**Final Checks**

It should be noted that for the sake of readability, in the paper we have used shorter IDS. Here we report the mapping.

|  |  |
| --- | --- |
| ID from this doc | ID in the paper |
| System-level.Data-quality.DCA9 | SU1 |
| System-level.Usability.DDD1.1 | SU2 |
| Component-level.Deployability.DCA10 | SDQ1 |
| System-level.Usability.DCA8 | SDQ2 |
| Architecture-process.DDA2 | SDQ3 |
| System-level.Data-quality.DCA7 | SDQ4 |
| Component-level.Driftrate.DCE2 | SDQ5 |
| System-level.Availability.DCQA6 | SU1 |
| System-level.Testability.DDQA2 | SU2 |
| Architecture-process.DDE1 | SMd1 |
| Architecture-process.DDA6 | SMd2 |
| System-level.Degree-of-modularity.BPA4 | SMd3 |
| Architecture-process.DDA1 | SMd4 |
| System-level.Monitorability.BPA6 | SMn1 |
| System-level.Monitorability.BPQA4 | SMn2 |
|  | SMn3 |
| Component-level.Deployability.DCSDLC4 | SDe1 |
| Component-level.Deployability.DDHP1 | SDe2 |
|  | SDe3 |
| System-level.Availability.DCA1 | SA1 |
| System-level.Availability.DCA5 | SA2 |
| System-level.Degree-of-modularity.BPA3 | SA3 |
| System-level.Data-quality.DCA6 | SR1 |
| System-level.Availability.BPQA2 | SS1 |
| Architecture-process.DDA7 | SS2 |
| SystemEnvironmentConcern.BPQA1 | SS3 |
| Project-process.DDQA1 | SS4 |
| System-level.Security.BPQA3 | SS5 |
| - | SS6 |
| Architecture-process.DDSDLC3 | SP1 |
| SystemEnvironmentConcern.DCQA7 | n PD1 |
| Project-process.DDSDLC8 | s PH1 |
| System-level.Testability.BPQA5 | PT1 |
| System-level.Degree-of-modularity.DDHP3 | PSp1 |
| System-level.Degree-of-modularity.DDSDLC4 | PSd1 |
| System-level.Degree-of-modularity.DCM2 | SML1 |
| System-level.Degree-of-modularity.DCM1 | SML2 |
| Component-level.Deployability.BPSDLC10 | SML3 |
| Project-process.DDM1 | SML4 |
| Project-process.BPM2 | SML5 |

**System-level concerns**

**ID:** System-level.Data-quality.DCA9

**Name:** Data visualisation

**Context:** Data visualisation is challenging as, e.g. bioinformatic information is not common, and current techniques are rarely applied.

**Check**: Do you have data visualisation techniques in place?

**Source:** Castellanos et al., 2020

**ID:** System-level.Usability.DDD1.1

**Name:** Visualisation techniques

**Context:** In large ML-based systems, Visualisation techniques help express the relationships between data and computing task

**Check**: Have you considered visualization techniques to identify or highlight relationships between data and computing tasks?

**Source:** [Panousopoulo](https://www.sciencedirect.com/science/article/pii/S0164121223002558#b19)u et al., 2018

**ID:** Component-level.Deployability.DCA10

**Name:** Data preparation

**Context:** data preparation is difficult especially making statistics of it

**Check**: Do you have strategies for data preparation and for making statistics on data?

**Source:** I8

**ID:** System-level.Usability.DCA8

**Name:** Data cleaning

**Context:** Data cleaning is cumbersome, but crucial for guaranteeing that the dataset is of good quality, not biassed and it does not lead the algorithm to strange or undesired behaviours.

**Check**: Is your dataset clean, of good quality, and free from potential bias?

**Source:** Castellanos et al., 2020

**ID**: Architecture-process.DDA2

**Name**: Dataset size

**Context**: Use micro-service architecture when developing an ML-based natural language processing system as they assist in perform better data cleaning.

