SOGETI

TESTING OF ARTIFICIAL INTELLIGENCE

AI QUALITY ENGINEERING SKILLS – AN INTRODUCTION

Table of contents

1 Executive summary

Artificial Intelligence (AI). It's something that many people only know about from Hollywood films, creating the impression that it will not impact their lives in the near future. In reality, AI is already changing our daily lives in ways that improve human health, safety, and productivity. Unlike in the movies, there is no race of superhuman robots on the horizon. And while the potential to abuse AI technologies must be acknowledged and addressed, their greater potential is, among other things, to make driving safer, help children learn, and extend and enhance people's lives. In fact, beneficial AI applications in schools, homes, and hospitals are already growing at an accelerated pace.

Five of the most valuable companies worldwide are acknowledged leaders in the field of AI. Across the automotive sector, all companies are being forced to adopt AI and must find smart solutions incorporating it. Deep learning, a form of machine-learning based on layered representations of variables referred to as neural networks, has made speech-understanding practical on our phones and in our kitchens. Further, its algorithms can be applied widely to an array of applications that rely on pattern recognition. Natural Language Processing (NLP) and knowledge representation and reasoning have enabled a machine to beat the Jeopardy TV quiz show champion and are bringing new power to Web searches.

Suddenly, a huge number of people jumped from their highly provincial lifestyles straight into a digital world, creating a tremendous demand for more, and increasingly innovative, software. Anyone responsible for producing software knows that the traditional ways of developing, testing and delivering software aren't adequate for meeting this new demand. Not long ago, most companies were releasing software annually, bi-annually, or quarterly. Now, iterations commonly last 2 weeks or less. We adopted Agile and DevOps to move beyond that acceleration plateau. Today, many organizations are talking about Continuous Testing and trying to implement it.

Nevertheless, when we look into the future, it's clear that even Continuous Testing will not be sufficient. That's where AI and Machine Learning enter game. They can, and will, take over the complex aspects of software development and testing. AI is perfectly able to advance software testing by automating tasks that involve self-learning, and which traditionally required human cognition.

This paper considers why we need to test AI and whether we should test it using well-known software testing skills, or with additional skills. What can and should be tested? We present general ideas, definitions and guidelines for the testing both 'of' Artificial Intelligence and 'with' the assistance of AI.

Everyone involved in IT projects will either have come across AI already, or will do so soon. Although a modern technology, we will still see patterns and models to which known testing skills and techniques can be applied. As such, we believe that in this new era the same solid base of testing knowledge still applies. But for testing AI and while using AI for testing, additional skills will be needed. It is an illusion to think that testers alone will be able to perform all testing tasks. Rather, testing must be a team effort and this paper provides an overview of the relevant skills in this new era.
Among the topics covered are:

- Terminology
- The six angles of quality for AI
- Traditional testing skills that remain relevant
- New quality engineering skills that are needed for the testing 'of' AI and/or testing 'with' AI
- Related fields of expertise that become relevant, such as sociology and psychology
- The importance of controlling input of learning machines because the output cannot easily be predicted.

This is the first in a series of papers focused on testing and AI. It draws on our many years of practical experience and theoretical knowledge in this fast-changing area of IT.

2 Setting the scene

2.1 Introducing this whitepaper

Modern information technologies and the advent of machines powered by Artificial Intelligence (AI) have already strongly influenced the world. Computers, algorithms and software simplify everyday tasks, and it is impossible to imagine how, in the near future, most of our life could be managed without them.

Software testing is an investigation process that validates and verifies the alignment of a software system's attributes and functionality with its intended goals. It is a labor intensive and costly process. Thus, unsurprisingly, automated testing approaches are desired to reduce cost and time. And we are convinced software testing can be further optimized with the help of machines powered by AI.

In this paper we present general ideas, definitions and guidelines for the testing of Artificial Intelligence, as well as for testing with the assistance of AI.

Developers are under relentless pressure to deliver innovation and software quickly to the market, whilst maintaining quality. Testing is a critical component of the Software Development Lifecycle and is expanding beyond its traditional definition. DevOps and Agile strategies are increasingly being adopted to deliver with speed and quality.

This need for increased speed and innovation is seeing the relationship between testing and development changing from a service to a partnership. The lines between testing and development have blurred. Developers are doing more testing, while testers are being involved much earlier in the lifecycle then before and participate in development activities. But what if some of the testing activities carried out by humans could be done by machines powered by Artificial Intelligence? Would that reduce the long-term testing costs? Would it increase the speed of testing? Would that be a clever strategy to adopt?
In our opinion, yes! Testing activities can be optimized by using AI for testing. But Artificial Intelligence itself must be tested too, to ensure that users can rely on the decisions taken by AI.

AI will have a fundamental impact on the global labor market in the coming years. A machine powered by Artificial Intelligence can work reliably, 24/7 – and it cannot be distracted by fatigue or other external circumstances. Another positive factor is that the level of accuracy is much higher than that of humans. In the decision-making process such systems can be guided by objective standards, so decisions can be made unemotionally, based on facts rather than feelings and opinions. To rely on the decisions, or to believe that decisions made by machines powered by Artificial Intelligence are correct, we need to test these systems. Such systems are already in use. Google, for example uses them to improve their products. The company is rethinking and has applied AI across all products to solve problems. For example, its Streetview automatically recognizes restaurants with the help of machine learning. Google is continuously testing and improving its machine learning using AI itself.

