A grayscale image of a robotic hand holding a chess piece on a chessboard. The hand is positioned in the center, holding a white chess piece. The chessboard is in the foreground, and other chess pieces are visible in the background. The overall scene is dimly lit, with a soft glow around the hand and the piece it holds.

Productionizing AI

Up to 80% of Artificial Intelligence (AI) projects never go into full production, and thus do not generate the envisioned value for the company that undertook them. This is why this paper summarizes, for the benefit of business stakeholders and project managers, our key learnings from successfully pushing AI projects past their go-live, both in large and smaller enterprises.

Realizing the Business Value of AI is Absolutely Feasible

Beyond the current hype, Artificial Intelligence (AI) holds the very real promise to businesses to improve decision making and to reduce cost – for applications as diverse as customer support, predictive maintenance, and general back-office automation, to name but a few. However, getting to these gains may prove challenging to companies, especially if they are just starting their AI journey. More often than not, AI projects get stuck at the very beginning, at the Proof of Concept (PoC) phase, and – even if a PoC was a success – some companies may never realize the full business value of AI and eventually fall behind in their industry.

For the benefit of business stakeholders and AI adopters, this paper summarizes the best practices that we found crucial to take an AI system from PoC to full production. They span aspects relating to the design and implementation of an AI system, the legalities of contracting with third-parties, training and testing, and then of course adoption. These points are complemented by a diverse set of voices from the AI community who share their own learnings and insights in taking AI from the drawing board to the core of business operations. The overall message is clear: AI projects may have their own peculiarities and quirks – and we list the most prominent ones in the following five sections – but these risks can be successfully navigated with awareness, clear communication, and a bit of foresight.

Prof. Dr. August-Wilhelm Scheer
Founder and CEO of the Scheer Group Innovation Network
Management of Expectations Page 2

Andreas Klug, CMO ITyX AG and AI Evangelist
President of Working Group ‘Artificial Intelligence’ at Bitkom
The PoC Barrier Page 4

Dr. Philipp Hartmann, Director of AI Strategy, appliedAI
It’s a Marathon Not a Sprint! Page 6

Britta Hilt, Managing Director and Co-founder at IS Predict GmbH
Unreliable Data in a Changing World Page 7

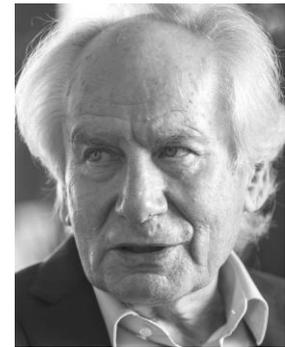
Peter Fach, Partner at Deloitte
Head of Deloitte Center for Process Robotics & AI as a Service
Size Matters — and Scale Will Get You There Page 9



Management of Expectations

Since the beginning of the 21st century, AI has been experiencing a new hype. The starting point was spectacular successes of AI victories in playing chess, Jeopardy and Go against human luminaries. At the same time, major advances had been made in speech recognition and translation, facial recognition, and autonomous systems such as driverless cars and robotics. However, with the quite justified hopes of being able to achieve significant progress through Artificial Intelligence, there is a danger of overestimating its performance. The management of expectations is important here.

Against this background, research into applications suitable for AI is of particular importance. The algorithms themselves are increasingly being made available almost free of charge by the major hyperscalers, so the focus is all the more on worthwhile applications. Here, we need to work closely with practical applications. This is also the goal of the non-profit August-Wilhelm Scheer Institute, AWSi, which I founded. For example, AWSi is conducting several AI projects in the field of predictive quality assurance. AI is an indispensable part of tomorrow's production, nonetheless, the secret of success lies in truly goal-oriented applications.



— Prof. Dr. August-Wilhelm Scheer
Founder and CEO of the Scheer Group Innovation Network

I. Design

AI may solve some business problems, but should not be applied to all of them

At times, some businesses are very keen on gaining their first experience in applying AI to their business problems. However, not all problems are nails in need of an AI-like hammer. Even though AI and specifically Machine Learning (ML) are very powerful technologies, more often than not other approaches may yield comparable or even superior results at less cost and definitely at less risk.

Companies should clearly differentiate between two very valid objectives: 1) Gaining experience with AI technology, and 2) solving a specific business problem. In the first case, go ahead and just make sure to

document the learnings. But if the objective is item #2, then be open-minded about not applying AI, but still solve the problem.

Don't try to solve the biggest problem on your first attempt

When a company makes the decision to implement its first AI solution, there is always a steep learning curve at the outset because people involved may not yet fully understand AI or ML. But not knowing the capability of AI technologies makes it difficult for companies to specify the details of the project and plan accordingly.

