Interstitial b-SHAP-centric Amalgam for the Enhancement of an AI-centric Construct Validity Approach

Steve Chan
VTIRL, VT/DE-STEA
Orlando, USA
Email: stevec@de-stea.tech

***Abstract*—This paper describes an Artificial Intelligence (AI)-based Construct Validity Verification Methodology (CVVM) being advanced. The proposed methodology includes an amalgam utilization of temporal-centric Finite-Change Shapley-Owen values along with, among others, Generic Shapley-Owen values and Variance-Based Shapley-Owen values (i.e., a bespoke SHAP amalgam or b-SHAP implementation), CRiteria Importance through Intercriteria Correlation (CRITIC), and Preference Ranking Organization Method for Enrichment Evaluation (PROMETHEE) for enhancing the interpretability of not only the machine learning constituent components of an AI system, but also the interstices (e.g., between/among individual components as well as amalgams/clusters of components locally/globally). This approach extrapolates upon and furthers current proposals for the utilization of SHAP in local, global, and glocal (a hybridized intermediary of local and global) contexts. It turns out that this Interstitial SHAP-centric Amalgam (ISA), by better correlating features with each interim intended construct, potentially segues to better interpretability and construct validity at the component, interstitial, and overall system level, particularly when ISA is conjoined with a well-counterpoised Multi-Attribute Decision-Making (MADM)/Multi-Objective Decision-Making (MODM) Subjective Measures (SM)/Objective Measures (OM) paradigm and a modified Constriction Factor (CF)-Particle Swarm Optimization (PSO)-Robust Convex Relaxation (RCR)-Long Short-Term Memory (LSTM)-Deep Convolutional Neural Network (DCNN) (CPRLD) metaheuristic architectural construct.**

Keywords-artificial intelligence systems; machine learning; construct validity; explainability; interpretability.

I. INTRODUCTION

The impact of AI within the industrial sector and business, in general, should not be underestimated. Subhadra and others underscore the "rise of AI in business and industry" [1]. As AI is a transformative technology, it is envisioned to spur innovation and revolutionize various industries [2]. Honeywell's *Industrial AI Insights* report notes that, for the majority of cases, the "C-Suite has already decided to expand AI use," and in 91% of the cases, new use cases are brought to light "during AI implementation" [3]. Hence, AI forays are begetting further AI forays. These implementations involve AI software engineering, which leverages Machine Learning (ML) models and techniques to automate various tasks. The ML models of these AI Systems (AIS) are being increasingly relied upon to process/interpret *Big Data* so as to put forth meaningful forecasts/posits, thereby enhancing and illuminating certain Decision Engineering (DE) pathways so as to inform Decision-Making (DM).

A. *The Criticality of Construct Validity*

To ensure that the AIS ML models are robustly depicting the Real-World Scenarios (RWS), which they are tasked to emulate, the notion of *construct validity* becomes central. Sjoberg depicts construct validity as being "concerned with whether one can justifiably make claims at the conceptual level that are supported by results at the operational level" [4]; Sjoberg had conducted a Software Engineering (SE)-centric Systematic Literature Review (SLR) for the years 2000 through 2019 and determined that over this period of time, the prominence of the construct validity term rose by "sevenfold" [4]. Zhou affirms the criticality of validity within the SE sector and noted, comparatively speaking, the lack of research regarding the challenges related to construct validity [5]. Hence, despite the "sevenfold" increase, Deets and others find that the notion of construct validity is still "underdiscussed" [6]. As the ML models for AIS evolve, construct validity becomes particularly important to ensure that the involved progression leads to the intended construct. For example, construct validity can help ensure that the feature set aligns with the intended construct (i.e., feature alignment); also, given the understood constraints of the Shannon-Weaver model in communications theory, consideration of construct validity can help to avoid misinterpretation of the AIS ML model's posits (i.e., more robust interpretation). In essence, failure modes/blindspots and bias can be more readily identified and mitigated against.

B. *Transparency, Explainability, and Accountability (TEA) Evaluation & Testing for Enhanced Construct Validity*

Evaluation/testing (which ensures that the ML model well handles unseen data) and fine-tuning (which ensures that the ML model is optimized for a winnowed subset of data or particular task) are both integral for the enhancement of the involved AIS. The evaluation/testing of ML models involves both *construct validity,* as well as *performance metrics* to capture the intended construct and generalize well upon unseen data, respectively. The distinction is often not made, but *evaluation* and *testing* are quite marked and disparate. For example, with regards to performance metrics, evaluation tends to encompass accuracy, precision, recall, F1 score (determined by the precision and recall scores), Area Under the Receiver Operating Characteristic (AUC-ROC), cross-validation, etc. However, these types of evaluation do not provide insight into particular behaviors and/or potential Root Cause Analysis (RCA), which resides more in the realm of testing; while evaluation tends to focus upon performance of the model in its entirety, testing tends to focus upon the performance intricacies of the constituent components of the ML model. In the case of this paper, it is posited that the testing paradigm should also be extended to the interstices (e.g., interstitial areas between/among individual/amalgam of components, particularly in a glocal context). In any case, the evaluation/testing and fine-tuning paradigms are complicated enough for a single AIS, but in a System-of-Systems (SoS) (wherein constituent systems support the overarching function of the larger system) paradigm (wherein the incorrect testing and/or fine-tuning of one AIS may adversely impact another AIS), the notion of construct validity is crucial. The improving of AIS TEA at the component/interstitial areas can lead to enhanced construct validity, as feature alignment, more robust interpretation, etc. can likely be more readily achieved.