Everyone involved in IT-projects will either have come across AI already, or will do so soon. Testing AI and while using AI for testing demands additional skills. It is an illusion to think that testers will be able to do all testing tasks. Rather, testing must be a team-effort and this paper provides an overview of the relevant skills. Additional papers soon will elaborate further on the specific skills.
Although testing is a profession, we don't believe there will be many "AI Testers" or "AI Quality engineers" working in projects. Most of the work will be carried out by common team members, such as Business Analysts, Data Scientists, Programmers, Operations and Maintenance people and End users. This paper aims to inform (and inspire) them about the skills they need to keep delivering IT systems that are fit-for-purpose and which deliver business value.

2.2 Terminology

This paper uses terms like "Artificial Intelligence", "Machine Learning", "Machine Intelligence", "Cognitive IT" and "Robotics". These new aspects of information technology are relevant in today's world of digital assurance and testing. The following describes in general terms how we define those terms.

2.2.1 Artificial Intelligence

There are multiple descriptions of AI, for example:

1. Artificial intelligence (AI) is a sub-field of computer science aimed at the development of computers capable of performing tasks that are normally done by people, in particular tasks associated with people acting intelligently.
2. A system, built through coding, business rules, and increasingly self-learning capabilities, that is able to supplement human cognition and activities and interacts with humans naturally, but also understands the environment, solves human problems, and performs human tasks.
3. AI is not required to learn, it could be using pre-programmed rules to handle all possible outcomes. However, for systems with more than basic complexity, this has proved to be a task too large and too complex to handle (it has been tried and failed multiple times since the 1960s).

2.2.2 Machine Learning

Machine Learning is one of the ways to achieve Artificial Intelligence. It contains different algorithms – each with its own strengths and weaknesses. The last major breakthroughs in the field of AI are based on machine learning or more specifically on "deep learning", which uses an artificial neural network. Other popular algorithms are: Bayesian networks, Decision Tree, K-Means Clustering and Support vector machines.
Each has its own strengths and weaknesses. These algorithms are often grouped into three categories:

- Supervised learning
- Unsupervised learning
- Reinforced learning

Although recognizing these differences[1], we have not differentiated the algorithms in this paper.

2.2.3 Machine Intelligence

Machine Intelligence (MI) is a unifying term for what others call Machine Learning (ML) and Artificial Intelligence (AI). We found that when we called it AI, too many people were distracted by whether certain companies were 'true AI,' and when we called it ML, many thought we weren't doing justice to the more 'AI-esque'-like aspects, such as the various flavors of Deep Learning. (Source: Machine Intelligence – Executive introduction, SogetiLabs). So, Machine Intelligence is a term that combines "Artificial Intelligence", "Machine Learning" and other related terms.

2.2.4 Cognitive IT

The word cognitive means "knowing and perceiving". Cognitive information technology is not just rule-based, but is able to react based on perception and knowledge.

Within Sogeti we use the term "Cognitive QA" for the use of cognitive IT to assist quality assurance & testing.

2.2.5 Robotics

What is a robot? It's a machine that gathers information about its environment by input from sensors and, based on this input, changes its behavior. Combined with Machine Learning and Machine Intelligence the robot's reactions over time become more adequate. The use of Internet of Things (IoT), Big Data Analytics and cloud technology make a robot versatile. Robots come in many different shapes and forms. It's not just the metallic man. Robots may equally be a smart algorithm on social media (for example a chatbot or a digital agent), an autonomous vacuum cleaner, or a self-driving car.

[1] http://aiso-lab.com/de/2017/07/28/ai-tutorial-1-artificial-intelligence-machine-learning-neuronale-netze-und-deep-learning-wo-sind-die-unterschiede/

2.3 Distinguishing "testing OF AI" and "testing WITH AI"

Artificial Intelligence can (and should) be tested. In this instance we talk about the testing 'of' AI. But Artificial Intelligence can also be used to make testing more effective and/or efficient. In that instance we talk of testing 'with' AI.

2.3.1 Testing OF Artificial Intelligence

The quality of Cognitive IT systems that use Artificial Intelligence needs to be assessed. The challenge in this case is in the fact that a learning system will change its behavior over time. Predicting the outcome isn't easy because what's correct today may be different from the outcome of tomorrow that is also correct. Skills that a tester will need for this situation are related to interpreting a system's boundaries or tolerances. There are always certain boundaries within which the output must fall.
To make sure the system stays within these boundaries the testers not only look at output but also at the system's input. Because by limiting the input we can influence the output.

2.3.2 Cognitive QA: Testing WITH Artificial Intelligence

As demand for the rapid delivery of software increases, strategies such as Agile and DevOps are already in common use. But how can this speed be boosted still further? The next big thing is testing helped by Machine Learning powered by Artificial Intelligence.

"Classical" testing was designed for software delivery cycles that span months (or sometimes even a year). Agile has made 2-week development iterations the Norm. Today, the vast majority of organizations are talking about Continuous Testing and trying to implement it. Nevertheless, when we look into the future, it's clear that even Continuous Testing will not be sufficient. We need help. We need "digital testing" to achieve further acceleration and meet the quality needs of a future driven by IoT, robotics, and quantum computing.