Companies should thus avoid trying to solve their biggest problem on their first attempt, but rather try to gain experience through proof of concepts to better under-

stand the capabilities and limitations of different AI technologies.

Less training data is OK, if you ask the right questions

While companies generally have sufficient data to train an ML system, certain less frequently taken decisions will almost certainly be found lacking. ML-based solutions are generally assumed to require large amounts of training data for all relevant use cases.

In practice, it turns out that classifiers for smaller, well-defined problems can be trained with as few as one hundred samples and still reach acceptable quality and accuracy levels for all practical purposes. Hence, the problem of insufficient training data can in part be mitigated by good problem engineering.

Don't believe the hype — AI is no silver bullet

Given the large amount of hype, it often occurs that businesses overestimate the capability of AI – they think it can solve any and every problem. Companies that overestimate the capability of AI will have expectations that likely can't be met, which sets any project up to fail from the get-go. To overcome unattainable expectations of AI, demonstrate where the technology fails, particularly on those tasks that seem to be so easy for humans.

Without reliable effort estimation, you'll miss value-adding quick-wins

The opposite of overestimating the capability of AI is also true: When starting to implement AI, in some cases, particularly if a task is difficult for a human, companies often fall into the trap of underestimating the capability of ML solutions. Not knowing where AI technologies shine can lead to missing low-risk, high-value opportunities. Hence, AI use cases should be prioritized not only by business impact, but, also by

implementation effort to catch value adding quick-wins. This may seem obvious, however, many companies carry out the selection process without an ML expert in the room and wildly misjudge effort.

Get audit-proof results with more granular classifiers

Businesses require the output of any AI solution to be understandable, up to having an auditable paper trail for compliance purposes. But some AI technologies and approaches, e.g., neural networks and deep learning, do not allow easily inspecting, comprehending, and fine-tuning the reasoning process.

Rather than tackling the client's use case with one big 'black box' AI component, e.g., a big neural network classifier, split the classification into multiple smaller sub-problems, solve each of them using the most appropriate AI technology, and then document how all sub-decisions led to the global outcome.

Problem owners and AI implementers are from different worlds — expect misunderstandings!

Most of the time, the problem owners of an AI use case are from the business side of things and, therefore, have trouble understanding, and agreeing to, the specifications of a proposed solution. If the problem owner has trouble understanding the specifications of a proposed AI solution, then this will not only cause delays in decision making but even worse, it could result in an implemented solution that fails to meet expectations.

It is imperative that key decisions about an AI solution can be made with confidence to avoid delays and, most importantly, ensure expectations are met. This can be achieved through an 'AI Translator' that understands both the business needs and how AI works, in order to bridge the gap between the business and the scientists.

The PoC Barrier

Artificial Intelligence and Intelligent Automation are at the core of our digital economy, heralding a new age of industrialisation. I am convinced that by the end of our decade, all relevant banks, insurance companies, energy suppliers and mobility service providers, among others, will be AI companies. AI has proven that it works and that it enables new business models and value chains. But despite this, discussions about regulation and data sovereignty hinder agile further development in the economy.

The worst obstacle, however, is in our heads: We lack the right attitude towards digitalisation in many places. This is the reason why we conduct Proof-of-Concept (PoC) projects. For many organisations, the first step into AI is a PoC. That may make sense if there is a lack of definition of an actual deployment scenario for Intelligent Automation – but certainly not to determine whether algorithms work! After many years of working with AI, I have experienced that it is not a successful or unsuccessful PoC that determines the successful use of AI in a company, but the actual use of AI. At the same time, and especially now, it should be part of our instincts to renew and adjust; to throw old thought patterns overboard.



— Andreas Klug, CMO ITyX AG and AI Evangelist
President of Working Group ‘Artificial Intelligence’ at Bitkom

II. Implementation

Don’t roll out without a pragmatic subject matter expert

When it comes to rolling out AI in a business, companies may at times be tempted to follow a preconceived waterfall-like plan of how AI should be used. In practice, AI roll-outs, like most other projects, will run into many trade-offs, big and small, as the roll-out progresses. Maybe one delivery of training data is not available in time, maybe one particular sub-decision turns out to be particularly difficult for a neural network or another ML component.

Roll-out projects are served best if led by a pragmatic subject matter expert, who can make informed trade-offs on how changes to the AI setup may affect the target use

case. This way, little adjustments can be decided upon as the roll-out progresses without becoming roadblocks.

