaries between these regions depend on the model used and the sentences compared. Employing statistical methodologies such as clustering can provide an estimate of these values, supporting the interpretation of semantic similarity across the dataset.

Hence, semantic similarity can provide a model of the skill demonstrated by the operator in executing this particular procedure. Furthermore, this data, combined with the communicating person's role, enable a qualification assessment, determining whether the demonstrated skill aligns with the required qualifications for the role. This analytical process can be scaled up to encompass all communications from a worker and all procedures outlined in the training materials pertaining to a specific qualification, grouping the skills models extracted for each of the communications sent by a person. In this case, the cosine-distance representations will be two-dimensional arrays rather than vectors, but the rationale of the methodological approach remains the same.

5. Results

5.1. Semantic Similarity between Training-Material Documents

The objective in this section is to analyse the semantic similarity between training-material documents, applying the NLP pipeline described in the previous section to compare them in pairs. As mentioned previously, the selected transformer requires that the encoded texts be similar in length (symmetric semantic search), so the content is divided into sentences to ensure that the length of the texts is approximately heterogeneous. Hence, S-BERT transformers are used to obtain sentence embeddings for each sentence from a document. Then, for every pair of files, the cosine distance between the sentence embeddings of each sentence is calculated and the minimum distance is selected to compute the semantic similarity of the documents. The result is a MxM matrix C, where M is the number of qualification tests and element c_{ij} represents the semantic similarity between training-materials documents of qualification tests i and j. The matrix is symmetric because the cosine-distance function is commutative. The values of the obtained matrix are shown in the heatmap in Figure 10. The X and Y axes represent the document index, a unique identifier of each document (ranging from 0 to $\mathrm{N} - 1$). The qualification tests are ordered by module, and the labels in both axes show the module identifier to facilitate the analysis of the result. Values of c_{ij} that are close to 0 (represented in yellow in the figure) indicate that documents i and j have very high semantic similarity, whereas values close to 1 (represented in blue in the figure) indicate that the documents have little semantic similarity.

The results show that in general, the training-material documents have similar content, according to this pipeline. Notably, the similarity is high in the regions near the diagonal that are shown as yellow squares in Figure 10. These regions represent the semantic similarity of documents from training materials in the same training module. Outside the diagonal, similar shapes can be identified, showing that training materials from different modules are also similar, except for documents in module 16 (hygiene, food handling and allergens management), which have lower similarity compared to the rest of modules.

Semantic Similarity of Communications and Training Materials

To obtain the results presented in this section, sentences in communications are compared to sentences in training-material documents. As in the previous section, S-BERT sentence transformers are applied to obtain sentence embeddings for sentences in communications and sentences in training-material documents. The cosine-similarity function is applied to the embeddings to estimate the semantic distance between pairs of sentences. Then, coefficients are grouped by qualification and by person and the minimum grouping distance is used to estimate the semantic distance between a qualification and a person. The result is a *MxN* matrix *C* (where *M* is the number of qualifications and N is the number of persons in the study group) where element c_{ij} represents the semantic distance between training-material documents of qualification test *i* and person *j*. The values of the matrix are shown in the heatmap in Figure 11, using the same colour map and scale (values in yellow represent high similarity between sentences). The training-material documents are ordered by the corresponding training-module identifier, and the Y-axis labels indicate the module identifier. The X axis represents the index of personnel and is ordered by the number of qualifications acquired.

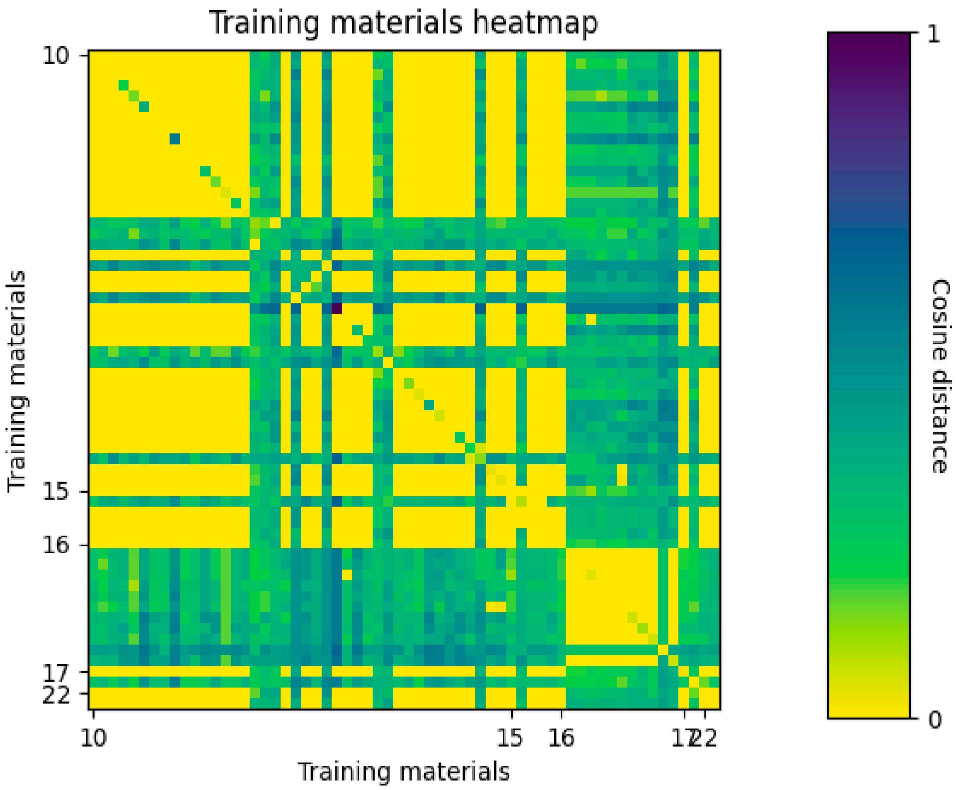

Figure 10. Heatmap of the cosine similarity across training-material documents for qualifications.