C. Enhancing TEA for Enhanced Construct Validity

Pathways for the advancement of System TEA (STEA) include a better understanding of the influence of Higher-Order Network (HONs), a finer-tuned Dynamic Assessment and Weighting System (DAWS) (wherein more apropos weights can be derived), as well as a more understandable/interpretable corpus of experience such that it can be better leveraged in a Lower Ambiguity (wherein the repertoire of experience suffices) Higher Uncertainty (LAHU) situation (given a sufficient repertoire of experience, the tolerance for uncertainty is higher, such that a decision can be made without, necessarily, the need for more *Big Data*) when time is of the essence. In addition, STEA-related SoS boundary areas also need to be taken into consideration as ML of ML becomes increasingly prevalent. After all, ML algorithms have a propensity to spawn “non-monotonic, non-polynomial [unable to be captured as a summation of terms], and even non-continuous functions” [7]. This is not dissimilar to the paradigm, wherein the transformation of “non-convex Mixed Integer Non-Linear Programming (MINLP) to convex problems, often spawn[ed] further non-convex MINLP problems” that necessitated further handling [8]. The enhancement of STEA can lead to better discernment of problematic constituent components (e.g., those exhibiting issues with *feature alignment*, *robust interpretation*, selection bias, etc.); this segues to enhanced *construct validity*.

Accordingly, this paper describes an AI-based Construct Validity Verification Methodology (CVVM) (i.e., the extent to which the AIS is accurately gauging the actual underlying concept/intended theoretical construct) being advanced. To assist the reader, a table of acronyms is provided in Table I as follows.

TABLE I. TABLE OF ACRONYMS

Acronym	*Full Form*
ACM	Association for Computing Machinery
ADMB	Automatic Differentiation Model Builder
AI	Artificial Intelligence
AIS	Artificial Intelligence System
AUC-ROC	Area Under the Receiver Operating Characteristic
c-SHAP	Classical Shapley Additive exPlanation
C2	Command and Control
CF	Constriction Factor
CNN	Convolutional Neural Networks
CPRLD	Constriction Factor-Particle Swarm Optimization-Robust Convex Relaxation-Long Short-Term Memory-Deep Convolutional Neural Network
CRITIC	CRiteria Importance through Intercriteria Correlation
CVVM	Construct Validity Verification Methodology
CWT	Continuous Wavelet Transform
DAWS	Dynamic Assessment and Weighting System
DCGAN	Deep Learning Convolutional Generative Adversarial Network
DCNN	Deep Convolutional Neural Network
DE	Decision Engineering
DeepLIFT	Deep Learning Important FeaTures
DL	Deep Learning
DM	Decision-Making
E	Execution Time
ELECTRE	ÉLimination Et Choix Traduisant la REalité
FCSO	Finite-Change Shapley-Owen
GAN	Generative Adversarial Network
GNU	GNU’s Not Unix
GPL	General Public License
Grad-CAM	Gradient-weighted Class Activation Mapping
GSO	Generic Shapley-Owen
HON	Higher-Order Network
I	Interpretability
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
IPOPT	Interior Point OPTimizer
ISA	Interstitial SHAP-centric Amalgam
ISO	International Organization for Standardization
LAHU	Lower Ambiguity Higher Uncertainty
LIME	Local Interpretable Model Agnostic Explanations
LSTM	Long Short-Term Memory
MA	Model Agnostic
MADM	Multi-Attribute Decision-Making
MINLP	Mixed Integer Non-Linear Programming
ML	Machine Learning
MODM	Multi-Objective Decision-Making
MS	Model Specific
NP-hard	Non-deterministic Polynomial-time Hardness
OM	Objective Measure
OSNS	Optimal Shapley-Nondominated Solution
OSONS	Optimal Shapley-Owen-Nondominated Solution
PROMETHEE	Preference Ranking Organization Method for Enrichment Evaluation
PSO	Particle Swarm Optimization
RCA	Root Cause Analysis
RCR	Robust Convex Relaxation
RR	Rank Reversal
RWS	Real-World Scenarios
S	Sensitivity
SDP	Semi-Definite Programming
SE	Software Engineering
SHAP	Shapley Additive exPlanation
SLR	Systematic Literature Review
SM	Subjective Measure
SNOPT	Sparse Nonlinear OPTimizer
SoS	System-of-Systems
SQP	Sequential Quadratic Programming
STEA	System Transparency, Explainability, and

