A Geometric Perspective on ML Safety Assurance

Emmanuel Ledinot, Gassino Jean, Ricque Bertrand, Mekki-Mokhtar Amina, Serratrice Franck, Philippe Quere

A Geometric Perspective on ML Safety Assurance

Ledinot Emmanuel
System & Software Engineering Laboratory.
THALES Research & Technology
Palaiseau, France
emmanuel.ledinot@thalesgroup.com

Gassino Jean
Nuclear Safety Department
Institut Radio Protection et Sûreté Nucléaire
Fontenay aux roses, France
jean.gassino@irsn.fr

Ricque Bertrand
Optronics and Defence Division Customer Support
SAFRAN Electronics & Defense
Massy, France
bertrand.ricque@safrangroup.com

Mekki-Mokhtar Amina
Safety Expertise Department
ANSYS
Toulouse, France
amina.mekkimokhtar@ansys.com

Philippe Quere
Software Safety Expert
STELLANTIS
Poissy, France
philippe.quere@stellantis.com

Serratrice Franck
Embedded Software and Connected Services Quality Department
RENAULT SAS
Guyancourt, France
franck.serratrice@renault.com

***Abstract*— Some people claim AI-ML suffers from a reliability glass ceiling effect, around ~10^{-2}/inference, that makes it incompatible with safety-criticality by several orders of magnitude. Others advocate that safety nets and development assurance will overcome this gap so that there is no real concern indeed. We propose an explanation to the reliability plateauing phenomenon based on geometry of approximant adjustment, and on ML verification practices. We advocate the need for a new field we coined as HR ML (Highly Reliable) and UHR ML (Ultra Highly Reliable). Relying on Topological Data Analysis in high dimensions, its aim is to supplement data-science point-based verification with *volume*-based verification in order to meet the needed 10^{-5} / inf. error rates (and beyond). We argue that process-based ML assurance and safety monitors alone will not overcome the reliability barrier. Our HR-ML concept for safety-related applications is a research proposition at the confluence of ML assurance and system assurance.**

Keywords— machine learning, ML reliability, safety assurance, ML assurance, latent manifold, extensional coverage analysis, Topological Data Analysis (TDA).

I. INTRODUCTION

Data analysis and statistics developed over centuries to extract synthetic information from population data as *insights* on complex phenomena. Inferential statistics focused on explanatory models of past observations, then used as predictors. Accuracy and consistency of estimators were the guarantees mathematics had to lay down. Never, until recently, had statistics to address safety-sensitive 'control'. We use 'control' in the broad sense of OODA loops (Observation, Orientation, Decision, Action), or any part thereof, where physics is involved with life, goods or environment exposition.

Machine Learning, especially Deep Learning (DL), opened a new era: unprecedented performances in machine vision and problem solving in higher dimensions, while being plagued by severe brittleness issues. Could DL, however, if supplemented with development assurance and fault-tolerance, meet the requirements of safety-sensitive 'control'? We investigate this question. We consider safety-sensitiveness (low to medium severity) and safety-criticality (catastrophic consequences). Our focus is limited to ML *reliability*, ML *verification coverage*, and to safety assurance of ML-dependent systems.

To our knowledge [1], the best image classification score obtained by ML on the MNIST benchmark is 3.10^{-3}. From a safety assurance perspective, this reliability score is poor. To cope with this matter of fact, [1] screened the techniques amenable to improve ML reliability. They questioned the feasibility of reaching the levels required by the higher assurance levels and concluded negatively. After some scoping and terminological preliminaries, we summarize this survey. Then we propose geometric reasons to explain why the reliability enhancement methods uniformly failed (sections II, III, IV).

Are there any solution to this problem? We explain why software assurance will have no impact on it (section V), and why fault-tolerant architectures solve only the easy cases (section VI). At this stage, we conclude that for true ML-dependent safety-sensitiveness and safety-criticality, there is no escape from *improving ML reliability* by orders of magnitude (3 to 5). Is it possible? From a geometric perspective, the complexity of the task is so high that there are many reasons to be hopeless. However, thanks to recent advances in Topological Data Analysis (TDA) in higher dimensions we propose to control sampling and adjustment more tightly. We coin 'HR-ML' (Highly Reliable) and 'UHR-ML' (Ultra Highly Reliable) this TDA-augmented and dependability-oriented variant of Machine Learning. As of writing this paper, we have no evidence to back the feasibility of our (U)HR-ML proposal. It is our best constructive and research perspective to overcome the reliability glass ceiling phenomenon.