AI, imitating intelligent human behavior for machine learning and predictive analytics, can help us get there. To meet the challenges presented by accelerating delivery speed with increasing technical complexity, we need to follow a very simple imperative: ***Test smarter, not harder***

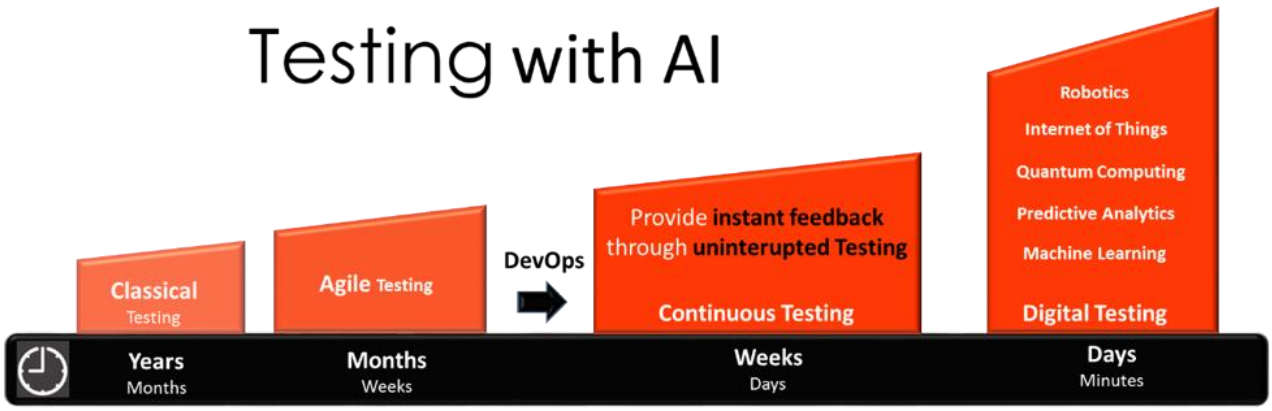


Diagram 1: Beyond Continuous Testing with AI [2]

With Cognitive QA, we enable our clients to achieve accelerated and optimized quality by using an intelligent approach to QA. This leverages self-learning and analytical technologies for Predictive QA, Dashboards, Smart Analytics for QA, Intelligent QA Automation and Cognitive QA Platforms.

This enables smart quality decision making based on factual project data, actual usage patterns and user feedbacks to deliver quality with speed in a complex connected world at optimized cost. For more information please refer to www.cognitive-qa.com.

2 Diagram1 based on "Beyond Continuous Testing with AI" by Tricentis

3 Testing of Artificial Intelligence

3.1 Six angles of quality for Artificial Intelligence

The illustration below depicts the six different angles that are used for digital assurance and testing of modern technology such as Artificial Intelligence, Robotics, Machine Intelligence and Cognitive IT.
The first two angles (Mechanical and Electrical) only apply to physical robots and other smart devices/machines. Methods and techniques for assurance and testing of the mechanical and electrical aspects of machines have existed for a long time and are not particularly different for new technology.
The third angle (Information Processing) relates to traditional IT-functions and systems. For this angle we have methods such as the TMap Suite available. The TMap suite is well-documented in books like "TMap NEXT", "TMap HD" and "IoTMap". And, of course, the website www.TMap.net gives a wealth of knowledge on testing. The new angles are quality assurance for Machine Intelligence and for the Business- and Social impact this new technology can have.