The figure shows that in general (Figure 11), the communications of personnel with more qualifications (on the left side of the figure) have on average a higher semantic similarity to training-material documents than do the communications of personnel with fewer qualifications (on the right side of the figure). This pattern can be seen in the fact that the heatmap is predominantly yellow to on the left side of the figure. Still, there are areas on the left side of the figure that show low semantic similarity. Hence, although these operators are qualified, they do not send any internal communications that are semantically related to the training-material documents. This result indicates that these persons need training to reinforce knowledge of the corresponding skills. Similarly, on the right side of the figure, there are areas in light green or yellow that show that personnel express themselves in internal communications using language similar to that found in training-material documents, although they have not formally acquired the qualification (or at least, they have not received formal training or passed the corresponding qualification tests). Therefore, these individuals should be encouraged to formally acquire the corresponding qualifications, fostering internal promotion. The following subsection describes specific recommendations, using the actual qualification tests to split the semantic-similarity results and clustering the values to better differentiate among the results.

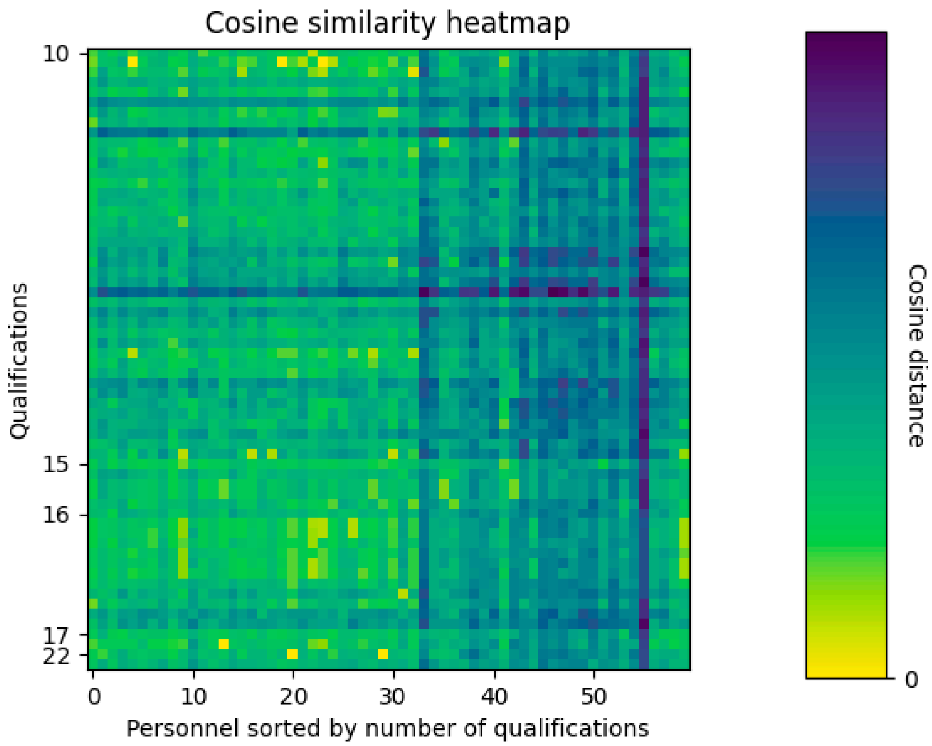

Figure 11. Heatmap of the cosine similarity between qualifications and personnel internal communications.

5.2. Recommendations

To derive the recommendations, the semantic cosine-similarity results are split into two different arrays, one containing the values corresponding to the qualifications acquired by workers, noted as q, and another containing the values corresponding to qualifications not acquired by workers, noted as n. The values are then clustered with Kmeans using the Scipy library [89], as described in previous sections. The clustering of q results in the clusters shown in Table 2.

Table 2. Values-clustering table.

Cluster Name	Centroid	Range	Recommendation
High Similarity	0.185	[0–0.306]	No recommendation
Medium Similarity	0.421	[0.307–0.586]	Re-training recommended
Low Similarity	0.744	[0.588–1]	Re-training highly recommended

Figure 12 shows the results obtained by the recommender system for all personnel. In more detail, it shows the qualifications that need to be reinforced with re-training. In cases in which these qualifications are critical (e.g., due to security or regulations), the information is highly valuable both for the organization and for individual employees, as it helps reinforce key company policies, concepts, and skills, thus preventing potential issues and manufacturing defects. The figure also allows the identification of qualifications for which the system recommends re-training for all employees. In these cases, the company should also consider reviewing and reinforcing the associated training materials to ensure that the low level of similarity is not due to a lack of adequate training materials. A similar analysis can be performed on the vertical axis, identifying employees who exhibit a low level of similarity for all qualifications. In such cases, it is recommended that the cases be analysed individually. Note that the methodological framework relies on internal personnel communications, so employees that do not engage in internal communications will likely exhibit low levels of similarity for all qualifications. The clustering of array *n* results in the clusters described in Table 3. Figure 13 illustrates the final upskilling-recommendation results.