Contribution: we propose a diagnosis on the ML-reliability plateau. We ground it at the confluence of data science, topological data analysis and system safety assurance. We propose orientations to supplement the classical *point*-based approach with a more progressive and *volume*-based approach to sampling coverage and adjustment verification.

Disclaimer: The views expressed in this paper are those of the authors as members of the Embedded France Working Group on safety assurance standards [24]. They do not reflect the opinion of their affiliations.

II. SCOPING ML-DEPENDENT SAFETY

A. System Perimeter

We address safety-sensitive and safety-critical embedded systems. Our prototypical use case in automotive is pedestrian detection systems coupled to automatic-braking systems. In aeronautics, autonomous flying cabs and drones are the ML-dependent examples we have in mind. More generally, we consider ML-dependent vehicle control, formation control, product health monitoring (PHM), and all kinds of operational technologies (OT). We question the possibility of transitioning from ML-dependent *advisory* mode to ML-dependent *full-authority* mode, by relying only on assurance and safety monitors to overcome the ML- reliability gap.

B. ML Perimeter

We consider off-line supervised learning in high to very high input-space dimension (e.g. the pixel count of the image feed to ML-classifiers, i.e. typically 10^4 to 10^6 and beyond). We exclude continuous learning. We exclude ML developments like ChatGPT. Regarding the ML-safety survey [5], we only address Robustness and Monitoring. Transformers, Q&A systems, representation learning, ethics, or Alignment are important issues out of scope of this paper.

III. TERMINOLOGICAL PRELIMINARIES

We recall a few definitions used in the sequel.

A. Machine learning

- *Approximant***,** any function $\mathbb{R}^n \rightarrow \mathbb{R}^p$, estimator of an underlying function specified by texts and datasets. We use 'ML-model' (after adjustment) as synonymous of fitted approximant.
- *Inference***,** and generalization, are used as synonymous: activation of the approximant on an input vector not seen during the training, validation and testing.
- *Ambient space***,** also named embedding space: the space where spread the vectors (or points) of the datasets. Depending on the context, we use "ambient space" for input only (nD), input-output $((n+p)D)$ or output only (pD) space.
- *Latent space* or latent manifold, the regions of ambient spaces where the sample points concentrate. Latent space has its own dimension named *latent dimension*.
- *Dimension reduction***.** The classical interpretation is the process of identifying the input space features that are salient in conditioning the form of the output latent manifold (i.e. sensitivity analysis). It is used to select the prominent ones (e.g. PCA). We never use this meaning. We only consider ambient to latent dimensionality reduction, if any …, by shifting from external view of the latent manifold, to internal and *intrinsic* view of it.

B. Logics

- *Extensional***,** qualifies extension as defined in "Extension Theory" [6], i.e. vector encoding of magnitudes for geometric and algebraic calculation. Two approximants are extensionally equivalent when they have the same external ("black-box") behavior, i.e. they form the same output space when exercised over their common input domain.
- *Intensional***,** qualifies objects or sets defined by indicator functions. Intensional equality is Leibnitz equality : "no discriminative property". Intensional equality is 'white box' and implies extensional equality. The converse is false.

IV. ML-RELIABILITY GLASS CEILING

A. Reliability augmentation techniques

In [1], researchers investigated the means to improve ML reliability. Though ML made major progress over the last decade (1 to 2 orders of magnitude in accuracy), 10^{-3}/inference is poor from safety perspective. The paper reviews quantitative results obtained by model diversification, monitoring (ODD, robustness, I/O consistency), by robustness enhancement techniques (model stability and training stability), by selective classification, conformal prediction, and temporal redundancy on sequences. The main conclusion is the following: all the methods that tried to increase reliability by redundancy of "independent" models, i.e. by independent hyper-parameters, independent datasets and independent optimization processes, succeeded marginally. Reliability stayed in the range of 10^{-2} / inference instead of the expected $10^{-4} = 10^{-2} * 10^{-2}$ or $10^{-6} = 10^{-2} * 10^{-2} * 10^{-2}$. Moreover, all the listed techniques improved the reliability performance marginally at the expense of availability losses.