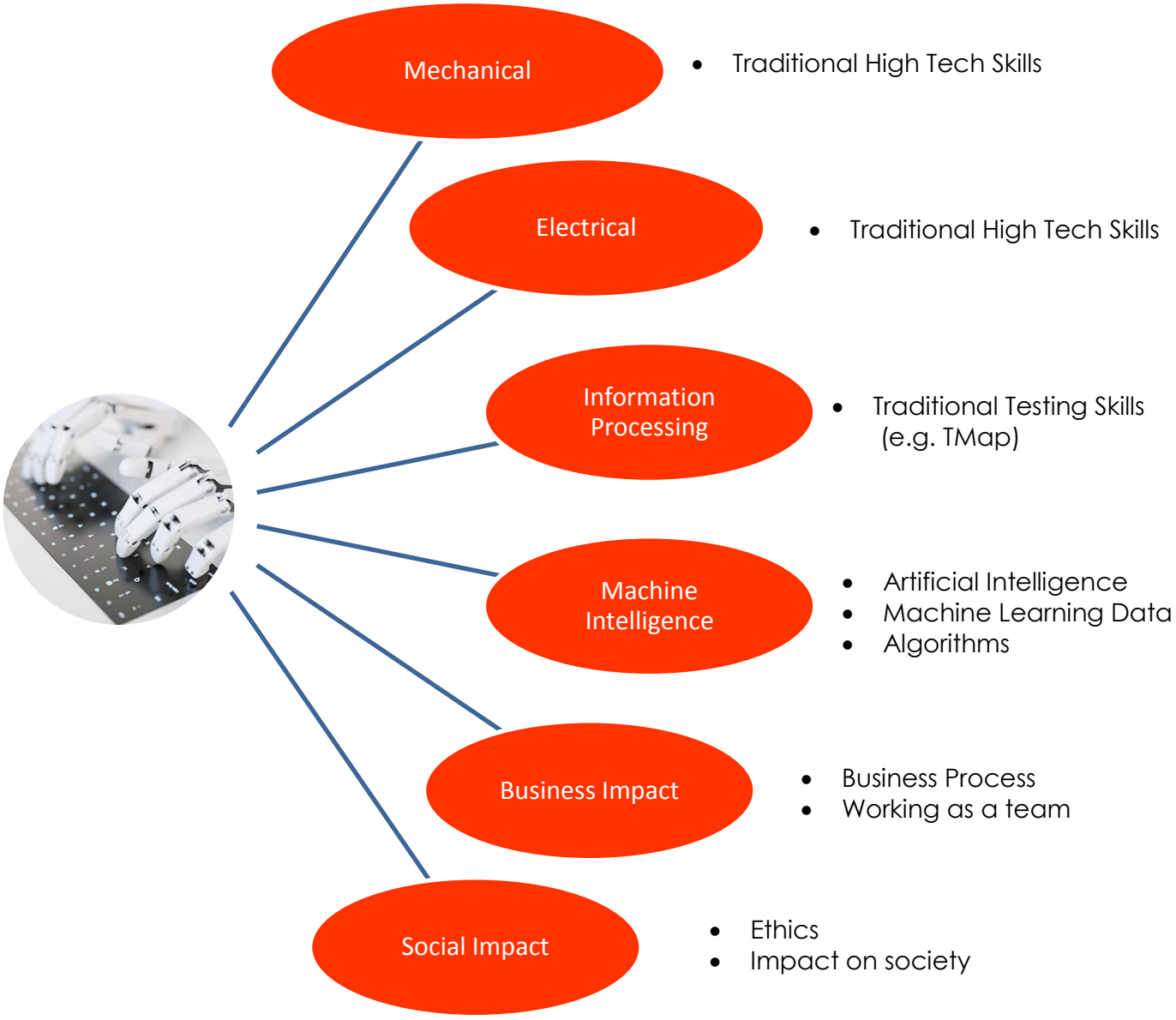


Diagram 2: Six angles of quality for AI

3.1.1 Quality Assurance & Testing for Machine Intelligence

The main difference between intelligent machines and traditional IT-systems is that it is very hard to predict the output generated by AI systems. In traditional IT a tester can use the defined rules to predict the exact outcome and resulting status of the system and can compare this with the actual outcome. Intelligent machines have machine learning capabilities that will result in different outcomes if the same function is called at different moments in time. This demands new testing skills.
One thing a tester can do is to define the tolerances for the outcome that sets boundaries between which an outcome is considered to be correct.

Furthermore, it is very important to control the input. A learning machine uses its input to discover new patterns and/or new options. So, if we can control the input we can also to some extent control the output. More about controlling the input can be found in sections 4.4 and 4.7.

3.1.2 Business Impact

The impact of intelligent technology on business processes and business results (such as profit) may be very different from the business impact of traditional IT-systems. Further, it can be difficult to anticipate. Still this is very important since the business impact or outcome is the reason for having IT systems.

When robotics first came into use the companies applying new technology also created it themselves and thus naturally had an eye for all consequences. Now that new technology has become generally available as off-the-shelf solutions, the challenge will be how to ensure that the resulting new business process will provide the expected benefits to a wider market. Digital assurance must therefore not only be active during the project phase, where the new solution is built and implemented, but also during the first use. This type of monitoring of the effect of new technology will ensure that both desired and unwanted effects are identified as soon as possible so that necessary corrective actions are quickly taken.

3.1.3 Social impact

Intelligent machines can have huge effect on the social environment of its users. For example, will taxi drivers become unemployed because of the introduction of self-driving cars? But also consider the positive social effects, such as nurses having time to pay more personal attention to elderly patients because robots are there to do the basic care-taking activities, such as distributing the right pills or food and to clean the home.

It is hard to define tests for this type of effect. But digital assurance is far more than just testing. With digital assurance we can also observe effects in another way, such as the difference between old and new situations. This will require the ability to compare these situations and to pick up small signs of changes in the effects new technology has on individuals, organizations and on society as a whole.
For this, the testers will need skills in areas such as sociology and psychology.

3.2 What needs to be tested?

When developing a machine-learning solution the testing task is a vital activity. Whoever performs this task must ensure that the functional and non-functional requirements are fulfilled. What´s changed is, that the behavior is much harder to predict. And because of the complex functioning of an AI system many more input values have to be tested to verify a robust solution.

Some examples of the questions we could ask when testing AI, are:

- What are the acceptance criteria?
- How can we design test cases that test those criteria with minimal effort?
- Are there different datasets for training-data and test-data?
- Does the test-data adequately represent the expected data well enough?
- Is the test- and training-data compliant with the legal framework?
- Is the training-data biased (see 3.3)?
- What is the expected outcome of the model?
- Is the model under- or overfitted (see 4.6)?
- Does the solution behave unethically (see 4.6)?
- Is the performance and robustness of the solution good enough?

3.3 Test objects in Machine Learning Solutions

Machine learning solutions contain many components, just like classical software solutions. But the components and especially their role and properties, are different. Here is an overview and a brief description of the test objects in a machine learning solution:

Dataset
The dataset contains all data that is available for the machine learning solution. Often it is a company´s historical data and needs to be processed to be viable.

