AI Implementation and Capability Development in Manufacturing: An Action Research Case

Jon Eklöf
University of Copenhagen,
GKN Aerospace
jon.eklof@gmail.com

Ulrika Lundh Snis
University West
ulrika.snis@hv.se

Thomas Hamelryck
University of Copenhagen
thamelry@bio.ku.dk

Alexander Grima
GKN Aerospace
alexander.grima@gknaerospace.com

Ola Rønning
University of Copenhagen
ola@di.dk.ku

Abstract

This action research article presents a case study of a global manufacturing company deploying artificial intelligence (AI) to develop capabilities and enhance decision-making. This study explores considerations and trade-offs involved in introducing AI into daily operations, leading up to the decision to develop AI capabilities in-house or outsource them.

The case study offers in-depth technical descriptions of model selection, dataset creation, model adoption, model training and evaluation while addressing organizational obstacles and decision-making processes. The study's findings highlight the importance of collaboration between technical experts, business leaders, and end-users, as well as the interaction and collaboration between AI systems and human employees in the workplace.

The article contributes a practical perspective on AI implementation in manufacturing, emphasizing the need to balance in-house capability development with external acquisition. Although the case study company managed to create an in-house model, factors such as implementation, debugging, data requirements, training time, and performance led to outsourcing the capabilities. However, making this informed decision required capabilities and insights that were acquired through practical work. Consequently, although in-house development can be challenging, it can also enhance organizational capabilities and provide the necessary knowledge to make informed decisions about future development or outsourcing.

Keywords: AI capability development, AI implementation, AI in manufacturing

1. Introduction

The adoption of AI in manufacturing has received significant attention in both academia and industry in recent years (Zeba, Dabic, Čičak, & Daim, 2021). Most of the research on the implementation of AI in manufacturing has been conducted independently from daily operations (Arinez, Chang, Gao, Xu &

Zhang, 2020). Further investigations are needed to better understand the implications of integrating AI into daily operations and manage AI project and deployment risk. It also raises questions on how organizations should manage, staff and coordinate AI development. While contemporary AI research typically describes a computational solution to a specific problem, this article investigates the practical process of developing such a solution, the obstacles encountered, and the considerations made during the process. This approach is supported by (Amabile, 1996) and (Govindarajan & Trimble, 2012), who argue that the solution-finding process is important in enhancing creativity and innovation and overcoming complex problems.

When considering integrating AI into their operations, companies must decide whether to develop capabilities in-house or outsource them to external service providers (Ransbotham, Gerbert, Kiron, & Reeves, 2017). Outsourcing AI capabilities allows companies to quickly acquire necessary expertise but may result in a loss of competitive advantage provided by AI solutions, only gaining access to technology and competence available in the open market (Govindarajan & Immelt, 2019; Teesce, 2014). In-house development provides organizations with greater control over the development process, but this comes with a cost. Companies must therefore weigh the resources and time needed against the benefits of acquiring the same capability externally. There is no clear guidance on the decision-making process or criteria for making such decisions (Ransbotham, Gerbert, Kiron, & Reeves, 2017; Govindarajan & Immelt, 2019; Teesce, 2014).

This paper aims to investigate the implications, considerations, and trade-offs of introducing AI into daily operations of a manufacturing company, leading up to the decision of whether to develop AI capabilities in-house or outsource them and the factors that influence this decision. The case study focuses on the in-house development of an AI model for defect detection in X-rays of welds of aerospace components (the In-House Model). It provides IS professionals

accounts of managing challenges faced by contemporary corporations and organizations. Although the current study investigates capability development in one company only, it aims to generalize within the specific setting, focusing on theoretical abstraction and insights that are well-grounded in the selected case.

2. Identifying the problem domain

2.1. The case

In the current case, an aerospace component manufacturing company (the Company) was increasing production of a critical component, the turbine exhaust case (TEC), the final stage of a commercial jet engine. The TEC is vital for heat dissipation and performance. It is a complex load-bearing structure designed to withstand high temperatures and loads while also maintaining aerodynamic performance, low weight, and cost-effectiveness. The TEC has a diameter of approximately 1 meter and weighs roughly 75 kgs. It is built by welding several segments together, and each TEC comprises more than 100 welds that undergo inspection using X-rays, a method commonly used to detect defects such as gas pores and cracks in welds (Tyystjärvi et al., 2022).

Figure 1: Image of the turbine exhaust case - the component subject to weld inspection.

As the production rate was ramping up, an increasing amount of welds required inspection. The X-ray images of the TEC were digitally scanned, and each scan contained between two and four X-ray images of welds. Three operators spent thousands of hours inspecting hundreds of thousands of images per year. As pointed out by Bertović, such a manual inspection process is time-consuming, fail-prone, and operator-dependent (Bertović, 2016). In the current case, the process of training and certifying these inspectors was expensive and cumbersome, taking up to two years of training and practical experience for operators to become certified for inspection. Although minor defects were present, critical defects were rare and did not often render components defective. As a

result, the monotony of inspecting welds and the rarity of critical defects increased the risk of human-related errors.

The Company had previously assessed automated inspection methods utilizing machine learning, but faced obstacles such as low or inconsistent contrast and brightness, and defect-like anomalies or geometries producing excessive false positives. These issues have been identified as recurring problems with automated inspection (Nacereddine, Zelmat, Belaïfa, & Tridi, 2005; Ronneberger, Fischer, & Brox, 2015). Despite having a database of hundreds of thousands of X-ray images, the Company lacked labeled data suitable for training AI models.