B. Common Cause Analysis

Strong correlation of inference errors between independently developed ML-models is an experimental fact evidenced by the studies reported in [1]. What could be the explanation? Our working hypothesis is that the complexity of the latent manifolds' forms to be fitted is the common mode between the so-called "independent" redundancies. Depending on the specific nature of the ML application to be developed, the adjustment problem and its associated assurance policy may be addressed in a deterministic geometric setting (smooth latent manifold) or[1] in a noisy aleatoric setting (statistical latent manifolds).

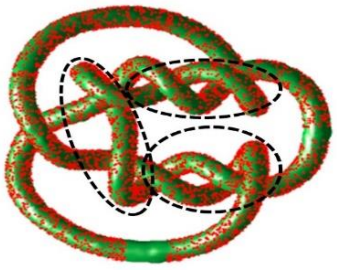

Fig. 1. 2D smooth manifold reconstruction (green surface), in 3D ambient space from red sample points. The picture is based on courtesy of [7]. We use this form as an example of regions hard to fit (surrounded by dotted-ellipses) and likely to cause poor inference reliability. This adjustment difficulty is (far) higher in higher dimensions. The ML-model redundancies are *intensionally* independent. Our interpretation of the marginal gain is that the redundant approximants are *extensionally* correlated by the '*shape of the problem'* to solve, i.e by the shape of the latent IO-manifold to adjust with.

[1] Depending on this choice TDA uses different types of point to point-set distances: off-sets in the deterministic case, distributional (KL-divergence, Wasserstein, earth-move) in the statistical case.

C. Plateauing performance

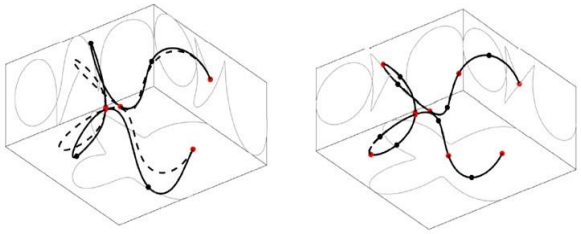

Fig. 2. 1D latent manifolds in 3D ambient space. Limiting generalization errors to extremelly small quantities (10^{-k} $k \geq 5$).requires controlling regularization with extreme precision, first locally, and then globally. Notice the impact of fitting variability on the projected curves (picture is courtesy of [8]).

When the approximant space is defined by the solution to (n – 1) polynomial equations (n-variable-polynomials) the ambient space is nD, and the latent space is 1D. Given k points in the nD space, the existence of an interpolation curve linking these k points is still an open mathematical problem. By 2022, a proof of existence was published on the Web and is under peer-review [9]. In case of confirmation, more than a century will have been necessary to solve the (1-latent, n-ambient) case for an intensively investigated approximant space: the algebraic curves.

This example substantiates the *complexity* of function fitting in high dimension. Collision avoidance or drone control, or any real life ML application requires solving even more complex adjustment problems. We suggest that high reliability of adjustment necessitates tighter control of the complex geometric solving process than with current black-box optimizers.

D. Zero-measure verification

Figure 2 visually suggests the difficulty of point-cloud fitting when only a few points are available. Information contained in dataset specifications is poor. Meeting failure rates as low as 10^{-k} /inference, $k \geq 5$, is highly demanding. Sample-oriented by nature, statistical estimation of functions naturally relies on *point*-based verification. Verification coverage of an estimated continuous function (regressor or separator) by means of some N-sample-dataset is N*0=0 in volume. At the opposite, the nD volume over which the estimated function has to generalize with high reliability is gigantic. Even worse, the definition domain of approximant is … *undefined*. There is a great discrepancy between the limited control of adjustment, the absence of formal input-domain definition, and the inference reliability levels required by safety.