Training data
The training data is a subset of the dataset. This data is used to train the model in the development process.

Test data
The test data are another subset of the dataset.
It is used to verify, that the model works as intended. It is essential not to use training-data as test data because verifying whether the AI has learned what it needs to perform its decision-making requires different data to the training data.

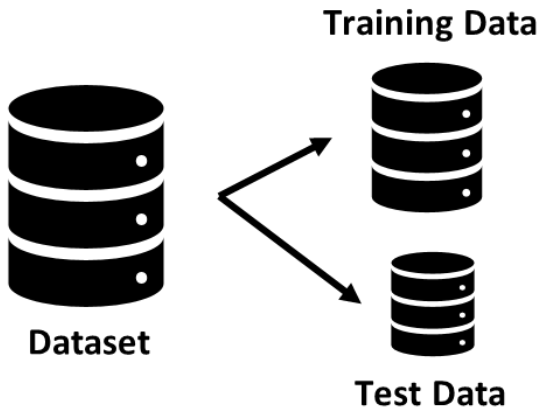


Diagram 3: Machine Learning Datasets

Model
The model consists of the algorithms being used, from which the AI learns from the given data.

Training (phase)
The training is the process in which the algorithm learns from data and makes predictions.

Inference (phase)
After the model is trained, it can make inferences based on the input data.

Source code
A machine learning solution has much fewer lines of code than the classical solutions. Nevertheless, there is source code which can have errors and therefore must have unit tests and other relevant tests.

Infrastructure
The infrastructure is subject to non-functional requirements, which must be checked and tested.

Requirements
The requirements in this field should be inspected carefully. The technology is new, so it´s possible that the expectations are unrealistic or outside of legal or ethical boundaries.

Input/Output Values
The most basic test-objects are the input- and output values. This is where the acceptance criteria are verified. The input values in machine learning solutions are crucial because it is unknown how the data is processed.

3.4 Fitting Data

Since machine learning is still new, the starting point for new projects is often the data. How can value be generated with the data? How can it be used in a machine learning solution?

Expectations and the technical capabilities are not always the same. Other questions regarding the data are of legal nature. After the data is collected and stored in a central database there must be assurances that the collection and processing of this data are within the legal constraints. As an example, are they compliant with the requirements of the EU's General Data Protection Regulation. In the development phase, statistical quality aspects of the data become relevant, especially in combination with the chosen model. How good is the data quality? Are the predictions accurate and precise enough? These quality aspects are incorporated in the development phase.

4 AI Quality Engineering Skills

4.1 Why do we need new skills?

A software Tester has experience in how to test software using different guidelines and test approaches. For professional and structured testing, there are standard certifications that are around for a long time, such as TMap and ISTQB. And other certifications in the fields of Agile, Requirements Engineering and Mobile testing are also valuable to a tester. But now the skill set will need to further grow as AI plays an increasingly major role.

Testing of AI embraces machine learning, mathematics & statistics, Big Data analysis and much more. The team needs to have many of these skills (discussed further below in detail) to successfully carry out their AI-related testing tasks.

4.2 How to use the existing skill sets to attack the new challenges

There are a number of well-known and established test design techniques and practices that can still be applied in this new era of testing. Patterns and models to which known skills and techniques can be applied are evident in this modern AI-led landscape. So, we believe that the same solid base of testing knowledge already residing in the test operation still applies.

Our vision on this is supported by the fact that in other recent developments in IT, such as the rise of DevOps approaches, also many people are educated in the basic testing skills. For example the activities, as defined by TMap, for Planning, Control,

Preparation, Specification, Execution, Completion and Test infrastructure. When testing new technology, including Artificial Intelligence, we still need to organize these basic testing activities. Naturally, there will be some different approaches for several activities, but the profession of testing does not change dramatically just because there is a new kind of system under test.

Of course, building on top of the well-known skills mentioned above, new skills will have to be acquired to be able to effectively test systems that include machine intelligence.

4.3 New Skills needed for AI Quality Engineering

To ensure quality in a machine learning project, the team's AI Quality Engineering needs an extended set of skills. On Top of the above-mentioned skills from TMap and ISTQB, the team should have expertise in A/B testing and metamorphic testing, amongst other techniques that have gained new importance.

Strong programming skills in the most prominent machine learning languages, such as Python, Scala, R, Spark, are required, as well as in languages such as Go and C++, and with open-source software libraries like Tensorflow. This isn't just about understanding the developed software, but is also about creating a custom toolset for specific tests. These skills need to be extended by a strong understanding of the new technologies: machine learning, Big-Data and cloud computing.

Strong mathematical skills, especially in statistics, calculus, linear algebra and probability are the core to understanding machine learning. Knowledge about computer-hardware-architectures is crucial to determine the performance of a chosen model.

Disciplines that used to be irrelevant for IT projects now gain a strong foothold in the world of AI. Biology, economics, social science and psychology have many use cases for the data scientist. Knowledge in these fields will be helpful. Philosophy and ethics also increase in importance when we create a new world with intelligent software.

And as soon as physical robots become involved too, team members need additional skills in the field of mechanics and electronics.

AI Quality Engineering is active, flexible and broad. AI Quality Engineering makes a difference, by contributing strong technical and functional skills to the software development project. A team member who is testing often will be the first one to see problems with performance, legal constraints, or problematic requirements.