**Check**: Are you concerned about data cleaning in NL processing?

**Source**: I9

**ID:** System-level.Data-quality.DCA7

**Name:** Dataset size

**Context:** Data management, e.g., pre-processing and preparation, may be problematic for ML systems with consequences lasting even after the system development. For example, to infer contextual information on extensive data as scientific imaging is challenging and time-consuming as a large amount of similar data is needed as input to the ML training system

**Check**: Do you have a well-sized dataset for training the ML component?

**Source:** Castellanos et al., 2020

**ID:** Component-level.Driftrate.DCE2

**Name:** Concept drift

**Context:** Updatability is one of the main challenges. Input data changing over time would require updating ML-based systems since updates in models can impact system outcomes and its performance.

**Check**: Are you engineering your ML-based system so as to adapt to input data changes, also known as concept drift?

**Source:** ([Baylor et al., 2017](https://www.sciencedirect.com/science/article/pii/S0164121223002558#b4)) and I9

**ID:** System-level.Availability.DCQA6

**Name:** System correctness

**Context:** Deductive verification and model checking would be very useful for increasing qualities such as safety, robustness, or dependability. However, because ML is fundamentally probabilistic and non-linear in nature, methods to ensure system correctness are barely applicable and only marginally relevant.

**Check**: Do you have proper techniques for ensuring system correctness?

**Source:** [Scheerer et al., 2020](https://www.sciencedirect.com/science/article/pii/S0164121223002558#b20)

**ID:** System-level.Testability.DDQA2

**Name:** Model validation

**Context:** Model validation is an essential phase to predict how a learning algorithm will behave on new data

**Check**: Are you performing the validation of the model, e.g., to predict how a learning algorithm will behave on new data? Are you combining model validation with data validation to better detect corrupted training?

**Source:** ([Baylor et al., 2017](https://www.sciencedirect.com/science/article/pii/S0164121223002558#b4))

**ID**: Architecture-process.DDE1

**Name**: Independent upgradability

**Context**: Use a multi-node architecture approach as it allows distributed ML system components to be upgraded independently.

**Check**: Are you building a component-based distributed system where parts may need to be upgraded?

**Source**: Burns and Oppenheimer, 2016

**ID**: Architecture-process.DDA6

**Name**: High cohesion and low coupling

**Context**: When building object recognition and image processing systems, if low coupling, high cohesion, and security are important for your object recognition/image processing system, consider using a client–server architecture.

**Check**: Are high cohesion, low coupling important?

**Source**: I10

**ID:** System-level.Degree-of-modularity.BPA4

**Name:** Microservice

**Context:** Microservice architectures (i) enable engineers to concentrate on building the business functionalities rather than writing glue code, (ii) are easier to maintain as compared to monolithic architectures as they tend to use smaller and independent components, and (iii) the low cohesion among these components helps in increasing the modifiability of ML systems.

**Check**: If you are interested in maintainability and modifiability, did you consider using a microservice architecture?

**Source:** I3, I5, I10, I12

**ID**: Architecture-process.DDA1

**Name**: Discrete service

**Context**: When building a system with multiple logical service, use micro-service architecture as a way of decomposing a big service into discrete services that can help reduce the system coupling and provide more flexibility.

**Check**: Can you decompose your system into discrete services?

**Source**: Jin et al., 2020

**ID:** System-level.Monitorability.BPA6

**Name:** Modeling intrinsic uncertainty

**Context:** It is important to explicitly model the intrinsic uncertainty of ML components and assess how it propagates and impacts other elements in the system at the designing stage.

**Check**: Can you explicitly model the intrinsic uncertainty of ML components and assess how it propagates and impacts other elements in the system at the designing stage?

**Source:** (Serban et al., 2020a)

**ID:** System-level.Monitorability.BPQA4

**Name:** Time predictability

**Context:** Static, monitor, a-posteriori analysability, and non-analysability are the best practices for time predictability and they help providing confidence to the design process for ML systems.

**Check**: Do you have proper mechanisms, like monitoring and a-posteriori analysability for time predictability?

**Source:** ([Biondi et al., 2019](https://www.sciencedirect.com/science/article/pii/S0164121223002558#b7))

**ID:**

**Name**: Monitoring Drift

**Context:** In a system that is working as intended, it should usually be the case that the distribution of predicted labels is equal to the distribution of observed labels. This simple check can help to detect cases in which the world behavior suddenly changes, making training distributions drawn from historical data no longer reflective of current reality.