E. 'One-shot' global adjustment

Complexity of ML is high and may be compared to that of non-linear control. Shape complexity of phase spaces (differential geometry and differential topology) compels control engineers to start by designing *local* control laws (dynamic regimes), and to progressively 'glue' them by scheduling and switching up to covering the global reachable state space. Our guess is that ML ensuring high levels of inference reliability can't afford monolithic, "one-model-fits-all" global trade-off over a gigantic domain. Like for non-linear control, (U)HR-ML is likely to necessitate a local-to-global approach: the global model would be a scheduler of precisely fitted local models. Because of varying curvature, varying dimensionality, and complex varying topologies as illustrated on fig. 1, fitting reliably a single approximant, even over-parameterized and with high capacity, might be illusory. Possibly, for a situation like that of fig. 1, 10^{-1} is reachable with a single ML-model, whereas one would need to schedule five local ML-models to meet a global 10^{-4} target: three models for the three difficult twistetd and entangled regions, plus one model for each of the two remaining 'easy' parts.

V. ML-Reliability gap filling by software assurance?

Some people advocate that reliability of 10^{-2} / inf. at ML model level could become 10^{-5} / inf. after implementation, provided the software is developed under DAL A assurance. The reason would be that "DAL A development delivers high integrity software" and "high integrity software is 10^{-k} /h reliable".

The goal of software assurance is to ensure *fidelity* of the transformation process that converts a specification such as an ML-model (e.g. TensorFlow mathematical equations) into binary code instructions. Fidelity, also named *compliance*, means ensuring extensional equivalence between the ML-model and its executable object code. In other words, regarding reliability of inference, DAL A ensures high trust on *reliability invariance* i.e. "garbage in, garbage out", not magic reliability augmentation during the translation flow.

Notice that explaining why there is no reliability augmentation with assured software is not bringing discredit on the value of software assurance. One may found more information on the link between qualitative and quantitative aspects of development assurance in [10].

VI. ML reliability gap filling with fault tolerance?

We claim that safety nets can handle only the easy cases of ML-dependent unreliability in safety-sensitive or safety-critical systems. In other words, the cases where risk does not depend on the performance premium uniquely provided by ML techniques, especially Deep Learning.

Let us consider our pedestrian collision avoidance or GPS-denied drone landing examples. Deep Learning systems have by far outperformed any other classical certifiable approach in machine vision. If some classical and underperforming vision monitor is sufficient to ensure controllability, then the DL-dependent channel provides just perception performance bonus (advisory mode, no true ML-dependent criticality). Otherwise, i.e. when ML-capability is mandatory to reach the required performance level, the monitor must have equivalent or just slightly lower performance, and thus cannot be implemented with classical underperforming machine vision. And in this case we face the above-mentioned issue that combining two independent ML models may not significantly improve reliability.

Safety nets alone are not a solution to circumvent the reliability glass ceiling problem when one needs COM+MON dependence on ML's distinctive (but brittle) performance superiority.

VII. TDA-enabled (U)HR ML Proposal

Let us recap where we are at this stage. We have proposed an explanation of the ML reliability gap (10^{-2} /inference .vs. 10^{-5} /inference or better): adjustment failure on higher dimensional difficult topological regions that correlate any

number of so-called “independent” approximants. We have justified why reliability of generalization must improve to meet ML-dependent dependability requirements in the difficult cases. We are facing a single point of failure that cannot be prevented by development assurance, nor be passivated by monitors in the true ML-dependency cases. We advocate that improving ML reliability, i.e. shifting from ML to (U)HR-ML, is the only path to reach full-authority safety-sensitive ML-dependent ‘control’ (DAL C and beyond).

Our last two sections are prospective. We propose means drawn from Computational Geometry (CG) and Topological Data Analysis (TDA) in higher dimensions [11] to support High Reliability Machine Learning. In case of success, UHR-ML would follow along similar lines, but with ever-tighter sampling and adjustment control. Better awareness of adjustment’s geometric (local) and topological (global) complexity is to our opinion the ‘missing link’ to meet the needed higher reliability levels on point-cloud generalization.