2.2. Related research

The study builds on previous research suggesting that machine learning techniques, including deep learning, can be used to detect defects in X-ray images (Bertovic & Virkkunen, 2021). To explore this further, a literature search was conducted on Google Scholar using the inclusion criteria "*defect detection,*" "*X-ray,*" and "*welds,*" in conjunction with the keywords "*machine learning*" and "*deep learning*" respectively. The search identified 18 articles investigating various AI techniques for detecting defects in X-ray images. These techniques include traditional machine learning methods such as support vector machines (Wang Y. e., 2008) as well as deep learning techniques such as segmentation networks (Tyystjärvi et al., 2022), generative adversarial networks (GANs) (Akcay, Atapour-Abarghouei, & Breckon, 2019; Guo et al., 2021), autoencoders (AEs) (Presenti et al., 2022), variational autoencoders (VAEs) (Banko et al., 2021; Lindgren & Zach, 2021), and, to the greatest extent, convolutional neural networks (CNNs) (Jiang et al. L. , 2021; Yaping & Weixin, 2019; Wang, Shi, & Tong, 2019; Yang et al. D. , 2021; Deng et al., 2021; Naddaf-Sh et al., 2021; Ajmi et al., 2020; Wen-ming, 2019; Jiang et al. H. , 2021; He et al., 2017) (Yang et al. L. , 2021). However, to fully leverage the benefits that deep learning techniques provide, it is typically necessary to have a large data base of labeled examples (Presenti et al., 2022). Creating labeled training data for image analysis in non-destructive evaluation can be challenging due to the complex and often subtle nature of the features that need to be identified. Because manual labeling is labor-intensive and susceptible to errors or inconsistency between labels, it becomes increasingly problematic when labeling large data volumes. As a result, there are limited public data resources available for training machine learning models to perform this task (Tyystjärvi et al., 2022; Mery et al., 2015).

More recently, semi-supervised VAEs (SS-VAEs) have emerged as a tool for image classification in instances where, as in the current case, there are low amounts of labeled data available, but a fair amount of

unlabeled data (Kong & Ni, 2020). However, our literature search revealed no studies investigating use of SS-VAEs for defect detection in X-ray images.

2.3. About the SS-VAE

The SS-VAE builds on the VAE, which, in turn, builds on the AE architecture. The AE contains an encoding function that maps an input to a compressed latent space representation and a decoding function that maps from the latent space back into the original space (Maheshwari, Mitra, & Sharma, 2022).

The VAE maps an input to a probability distribution over the latent space, allowing it to model complex and multimodal data distributions and generate new data by sampling the distribution (Kingma & Welling, 2014).

VAEs have been used in various domains, including image recognition and anomaly detection (Maheshwari, Mitra, & Sharma, 2022). In instances where labeled data is limited, it is possible to estimate the label using a SS-VAE. The SS-VAE decoder (i.e. the probabilistic model) represents missing labels as latent variables sampled from a prior distribution while the encoder (i.e. the approximating distribution or guide) infers missing labels from the image data. This approach has been used to improve classification in various domains with a limited amount of labeled data (Kingma & Welling, 2014; Wu et al., 2021).

3. Methodological framework

3.1. Action research

The study uses Mathiassen et al.'s action research model, which involves collaboration between practitioners and researchers to address real-world issues (Mathiassen, Chiasson, & Germonprez, 2012). Put simply, action research is a "learning by doing" process where a group of individuals identifies a problem, plans how to tackle the problem, takes action to resolve it, evaluates the success of their efforts, and iterates as necessary (Susman, 1983; O'Brien, 1998). The distinguishing feature of action research compared to professional practices is its scientific approach, wherein the problem is systematically studied, and the intervention is grounded in theoretical considerations (O'Brien, 1998). Through collaboration between practitioners and researchers, this study seeks to contribute to solving the Company's problem of establishing necessary capabilities while addressing the needs identified in current research, how to identify necessary capabilities and optimal ways to develop them.

Furthermore, this study adopts a practice-based research approach that leverages the main author's expertise in leading digital transformation, enabling the development of necessary capabilities and

constant reflection on and refinement of the practice (Candy, 2006). By adopting this approach, the study aims to not only address the Company's problem but also to contribute to the academic literature on capability development in practice-based contexts.

3.2. Agile approach with experiment-first-mindset

Rather than adhering to a strict plan, setting out each step and the requirements of the end product, the study utilized an agile approach, emphasizing experimentation and creativity in the project. The project was executed as a series of iterative experiments, refining the project design based on continuous feedback, collaboration, and flexibility, which are important components both in action research and in addressing the challenges of AI projects (Sousa, 2019; Mukherjee, 2020; Steiber, Alänge, Ghosh, & Goncalves, 2020; Curcio et al., 2018).

3.3. Action Planning

In this case, the Company's senior leadership requested their in-house digital innovation team (the Team), led by the main author, to develop a model in-house (In-house Model) that could be used as a benchmark against a proprietary model (Proprietary Model) from an external service provider specializing in detecting defects in images with limited labeled data. The Team had no insight into the Proprietary Model other than that it was partly trained in an unsupervised manner. The Company's goal was to make a decision within a few months, so the Team was reminded of the importance of time and tasked with creating a model that allowed them to benchmark the Proprietary Model. The Team had limited experience of developing advanced AI models in general and models for defect detection in particular and therefore needed to develop these capabilities.