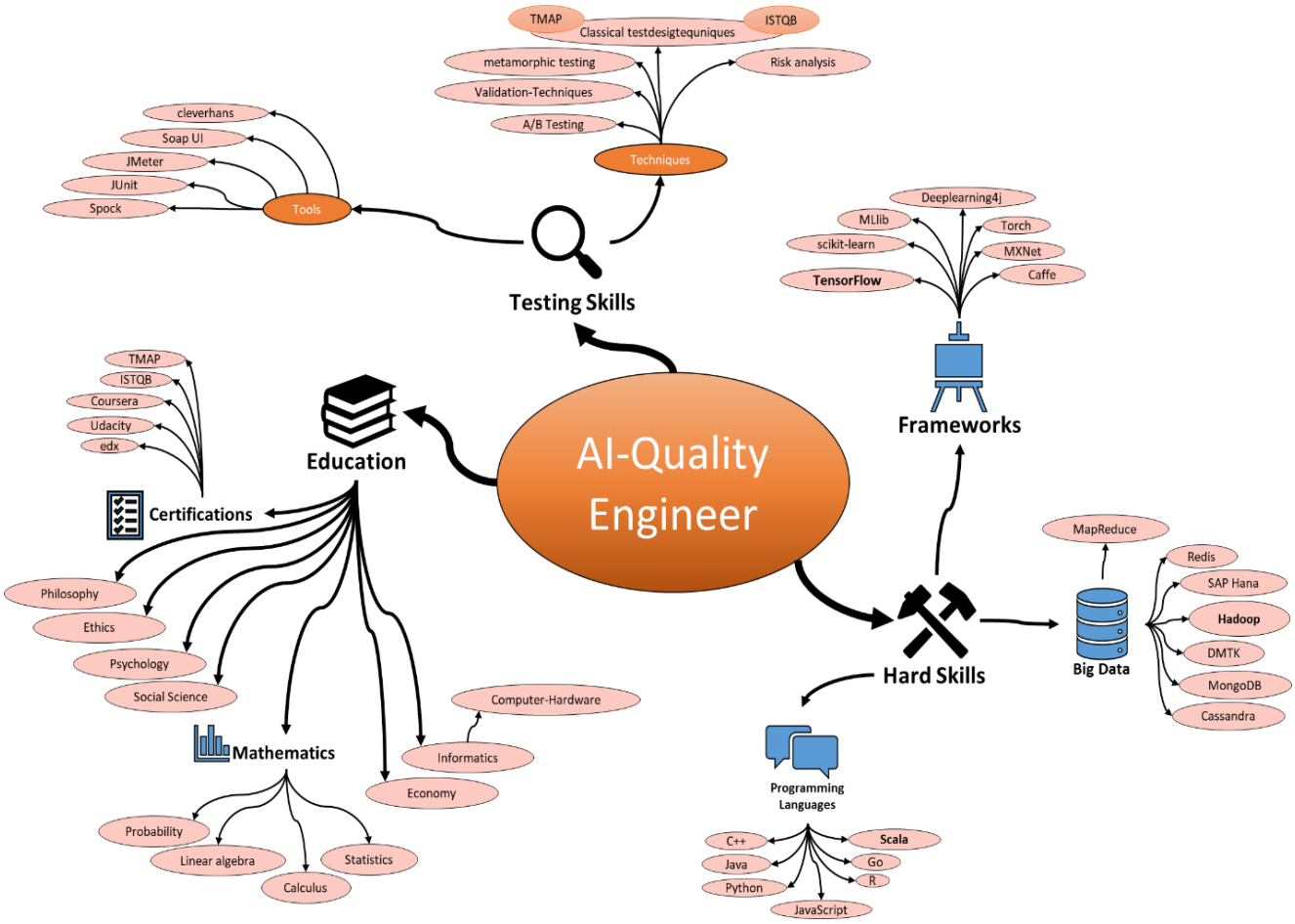


Diagram 3: AI-Quality Engineering skills

4.4 How to test Machine Intelligence?

Since machine learning solutions are quite new, experience in this field is scarce. Nevertheless, AI Quality Engineering has to formulate or evaluate complete and strong acceptance criteria that verify the outcome. The outcome is determined by the input data and the trained model. Testing those is the core activity.

A/B testing

To test the end-user experience, many big software companies use A/B testing. They deliver two different versions of their software and test the user's reaction. This approach is based on the scientific method and should be utilized when developing AI capabilities.

Testing the input values
This is where AI Test Engineering has to be creative and diligent. Different kinds of input values can lead to expected and unexpected behavior. The input values are relevant for the functional integrity, security, robustness and performance, as well as of how the data is processed – in the present and in the future.

Feeding specific input data to see how the AI learns
Different kinds of input values can lead to expected and unexpected behavior. An AI-system keeps on improving the more data it collects. An AI tester could try to spoon-feed the AI with specific data to change its behavior, for example by making a "ticket selling AI" lower the prices as much as possible. Depending on the project's goal, this can have serious consequences. AI test engineering could even apply another AI to interact with the "ticket selling AI", with the goal of finding weaknesses.