A. *Semantics of emptiness*

High dimensional void is the ambient space of training and test datasets. Emptiness around samples may result from principled design choices or from loopholes. Emptiness may be full of missing information that prevents from meeting the reliability target. We distinguish four types of voids:

1) *Causal impossibility*

Physics, scene or environment evolution laws, operational concepts or ODD constraints may prevent the generation of samples in definite regions of the input space. It leads to distant clusters i.e. to disconnected sums of sub-manifolds in topological language.

2) *Sample incompleteness*

The sampling plan or data-collection process, compliant with the ODD and with the ML-model’s textual specification, may overlook some input space regions. Depending on local regularity and on approximation sparsity, these sampling lacuna may or may not constitute potential sources of inference errors.

3) *Designed separability*

In classification, separation and separators (i.e. intended void regions) may be looked after and engineered (e.g. SVMs).

4) *Designed sparsity*

When variance is assumed bounded (e.g. Lipchitz) and regularity is well understood, sampling and approximant structure may be sparse. Computation tractability and energy saving are the intended benefits of extensional and intensional sparsity. In that case, the intended voids are not sampling lacuna, there is no missing information.

In summary, regarding sampling coverage analysis, our TDA-enabled (U)HR-ML proposal deals with exploration of *point-cloud forms* to identify the shape of the higher dimensional voids (see fig. 5). In other words, it would consist in detecting the unintended informational lacunas as potential inference unreliability sources, or adjustment complexity sources.

B. *Design of sample-interpretation hypotheses*

TDA offers a portfolio of algorithms to analyze point clouds in 2D, 3D, and higher dimensions. We focus on persistence homology (PH) which plays a central role in TDA. In our context, Persistent Homology may be seen as a means to mesh the latent manifolds. It is used in ML for clustering and for feature engineering (e.g. [12], [15]). We propose a new application of PH to machine learning, in order to overcome the low reliability barrier: *sampling coverage* analysis and adjustment *verification coverage* analysis.

Roughly, PH computes a growing sequence of balls centered on each point of the dataset. For each ball radius of the sequence, named filtration, it computes the ball intersections and creates edges between the centers of intersecting balls (see the four filtration steps of fig. 5). These edges constitute a mathematically well-founded nested mesh (simplicial complex) that enables reasoning on a discrete *approximation* of point cloud shapes, i.e of latent manifolds. In particular, PH focuses on the creation of cycles, of cavities and holes, and disappearance thereof as ball radius grows during the filtration process. It ends when the radius is so large that all balls intersect and no inner void is left within the point cloud.

Figure 5 shows an example of a 2D point cloud. We propose to use PH filtration as a (U)HR-ML data engineering practice to *design* the best ODD-compliant interpretation of the training point cloud. Output of this task would be the ID-OoD [2] oracle of the approximant. For computational tractability, latent dimension must be far lower than ambient dimension. Geometrical Deep Learning [4] is pivotal here: latent dimension *may* collapse in datasets only if sampling is performed for the quotiented latent I-manifold by the ODD-compliant symmetries [16], [21].

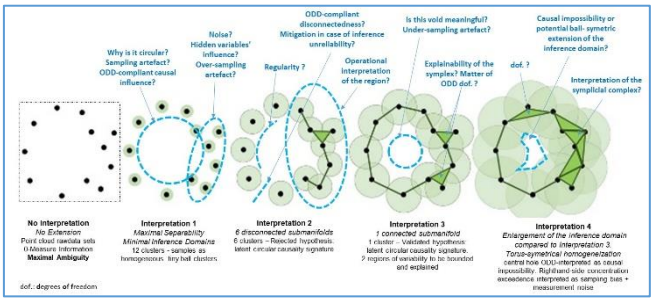

Fig. 3. This figure illustrates a process that would be distinctive of (U)HR-ML: design of a “meaning” to point clouds. Four steps of persitence homology filtration are represented. In the upper part of the figure are examples of typical questions to interpret the filtration step (the green 2D-balls centered on the sample-points). At bottom we wrote examples of interpration decisions that could lead to select this step. PH does not compute a model, it computes a sequence of models. The selection of an interpretation would start the task of explicit definition of the approximant’s domain.

