How can functional AI be safely integrated into a machine?

Sarrey M[1] (michael.sarrey@inrs.fr)

1 Institut National de Recherche et de Sécurité (INRS)

KEYWORDS: artificial intelligence, machine, safe design

ABSTRACT

Integrating machine-learning (ML) into automatic machines brings new capabilities and new constraints. This article highlights the specificities of these techniques and aims, thanks to a relevant case, to guide the machines designer, integrator and end-user through a safe practice. The whole approach is oriented towards the prevention of occupational risks related to machines. In other words, what are the safety impact of the learned automation compared to conventional automation?

An online quality control application of organic products (such as fruit or wooden plank type) is used as an example for the design and safety integration process; advice and best practices for maintenance and operation are also provided.

This article is based on (i) studies of safe design, including ML, carried out at the French occupational safety and health national institute (INRS), (ii) GEMMA, French guide of study of the operating and stop modes and (iii) experiments carried out at the machinery and automation safety laboratory of INRS.

These topics are developed as followed: (i) the needs analysis, the proof of concept and the choice of the technical solution, (ii) the machine's life cycle phase related to the creation of a data bank, (iii) the "learning" operating mode, (iv) the performance analysis and validation, (v) the maintenance and (vi) the technical documentation.

This guide emphasizes the study and the realization of the human/machine interface (with learned automation). This interface is a key factor of the success of safety integration. Nowadays, learning techniques do not directly integrate safety functions, but it is well-known that the lack of functional reliability of machines leads to human interventions that can be dangerous.

1 INTRODUCTION

The spread of « machine learning » algorithms has shown new opportunities in machinery automation. Today, automation can find its way in workshops and fields that were inaccessible before. The best example is the analysis of living things such as fruit and vegetables quality assessment on packaging line, and crop and weeds discrimination in the fields.

Integrating this kind of technology into machine leads to deep modifications during design, integration and operation steps. This article aims to provide advice and warnings to integrate successfully AI embedded into machine. On following, to clearly distinguish machine learning based automation from conventional automation, these systems will be referred to as a learned automation (LA) control system. This document focuses on machine learning applied to computer vision onto automatic machinery.

2 NEEDS ANALYSIS AND TECHNICAL SOLUTION.

2.1 Needs analysis

AI into machine is not a commercial argument! Even if it seems to be a token of high technology and high performance, embedding AI system into machine is first of all an answer to a technical need. Like any other functions of the machine, its use must be the result of a needs analysis and this technology must match the required function.

Like any other function, the implementation of this algorithm is the outcome of a design process. This process is standardized in the ISO12100 [1] standard, which defines the general principles of machine design to reduce risks and thus ensures better safety. In this context, in 2021, ISO's work focused on the specific case of machines that embedded machine learning. This work is included in ISO/TR22100-5 [2]. This technical report concludes that the design method described in ISO12100 is suitable for machine learning except that AI does not evolve freely and is not implemented in the safety-related control system.

Machine learning is a technique that allows an algorithm to infer behavior by aggregating a multitude of examples describing that behavior. Collecting these samples can be the first hurdle for this application. Especially in the case of supervised-learning, which involves labeling the training data.
In this case, the obstacle of data collection can be expressed in terms of endeavor (cost). Then, the ratio of costs between the technical solution "Learned Automation" and "Conventional Automation" will be a key factor when choosing the technical solution. Indeed, when it is possible to carry out the program of the machine indifferently either by a program resulting from automatic learning or by a design approach of the automation engineer, the choice of the solution will go towards the easiest.

2.1.1 Comparison between learned-automation and conventional-automation endeavor

Overall, the LA solution requires providing a learning algorithm with input data associated with the expected outcome. At this stage, two sources of costs can be assessed: data production (collection if they already exist) and labeling (reward in the case of reinforcement learning). The comparison with conventional automation requires costs quoted in hours of labor, in any subcontracted resources or any other units.
Then, we need to add the costs associated with automation equipment. This equipment probably comes from the information technology (IT). Accelerators of artificial neural network [3] (ANN) are available on the market. These computers with a massive parallel architecture will make carrying out learned automatisms possible.
Thus, the ANN accelerator will be an integral part of the machine control system, but this device does not belong to operational technology (OT) but to IT.
These two technologies are often difficult to interface, even if they have been getting closer for a few years. The use of middleware, such as Robot Operating System (ROS), or a field network on the Ethernet medium, such as Modbus TCP/IP, allows linking these two technologies. This new cost will be borne by the LA solution.
Figure 1 shows the breakdown and comparison of the costs of each solution. The example presented shows that the distribution of automation design costs is radically different between the two solutions. These costs should be well assessed before choosing to develop a learned automation application. Of course, this comparison is only possible when both automation modes meet the specification requirements.
The large dominance of the data production costs of the learning database shows the importance of automating data collection and labeling when possible. From the point of view of reducing machine risks, the simplest solution is also the safest because it requires less effort and less objects handling.