Metamorphic testing
Metamorphic testing is a software testing technique that attempts to alleviate the test oracle problem. A test oracle is the mechanism by which a tester can determine whether a system reacts correctly. A test oracle problem occurs when it is difficult to determine the expected outcomes of selected test cases or to determine whether the actual outputs accord with the expected outcomes.

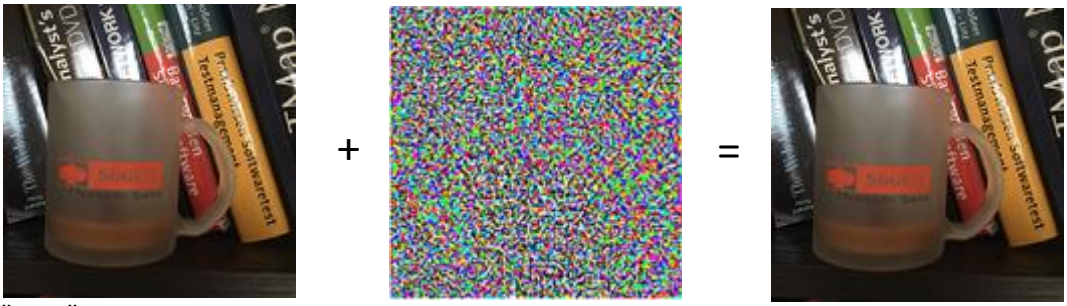
Testing the nonfunctional requirements
Non-functional requirements like performance, security and privacy gain significance under the veil of the new technologies. They must be addressed and tested in different ways. The AI quality engineer has to be able to address and test them in proper manner. This includes a strong understanding of the underlying hardware.

4.5 Security issues regarding AI

Since machine learning is part of the information technology, it is exposed to hackers. It is a new field, but there are already some security issues which should be considered and old ones which shouldn´t be ignored.

Data poisoning
It is possible to manipulate the training data to teach a machine learning model something that the attacker wants to. When that succeeds, the model will make predictions that the attacker intends. This can have serious consequences.

Adversarial examples
The input values for a model can be manipulated in a way which leads to wrong predictions. For example[3], you see on the picture a mug but the machine learning model classifies it as a skyscaper. The reason is, that the noise in the middle was added.



"Mug"
63,4% confidence

"nematode"
9,3% confidence

"skyscraper"
99,2% confidence

Diagram 5: Adversarial pictures

A human eye does not see any differences, but the machine does. Self-driving cars that rely on the right classification of their environment could be negatively influenced by this type of attacks[3].

These are just two examples of new attacks against systems that use machine learning. It is expected that there will be more.

4.6 Privacy, Big Data and AI

Data is the fuel of the new AI engines. Today there is more data collected than ever before. This data is not only collected from computers, but also comes from mobile devices and the Internet of Things (IoT). This includes financial, positioning, health and behavioral-data. The new flood of data, combined with national laws, customer preferences, hackers, state intelligence, industry standards and competitors all over the world make questions about privacy and ethics very urgent. The consequences of the answers can be far reaching. The next few paragraphs provide some examples of these challenges.

Findfaces and VKontake
Findface is a new facial recognition service that uses a photograph of a person to find information about them. This service searches the internet, but gets most of its data from VKontakte – a competitor of Facebook.

Live face recognition
Pictures taken by a video-camera can be used in real time to find other information about that person. This camera was in a train-station in berlin. There are several startups which use machine learning to find out more about a person's mood.

Filter-bubble
By delivering customized news or content, internet companies created what's known as the filter bubble. One person's views are enhanced, and others just ignored. This environment contains danger, such as fake news and supports radical views.

Sexism and Racism
There are several computer programs, based on machine learning, which reproduce or even enhance unwanted behavior, including racism and sexism. One reason for this is societal prejudices that can be found in the existing data. AI-solutions take that data and learn from it.

Data can be breached
Data breaches occur. That's a fact. This gives hackers access to sensitive security information, such as that contained in email attachments, which should not be there in the first place.
Privacy considerations now have a bigger scale and impact. They should be handled carefully, not only because of the social responsibility, but because legislation, like GDPR, must be complied with.

4.7 Monitoring the input of the AI

In 2016 a new chatbot started to express fascist ideas, causing a lot of concerned comment. The owner of the chatbot was quick to take the chatbot offline. Investigation showed that people had deliberately supplied ultra-right texts to the chatbot and the machine learning algorithm simply did as it should: it learned based on its input. So, the chatbot itself wasn't wrong; the trouble came through the input.
Thus, we can clearly see this as part of digital assurance.

The first step of any AI-project is to see how the machine learns and how it reacts to certain types of input. Based on this, the learning algorithm's behavior can be influenced so that the learning is improved.

3: https://en.wikipedia.org/wiki/FindFace

The second step, during live operation, is to monitor the actual behavior of the learning machine and to see whether its output stays within boundaries that were set up-front.

As well as monitoring the results, it is equally important to monitor the input and to see whether this stays within tolerances for normal input. Since the input for chatbots and similar systems can be massive, machine intelligence could be applied to support the monitoring activity. Of course, in this situation we must all be aware that we use one system with an unpredictable outcome to monitor another system with an unpredictable outcome.

When working on controlling the input, the tester has to observe the effects of both overfitting and underfitting in machine learning – see below.