C. *Formal definition of domains or supports*

To our opinion, HR-ML’s approximants will require formal and executable definition of their domain (i.e. of their precondition in formal method language). PH offers means to define ID-OoD oracles in a way that does not depend on distributional assumptions or ML techniques like VAEs, GANs etc. [22].

[2] In Distribution - Out of Distribution

D. Extensional verification coverage analysis

We envision PH-based construction of latent space meshes as means to guide scrutiny of generalization reliability. Datasets augmented with a few "interpretation meshes" (Fig. 5) selected from the filtration could support verification coverage criteria. We name *extensional coverage* analysis this (hyper-) *volume*-based verification activity. It would be the extensional counterpart of structural coverage analysis in software (e.g. DC, MC/DC criteria). Such latent-space oriented verification coverage ideas are being explored in [23].

Fig. 4. Filling the ML reliability gap by better verification coverage techniques. Extensional verification should ensure non-zero measure coverage. PH's simplicial complexes are volumes (hyper-tetrahedrons) within which approximants could be massively tested or formaly proved when possible (e.g. Reluplex). This would be principled extensional verification approach to mastery of complexity [17].

E. Dependent directional statistics in higher dimensions

Entanglement of deterministic and stochastic processes generate the samples of ML applications. The deterministic part of sample dependences invalidate the pivotal *i.i.d* hypothesis of probability and statistics. ML and (U)HR-ML require sophisticated statistics on manifolds [18], [19], [13]. Our proposed use of topological inference to support awareness of adjustment complexity is also intended to support high-dimensional non-asymptotic statistics (e.g. by aggregating empirical concentration information during the filtration process).

VIII. (U)HR-ML ASSURANCE

Our orientation is risk-based and product-based assurance, as possibly contributing process-based [2] and property-based assurance [3].

A. Geometric perspective on ML risks

On basis of our preceding diagnoses and proposals, our top 3 risks on ML-dependent 'control' programs are:

1) Information shortage at specification stage

Dataset-specifications ensure that almost nowhere we have information on the program's intended behavior. Almost everywhere, the unreliability risk is looming. This specification information shortage is an essential difference with classical software engineering. It is specific to machine learning. It could invalidate the assurance concept of potentially realizable "perfect development" that prevails in software assurance. This concept grounds the *Fault Elimination* policy in development (Dev). Known bugs that may be fixed should be fixed (no hidden defects).

2) Undefined inference domains

In high dimensions, explicit definition of distribution supports and of approximants' ID-OoD oracles are still open problems. It is a barrier to HR and UHR ML. We propose a PH-based approach to try overcoming this problem.

3) Adjustment complexity

Defined in VII.D, we proposed *volume*-based extensional verification activities as mitigation means.

B. Assurance policy on known residual faults by end of development

For high criticalities, if Authorities decide that a fault-free development should exist and should be approached as far as reasonably possible (ALARP), like for software for instance, then DAL A ML-models transitioning from Dev to Ops should be free of *known* faults (*Eliminative* assurance policy). Under this policy, exposing life and goods to failures originating from remaining known and fixable faults is unacceptable. On the contrary, if the remaining known faults are *unavoidable*, i.e. there is no notion of possible "perfect" development; *Quantitative* ML assurance policy is acceptable. Because of the information shortage problem specific to ML, Eliminative policy could be inapplicable and Quantitative policy on ML-software might become acceptable (i.e. sufficient rareness of failure occurrences). Such a decision seems under way in aeronautics as assurance concept [2].

The quantitative policy assumes existence of engineering capabilities to predict the future Ops-failure rates from the failure rates observed at Dev-time. It also assumes in-service data logging and mining. In a continuous development / continuous assurance setting (e.g. [2]), the Dev-failure events and Ops-failure events are altogether continuously managed. For high criticality levels, and from a "classical software" assurance perspective, introducing reliability quantification on software-implemented ML-models is a significant paradigm shift. Figure 5 proposes a SWOT analysis of this shift.