The Team had a proved ability to implement digital solutions into various areas of the Company. However, given the Team's limited experience working with AI solutions, the study adopted an action research approach that involved forming a collaborative group of both external and internal stakeholders to provide a broader range of expertise and perspectives (Reason & Bradbury, 2001). The Team collaborated with external experts in deep probabilistic programming (the Experts) and internal X-ray inspection operators (the Operators) who provided insights and recommendations based on their expertise. The Experts and Operators also served as a benchmark for evaluating the effectiveness of the interventions implemented during the research process.

As the Team leader, the study's main author was responsible for AI capability development within the

Company. In this study, he worked alongside an industry programmer to perform the practical work, enabling practice-based research. Simultaneously, the main author researched leadership aspects of AI implementation in the manufacturing industry allowing for the immediate application of research findings in the field.

To further conceptualize the results, ensuring they were research-grounded and contributed to the academic discourse, the main author collaborated with the second author and engaged with the research community. This approach resulted in insights that enhanced the practical experiences gained in the manufacturing industry, making the results both practically and scientifically grounded.

4. Taking action

This section describes the chronological order of the different stages of the practical work being undertaken, namely model selection, dataset creation, model development, model training, and model evaluation.

4.1. Selecting the In-house Model

The Company's senior leadership had requested the Team to develop an In-house Model that could serve as a benchmark against the Proprietary Model. However, the leadership did not provide explicit criteria for the In-House Model, apart from its role as a benchmark. Faced with a lack of clear guidelines, the Team collaborated with the Experts to determine the properties the model should possess to function as a benchmark. Considering the Proprietary Model's ability to operate with limited labeled data, and the fact that there was a fair amount of unlabeled data available, the Experts advised the Team to develop a model that could efficiently utilize unlabeled data without relying on extensive labeled data. With this in mind, they recommended the Team to construct a SS-VAE. For a comprehensive description of the SS-VAE, please see Kingma and Welling, 2014).

While creating a model from scratch may have resulted in better performance, the Team faced time constraints. To expedite the development process and meet the leadership's goal of making a decision soon, the Experts recommended that the team adopt a publicly accessible model (the Baseline model) implemented in the deep probabilistic programming language Pyro (https://pyro.ai/examples/ss-vae.html). This approach, they suggested, would provide an adequate benchmark for comparison purposes.

4.2. Creating the dataset

The inspection criteria for the TEC state that several minor defects, such as gas pores located within a specific area, could render the component defective. In consultation with the Operators, the Team identified that the In-House Model needed to be able to process 256x256 pixel images to capture all known types of defects, based on the inspection criteria.

According to the external service provider, around 20 images per class were required to effectively showcase the capabilities of the Proprietary Model. The Experts advised that for an SS-VAE, 20 labeled images per class was low. The Baseline model was trained using a varied range of labeled MNIST images (28x28 pixel black and white representations of the handwritten digits 0-9), spanning from 100 to 3000 instances (Pyro, 2017). This means that there were approximately 10 to 300 training images available per individual handwritten digit. It is worth noting that the MNIST images were considerably smaller compared to the 256x256 pixel images intended for use in the In-house Model.

Although the lower number of labeled training images in the Baseline model potentially could have an adverse effect on the accuracy of the In-house Model, the Team deemed that an important aspect of benchmarking the Proprietary Model would be to use the same amount of labeled training data. The Team therefore decided to use 20 around labeled images per class for training.

Together with the Operators, the Team generated a labeled dataset from 10 full-size scans (Scans), containing between four to five X-ray images of welds (Weld-images) of the TEC. For each Weld-image, the regions that could contain defects and the defects in each region were marked. From these regions, the Weld-images were divided into 256x256 pixel images (Sub-images).
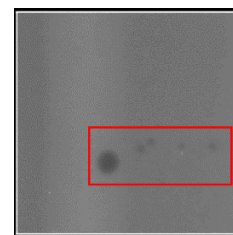

Figure 2: An example of a 256x256 pixel Sub-image with marked defects.

From these Sub-images the Team created a labeled training dataset containing 21 defect and 21 non-defect Sub-images. This was the dataset that was provided to the external consultant under a confidentiality agreement to train the Proprietary Model. Due to the low amount of labeled training data, the Team augmented the Sub-images to increase the number of data points in the training dataset. By flipping the Sub-images, a commonly known method to increase the number of data points (Shorten & Khoshgoftaar., 2019), the Team generated an additional three images per labeled image. This resulted in a total of 168 Sub-images for the training

dataset - 84 Sub-images per class. The Team decided together with the Experts to leverage the SS-VAE's ability to use unlabeled data for training.

The Baseline model used a ratio of 1:500 to 1:16 between labeled and unlabeled images. The Experts initially recommended similar ratios for the project. However, due to the differences in size between 28x28 pixel MNIST images used in the Baseline model and the 256x256 pixel Sub-images in the current project, the Team recognized that using the same ratios as the Baseline model would significantly increase the training time of the model, to the point of being infeasible (for example, a ratio of 1:500 would have resulted in a training time of over eight years).