**Check**: Do you have tests that monitor changes in input distributions?

**Source:** Sculley et al., 2015a

**ID:** Component-level.Deployability.DCSDLC4

**Name:** Continuous integration

**Context:** Continuous integration techniques are not generally used, even though they may decrease the complexity of the system development and integration.

**Check**: Can you use continuous integration techniques for the development of your system?

**Source:** I10

**ID:** Component-level.Deployability.DDHP1

**Name:** Infrastructure as Code

**Context:** One effective design decision when building ML systems associated with databases, servers, and other IT infrastructure is to employ Infrastructure as Code (IaC) so as to use the same structures and rules used for code development. This can reduce cost, time and risks associated with the IT infrastructure.

**Check**: Do you have a method to manage the entire IT infrastructure, e.g. databases, servers, etc., that is needed to build your ML system?

**Source:** ([Castellanos et al., 2020](https://www.sciencedirect.com/science/article/pii/S0164121223002558#b9))

**ID:**

**Name**: Blue/Green, Canary testing

**Context**: Of all of the deployment methods in a Machine Learning Operative's arsenal, there are two that can be used specifically for the purpose of bias mitigation...: Canary Deployment and Blue/Green deployment.

**Check**: Are you including Blue / Green or canary testing in your standard MLOps pipelines?

**Source**: Expert - [https://arxiv.org/pdf/2301.05775](https://eur01.safelinks.protection.outlook.com/?url=https%3A%2F%2Farxiv.org%2Fpdf%2F2301.05775&data=05%7C02%7Calessio.bucaioni%40mdu.se%7C54d91c3f5324451fc44308dce57175f0%7Ca1795b64dabd4758b988b309292316cf%7C0%7C0%7C638637522704240670%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C0%7C%7C%7C&sdata=Joc7cR6Qw4t2wS1m5Adl62TW9MEFlGF%2BSmT8N9c%2B10k%3D&reserved=0)

[https://blog.devops.dev/machine-learning-deployment-strategies-in-kubernetes-canary-blue-green-and-a-b-testing-3203c6895450](https://eur01.safelinks.protection.outlook.com/?url=https%3A%2F%2Fblog.devops.dev%2Fmachine-learning-deployment-strategies-in-kubernetes-canary-blue-green-and-a-b-testing-3203c6895450&data=05%7C02%7Calessio.bucaioni%40mdu.se%7C54d91c3f5324451fc44308dce57175f0%7Ca1795b64dabd4758b988b309292316cf%7C0%7C0%7C638637522704259422%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C0%7C%7C%7C&sdata=N8mcHwAYmZ7e4aZoONkFwyqNkAi9Mu7LGojLnUVXzF0%3D&reserved=0)’

**ID:** System-level.Availability.DCA1

**Name:** Failure recovery strategy

**Context:** Microservice architectures are particularly challenging when used for developing ML systems with respect to failure recovery as the failure of a service may be propagated to other services

**Check**: Did you consider failure recovery strategies to avoid propagation of failures?

**Source:** [Jin et al., 2020](https://www.sciencedirect.com/science/article/pii/S0164121223002558#b12)

**ID:** System-level.Availability.DCA5

**Name:** Domain knowledge

**Context:** Specific domains such as mobile robots can require specific knowledge, like mobile robotics can require knowledge in computer algorithms or probability theory to select and design an architecture that satisfies both functional and [quality attribute requirements](https://www.sciencedirect.com/topics/computer-science/quality-attribute-requirement)**.**

**Check**: Do you have the required level of domain knowledge to take availability decisions?

**Source:** [Muzaffar et al., 2015](https://www.sciencedirect.com/science/article/pii/S0164121223002558#b18), [Bhat et al., 2018](https://www.sciencedirect.com/science/article/pii/S0164121223002558#b6), I4

**ID:** System-level.Degree-of-modularity.BPA3

**Name:** Layered/tiered architecture

**Context:** The use of three layers architectures brings some benefits like (i) facilitates isolating failures, (ii) splits the business logic part from ML components, and (iii) allows engineers to rollback the inference engine independently of the business logic when the inference engine encounters some issues.