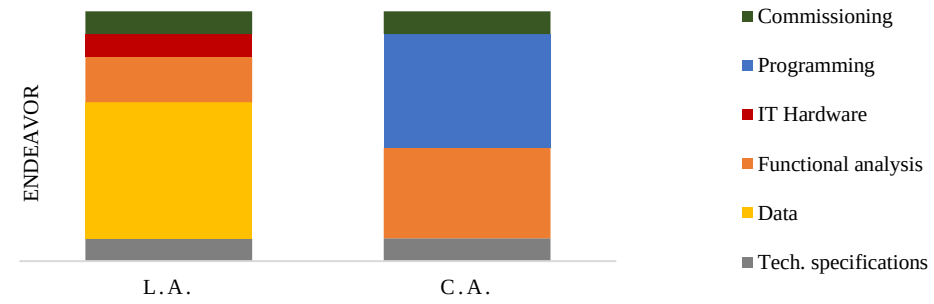

Figure 1. Costs breakdown

2.1.2 Asymmetry of expected outcome

For a classification process, the confusion matrix is the only assessment tool of LA solution performances. In the case of two classes, for example: {"Good part"; "Bad part"}, this matrix is constructed as follows:

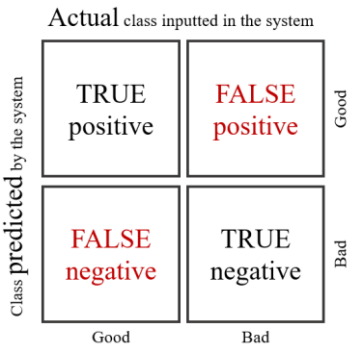

Figure 2. Confusion matrix.

The system can make two different types of mistakes:

- Predicting that a part is good when it is actually bad (false positive). In the case of a quality control application, such as detecting a foreign object in a food product, this error can have significant implications. This type of mistake is evaluated by the precision of the system (see Metrics).
- Predicting that a part is bad when it is actually good (false negative). In the context of quality control, in production line of manufactured parts, this error can lead to the rejection of good pieces, resulting in important economic consequences. This error is rated by the sensitivity of the system (see Metrics).

It is crucial to be aware of the process specificities that the algorithm is meant to handle; mistakes can generate risks. In the above examples, these risks are of a sanitary or economic nature.
Note that precision and sensitivity of the algorithm are two independent indicators; the system can be optimized to improve one or the other.
This issue is purely functional; however, a functional failure of a machine can generate operator's actions to compensate the function. These can lead to risks for people.

2.1.3 Proof of Concept (POC): How to ensure feasibility without a fully trained model?

Demonstrating the feasibility of the learned automation at first is complex. Without providing a concrete proof of concept, series of experiments can be implemented to obtain a first idea of the performance of the learned function.

A. Research by analogy

Numerous applications based on machine learning are available. Their study can provide clues about the LA solution feasibility even if the field of application is different and if the destination of this application is not automation.

B. Training data sample

Evaluating the quantity of images required to achieve the expected performance is complicated.
Of course, this quantity of images depends on the variability of good parts and how big is the fault on the picture to be detected. An experiment can be performed by simplifying the problem: (i) Select a small number of good parts with very low variability, (ii) also select a small number of bad parts, (iii) train the model with this simplified and reduced data set, (iv) create the confusion matrix and evaluate the performance of the system on this sample.
Indeed this experiment does not represent the complexity of the problem as a whole. Nevertheless, the results of this trial will provide a start to feasibility assessment. Interpolating the results of this experiment to the full system is not linear. However, it is a first approximation.

C. The field experience

It would provide relevant lesson to compare the envisaged system with the professional practice carried out by an experienced operator. If the human picture observation cannot detect the non-conformity, the LA system will not be able to do so. This is a key factor for a good performance. Lighting (highlighting the non-conformities by contrast), adapted image resolution, wavelengths of captured light, framework, background plan, etc., are factors that - when optimized - will improve the performance of the system. Thanks to the human sense of visual perception, these topics can be finely adjusted.
However, computer vision systems based on machine learning, unlike humans, have difficulty focusing their processing on the useful part of the image presented to them. The creation of a third class of data can be a solution: the "Background" class. This involves creating a training image database as follows: {"good parts"; "bad parts"; "Background"}. This last class will contain photos of the conveyor belt without any parts. The model thus trained will better distinguish the conformity of parts by suppressing the background variability. These practices can be implemented for the feasibility studies.

2.2 Technical solution

The technical solution of the AI system is evaluated as follows:

- First in terms of relevance, is it appropriate to use this technique for this problem?
- What is the cost endeavor required compared to a conventional solution?
- Is the system's reliability compatible with the expected function?
- Then, it is necessary to study feasibility, especially by evaluating the ability of the future model trained before having all the training data.