The Team thus faced a trade-off between model optimization and speed of execution. Using more labeled images would likely increase the accuracy of the In-House Model but also increase the training time. In consultation with the Experts, the Team ultimately opted for a ratio of 1:10 between labeled and unlabeled Sub-images. This decision was motivated by the fact that the labeled dataset already contained a sufficient number of training examples per class, and that a ratio of 1:10 was deemed to provide enough information to be able to evaluate the In-House Model.

The Team thus generated an unlabeled dataset of 1680 Sub-images using the same process as the labeled training data, excluding augmentation.

4.3. Reducing dimensionality

The Sub-images initially had a 16-bit grayscale depth. The Experts advised mapping the pixel values to a categorical distribution. Instead of representing each pixel's intensity as a numerical value, it would be transformed into a probability distribution across categories, where each category represented a specific intensity level. Unfortunately, this approach caused a memory leak on the GPU due to a bug in Pyro - the software framework used to build the In-house Model (Pyro, 2022). The Team spent weeks trying to solve this issue. Eventually, the Team sought help from the Pyro development team who suggested using a previous version of Pyro, which inexplicably solved the memory leak.

However, after this issue was solved, the Team realized that, despite using powerful GPUs, training the In-House Model on the 16-bit grayscale Sub-images would take over six months, which was unfeasible given the timeline provided by the Company leadership. Consequently, the Team explored various methods of reducing the dimensionality of the images.

Hou et al. have emphasized the importance of pre-processing high-dimensional data to reduce the complexity of classification tasks and computational costs (Hou et al., 2020). To determine the necessary information level for Operators to detect potential defects in the Sub-images, the Team consulted with them and concluded that black and white images would suffice for capturing all defects. To achieve this, the Team used OpenCV, a Python framework, and the CV2 module, along with an adaptive binary Gaussian model, to reduce the dimensionality of the images and convert them to black and white (OpenCV, 2022).

The adaptive binary Gaussian algorithm sets pixel thresholds using surrounding regions, accommodating images with varying illumination. However, given the variability in optimal thresholds for different images, the Team adjusted them according to the weld inspection criteria.

Noise and artifacts in images complicated the conversion process, potentially impacting the model's accuracy. Thus, the Tam and Operators jointly reviewed all training images to ensure no defects were missed during conversion.
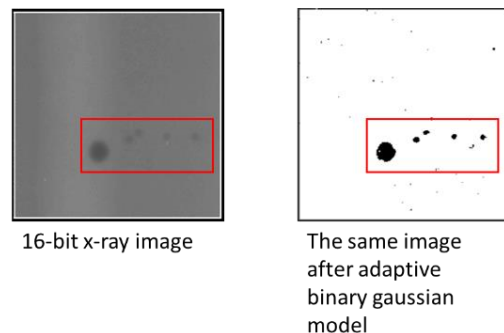

16-bit x-ray image

The same image after adaptive binary gaussian model

Figure 3: A comparison of a Sub-image with 16-bit grayscale depth and its conversion to black and white using an adaptive binary Gaussian model.

4.4. Adopting and training the In-house Model

The Team adopted the Baseline model, a semi-supervised variational autoencoder using the Python programming framework Pyro. Python code for the Baseline model can be found at https://github.com/pyro-ppl/pyro/blob/dev/examples/vae/ss_vae_M2.py. The In-house Model consisted of an encoder, a latent space, a classifier and a decoder.

The encoder utilized four tiers of convolutional and max pooling layers, each with 4, 8, 16, and 16 channels, respectively. These layers were defined by a kernel size of 3, a stride of 1, and a pooling size of 2. Following these layers, three fully connected layers were employed, having 3136 and 1000 nodes, and a pair of parallel layers each with 75 nodes. Softplus was used as the activation function for both the convolutional and fully connected layers. The encoder configured a 75-dimensional latent space, where each dimension was represented by a Gaussian distribution.

The classifier, sharing the same architecture as the encoder, differed by using 500 hidden units in the second fully connected layer. The output, rather than being mapped to a 75-dimensional space, was linked to two separate classes. The final layer of the classifier implemented a Softmax activation.

As for the decoder, it maintained a structure inverse to that of the encoder. It was composed of three sets of fully connected layers, followed by four tiers of upscale and transpose convolutional layers. The Sigmoid function was used for the final activation of each value, with the probability distribution of a single pixel denoted by a Bernoulli distribution.

In the adoption process, the latent space was a key consideration. Contrary to the approach suggested by e.g. Li et al., which advocates for meticulous optimization when configuring AI models (Li, Swersky, & Zemel., 2015), the Team, in collaboration with the Experts, opted for a less laborious approach. Instead of conducting an exhaustive search, they made an informed decision to expand the latent space from 50 as used in the Baseline model to 75 dimensions. This decision was influenced by considerations such as the size and complexity of the images.

To enhance the model's training time and ensure high defect detection accuracy in the Sub-images, convolutional neural networks (CNNs) were incorporated. Recognized for their proficiency in image analysis and pattern recognition tasks, CNNs can assign significance to spatial relationships within the data, capturing image-specific features while minimizing the parameter count (O'Shea and Nash 2015). The Team experimented with various CNN configurations, eventually integrating four sets of convolutional and max pooling layers, followed by three fully connected layers into the encoder. This decision was informed by the need to balance model complexity, computational efficiency, and image resolution.