**Check**:Can you cleanly split business logic from ML components? If so, did you consider using a layered/tiered architecture?

**Source:** [Amershi et al., 2019](https://www.sciencedirect.com/science/article/pii/S0164121223002558#b2)

**ID:** System-level.Data-quality.DCA6

**Name:** Uncertainty

**Context:** ML components may introduce uncertainty with respect to the reliability of the software architecture design. In addition, design smells are challenging to find.

**Check**: Do you complete information on the uncertainty of the ML components used at design time?

**Source:** [Serban et al., 2020](https://www.sciencedirect.com/science/article/pii/S0164121223002558#b24), [Sculley et al., 2015](https://www.sciencedirect.com/science/article/pii/S0164121223002558#b22)

**ID:** System-level.Availability.BPQA2

**Name:** Fail safe

**Context:** Safety-critical systems developed using ML should not reach hazardous conditions; to this end, components may be switched off to reach a safe state quickly.

**Check**: Do you have proper techniques for reaching safe states quickly when needed?

**Source:** [Biondi et al., 2019](https://www.sciencedirect.com/science/article/pii/S0164121223002558#b7), [Serban, 2019](https://www.sciencedirect.com/science/article/pii/S0164121223002558#b23)

**ID**: Architecture-process.DDA7

**Name**: Safety evaluation

**Context**: Include an evaluation process of architectural safety methods before moving to the development stages when developing safety-critical ML systems.

**Check**: Have you included an evaluation process for architectural safety design choices?

**Source**: Serban, 2019

**ID:** SystemEnvironmentConcern.BPQA1

**Name:** Coding standards

**Context:** When developing safety-critical ML components it is important to follow strict and certified coding standards.

**Check**: Do you use strict and certified coding standards when developing safety-critical ML components?

**Source:** ([Biondi et al., 2019](https://www.sciencedirect.com/science/article/pii/S0164121223002558#b7))

**ID:** Project-process.DDQA1

**Name:** External certifications

**Context:** Creators of safety-critical systems shoulduse of strict coding standards and the use of (safety) certification from authorised bodies.

**Check**: Are you having your system safety-certified by an external body?

**Source:** Biondi et al., 2019

**ID:** System-level.Security.BPQA3

**Name:** Design to defend

**Context:** -

**Check**: Are you explicitly designing and developing your ML system to defend vulnerable sections of the code that may be subject to cyber-attacks?

**Source:** [Biondi et al., 2019](https://www.sciencedirect.com/science/article/pii/S0164121223002558#b7)

**ID:** -

**Name:** Safety and fairness

**Context:** FMEA can help companies identify safety and fairness risk in multiple failure modes of an AI system.

**Check**: Do you have a way of systematically ensuring safety and fairness in your system?

**Source**: [https://link.springer.com/article/10.1007/s43681-022-00145-9](https://eur01.safelinks.protection.outlook.com/?url=https%3A%2F%2Furldefense.com%2Fv3%2F__https%3A%2F%2Flink.springer.com%2Farticle%2F10.1007%2Fs43681-022-00145-9__%3B!!PvDODwlR4mBZyAb0!WEyzxo-hrMrFZdGVhiy2L-qcX7OLxPVynCg-nfPWjevl8Eo9jeXa-kem6VBunX2lgxvzwO9KNKR9orTUatUjdIcfA-M%24&data=05%7C02%7Calessio.bucaioni%40mdu.se%7C54d91c3f5324451fc44308dce57175f0%7Ca1795b64dabd4758b988b309292316cf%7C0%7C0%7C638637522704272353%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C0%7C%7C%7C&sdata=82476sQkLM1lzauLBgI5JxhmbXROBfOh7yIyIMUCCkw%3D&reserved=0)

**ID**: Architecture-process.DDSDLC3

**Name**: Data loss

**Context**: When building distributed ML-systems, if you need to reduce data loss as well as improving the preservation of data privacy, use federated learning.