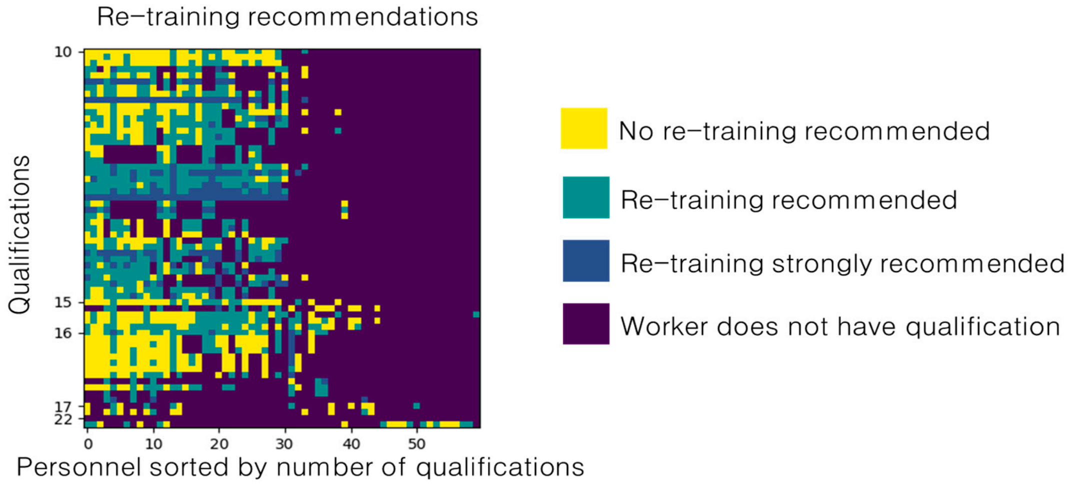

Figure 12. Retraining-recommendation results.

Table 3. Clustering of array n results.

Cluster Name	Centroid	Range	Recommendation
High Similarity	0.186	[0–0.303]	Upskilling highly recommended
Medium Similarity	0.420	[0.303–0.555]	Upskilling recommended
Low Similarity	0.690	[0.556–1]	No recommendation

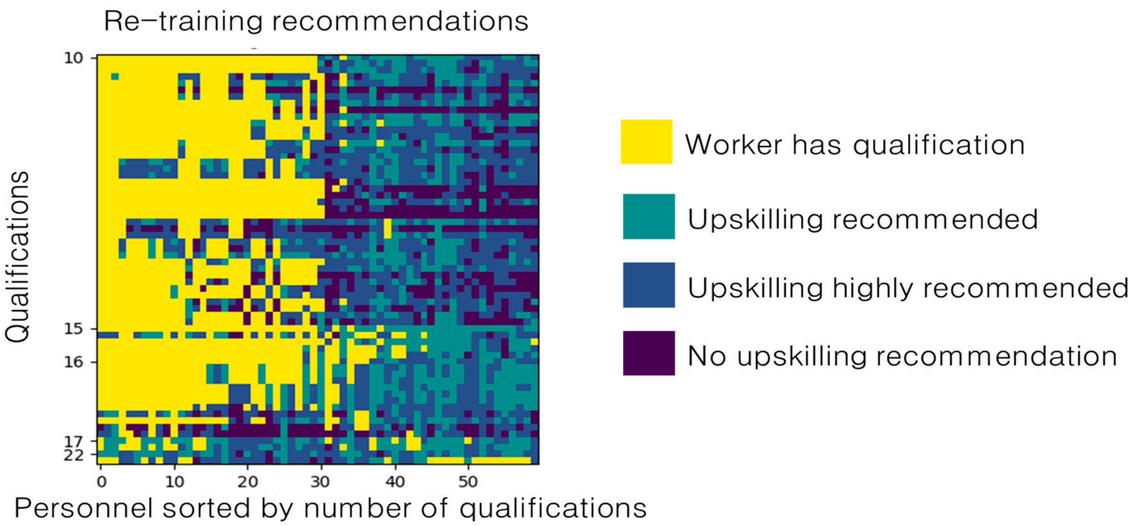

Figure 13. Upskilling-recommendation results.

6. Conclusions

This paper presents a methodological framework based on LLMs to design personalised training programs to facilitate personnel upskilling in Industry 5.0. The methodological framework shows how LLMs and natural-language interactions with operators can be used to extract models of the operators' skills and guide personnel-training programs based on the extracted knowledge. Despite significant efforts to capture, develop, and retain talent, many organisations are currently facing a shortfall of personnel qualified to cope with exponential technological and societal change. The acquisition and engagement of suitable personnel is becoming increasingly difficult. While automation, robotics, and artificial intelligence can automate and streamline many processes, their adoption also requires human expertise. Industry 4.0 is transforming (rapidly) the roles and workplaces of workers, rather than removing workers. Persons who are not able to receive training and upskilling see themselves displaced from their workplace, and companies are not able to cope with increasing pressures related to personnel shortages. While it is clear to organisations that they need Industry 4.0 to scale up their businesses and remain competitive, it is also clear that their workers are a necessary asset and that it is thus essential to achieve high levels of commitment and engagement. Personalised training can certainly be a cornerstone supporting this challenging transition, and the presented framework shows a possible pathway to implementing such a system.