Assuming all this is in place, the machine design can begin. An indisputable proof of concept will provide a solid base for the safe design of the machine. It is interesting to note, this technology will modify some phases of machine life.

3 Phases of machine life

The phases of machine life, or more precisely, the phases of the life cycle of a machine as described in ISO12100 [1], will not be fundamentally modified by the implementation of the LA system. However, they may be impacted, like in the "Setting / Learning / Programming" phase which will have to integrate a new task: building the data bank.

Programming Phase - Data Bank Building Task

As mentioned in the section Comparison between learned-automation and conventional-automation endeavor the task of building the data bank represents a significant work in the machines programming phase. This work may require handling many parts, setting up and carry out many hours of shooting. Additionally, labeling the collected images will further increase the workload.

However, the future machine will have the ability to handle and take pictures of parts itself. Using this future ability will automate the data bank building task. Thus, it may be interesting to create a machine designed to train the machine. Of course, these two machines, defined for two different applications, will become one! The life phases will be different. In short, only one machine for two different defined applications.

The example of a vision quality control machine for parts handled on a conveyor belt illustrates this principle as follow: This machine will have at least a feeder, a computerized imaging device, and a sorting device at the conveyor output to separate conform from non-conform parts. All this will ensure quality control and then thanks to the creation of a specific function, building the data bank. For this, the LA system will incorporate a new automatic mode: the learning mode.

4 Operating Modes

The study guide for operating and stop modes (GEMMA) [4] proposes a technique to analyze machine operating on and off modes. These modes are represented in the shape of a block diagram in two main parts: the control part (CP) and the operative part (OP). The control part gathers pre-actuators (such as hydraulic and pneumatic valves and power stages of the frequency inverters). This part controls the energies and the movements of the machine. The OP, on the other hand, manages the logic of the machine, its status, and its sequences.

For the OP, GEMMA groups the status (procedures) into three families: F: operation, A: stops, D: failures.

In this paper, Fig. 3 illustrates the families of procedures for an automated machine. Specifically, the family of operating procedures includes a sub-family of tests and verifications, including the following procedures:

F4: operating verification without order. Commonly referred to as manual mode, this mode is used to check the operating status of one or a few components. It is considered as servo-controlled if the controls (combinatorial only) of the actions are carried out by the OP. This mode cannot be activated during production.

F5: Operating verification with order. This "step-by-step" mode allows the machine's functions to be checked in the order of the production sequence but without automatically chaining them.

F6: Test run. This mode is designed to isolate a component from normal production operations in order to test, adjust, calibrate, etc.

Fig. 3 shows the operating procedures as defined in the GEMMA guide. Procedure F7 has been added to the Tests and Verifications sub-family. Under certain conditions, this procedure (mode) can be activated during production. However, it does not belong to a normal operating mode.

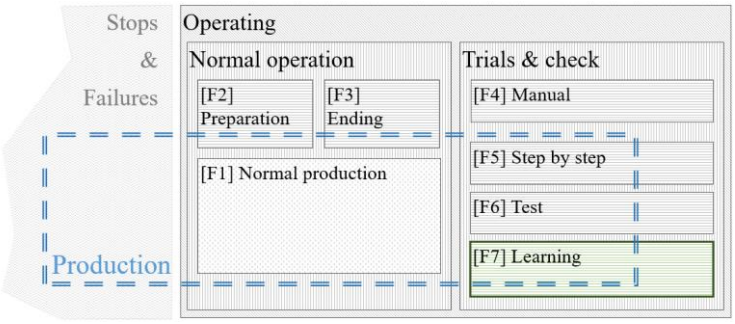

Figure 3. GEMMA - Learning mode in the operating part.

4.1 Learning mode

The Learning mode intends to provide the ML algorithm with the necessary data for training. Its integration into the machine's operating modes enables at least partial automation of the training data collection.

The previously described quality control machine could have this mode. This is an automatic mode where the machine manipulates the parts and takes the pictures. The resulting images are simply recorded to complete the database that will be used during the algorithm's training phase.

This automatic mode is a major contribution to the safety of the machine operators. Indeed, if these tasks were carried out manually, they would generate a lot of part handling and non-productive actions.
Of course, when using the learning mode during a production phase, the images can also be used to search for non-conformities. However, this use does not allow automatic image labeling.

4.2 A priori labeling

When possible, using the learning mode on part batches with previously known conformity will automatically create a labeled database.
During the operating preparation procedure (F2 of the GEMMA), the operator will be prompted to select the class of parts presented to the machine. This is called a "prior labeling". The learning cycle launches that way, will directly record the images in the correct label (class) of the database.
Once all the parts of the class are recorded in the database, the end of the cycle will enable the training procedure button.
Training is not a mode but a computer procedure that will generate the LA system inference engine.