**Check**: If you need to reduce data loss as well as improving the preservation of data privacy, one way is to use federated learning.

**Source**: Li et al., 2020

**Process-level concerns**

**ID:** SystemEnvironmentConcern.DCQA7

**Name:** Documentation

**Context:** A lack of documentation can affect the quality of ML systems and can negatively affect the reuse of processes. Contrariwise, proper documentation increases the efficiency, reusability, reproductivity, and shareability.

**Check**: Do you have proper documentation or a plan to document your ML system?

**Source:** I10 and [Wan et al. (2019)](https://www.sciencedirect.com/science/article/pii/S0164121223002558#b26), (Anjos et al., 2017)

**ID:** Project-process.DDSDLC8

**Name:** Heterogeneous teams

**Context:** Consider building heterogeneous teams for maximising knowledge sharing. This leverages the practical experience of ML developers as a complement to the expertise of architects.

**Check**: Do you have heterogeneous teams mixing ML developers, data engineers, architects

**Source:** I3

**ID:** System-level.Testability.BPQA5

**Name:** Test-driven

**Context:** test-driven development strategies are good means for quality assurance.

**Check**: Do you have a test-driven development strategy for your QA and, overall, is your testing process standardised?

**Source:** I9 and I6

**ID:** System-level.Degree-of-modularity.DDHP3

**Name:** Separate pipelines

**Context:** In large scale ML systems, to avoid the so-called pipeline jungle, it is a good idea to separate the branches for the training of the pipelines from the training of the model.

**Check**: Do you separate the branches for the training of the pipelines from the training of the model?

**Source:** [Abadi et al., 2016](https://www.sciencedirect.com/science/article/pii/S0164121223002558#b1), I12

**ID:** System-level.Degree-of-modularity.DDSDLC4

**Name:** Separate models

**Context:** Configurations of ML-based systems are hard to modify and configuration mistakes can be costly in terms of time, computing resources, etc. This also helps to detect unused or redundant models in the ML-based systems.

**Check**: If you have more than one model to develop, did you plan to develop them separately?

**Source:** Sculley et al., 2015a

**ID:** System-level.Degree-of-modularity.DCM2

**Name:** Model customisation and reuse

**Context:** Model customisation and reuse may be difficult considering that some of the required expertises are not commonly found in software teams.

**Check**: Do you have expertise to customise and reuse models?

**Source:** [Amershi et al., 2019](https://www.sciencedirect.com/science/article/pii/S0164121223002558#b2)

**ID:** System-level.Degree-of-modularity.DCM1

**Name:** Managing and versioning

**Context:** ML models are more difficult to manage as separate modules than other software components because they can get entangled in complicated ways and exhibit non-monotonic error behaviours. This issue is worsened by the lack of knowledge of the problems and best practices for ML model maintenance, due to ever-evolving research in both ML infrastructure and algorithms, increasing complexity of models and their opacity.

**Check**: Do you have managing and versioning techniques in place?

**Source:** [Schelter et al., 2018](https://www.sciencedirect.com/science/article/pii/S0164121223002558#b21), [Amershi et al., 2019](https://www.sciencedirect.com/science/article/pii/S0164121223002558#b2), [Serban, 2019](https://www.sciencedirect.com/science/article/pii/S0164121223002558#b23)

**ID:** Component-level.Deployability.BPSDLC10

**Name:** ML infrastructure for deployment

**Context:** It is important to set up a proper ML infrastructure, as well as suitable training and deployment processes. For example, Federated Learning requires a distributed architecture hence the team needs to design a distributed and dynamic system.

**Check**: Have you defined an ML infrastructure and deployment processes?

**Source:** I7

**ID:** Project-process.DDM1

**Name:** Model selection

**Context:** Consider putting more attention to your choice of model, depending on the domain type.

**Check**: How did you go about choosing an ML model? Based on what criteria?

**Source:** I6, Amershi et al., 2019

**ID:** Project-process.BPM2

**Name:** Model training

**Context:** Train the model may enhance the performances.

**Check**: Are you taking care of training and performance of the model?

**Source:** Amershi et al., 2019, I4