Based on this framework, the paper presents a use case based on a dataset from a food-processing company. The use-case implementation uses S-BERT transformers, and the results clearly demonstrate how the methodological framework can support the development of personalised training and upskilling programs. This implementation should be regarded as a first approach and obviously has some limitations. The main advantage is that the natural-processing methods, techniques, and models used in the data pipelines of this early prototype are industry-standard open-source tools. In this sense, the main objective of this early implementation is to showcase the feasibility of implementing the methodological framework in a real use case. The main limitations are that the selected model does not consider sentiment (confidence, hesitation) or other important elements of communication and that it does not take into account the conversation in full, instead using tokenized sentences, so much valuable context information is missed. Most ambitious future work based on state-of-the-art very large LLMs such as GPT, BLOOM, or LLaMa, can clearly overcome these limitations. These models are instruction-tuned LLMs (fine-tuned to carry out very general kinds of activities) and are able to perceive more subtle nuances in natural language and discern the required information with a higher degree of confidence. While the presented results are good, using instruction-tuned LLMs as the backbone of the methodological approach presented in this paper can give extraordinary results with much less effort. For this reason, while the core contributions of the research work herein described are the definition of the methodological framework and the practical implementation, future work will aim in two directions: firstly, the incorporation of novel techniques, leveraging state-of-the-art models to improve results; secondly, the implementation of the continuous improvement stage. This stage will involve the integration of stakeholder feedback, which will serve as a validation mechanism.

Author Contributions: Conceptualization, F.F. and F.P.; Methodology, F.F. and F.P.; Software, F.F.; Formal analysis, F.F.; Writing—original draft, F.F.; Writing—review & editing, F.P. and F.A.; Visualization, J.J. All authors have read and agreed to the published version of the manuscript.

Funding: The presented work was partially supported by the projects i4Q, TALON, EU H2020 projects under grant agreements No 958205, 101070181 accordingly. The paper reflects the authors' views, and the Commission is not responsible for any use that may be made of the information it contains.

Data Availability Statement: Available on request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Psarommatis, F.; Bravos, G. A Holistic Approach for Achieving Sustainable Manufacturing Using Zero Defect Manufacturing: A Conceptual Framework. *Procedia CIRP* **2022**, *107*, 107–112. [CrossRef]
2. Mourtzis, D.; Vasilakopoulos, A.; Zervas, E.; Boli, N. Manufacturing System Design Using Simulation in Metal Industry towards Education 4.0. In *Proceedings of the Procedia Manufacturing*; Elsevier: Amsterdam, The Netherlands, 2019; Volume 31, pp. 155–161.
3. Kuhn, C.; Lucke, D. Supporting the Digital Transformation: A Low-Threshold Approach for Manufacturing Related Higher Education and Employee Training. *Procedia CIRP* **2021**, *104*, 647–652. [CrossRef]
4. Filz, M.-A.; Bosse, J.P.; Herrmann, C. Digitalization Platform for Data-Driven Quality Management in Multi-Stage Manufacturing Systems. *J. Intell. Manuf.* **2023**. [CrossRef]