ML-Assurance QUANTITATIVE Policy	
STRENGTH	***WEAKNESS***
Assurance Rationale ML Statistical Foundations **CONSISTENCY**	ML Reliability Computation **DIFFICULTY**
OPPORTUNITY	***THREAT***
Known faults @Dev-time Unknown faults @Ops-time Statistical Quantification **UNIFORMITY**	Reducing Trust to Dubious Numbers Eliminative Policy @Dev-time **REGRESSIVITY**

Fig. 5. SWOT of the "pan-statistical" option in ML assurance. Difficulty to get valid risk quantifications and possible regressive enforcement of fault elimination policy are its drawbacks.

IX. OPEN SCIENCE EXPLORATION

Accessibility of foundational principles, availability of mature and affordable tools, and availability of skills are the prerequisites for acceptance of new engineering practices and assurance rationales in certification. Our next steps are oriented toward meeting some of these requirements. Short- to mid-term actions are development of a few use cases to test the feasibility of (U)HR-ML in "low-higher dimensions" (latent-D < 100). The first milestone is 10^{-4} on MNIST.

X. CONCLUSION

Our starting point was the following question: in spite of ML-reliability plateauing performance, could current ML best practices, supplemented with safety monitoring and ML assurance, altogether meet the reliability requirements of ML-dependent safety? We proposed a group of reasons, centered on geometric complexity of approximant adjustment in higher dimensions, to explain the ML reliability glass ceiling, and the insufficient effectiveness of independent redundancies. Point-cloud forms may be too complex to fit, dataset geometric information too poor, and goodness-of-fit too loosely verified to meet the required 10^{-k} levels. Safety nets and ML assurance alone are unable to overcome this intrinsic and experimentally established barrier when no impact on function's availability is required. Our answer to our initial question is "no"; something is missing.

We argued that optimizer-synthesis of safety-sensitive 'control' programs from datasets requires more than pure data-science and statistics. We coined (U)HR-ML our proposal of ML augmentated engineering. We put forward computational geometric and topological inference in higher dimensions as our privileged supplement. We focused on application of persistence homology in TDA to formally define the approximants' inference domains, and to tightly control sampling and adjustment quality.

Finally, we discussed the two possible assurance policies applicable to (U)HR-ML: Eliminative and Quantitative. We analyzed ML-fault elimination at development-time, and ML-failure quantification at operation-time. Considering the DevOps, MLOps and continuous assurance trends, we advocated flexible mixes of the two assurance policies.

ACKNOWLEDGMENT

The authors pay tribute to the members of Embedded France's Working Group on safety standards (NSL) for their peer review of this paper, especially to Hugues Bonnin who initiated this work, and to Jean Paul Blanquart who was pivotal in analyzing the rationales of development assurances. E. Ledinot is indebted to Stephane Mallat [21], Marc Mezard [14], Stephane Herbin and Jérôme Lacaille [20] from BNAé 's VAML WG, and to Pierre-Yves Lagrave [16] for their insightful perspectives on Machine Learning.

REFERENCES

[1] Lucian Alecu, Hugues Bonnin,Thomas Fel, Laurent Gardes, Sébastien Gerchinovitz, Ludovic Ponsolle, Franck Mamalet, Éric Jenn,Vincent Mussot, Cyril Cappi, & al. "Can we reconcile safety objectives with machine learning performances?". ERTS2022, Jun2022, Toulouse, France.

[2] "EASA Concept Paper : First usable guidance for level 1 & 2 machine learning applications" february 2023. Proposed issue 02.

[3] Morayo Adedjouma, Christophe Alix, Loic Cantat, Eric Jenn, Juliette Mattioli, et al.. Engineering Dependable AI Systems. 17th Annual System of Systems Engineering Conference (SOSE), IEEE, Jun 2022, Rochester, United States.

[4] Michael M. Bronstein, Joan Bruna, Taco Cohen, Petar Velickovic. "Geometric Deep Learning Grids, Groups, Graphs, Geodesics and Gauges. arXiv:2104.13478v2 [cs.LG] May 2021.