Underfitting data
Underfitting refers to a model that can neither model the training data nor generalize to new data. An underfit machine learning model is not suitable and this becomes obvious in the poor performance of the training data. Underfitting is easy to detect given a good performance metric. The remedy is to move on and try alternative machine learning algorithms. Nevertheless, it does provide a good contrast to the problem of overfitting.

Overfitting data
Overfitting refers to a model that models the training data too well. This happens when a model learns the detail and noise in the training data to the extent that it negatively impacts the performance of the model on new data. This means that the noise or random fluctuations in the training data is picked up and learned as concepts by the model. The problem with this is that these concepts do not apply to new data and negatively impact the model's ability to generalize.

Overfitting is more likely with non-parametric and non-linear models that have more flexibility when learning a target function. As such, many non-parametric machine learning algorithms also include parameters or techniques to limit and constrain how much detail the model learns.

5 Conclusion

We're fast approaching a time when even "Continuous Testing" will be unable to keep pace with shrinking delivery cycle times, increasing technical complexity, and accelerating rates of change. We're already starting to use basic forms of AI, but we need to continue the testing evolution to achieve the efficiency needed for testing of robotics, IoT, and so forth.

We need to learn how to work smarter, not harder. Ensuring quality in an era where software will be processing an unimaginable number of data points in real time, for example on the Internet of Things and while literally driving "self-driving" cars, has become non-negotiable.

As more and more Artificial Intelligence comes into our lives, the need for testing both OF and WITH AI is increasing. Take the self-driving cars as an example: if the car's intelligence doesn't work properly and it makes a wrong decision, or the response time is slow, it could easily result in a car crash and put human life in danger.

Companies are clamoring for employees who can take huge amounts of information and analyze it for insights. The demand for people skilled in Artificial Intelligence, data analysis and machine learning in 2017 has significantly increased. Programming languages such as R, Python, SAS, Scala, Go and C++ have gained in importance faster in 2017 than in the last few years. And new machine intelligence libraries like Tensorflow have been introduced.

In this paper we have discussed the skills needed to cope with new AI and machine learning tasks in the context of quality assurance. It is clear that it's not just about testing: it is far more than that. It might not be easy finding employees to create the cross-functional teams with all the required skills. So, it's time to rethink and start investing in employees to learn and polish their existing skills and develop the new skills needed.

As a conclusion, the following summarizes the so called "must have" skills and knowledge-sets for testing in this world of AI, machine learning and robotics:

- Educational background in software engineering, informatics, applied statistics or comparable field
- Experience in implementing analytical solutions using programming languages, such as R, Python, C++, Java and more, for solving analytics problems within engineering
- A deep understanding of statistical and predictive modeling concepts, machine-learning approaches, clustering and classification techniques, and recommendation and optimization algorithms
- Ability to define key business problems to be solved, formulate mathematical approaches and gather data to solve those problems, develop, analyze/draw conclusions and present
- A keen desire to solve business problems, and to find patterns and insights within structured and unstructured data
- Insight into the six angles of quality for AI and how these impact the testing activities
- Able to propose analytics strategies and solutions that challenge and expand the thinking of everyone around them
- Data analytics capabilities
- Complete understanding of quality assurance throughout the software development lifecycle
- Strong knowledge of diverse test varieties like unit testing, system testing, system-of-systems testing, regression testing, performance testing, security testing, etc.
- Good knowledge of testing methods, strategies, business processes, testing tools and test automation
- Good understanding of computer chip architecture and its impact on the performance on different machine learning approaches.

We will continue working on this subject, elaborating on topics like "Quality Attributes for Testing of AI & Cognitive IT".

6 Acknowledgments

The authors Humayun Shaukat, Toni Gansel and Rik Marselis would like to thank everybody who has read this paper. We welcome your feedback and would be delighted to discuss how we approached the topics covered – and how we might take the subject forward.

We would also like to thank our reviewers, who invested their valuable time to read, review and provide valuable feedback. In no particular order, we thank Andrew Fullen, Mark Oost, Carlos Ribeiro Simoes, Jeroen Franse, Bahadir Kasap, Robiel Nazirugli, Rakesh Partapsing and Bartek Warszawski.

Further insight, useful tips, content and ideas were gratefully received from Mark Buenen, Stefan Gerstner, Rob Crutzen and Tom van der Ven.

Finally, our special thanks to Gregory Biernat (Head of Quality Assurance & Testing at Sogeti Germany) for giving us the inspiration, support, time and resources we needed to write this paper.

For More Details, Contact:

Humayun Shaukat
Senior Consultant Quality Assurance Digital and Artificial Intelligence
humayun.shaukat@sogeti.de

Rik Marselis
Management Consultant Quality & Testing at Sogeti Nederland B.V.
rik.marselis@sogeti.nl

About Sogeti

Sogeti is a leading provider of technology and engineering services. Sogeti delivers solutions that enable digital transformation and offers cutting-edge expertise in Cloud, Cybersecurity, Digital Manufacturing, Quality Assurance & Testing, and emerging technologies. Sogeti combines agility and speed of implementation with strong technology supplier partnerships, world class methodologies and its global delivery model, Rightshore®. Sogeti is a wholly-owned subsidiary of Capgemini SE, listed on the Paris Stock Exchange.

Learn more about us at
www.sogeti.com

FOLOW US @SOGETI