4.5. Model training

Further delays were incurred by the global semiconductor shortage, which resulted in a lack of powerful hardware for model training, as well as specific IT department hardware and software requirements, before the Team was able to start training the In-house Model. The encoder was trained using mini-batches of 300 Sub-images. The encoder transformed the Sub-image's 65,536 pixels into the 75 dimension latent space. The classifier then categorized the Sub-image as defective or not, leveraging both labeled and unlabeled Sub-images for training and label estimation, respectively. The decoder, using the latent representation of a Sub-image and its estimated label, reconstructed the Sub-image, outputting 65,536 pixel values of 0 or 1 due to the Sub-images' conversion to black and white.

Prior to the In-house Model training, the Team conducted initial experiments using a fully supervised VAE with the labeled training dataset of Sub-images. Given that the VAE converged after approximately 85,000 epochs, the Team decided to train the In-house Model for the same number of epochs.

4.6. Evaluating performance and training time

The common evaluation measurements for these kinds of models are accuracy $\frac{true\ positives + true\ negatives}{total\ number\ of\ predctions}$, precision $\frac{true\ positives}{true\ positives + false\ positives}$ and, recall $\frac{true\ positives}{true\ positives + false\ negatives}$ (Tyystjärvi et al., 2022). Hence, these were used as evaluation measurements of the In-house Model after 85 000 epochs. A summary of the results can be found in Table 1. The results on accuracy, precision and recall of the Proprietary Model were received from the external consultant within a week after providing the training data.

Table 1: A comparison of training time, accuracy, precision, and recall between the In-house Model and the Proprietary Model

Model	Accuracy	Precision	Recall
In-house Model	75%	96%	46%
Proprietary Model	87%	100%	100%

As can be seen in Table 1, the In-house Model achieved an accuracy of 75% and a precision of over 90%, but had a recall of only 46%, indicating that it often misclassified images as not containing defects. The acceptance criteria for the TEC requires the detection of all defects. A recall as low as 46% would render it unsuitable for use in production. Even with additional data and training, it was deemed unlikely that the recall of the In-house Model would reach the required 100%. The Proprietary Model achieved an accuracy of 87% with 100% precision and recall.

The Baseline model used the MNIST dataset consisting of 60 000 28x28 pixel images. However, when the Team applied the In-house Model to its Sub-image dataset of 256x256 pixel images, even though it consisted of only 1848 images, the training time increased significantly. Despite training the In-house Model for nine weeks and 85,000 epochs, the reconstruction loss had not yet converged. The external consultant provided the results of the Proprietary Model within a week after receiving the training data. Even though the Proprietary Model was not perfect and misclassified some images as

containing defects, the Team concluded that both the performance and training time of the Proprietary Model was superior to the In-house Model and that further work on the In-House Model therefore should be discontinued.

5. Lessons learned

In the work of identifying and developing necessary capabilities of AI implementation, the hands-on work of developing a solution and dealing with the obstacles and considerations encountered along the way eventually proved to be more important than the solution itself. As Govindarajan and Trimble and have emphasized, the process of finding solutions is crucial for enhancing creativity, innovation, and solving complex problems (Govindarajan & Trimble, 2012).

The development of the In-house Model revealed that a range of capabilities were required that were not anticipated when starting the project. This illustrates the difficulty of predicting the necessary capabilities beforehand, the importance of hands-on work for capability development and the need of investing in diverse skill development to cultivate a workforce capable of addressing the challenges and complexities of AI development. In the current case, these unforeseen capabilities were developed through a continuous learning process in close collaboration with the internal and external stakeholders as well as through formal and informal course work.

Below, we discuss key lessons learned in terms of capability development.

5.1. Lessons learned for the Team

To begin with, the Team needed to create an underlying data structure that could efficiently store and label the Sub-images. Once the defects in the Sub-images were marked, the Team had to develop a script that converted the marked regions of the Sub-images that contained defects into a dataset that could be merged with the remaining metadata from the Sub-images.

The Sub-images had a wide range of brightness levels, and their pixel values were represented in 16-bit grayscale. To convert them into black and white images, the Team employed thresholding, which involves setting a threshold value. However, determining the optimal threshold value for each image was challenging as it can vary depending on the image's characteristics and using a global thresholding approach could lead to information loss.

To tackle these challenges, the Team had to acquire new image processing skills to be able to handle the images' high dynamic range and the presence of noise and artifacts while identifying defects accurately. This required collaboration with Operators to gain a deep understanding of the inspection domain and from this

understanding develop appropriate thresholding strategies to capture the necessary information accurately.

Further, to be able build the In-house Model, the Team had to acquire a basic understanding of the theoretical foundations of deep probabilistic models, including Bayesian statistics, variational inference, and deep generative models. Additionally, they needed to develop practical capabilities in hyperparameter tuning and distribution selection, crucial components of configuring the SS-VAE architecture.

The In-house Model relied on unlabeled data to learn the underlying distribution of the input data. The Team had to make informed decisions and trade-offs regarding the size and ratios of the training dataset to ensure that the SS-VAE architecture was robust and generalizable to new data.