Boring but true — an AI roll-out is a project that needs to be managed

AI roll-outs are – boring but true – just projects that need to be managed between organisations and vendors. The results of AI projects, due to the nature of ML technology, can be vastly improved through iteration. This naturally fits an agile project setup, that some companies may still be learning how to get right.

There is no need to re-invent the wheel when it comes to good project management and governance. But it must be in

place, roles must be clearly assigned, and staffed with people that can comfortably fill them.

Success breeds success

AI projects can be complex, involving many moving parts, all of which have a chance of not working as expected. Hitting a roadblock on any project is never good as it slows down the project and makes stakeholders nervous, and since AI projects tend to be complex, one will likely hit more of them.

When implementing an AI solution, try to break the problem down into smaller and simpler components that can be tested independently. Not only does this make the project as a whole more manageable, but it provides more opportunities to share good news with stakeholders which will keep up project momentum, create implicit feedback opportunities and maintain trust.

If the data won't come to the AI, the AI needs to go to the data

Data security is important, not only because of legal regulations but also because a company's data is a key asset that needs to be protected and handled with care. AI cannot operate without data, and thus data needs to be made available to the AI system. More often than not, this gives rise to data security concerns that need to be addressed less they may derail an AI project – especially when considering Cloud-based AI solutions that may operate under a different legal regime.

The key observation here is that while AI and data need to be in the same place for any meaningful value creation to happen, there are many degrees of freedom as to where that could be. AI does not imply Cloud Computing, and it certainly does not require companies to ship their data overseas. Instead, the AI can come and meet the data at an established company location: its own on-premise data centre, on the servers of an established outsourcing provider, or even on the laptops of key employees.

III. Contracting

Guarantee accuracy, but share the cost

When setting up the legal framework for an AI PoC or roll-out, companies tend to take inspiration from IT contracting. Some key metrics used in IT contracting, notably guaranteed reliability and accuracy, are not applicable to ML technology, and either cannot formally be guaranteed (e.g., in the case of neural networks) or depend heavily on client-provided input (e.g., training data).

Rather than stipulate contractually agreed-upon reliability, specify reliability levels

and which contractual party needs to take which action to reach which level. For example, lower levels (<80%) are to be reached by the AI provider alone at its own cost, more advanced levels are to be reached at the joint expense, refinements towards the top-levels (e.g., +95%) are to be paid for by the client.

A stitch in time saves nine: Identify risks and mitigations up-front

Businesses starting to implement AI will likely not have the knowledge to provide a detailed specification of the solution up-

front. This may lead to two issues during implementation: 1) The project will incur delays because effort estimations were inaccurate, and 2) there may be an unforeseen issue that can't be solved within the bounds of the project.

Do not start any AI project before the specifications have been reviewed by a business expert, a technical expert, and

the solution implementer. Each review should include an assessment of the riskiest parts of the project – these items must be investigated first to ensure they are solvable within the bounds of the project. For a client-vendor relationship, these risk and mitigation steps should be included in the contractual terms.

It's a Marathon Not a Sprint!

By now, most companies understand that AI will impact their business. Still many companies struggle to realize value from it and are often stuck with a myriad of “successful” AI PoCs that do not create business value. Often, this struggle stems from a lack of understanding of how to select, develop and operate AI solutions. Thus, the very first step of applying AI should be to educate oneself on what AI is and how it works.

AI means a fundamental shift of how companies can operate and compete and thus requires a holistic approach. You should focus on the highest-value applications of AI, but you also must go beyond the status quo: How might AI affect your industry and your business model? Which fundamental sources of value might disappear? What is the new moat? At the same time, you have to create the necessary enablers to successfully build and deploy AI solutions at scale, including data and ML infrastructure, people and talent, organization and governance, and, ultimately, suppliers and partners.

Don't get discouraged, there is nothing magical about scaling AI and creating value from it, but there are no easy shortcuts either.



— Dr. Philipp Hartmann
Director of AI Strategy, appliedAI

Unreliable Data in a Changing World

It is a big step from an AI Proof-of-Concept (PoC) to a fully productionized AI system. Why? In our industry projects, we see two major challenges: 1) Providing data reliably with sufficient quality, and 2) an AI algorithm that understands a changing world. As for the data, a PoC differs from a production system in that way that it is often based on historical data that is cleaned, amended and validated. However, a production system does not get this luxury. We solve this by a) detecting whether data is missing; b) semantically validating the data in its context, e.g., confirming the correct production line on a factory floor; and c) transforming the incoming live data stream to optimize performance.