5. Kaiser, J.; McFarlane, D.; Hawkridge, G.; André, P.; Leitão, P. A Review of Reference Architectures for Digital Manufacturing: Classification, Applicability and Open Issues. *Comput. Ind.* **2023**, *149*, 103923. [CrossRef]
6. Maretto, L.; Faccio, M.; Battini, D. The Adoption of Digital Technologies in the Manufacturing World and Their Evaluation: A Systematic Review of Real-Life Case Studies and Future Research Agenda. *J. Manuf. Syst.* **2023**, *68*, 576–600. [CrossRef]
7. Carluccio, J.; Cuñat, A.; Fadinger, H.; Fons-Rosen, C. Offshoring and Skill-Upgrading in French Manufacturing. *J. Int. Econ.* **2019**, *118*, 138–159. [CrossRef]
8. Cirillo, V.; Rinaldini, M.; Staccioli, J.; Virgillito, M.E. Technology vs. Workers: The Case of Italy's Industry 4.0 Factories. *Struct. Chang. Econ. Dyn.* **2021**, *56*, 166–183. [CrossRef]
9. Arai, Y.; Ichimura, H.; Kawaguchi, D. The Educational Upgrading of Japanese Youth, 1982–2007: Are All Japanese Youth Ready for Structural Reforms? *J. Jpn. Int. Econ.* **2015**, *37*, 100–126. [CrossRef]
10. Psarommatis, F.; Sousa, J.; Mendonça, J.P.; Kiritsis, D. Zero-Defect Manufacturing the Approach for Higher Manufacturing Sustainability in the Era of Industry 4.0: A Position Paper. *Int. J. Prod. Res.* **2022**, *60*, 73–91. [CrossRef]
11. Wang, B.; Zheng, P.; Yin, Y.; Shih, A.; Wang, L. Toward Human-Centric Smart Manufacturing: A Human-Cyber-Physical Systems (HCPS) Perspective. *J. Manuf. Syst.* **2022**, *63*, 471–490. [CrossRef]
12. Wan, P.K.; Leirmo, T.L. Human-Centric Zero-Defect Manufacturing: State-of-the-Art Review, Perspectives, and Challenges. *Comput. Ind.* **2023**, *144*, 103792. [CrossRef]
13. Leng, J.; Sha, W.; Wang, B.; Zheng, P.; Zhuang, C.; Liu, Q.; Wuest, T.; Mourtzis, D.; Wang, L. Industry 5.0: Prospect and Retrospect. *J. Manuf. Syst.* **2022**, *65*, 279–295. [CrossRef]
14. Xu, X.; Lu, Y.; Vogel-Heuser, B.; Wang, L. Industry 4.0 and Industry 5.0—Inception, Conception and Perception. *J. Manuf. Syst.* **2021**, *61*, 530–535. [CrossRef]
15. European Commission. *Industry 5.0*; Elsevier: Amsterdam, The Netherlands, 2021.
16. Barata, J.; Kayser, I. Industry 5.0—Past, Present, and Near Future. *Procedia Comput. Sci.* **2023**, *219*, 778–788. [CrossRef]
17. Wang, B.; Zhou, H.; Li, X.; Yang, G.; Zheng, P.; Song, C.; Yuan, Y.; Wuest, T.; Yang, H.; Wang, L. Human Digital Twin in the Context of Industry 5.0. *Robot. Comput. Integr. Manuf.* **2024**, *85*, 102626. [CrossRef]
18. Leng, J.; Zhong, Y.; Lin, Z.; Xu, K.; Mourtzis, D.; Zhou, X.; Zheng, P.; Liu, Q.; Zhao, J.L.; Shen, W. Towards Resilience in Industry 5.0: A Decentralized Autonomous Manufacturing Paradigm. *J. Manuf. Syst.* **2023**, *71*, 95–114. [CrossRef]
19. Ghobakhloo, M.; Iranmanesh, M.; Foroughi, B.; Babaee Tirkolaee, E.; Asadi, S.; Amran, A. Industry 5.0 Implications for Inclusive Sustainable Manufacturing: An Evidence-Knowledge-Based Strategic Roadmap. *J. Clean. Prod.* **2023**, *417*, 138023. [CrossRef]
20. Psarommatis, F. A Generic Methodology and a Digital Twin for Zero Defect Manufacturing (ZDM) Performance Mapping towards Design for ZDM. *J. Manuf. Syst.* **2021**, *59*, 507–521. [CrossRef]
21. Psarommatis, F.; May, G.; Dreyfus, P.-A.; Kiritsis, D. Zero Defect Manufacturing: State-of-the-Art Review, Shortcomings and Future Directions in Research. *Int. J. Prod. Res.* **2020**, *7543*, 1–17. [CrossRef]
22. Jun, J.; Chang, T.-W.; Jun, S. Quality Prediction and Yield Improvement in Process Manufacturing Based on Data Analytics. *Processes* **2020**, *8*, 1068. [CrossRef]
23. Psarommatis, F.; Kiritsis, D. A Hybrid Decision Support System for Automating Decision Making in the Event of Defects in the Era of Zero Defect Manufacturing. *J. Ind. Inf. Integr.* **2021**, *26*, 100263. [CrossRef]
24. Kumar, P.; Maiti, J.; Gunasekaran, A. Impact of Quality Management Systems on Firm Performance. *Int. J. Qual. Reliab. Manag.* **2018**, *35*, 1034–1059. [CrossRef]
25. Cromwell, S.E.; Kolb, J.A. An Examination of Work-Environment Support Factors Affecting Transfer of Supervisory Skills Training to the Workplace. *Hum. Resour. Dev. Q.* **2004**, *15*, 449–471. [CrossRef]
26. Kumar Sharma, A.; Bajpai, B.; Adhvaryu, R.; Dhruvi Pankajkumar, S.; Parthkumar Gordhanbhai, P.; Kumar, A. An Efficient Approach of Product Recommendation System Using NLP Technique. *Mater. Today Proc.* **2021**, *80*, 3730–3743. [CrossRef]
27. Müller, M.; Metternich, J. Production Specific Language Characteristics to Improve NLP Applications on the Shop Floor. *Procedia CIRP* **2021**, *104*, 1890–1895. [CrossRef]
28. Garcia Fracaro, S.; Glassey, J.; Bernaerts, K.; Wilk, M. Immersive Technologies for the Training of Operators in the Process Industry: A Systematic Literature Review. *Comput. Chem. Eng.* **2022**, *160*, 107691. [CrossRef]
29. Abidi, M.H.; Al-Ahmari, A.; Ahmad, A.; Ameen, W.; Alkhalefah, H. Assessment of Virtual Reality-Based Manufacturing Assembly Training System. *Int. J. Adv. Manuf. Technol.* **2019**, *105*, 3743–3759. [CrossRef]
30. Srinivasan, R.; Srinivasan, B.; Iqbal, M.U.; Nemet, A.; Kravanja, Z. Recent Developments towards Enhancing Process Safety: Inherent Safety and Cognitive Engineering. *Comput. Chem. Eng.* **2019**, *128*, 364–383. [CrossRef]
31. Rosero, M.; Pogo, R.; Pruna, E.; Andaluz, V.H.; Escobar, I. Immersive Environment for Training on Industrial Emergencies. In *Augmented Reality, Virtual Reality, and Computer Graphics*; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2018; Volume 10851, pp. 451–466.
32. Nakai, A.; Kajihara, Y.; Nishimoto, K.; Suzuki, K. Information-Sharing System Supporting Onsite Work for Chemical Plants. *J. Loss Prev. Process Ind.* **2017**, *50*, 15–22. [CrossRef]
33. Brambilla, S.; Manca, D. Recommended Features of an Industrial Accident Simulator for the Training of Operators. *J. Loss Prev. Process Ind.* **2011**, *24*, 344–355. [CrossRef]
34. Nazir, S.; Totaro, R.; Brambilla, S.; Colombo, S.; Manca, D. Virtual Reality and Augmented-Virtual Reality as Tools to Train Industrial Operators. *Comput. Aided Chem. Eng.* **2012**, *30*, 1397–1401. [CrossRef]