5.2. IT infrastructure and capabilities

The project not only enhanced the capabilities of the Team but also provided valuable insights for the IT department within the Company. IT infrastructure and capabilities proved essential for the effective development of the In-house Model. The development required a robust IT infrastructure, comprising appropriate hardware and software tools.

As the Team developed the In-house Model, the IT department gained knowledge on the current hardware and software requirements. They had to gain expertise in selecting appropriate hardware for the project, requiring an understanding of the latest advancements in hardware technologies, such as GPU architectures.

In addition to high-performance hardware, the Team required specific programming frameworks - PyTorch, Pyro and OpenCV. The IT department thus had to acquire knowledge about these programming frameworks, ensuring compatibility with hardware and assessing support and documentation availability. The project provided an opportunity for the IT department to develop the necessary capabilities to support future AI projects.

6. Implications and generalizability

One of the limitations of practice-based research is its generalizability to wider populations. However, generalization across cases is not always necessary (Geertz, 1973). Instead, generalization can be achieved by generalizing within individual cases, identifying theoretical abstractions, and using these insights to generalize to theory (Lee & Baskerville, 2003). The current study investigates capability development in one company only. The intent of this in-depth engagement with a particular case, and with a particular technology, extends beyond the mere application of statistical generalizability across cases.

This research intends to provide theoretical insights that are applicable to the particular case under study and that contribute to a more generalized understanding of AI capability development. Below we discuss the implications and contributions based on the findings of the study and in light of current research.

6.1. Balancing cost and benefit of in-house solutions

The Company found that an in-house solution was not viable due to challenges with Sub-images, long training times, and the superior performance of the Proprietary Model. As highlighted by Ransbotham and others, this emphasizes the need to weigh the pros and cons of in-house AI development, considering expertise, resources, and time for model creation and validation (Ransbotham, Gerbert, Kiron, & Reeves, 2017; Govindarajan & Immelt, 2019; Teesce, 2014). In some cases, it may be more effective and efficient to leverage existing AI solutions or to outsource AI development to external service providers who can offer specialized expertise and resources. Although developing the In-House Model was challenging, such efforts can strengthen the capabilities of organizations, enabling them to better weigh the pros and cons of future in-house development or outsourcing opportunities.

6.2. Collaboration with internal and external stakeholders

The Team overcame many of its challenges through collaboration with the Experts and Operators. Firstly, this points to the importance of having a proactive mindset and the ability to recognize one's limitations to identify the need for further knowledge. Secondly, it shows the significance of seeking internal or external expertise to gain additional knowledge, overcome challenges, and advance the organization's capabilities. Collaboration is an essential component of building in-house AI capabilities (Govindarajan & Immelt, 2019; Mukherjee, 2020; Ancona, 2019). Encouraging open communication and knowledge sharing across departments, as well as with external partners, strengthens collaborative capabilities of employees, management, and external experts, which in turn can help enhance creativity and innovation, as well as allowing organizations to overcome complex problems (Govindarajan & Trimble, 2012).

6.3. Agile and experimental approach

By adopting an agile and experimental approach that prioritized flexibility and collaboration, the Team made progress and quickly adapted to changing requirements and unforeseen obstacles. This approach aligns with the works of several authors who have emphasized the importance of agility and experimentation in AI development (Sousa, 2019; Mukherjee, 2020; Steiber, Alänge, Ghosh, & Goncalves, 2020). Continuous feedback from all stakeholders enabled the Team to refine the project design and improve model performance iteratively, ensuring that the final product met the needs of the Company leadership. Moreover, by emphasizing creativity and experimentation, the Team was able to use innovative solutions to overcome complex problems, such as using OpenCV for dimensionality reduction.

6.4. Recognizing the value of human expertise in AI development

Subject matter expertise proved integral to the AI development process. The Operators' domain-specific knowledge about X-ray inspection was crucial in developing both the dataset and the model in a way that accurately detected defects. As suggested by Fountaine and Govindarajan, by leveraging subject matter experts' knowledge and expertise, organizations can develop AI models that are more accurate, effective, and useful, ultimately benefiting both workers and the organization as a whole (Fountaine, McCarthy, & Saleh, 2019; Govindarajan & Immelt, 2019).

6.5. AI Models as tools to augment human labor

Further, by involving such internal subject matter experts, organizations can alleviate concerns about obsolescence and demonstrate the value of workers in the AI development process (Babic, Chen, Evgeniou, & Fayard, 2020; Fountaine, McCarthy, & Saleh, 2019; Schepman & Rodway, 2020). In the present case, it soon became clear that neither modelthe In-house Model or the Proprietary Model fully could replace human capabilities in detecting defects in X-ray images. This experience illustrates the importance of recognizing the limitations of AI models and leveraging them as tools to augment human labor rather than replace it entirely. Incorporating AI into the workforce can improve job satisfaction by enabling workers to focus on higher-level tasks that require human expertise, rather than performing tedious and repetitive tasks (Babic, Chen, Evgeniou, & Fayard, 2020; Fountaine, McCarthy, & Saleh, 2019; Schepman & Rodway, 2020). By emphasizing the collaborative relationship between AI models and humans, organizations can foster a culture that values the contributions of both and maximize the benefits of AI in the workforce.