As for the second item, we often see complex production processes with a high number of variants, e.g., in the manufacturing, automotive and process industries. Consequently, processes and machine settings keep evolving. This is reflected in the data, which shows completely different patterns and anomalies. Any productionized AI system must thus learn along with this changing data and complexity – just as we humans do – to achieve reliable analyses in the long run. With these two points in mind, an AI PoC can reproducibly be turned into a stable productionized AI solution.



— Britta Hilt
Managing Director and Co-founder at IS Predict GmbH

IV. Training and Testing

Lack of training data? No problem!

Sufficient quality training data is a key prerequisite for properly training an AI classifier. Companies at times do not have enough training data, especially when it comes to corner cases that must be addressed but rarely occur in practice.

There are multiple well-established workarounds to deal with lacking training data: Synthetic data from a model or a simulation of the actual process is one possibility, which comes with certain overhead and biases. But in practice, higher accuracies may also be achieved by falling back to heuristic, statistical or rule-based classifiers and then add to the training data at run-time when rare events actually occur.

Annotations are critical and dependent on subject matter experts

To build robust AI classifiers, training data needs to be annotated by a subject matter expert. Subject matter experts tend to be in short supply and sufficiently annotated training data may only become available late in the project.

Data annotation will almost certainly be on the critical path of the project and thus needs to be managed accordingly. If it is delayed because of lack of subject matter expert availability, proceed iteratively with some best-guess annotations that can be refined later on. This requires training toolchains to support quick iterations and restarts from scratch.

KPI selection matters for AIs just as much as it does for business

Multiple different metrics can be used to evaluate the quality of an AI system, and many of them require a reasonable understanding of statistics. If suboptimal metrics are selected, and possibly even tied to project success or failure, then the resulting AI system may fail to optimally solve the target business problem, e.g., by overfitting to the training data.

Unless the client has a strong statistician on the team, spend some time to define and select key metrics by which to evaluate the AI system, and ultimately decide upon project success.

To beat a human, you need to know how good the human is doing

The correctness and accuracy of a newly introduced AI system are a key concern of businesses that integrate these technologies for the first time. More often than not, the business process that an AI is supposed to automate has not been evaluated quantitatively, and hence no data is available regarding the correctness or accuracy of the current (human) process. In the absence of data, clients may be tempted to set arbitrary requirements, e.g., “AI must reach an accuracy of 99.99%.” This is harmful to the project because the final fractional percentage points towards the ideal 100% accuracy are in virtually all cases the most expensive ones.

Before setting arbitrary requirements on the required accuracy of an AI solution, companies should spend a brief moment to roughly establish key metrics of their current process. We have seen existing (human) business processes operate at a day-to-day accuracy as low as 75% in certain cases – which was acceptable to all stakeholders! For the AI rollout, it can then be decided whether it should add value by exceeding current human levels, or whether it should merely match human levels and keep project cost low.

Sometimes, AIs need help, too

To-be-automated business processes can be split into multiple sub-decisions, some of which are easy for an AI, but some of which may not. For certain parts of a business process, an AI classifier may not reach sufficiently good accuracy levels for production-level use (even with enough training data). Certain kinds of classification problems, especially those that rely on implicit context knowledge and common sense, are just very difficult, even for advanced AI.

Rather than let an entire rollout fail due to difficulties in one sub-decision, companies may consider whether minor adaptations to the existing process could make it easier for the AI to reach the correct conclusion. E.g., can a certain piece of implicit context knowledge – that is potentially very obvious to a human – be made explicit in the input to the AI classifier?

Size Matters — and Scale Will Get You There

“AI in everything we do” – a growing number of companies are realizing this will be their ambition quite soon, with just a few that have already gone the step to become an AI-fueled organization. So what does it take to implement AI at scale? While companies invest in talent and skills, technology, and ways to utilize their own data as well as access external data, many are still struggling to really scale their algorithms. On the one hand, developers are lacking the mechanisms to provide their developed services to a larger audience, ‘advertise’ them, provision users and use charging mechanisms. On the other hand, potential users are missing the opportunity to have transparency on what might be available to them, get immediate access in a self-service manner and having ethical, legal and compliance topics handled up-front.

To solve this, a platform and process are required that manage the evolution from a working PoC to a scaling AI product or service. In addition to the core AI, this requires a (cloud based) run-time environment, service and access management (incl. SSO, reverse proxy, user provisioning, etc.), measurability as tracking and charging baseline and, finally, the ability to book or ‘purchase’ the AI service. This way, true scaling and (if externally used) monetization of AI models become a reality.