35. Weritz, P. Hey Leaders, It's Time to Train the Workforce: Critical Skills in the Digital Workplace. *Adm. Sci.* **2022**, *12*, 94. Available online: https://www.mdpi.com/2076-3387/12/3/94 (accessed on 26 October 2023). [CrossRef]
36. Fake, H.; Dabbagh, N. The Personalized Learning Interaction Framework: Expert Perspectives on How to Apply Dimensions of Personalized Learning to Workforce Training and Development Programs. In Proceedings of the Ninth International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM'21), Barcelona, Spain, 26–29 October 2021; Association for Computing Machinery: New York, NY, USA, 2021; pp. 501–509.
37. Nair, S.; Kaushik, A.; Dhoot, H. Conceptual Framework of a Skill-Based Interactive Employee Engaging System: In the Context of Upskilling the Present IT Organization. *Appl. Comput. Inform.* **2020**, *19*, 82–107. [CrossRef]
38. Shemshack, A.; Spector, J.M. A Systematic Literature Review of Personalized Learning Terms. *Smart Learn. Environ.* **2020**, *7*, 33. [CrossRef]
39. Kucirkova, N.; Gerard, L.; Linn, M.C. Designing Personalised Instruction: A Research and Design Framework. *Br. J. Educ. Technol.* **2021**, *52*, 1839–1861. [CrossRef]
40. Kluge, A.; Nazir, S.; Manca, D. Advanced Applications in Process Control and Training Needs of Field and Control Room Operators. *IIE Trans. Occup. Ergon. Hum. Factors* **2015**, *2*, 121–136. [CrossRef]
41. Tatić, D.; Tešić, B. The Application of Augmented Reality Technologies for the Improvement of Occupational Safety in an Industrial Environment. *Comput. Ind.* **2017**, *85*, 1–10. [CrossRef]
42. Leder, J.; Horlitz, T.; Puschmann, P.; Wittstock, V.; Schütz, A. Comparing Immersive Virtual Reality and Powerpoint as Methods for Delivering Safety Training: Impacts on Risk Perception, Learning, and Decision Making. *Saf. Sci.* **2019**, *111*, 271–286. [CrossRef]
43. Avveduto, G.; Tanca, C.; Lorenzini, C.; Tecchia, F.; Carrozzino, M.; Bergamasco, M. Safety Training Using Virtual Reality: A Comparative Approach. In *Augmented Reality, Virtual Reality, and Computer Graphics*; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2017; Volume 10324, pp. 148–163.
44. Tay, J.; Goh, Y.M.; Safiena, S.; Bound, H. Designing Digital Game-Based Learning for Professional Upskilling: A Systematic Literature Review. *Comput. Educ.* **2022**, *184*, 104518. [CrossRef]
45. Pedota, M.; Grilli, L.; Piscitello, L. Technology Adoption and Upskilling in the Wake of Industry 4.0. *Technol. Forecast. Soc. Chang.* **2023**, *187*, 122085. [CrossRef]
46. Doyle-Kent, M.; Shanahan, B.W. The Development of a Novel Educational Model to Successfully Upskill Technical Workers for Industry 5.0: Ireland a Case Study. *IFAC-Papers* **2022**, *55*, 425–430. [CrossRef]
47. Yin, T.; Zhang, Z.; Wu, T.; Zeng, Y.; Zhang, Y.; Liu, J. Multimanned Partial Disassembly Line Balancing Optimization Considering End-of-Life States of Products and Skill Differences of Workers. *J. Manuf. Syst.* **2023**, *66*, 107–126. [CrossRef]
48. Apostolopoulos, G.; Andronas, D.; Fourtakas, N.; Makris, S. Operator Training Framework for Hybrid Environments: An Augmented Reality Module Using Machine Learning Object Recognition. *Procedia CIRP* **2022**, *106*, 102–107. [CrossRef]
49. Checa, D.; Bustillo, A. A Review of Immersive Virtual Reality Serious Games to Enhance Learning and Training. *Multimed. Tools Appl.* **2020**, *79*, 5501–5527. [CrossRef]
50. Gajek, A.; Fabiano, B.; Laurent, A.; Jensen, N. Process Safety Education of Future Employee 4.0 in Industry 4.0. *J. Loss Prev. Process Ind.* **2022**, *75*, 104691. [CrossRef]
51. Kulkarni, P.M.; Appasaba, L.V.; Gokhale, P.; Tigadi, B. Role of Digital Simulation in Employee Training. *Glob. Transit. Proc.* **2022**, *3*, 149–156. [CrossRef]
52. Martínez-Gutiérrez, A.; Díez-González, J.; Verde, P.; Perez, H. Convergence of Virtual Reality and Digital Twin Technologies to Enhance Digital Operators' Training in Industry 4.0. *Int. J. Hum. Comput. Stud.* **2023**, *180*, 103136. [CrossRef]
53. Monetti, F.M.; de Giorgio, A.; Yu, H.; Maffei, A.; Romero, M. An Experimental Study of the Impact of Virtual Reality Training on Manufacturing Operators on Industrial Robotic Tasks. *Procedia CIRP* **2022**, *106*, 33–38. [CrossRef]
54. Dantan, J.Y.; El Mouayni, I.; Sadeghi, L.; Siadat, A.; Etienne, A. Human Factors Integration in Manufacturing Systems Design Using Function–Behavior–Structure Framework and Behaviour Simulations. *CIRP Ann.* **2019**, *68*, 125–128. [CrossRef]
55. Kolus, A.; Wells, R.; Neumann, P. Production Quality and Human Factors Engineering: A Systematic Review and Theoretical Framework. *Appl. Ergon.* **2018**, *73*, 55–89. [CrossRef] [PubMed]
56. Hosseini, M.; Malric, F.; Georganas, N.D. A Haptic Virtual Environment for Industrial Training. In Proceedings of the HAVE 2002—IEEE International Workshop on Haptic Virtual Environments and their Applications, Ottawa, ON, Canada, 17–18 November 2002; pp. 25–30. [CrossRef]
57. Lin, F.; Ye, L.; Duffy, V.G.; Su, C.J. Developing Virtual Environments for Industrial Training. *Inf. Sci.* **2002**, *140*, 153–170. [CrossRef]
58. Bao, L.; Tran, S.V.-T.; Nguyen, T.L.; Pham, H.C.; Lee, D.; Park, C. Cross-Platform Virtual Reality for Real-Time Construction Safety Training Using Immersive Web and Industry Foundation Classes. *Autom. Constr.* **2022**, *143*, 104565. [CrossRef]
59. Zhu, Q.-X.; Zhang, H.-T.; Tian, Y.; Zhang, N.; Xu, Y.; He, Y.-L. Co-Training Based Virtual Sample Generation for Solving the Small Sample Size Problem in Process Industry. *ISA Trans.* **2023**, *134*, 290–301. [CrossRef]
60. Garcia Fracaro, S.; Chan, P.; Gallagher, T.; Tehreem, Y.; Toyoda, R.; Bernaerts, K.; Glassey, J.; Pfeiffer, T.; Slof, B.; Wachsmuth, S.; et al. Towards Design Guidelines for Virtual Reality Training for the Chemical Industry. *Educ. Chem. Eng.* **2021**, *36*, 12–23. [CrossRef]
61. Kazemi Kheiri, S.; Vahedi, Z.; Sun, H.; Megahed, F.M.; Cavuoto, L.A. Human Reliability Modeling in Occupational Environments toward a Safe and Productive Operator 4.0. *Int. J. Ind. Ergon.* **2023**, *97*, 103479. [CrossRef]