7. Conclusions and future research

This action research article explores the process of implementing AI in the manufacturing industry and developing AI capabilities. It offers practical insights into the technical aspects and trade-offs between training time and optimal performance. Given the time-sensitive nature of many projects, efforts to reduce training time of AI models can lead to suboptimal solutions. However, in a business context, such trade-offs are sometimes necessary to meet operational targets. Further, while external acquisition can offer quick access to expertise, in-house development can provide control and potential competitive advantages, Even though in-house development can prove challenging, we argue that such efforts can strengthen organizational capabilities and enable informed decisions about future in-house development or outsourcing.

The study also contributes to theory on AI implementation, confirming the need for a balanced evaluation of in-house versus outsourced solutions, considering costs, expertise, and performance. It emphasizes the importance of collaboration with internal and external stakeholders as well as researchers, agile and experimental methodologies, and the integration of human expertise in AI development. Furthermore, the study highlights the role of AI as a tool to augment human labor, adding to the discourse on human-machine collaboration, organizational strategy, and AI capability development.

Future research considering these theoretical implications in other organizations would allow for generalizations both within and across cases. In particular, we suggest further research investigating the relationships between in-house and outsourced AI development, exploring how different industries, organizational sizes, or technological complexities influence the decision-making process. Additionally, studies examining human-AI collaboration across various sectors could provide insights into optimizing the blend of human expertise and AI, potentially leading to new models for organizational efficiency, innovation, workforce satisfaction, and capability development.

8. References

Ajmi et al., C. (2020). Deep learning technology for weld defects classification based on transfer learning and activation features. *Advances in Materials Science and Engineering* .

Akcay, S., Atapour-Abarghouei, A., & Breckon, T. P. (2019). Ganomaly: Semi-supervised anomaly detection via adversarial training. *Computer Vision–ACCV 2018: 14th Asian Conference on Computer Vision, Australia, December 2–6, 2018, Revised Selected Papers, Part III 14.* Perth: Springer International Publishing.

Amabile, T. M. (1996). *Creativity and innovation in organizations.* Boston: Harvard Business School.

Ancona, D. (2019). Five Rules for Leading in a Digital World. *MIT Sloan Management Review, 61*(1), 1-4.

Arinez, J. F., Chang, Q., Gao, R. X., Xu, C., & Zhang, J. (2020). Artificial Intelligence in Advanced Manufacturing: Current Status and Future Outlook. *Journal of Manufacturing Science and Engineering, 142.*

Babic, B., Chen, D. L., Evgeniou, T., & Fayard, A.-L. (2020). A better way to onboard AI. *Harvard Business Review, 98*(4), 56-65.

Banko et al., L. (2021). Deep learning for visualization and novelty detection in large X-ray diffraction datasets. *Npj Computational Materials, 7*(1), 104.

Bertović, M. (2016). *Human factors in non-destructive testing (NDT): risks and challenges of mechanised NDT.*

Bertovic, M., & Virkkunen, I. (2021). NDE 4.0: New Paradigm for the NDE Inspection Personnel. In *Handbook of Nondestructive Evaluation 4.0* (pp. 1-31). Springer.

Candy, L. (2006). *Practice Based Research: A Guide.* Sydney: Creativity & Cognition Studios, University of Technology.

Curcio et al., K. (2018). Requirements engineering: A systematic mapping study in agile software development. *Journal of Systems and Software, 139*, 32-50.

Deng et al., H. (2021). Industrial laser welding defect detection and image defect recognition based on deep learning model developed. *Symmetry, 13*(9), 1731.

Fountaine, T., McCarthy, B., & Saleh, T. (2019). Building the AI-Powered Organization. *Harvard Business Review, 97*(4), 62-73.

Geertz, C. (1973). *The interpretation of cultures* (Vol. 5019). Basic books.

Govindarajan, V., & Immelt, J. R. (2019). The Only Way Manufacturers Can Survive. *MIT Sloan management review, 60*, 24-33.

Govindarajan, V., & Trimble, C. (2012). *Reverse innovation: Create far from home.* Boston: Win Everywhere.

Guo et al., R. (2021). Weld defect detection from imbalanced radiographic images based on contrast enhancement conditional generative adversarial network and transfer learning. *IEEE Sensors Journal, 21*(9), 10844-10853.

He et al., K. (2017). Mask r-cnn. *Proceedings of the IEEE international conference on computer vision.*

Hou et al., W. (2020). Review on computer aided weld defect detection from radiography images. *Applied Sciences, 10*(5), 1878.

Jiang et al., H. (2021). Convolution neural network model with improved pooling strategy and feature selection for weld defect recognition. *Welding in the World, 65*, 731-744.

Jiang et al., L. (2021). Casting defect detection in X-ray images using convolutional neural networks and attention-guided data augmentation. *Measurement 170*, 108736.

Kingma et al., D. (2014). Semi-supervised learning with deep generative models. *Advances in neural information processing systems, 27.*

Kingma, D., & Welling, M. (2014). Auto-Encoding Variational Bayes. *arXiv preprint arXiv:1312.6114* .

Kong, Y., & Ni, D. (2020). A semi-supervised and incremental modeling framework for wafer map classification. *IEEE Transactions on Semiconductor Manufacturing, 33*(1), 62-71.

Lee, A. S., & Baskerville, R. L. (2003). Generalizing generalizability in information systems research. *Information systems research, 14*(3), 221-243.