— Peter Fach, Partner at Deloitte
Head of Deloitte Center for Process Robotics & AI as a Service

V. Adoption

End-users won't read the manual

Great documentation always helps software adoption, including AI. However, if the documentation is hidden away in a shared folder, then no-one will ever read it. Great documentation helps software adoption, however, if it requires the need to search for, it will, in most cases, never be found and read by end-users. This can lead to user frustration and, ultimately, rejection of the solution.

If possible, integrate the documentation into the end-users' workflow to guide them when needed. For example, one of our AI projects – automating travel expense appraisals – provides a hint on how to submit higher quality images in the future within

the expense appraisal tool rather than pointing users to the handbook on a shared directory.

Keep it simple: End-users don't need an overblown AI experience

AI has much potential for adding business value and, given the hype, end-users have great expectations from 'intelligent' solutions. Also, AI vendors tend to be guilty of over-marketing the AI in their solutions as it helps sales. However, for end-users, this has the effect of raising their expectations, in some cases, to an unattainable level.

Despite the hype and excitement around AI, in reality, end-users simply want their problem solved with zero fuss. This

means, what works well for marketing an AI might not align with the day-to-day needs of the end-users. Therefore, the KISS (Keep It Simple Stupid) principle should continuously be considered during the productionization of the solution.

Don't forget to check up on your AI

For any real-world business problem, an AI will make mistakes. Even with ample training data, an ML model will make incorrect decisions, which could be due to exceptional cases or because the underlying process deviates over time from the initial training.

Verifying an AI's performance over time is imperative to ensure high accuracy over the solution's lifespan. A thorough monitoring setup should contain automated monitoring of decision distribution and periodic manual verification on a relevant sample of decisions.

Don't break a working system

Businesses implementing AI are typically looking to optimize existing human processes to increase efficiency and accuracy. Thus, the focus for such projects is very much on ensuring high accuracy. However, it is critical that the solution can also be integrated into the existing business process without breaking it. Business processes have been designed and optimized over time for humans in the loop rather than machines. Therefore, integrating an AI component into such a process will likely be a complex, risky undertaking. No matter how well an AI solution performs, if it can't be seamlessly integrated into the existing business process, it is doomed to fail. Put as much effort into integration and functional testing as training and measuring the performance of the AI's decision-making. When you hit a wall, you may have to consider a pragmatic change to the current process or combine the AI component with robotic process automation (RPA) for greater flexibility.

If it feels like magic, then it won't be trusted

If an AI solution is making business-relevant decisions, which most solutions in production should be doing, then it is critical that the reasoning behind those decisions is transparent and understandable. Businesses will not trust an AI system without a transparent ML process and interpretable models. In some industries, healthcare or banking, it may even be a legal requirement.

An AI solution in production should be able to offer, to the greatest extent possible, a meaningful explanation for each decision it makes. This means the features used to train a model must be understandable and the reasoning behind the ML process must be accessible. In some cases, such as highly regulated industries, it may be necessary to choose a lower-performing algorithm if it has better interpretability.

Be nice, be friendly, behave!

Every business process handled by humans has its little implicit rules, its little grey zones – and some co-workers and maybe even clients may have grown accustomed to these. As part of reviewing a process for AI-based automation, these grey zones will likely be reduced. An AI that prohibits a practice that used to be allowed is likely to cause a backlash. And even if it is just that existing rules are now merely applied, a new AI may still get blamed for decisions that are perceived as unfair or breaking with habits.

When introducing a new AI system to automate an existing process, the best choice is to be transparent both about the AI and the tweaks made to the existing process. Furthermore, the AI should always strike the right tone when communicating about (recently) disallowed actions. A friendly word on why something isn't allowed, even if it comes from an AI, goes a very long way.

About Inspirient GmbH

At Inspirient, we help companies to unlock the value hidden in their data. We use Artificial Intelligence (AI) to automate data-centric business processes, from adaptive Anomaly / Fraud Detection to end-to-end Automated Analytics and Intelligent Process Automation (IPA). These fully automated approaches allow our clients to increase the effectiveness of their analytics setup by up to two orders of magnitude – thereby freeing up expert data scientists – and to discover new insights in their data beyond human intuition.

You may reach us by phone at +49 - (0)30 – 2007 4820, by email at contact@inspirient.com, or online at www.inspirient.com.



Published on 10 January 2021 with contributions from Dr. Guillaume Aimetti and Patrick Mertes. Edited by Dr. Georg Wittenburg.

Copyright © 2021 by Inspirient GmbH. All rights reserved.