62. Simeone, A.; Grant, R.; Ye, W.; Caggiano, A. Operator 4.0 Intelligent Health Monitoring: A Cyber-Physical Approach. *Procedia CIRP* **2023**, *118*, 1033–1038. [CrossRef]
63. Ciccarelli, M.; Papetti, A.; Germani, M. Exploring How New Industrial Paradigms Affect the Workforce: A Literature Review of Operator 4.0. *J. Manuf. Syst.* **2023**, *70*, 464–483. [CrossRef]
64. Gladysz, B.; Tran, T.; Romero, D.; van Erp, T.; Abonyi, J.; Ruppert, T. Current Development on the Operator 4.0 and Transition towards the Operator 5.0: A Systematic Literature Review in Light of Industry 5.0. *J. Manuf. Syst.* **2023**, *70*, 160–185. [CrossRef]
65. Özcan, A.M.; Akdoğan, A.; Durakbasa, N.M. Improvements in Manufacturing Processes by Measurement and Evaluation Studies According to the Quality Management System Standard in Automotive Industry. In Proceedings of the Lecture Notes in Mechanical Engineering, Antalya, Turkey, 24–26 September 2021; Springer Science and Business Media: Deutschland, Germany, 2021; pp. 483–492.
66. Psarommatis, F.; Prouvost, S.; May, G.; Kiritsis, D. Product Quality Improvement Policies in Industry 4.0: Characteristics, Enabling Factors, Barriers, and Evolution toward Zero Defect Manufacturing. *Front. Comput. Sci.* **2020**, *2*, 26. [CrossRef]
67. Psarommatis, F.; Fraile, F.; Ameri, F. Zero Defect Manufacturing Ontology: A Preliminary Version Based on Standardized Terms. *Comput. Ind.* **2023**, *145*, 103832. [CrossRef]
68. Mourtzis, D.; Vlachou, E.; Milas, N. Industrial Big Data as a Result of IoT Adoption in Manufacturing. *Procedia CIRP* **2016**, *55*, 290–295. [CrossRef]
69. Chien, C.F.; Hsu, S.C.; Chen, Y.J. A System for Online Detection and Classification of Wafer Bin Map Defect Patterns for Manufacturing Intelligence. *Int. J. Prod. Res.* **2013**, *51*, 2324–2338. [CrossRef]
70. Kuo, C.F.; Hsu, C.T.M.; Fang, C.H.; Chao, S.M.; Lin, Y.D. Automatic Defect Inspection System of Colour Filters Using Taguchi-Based Neural Network. *Int. J. Prod. Res.* **2013**, *51*, 1464–1476. [CrossRef]
71. Choi, G.; Kim, S.H.; Ha, C.; Bae, S.J. Multi-Step ART1 Algorithm for Recognition of Defect Patterns on Semiconductor Wafers. *Int. J. Prod. Res.* **2012**, *50*, 3274–3287. [CrossRef]
72. Psarommatis, F.; May, G.; Kiritsis, D. Predictive Maintenance Key Control Parameters for Achieving Efficient Zero Defect Manufacturing. *Procedia CIRP* **2021**, *104*, 80–84. [CrossRef]
73. Psarommatis, F.; May, G. A Practical Guide for Implementing Zero Defect Manufacturing in New or Existing Manufacturing Systems. *Procedia Comput. Sci.* **2023**, *217*, 82–90. [CrossRef]
74. Psarommatis, F.; Kiritsis, D. Comparison between Product and Process Oriented Zero-Defect Manufacturing (ZDM) Approaches. In *Proceedings of the IFIP Advances in Information and Communication Technology*; Springer: Cham, Switzerland, 2021; pp. 105–112.
75. Psarommatis, F.; May, G. A Systematic Analysis for Mapping Product-Oriented and Process-Oriented Zero-Defect Manufacturing (ZDM) in the Industry 4.0 Era. *Sustainability* **2023**, *15*, 12251. [CrossRef]
76. Ameri, F.; Sormaz, D.; Psarommatis, F.; Kiritsis, D. Industrial Ontologies for Interoperability in Agile and Resilient Manufacturing. *Int. J. Prod. Res.* **2022**, *60*, 420–441. [CrossRef]
77. Grevenitis, K.; Psarommatis, F.; Reina, A.; Xu, W.; Tourkogiorgis, I.; Milenkovic, J.; Cassina, J.; Kiritsis, D. A Hybrid Framework for Industrial Data Storage and Exploitation. In *Proceedings of the Procedia CIRP*; Elsevier: Amsterdam, The Netherlands, 2019; Volume 81, pp. 892–897.
78. Vaswani, A.; Brain, G.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention Is All You Need. *Adv. Neural Inf. Process. Systems* **2017**, *30*.
79. Ayadi, A.; Auffan, M.; Rose, J. Ontology-Based NLP Information Extraction to Enrich Nanomaterial Environmental Exposure Database. *Procedia Comput. Sci.* **2020**, *176*, 360–369. [CrossRef]
80. Lareyre, F.; Nasr, B.; Chaudhuri, A.; Di Lorenzo, G.; Carlier, M.; Raffort, J. Comprehensive Review of Natural Language Processing (NLP) in Vascular Surgery. *EJVES Vasc. Forum* **2023**. [CrossRef] [PubMed]
81. Ayadi, A.; Samet, A.; de Beuvron, F.d.B.; Zanni-Merk, C. Ontology Population with Deep Learning-Based NLP: A Case Study on the Biomolecular Network Ontology. *Procedia Comput. Sci.* **2019**, *159*, 572–581. [CrossRef]
82. Chung, S.; Moon, S.; Kim, J.; Kim, J.; Lim, S.; Chi, S. Comparing Natural Language Processing (NLP) Applications in Construction and Computer Science Using Preferred Reporting Items for Systematic Reviews (PRISMA). *Autom. Constr.* **2023**, *154*, 105020. [CrossRef]
83. Patel, A.S.; Merlino, G.; Puliafito, A.; Vyas, R.; Vyas, O.P.; Ojha, M.; Tiwari, V. An NLP-Guided Ontology Development and Refinement Approach to Represent and Query Visual Information. *Expert. Syst. Appl.* **2023**, *213*, 118998. [CrossRef]
84. Sala, R.; Pirola, F.; Pezzotta, G.; Cavalieri, S. NLP-Based Insights Discovery for Industrial Asset and Service Improvement: An Analysis of Maintenance Reports. *IFAC-PapersOnLine* **2022**, *55*, 522–527. [CrossRef]
85. Mourtzis, D.; Angelopoulos, J.; Siatras, V.; Panopoulos, N. A Methodology for the Assessment of Operator 4.0 Skills Based on Sentiment Analysis and Augmented Reality. *Procedia CIRP* **2021**, *104*, 1668–1673. [CrossRef]
86. *ANSI/ISA-95.00*; Enterprise-Control System Integration—Part 2: Object Model Attributes. American National Standard: Washington, DC, USA, 2001.
87. SCOR: Supply-Chain Reference Model. ILIM, Institute of Logistics and Warehousing ainia, Technological Centre. Available online: https://www.tecnoali.com/files/emensa/D11/Report%20Ilim.pdf (accessed on 1 September 2023).
88. *Explosion: Available Trained Pipelines for Spanish*; Explosion: Berlin, Germany, 2022.
89. Devlin, J.; Chang, M.-W.; Lee, K.; Google, K.T.; Language, A.I. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. *arXiv* **2018**, arXiv:1810.04805.