[5] Dan Hendriecks, Nicholas Carlini, John Schulman, Jacob Steinhardt. "Unsolved Problems in ML Safety". arXiv:2109.13916v5 [cs.LG] 16 Jun 2022.

[6] Hermann Grassmann, "Extension Theory" 1862. History of mathematics Vol. 19. American Mathematical Society 2000.

[7] Jian Liang, Frederick Park, and Hongkai Zhao. "Robust and Efficient Implicit Surface Reconstruction for Point Clouds Based on Convexified Image Segmentation. University of California, Irvine March 21st, 2011.

[8] Carlotta Giannelli, Lorenzo Sacco, Alessandra Sestini. "A local C2 Hermite interpolation scheme with PH quintic splines for 3D data streams". arXiv:2108.12948v1 [math.NA] 30 Aug 2021.

[9] Clémentine Laurens, "Un vieux problème de courbes enfin bouclé". Pour la Science N° 545, Mars 2023.

[10] Jean-Paul Blanquart, Philippe Baufreton, Jean-Louis Boulanger, Jean-Louis Camus,Cyrille Comar, Hervé Delseny, Jean Gassino, Emmanuel Ledinot, Philippe Quéré, Bertrand Ricque. "Software safety assessment and probabilities". DSN 2016 Toulouse June 28th-July 1st.

[11] Jean-Daniel Boissonnat, Frédéric Chazal, Mariette Yvinec "Geometric and Topological Inference" Cambridge Texts in Applied Mathematics 2018.

[12] Aditi S. Krishnapriyan1, Joseph Montoya, Maciej Haranczyk, Jens Hummelshøj, Dmitriy Morozov "Machine learning with persistent homology and chemical word embeddings improves prediction accuracy and interpretability in metal-organic frameworks Nature Scientific Reports 11:8888 2021.

[13] Frédéric Barbaresco, Frank Nielsen Editors, "Geometric Structures of Statistical Physics, Information Geometry, and Learning. SPIGL'20, Les Houches, France, July 27-31.

[14] Marc Mézard "Désordre et frustration … et au-delà" in Systèmes complexes, autour de Giorgio Parisi. Institut de France 11 octobre 2022 (unpublished communication).

[15] Mark Lexter D. De Lara, "Persistent homology classification algorithm" PeerJ Computer Science January 10, 2023.

[16] Simon Martin, Pierre Yves Lagrave, "On the benefits of SO(3)-Equivariant Neural Networks for Spherical Image Processing. 2022. Hal-03763121.

[17] Herbert A. Simon "The Architecture of Complexity: Hierarchical Systems" in The Sciences of the Artificial, MIT Press 1969.

[18] Martin J. Wainwright 'High dimensional statistics – A Non-Asymptotic Viewpoint". Cambridge Series in Statistical and Probabilistic Mathematics 2019.

[19] Kanti V. Mardia, Peter E. Jupp "Directional Statistics" Wiley Series in Probability and Statistics 1999.

[20] Karim Benmeziane, Patrick Fabiani, Stephane Herbin, Jérôme Lacaille, Emmanuel Ledinot "Trusting Machine Learning Applications in Aeronautics" IEEE Aerospace Conference, Yellowstone, March 4-11 2023.

[21] Stephane Mallat, "Cours 3 : Malédiction de la grande dimension," in L'apprentissage face à la malédiction de la grande dimension, Collège de France, 2018.

[22] Mohammadreza Salehi, Hossein Mizaei, Dan Hendrycs, Yixuan Li, Mohammad Hossein Rohban, Mohammad Sabokrou "A Unified Survey on Anomaly, Novelty, Open-Set and Out-of-Distribution Detection: Solutions and Future Challenges" arXiv:2110.14051v1 26 oct. 2021.

[23] Taejoon Byun, Sanjai Rayadurgam "Manifolf for Machine Learning Assurance" arXiv:2002.03147v1 8 Feb. 2020.

[24] Embedded France Groupe de travail - NSL Normes pour la Sûreté de fonctionnement Logiciel et système - Embedded France (embedded-france.org)