Lee, J., Suh, T., Roy, D., & Baucus, M. (2019). Emerging technology and business model innovation: the case of artificial intelligence. *Journal of Open Innovation: Technology, Market, and Complexity, 5*(3), 44.

Li, Y., Swersky, K., & Zemel., R. (2015). Generative moment matching networks. *International conference on machine learning*.

Lindgren, E., & Zach, C. (2021). Autoencoder-based anomaly detection in industrial x-ray images. *Quantitative Nondestructive Evaluation, 85529*.

Maheshwari, A., Mitra, P., & Sharma, B. (2022). Autoencoder: Issues, Challenges and Future Prospect. *Recent Innovations in Mechanical Engineering* (pp. 257-266). Springer.

Mathiassen, L., Chiasson, M., & Germonprez, M. (2012). Style Composition in Action Research Publication. *MIS Quarterly, 36*(2), 347-363.

Mery et al., D. (2015). GDXray: The database of X-ray images for nondestructive testing. *Journal of Nondestructive Evaluation, 34*, 1-12.

Mukherjee, A. S. (2020). 5 Musts for Next-Gen Leaders. *MIT Sloan Management Review, 61*(3), 1-4.

Nacereddine, N., Zelmat, M., Belaïfa, S. S., & Tridi, M. (2005). Weld defect detection in industrial radiography based digital image processing. *3rd International Conference: Sciences Of Electronic, Technologies Of Information And Telecommunications.*

Naddaf-Sh et al., M.-M. (2021). Defect detection and classification in welding using deep learning and digital radiography. *Fault diagnosis and prognosis techniques for complex engineering systems*, 327-352.

O'Brien, R. (1998). An overview of the methodological approach of action research.

OpenCV. (2022). *OpenCV Image Thresholding*. Retrieved Sep 26, 2022, from https://docs.opencv.org/4.x/d7/d4d/tutorial_py_thresholding.html)

Presenti et al., A. (2022). Automatic anomaly detection from X-ray images based on autoencoders. *Nondestructive Testing and Evaluation, 37*(5), 552-565.

Pyro. (2017, November 3). *The Semi-Supervised VAE*. Retrieved May 04, 2023, from https://pyro.ai/examples/ss-vae.html

Pyro. (2022). *Pyro Deep Universal Probabilistic Programming*. Retrieved Sep 16, 2022, from https://pyro.ai/

Ransbotham, S., Gerbert, P., Kiron, D., & Reeves, M. (2017). Reshaping Business With Artificial Intelligence. *MIT Sloan Management Review, 59*(1).

Reason, P., & Bradbury, H. (2001). *The SAGE Handbook of Action Research, Participative Inquiry and Practice.* Sage.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *In International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241). Springer.

Schepman, A., & Rodway, P. (2020). Initial validation of the general attitudes towards Artificial Intelligence Scale. *Computers in Human Behavior Reports*, 100014.

Shorten, C., & Khoshgoftaar., T. M. (2019). A survey on image data augmentation for deep learning. *Journal of big data, 6*(1), 1-48.

Sousa, M. J. (2019). Skills for disruptive digital business. *Journal of Business Research*, 257-263.

Steiber, A., Alänge, S., Ghosh, S., & Goncalves, D. (2020). Digital transformation of industrial firms: an innovation diffusion perspective. *European Journal of Innovation Management, 24*(3), 799-819.

Susman, G. I. (1983). Action research: a sociotechnical systems perspective. *Beyond method: Strategies for social research, 95*(113), 95.

Teesce, D. (2014). The Foundations of Enterprise Performance: Dynamic and Ordinary Capabilities in an (Economic) Theory of Firms. *The Academy of Management Perspectives, 28*(4), 328–352.

Tyystjärvi et al., T. (2022). Automated defect detection in digital radiography of aerospace welds using deep learning. *Welding in the World, 66*(4), 643-671.

Wang, Y. e. (2008). Detection of line weld defects based on multiple thresholds and support vector machine. *Ndt & E International, 41*(7), 517-524.

Wang, Y., Shi, F., & Tong, X. (2019). A welding defect identification approach in X-ray images based on deep convolutional neural networks. *Intelligent Computing Methodologies: 15th International Conference, ICIC 2019, Nanchang, China, August 3–6, 2019, Proceedings, Part III 15.* Springer International Publishing, 2019.

Wen-ming, G. U.-F. (2019). Welding defect detection of x-ray images based on faster r-cnn model. *Journal of Beijing University of Posts and Telecommunications, 42*(6), 20.

Wu et al., X. (2021). A hybrid classification autoencoder for semi-supervised fault diagnosis in rotating machinery. *Mechanical Systems and Signal Processing, 149*, 107327.

Yang et al., D. (2021). Deep learning based steel pipe weld defect detection. *Applied Artificial Intelligence, 35*(15), 1237-1249.

Yang et al., L. (2021). Inspection of welding defect based on multi-feature fusion and a convolutional network. *Journal of Nondestructive Evaluation, 40*, 1-11.

Yaping, L., & Weixin, G. (2019). Research on X-ray welding image defect detection based on convolution neural network. *Journal of Physics: Conference Series. 1237.* IOP Publishing.

Zeba, G., Dabic, M., Čičak, M., & Daim, T. (2021). Technology mining: Artificial intelligence in manufacturing. *Technological Forecasting and Social Change, 171*.