90. Guven, Z.A.; Unalir, M.O. Natural Language Based Analysis of SQuAD: An Analytical Approach for BERT. *Expert. Syst. Appl.* **2022**, *195*, 116592. [CrossRef]
91. Turchin, A.; Masharsky, S.; Zitnik, M. Comparison of BERT Implementations for Natural Language Processing of Narrative Medical Documents. *Inform. Med. Unlocked* **2023**, *36*, 101139. [CrossRef]
92. Jupin-Delevaux, É.; Djahnine, A.; Talbot, F.; Richard, A.; Gouttard, S.; Mansuy, A.; Douek, P.; Si-Mohamed, S.; Boussel, L. BERT-Based Natural Language Processing Analysis of French CT Reports: Application to the Measurement of the Positivity Rate for Pulmonary Embolism. *Res. Diagn. Interv. Imaging* **2023**, *6*, 100027. [CrossRef]
93. Xu, S.; Zhang, C.; Hong, D. BERT-Based NLP Techniques for Classification and Severity Modeling in Basic Warranty Data Study. *Insur. Math. Econ.* **2022**, *107*, 57–67. [CrossRef]
94. Cañete, J.; Chaperon, G.; Fuentes, R.; Ho, J.-H.; Kang, H.; Pérez, J. Spanish Pre-Trained BERT Model and Evaluation Data. *arXiv* **2023**, arXiv:1810.04805.
95. Reimers, N.; Gurevych, I. Sentence-BERT: Sentence Embeddings Using Siamese BERT-Networks. In Proceedings of the EMNLP-IJCNLP 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Hongkong, China, 3–7 November 2019; pp. 3982–3992. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.