4.3 Training

Training is not really a part of the program ; it's a model compilation to create a classification program: the inference engine. The machine automation will use this engine to perform its classification function. The operator will trigger the operating procedures during production.
It is conceivable to include on the control panel machine interface the inference engine buttons. However, the question of program validation remains open.

5 Performance analysis & validation

5.1 Performance Factors

Many factors can impact performance of the LA system. Among them, we can mention the number of epochs, the size of the data batches, the ratio between training data and test data. However, it is the constitution and size of the database that have a major impact on performance.
A database where each class contains the same number of elements improve performance compared to a database where one class is over or under represented. However, the quantity of data that has the greatest impact on performance. Indeed, the more examples of each class are available in the training, the better performance.
For a given number of parts, there are two ways to increase the training data sets.
- Data augmentation by a numerical treatment. This involves artificially increasing the number of shots by performing image processing operations (for example, cropping, rotation and filtering).
- Data augmentation through a process loop called Marseilles ducks. When possible, in the a priori labeling phase (described above), the parts can be presented in a loop by the machine's conveyor system. This technique makes it possible to multiply the number of part shots of a batch. These data will not represent a large variety of parts, but will increase the robustness of the LA system by showing the variability of the process.

5.2 Metrics

The following metrics can evaluate the LA system using the confusion matrix:

Precision: $Positive\ Predictive\ value = \frac{TruePos}{(TruePos + FalsePos)}$
Sensitivity (or recall): $TruePositive\ rate = \frac{TruePos}{(TruePos + FalseNeg)}$
Specificity: $TrueNegative\ rate = \frac{TrueNeg}{(TrueNeg + FalsePos)}$

As seen in the Asymmetry of expected outcome section, these metrics are intended to evaluate the performance of the training on a test set. Nevertheless, it is only a post-test; it is not a validation
The probability of prediction of the class returned by the inference engine during classification cannot be used as a criterion for validation either. Indeed, this probability simply indicates how close the presented part is to a learned class. For example, if a beech plank is presented to a system trained to distinguish oak from chestnut, the probability that the system identifies oak (or chestnut) may be very high! However, it is neither oak nor chestnut. This metric cannot be used to validate the system.

6 Setting & maintenance

LA system have its own means of programming. Indeed, the Learning mode is a way to program and optimize the performance of the machine. This means could be left under the responsibility of an experienced operator, in charge of showing how to identify the part classes.
However, the validation of the new training is still problematic. As during the initial commissioning of the machine, only trials on a test set can - in absence of validation - provide an indication of the performance of the LA system. A set of standard parts could be kept to compare the results between two new trainings. For perishable items (such as fruits or vegetables), a simple set of photos can be used as a set of test objects.
By design, the LA system does not evolve between two training sessions. However, the process, such as lighting, camera position, or objects in front of it, may change. The items themselves may change: such as the variety of fruits or vegetables whose quality does the machine test.

7 Documentation and Operators training

Like with any other machine, technical documentation and user training materials must be written. However, with the learning evolution the documentation will need regular updates. In addition to the timestamp and algorithm version information, the documentation should also contain the data used in performance testing. This data should be reusable for system verification or analysis.
Based on the type of job, the training for using this type of equipment should focus on the specific properties of machine learning. This training should lead the user to understand how to teach the machine to perform its task. The training manual should highlight the issues of bias in the data and performance evaluation.

8 Conclusion

The integration of a learning algorithm provides new functionalities to the machine. However, these learning functions require taking some precautions in order to make the machine as safe as possible throughout its life cycle.
As from the pre-project stage, a specific need assessment will enable the safest technical solution to be chosen; the one requiring the least effort. Then, the automation of the laborious and potentially dangerous task of building the database will reduce the handling and the labelling tasks. Because the unreliability of a machine leads to risky compensation behaviors, the evaluation of functional performances and the optimization of algorithms need to be well described and documented.
A machine that integrates a learning algorithm will have a more evolutionary control logic than a conventional one. Those in charge of operation production can have the responsibility of this learning. However, the outcome of the learning algorithms remain inexplicable, and therefore the human/machine interface must be designed with great care, which must give the user a good understanding of the procedures and performances of the system. This good understanding is also a guarantee of a better safety for the operators.

9 References

[1] ISO, *ISO12100 Safety of machinery - General principles for design - Risk assessment and risk reduction,* 2010.

[2] ISO, "TR22100-5 Safety of machinery - Relationship with ISO 12100 - Part 5: Implications of artificial intelligence machine learning.," 2021.

[3] J. G. &. Al., "Recent advances in convolutional neural networks," *Pattern Recognition,* no. 77, 2018.

[4] M. S. and P. E., Le GEMMA, PARIS: Educalivre, 1997.