	Accountability
TEA	Transparency, Explainability, and Accountability
TOPSIS	Technique for Order Preference by Similarity to Ideal Solution
U	Performance under Uncertainty
V	Validity
VBSO	Variance-Based Shapley-Owen
VC-dim	Vapnik-Chervonenkis dimension
XAI	Explainable AI

Section I provided an overview, which underscored the criticality of the notion of *construct validity*. The remainder of this paper is organized as follows. Section II reviews the notion of AI SoS ML on ML and the need for STEA (particularly *interpretability* as an actualizing agent for enhanced STEA) to facilitate viable ML of ML. Section III presents theoretical foundations, the experimental testbed, and the experimental construct for addressing the challenge of AI-based CVVM. Section IV provides some concluding remarks and puts forth some future work.

II. Background

A. AI System of Systems (SoS)

The notion of SoS is well-known; it is then axiomatic that an AI-related SoS is comprised of subordinate AIS. In theory, the involved ML at the top-tier AIS should be able to leverage the experiential base (e.g., lessons learned) of the lower-tier ML; in essence, the upper echelon ML should be able to enhance its efficacy by "learning" from the "successes" and "failures" of the lower echelon ML systems. This "learning" is effectuated by way of, among other types: (1) Collaborative learning (wherein ML systems collectively address a problem, such as in an ensemble and/or federated fashion, via *learning* from each other's discernments and approaches), (2) Multi-agent reinforcement learning (wherein ML system *learnings* can inform the subsequent pathways undertaken by other AIS to achieve more optimal results), (3) Coopetition (a portmanteau of "cooperation" and "competition") *learning*, such as in the case of Generative Adversarial Networks (GANs), wherein two AIS (e.g., generator and discriminator) engage in an "adversarial process" that segues to a win-win cooperative paradigm, (4) Transfer learning (e.g., wherein a pre-trained ML, with certain *learnings* already incorporated, can be fine-tuned and leveraged to undertake other tasks or wherein a distillation ML can transfer knowledge in a condensed form, thereby quickly enhancing efficacy and efficiency). However, to ascertain whether the *learnings* (e.g., employed approaches) are "effective" (or not) necessitates an AIS SoS ML on ML architecture that is more "white box" (e.g., wherein there is a higher degree of interpretability, such that the influencing variables are readily identifiable and the process — the involved model by which posits are generated — is more readily discernable) than "black box" (e.g., wherein opaqueness and/or translucency abounds); in other words, the desired "white box" AIS SoS ML on ML architectures need to have higher STEA (particularly interpretability).

B. AI-centric STEA and its Criticality for ML on ML

Along this vein, International Organization for Standardization (ISO)/International Electrotechnical Commission (IEC) 42001 focus upon AIS STEA; likewise, the Association for Computing Machinery (ACM) "Principles for Algorithmic Transparency and Accountability," Institute of Electrical and Electronics Engineers (IEEE) Standard for Transparency of Autonomous Systems (P7001), and others follow suit. Addressing the "T," a key factor for AIS architecture (e.g., "black-box," "gray-box," and "white-box") is in the form of transparency (e.g., opaque, translucent, and fully transparent). Addressing the "E," McKinsey portrays it as the "capacity to express why an AIS reached a particular decision, recommendation, or prediction" [9]; this tracks with prevailing definitions within the Explainable AI (XAI) field. Addressing the "A," it involves the prior "T" and "E," as the justification logic employed needs to be articulated; on this point, there is a nuance. While *explainability* and *interpretability* are often treated synonymously within the literature, perhaps they should be better distinguished. While *explainability* focuses upon *why* the AIS made certain posits, *interpretability* focuses upon *how* the AIS formulated its posits; restated, the latter delves into the AIS's DE/DM processes to derive insights into the pathways for the justification logic involved. Together, *interpretability* & *explainability* are referred to as I&E, and I&E is a lynchpin for operationalizing effective ML of ML.

C. Interpretability and AIS SoS ML on ML Architecture

For the dual pillars of I&E, interpretability turns out to be paramount. Yet, despite its criticality, interpretability tends to be challenged by the degree of complexity of the involved AIS architecture. For example, Table I presents degrees of interpretability (wherein green denotes high, yellow denotes medium, orange denotes medium/low and red denotes low) for various complexities; there is a column "Monotonic" denoting when the ML model is monotonically constrained (wherein a change at the input variable segues to a change at the response function output), and there is a row "Linear" to indicate when the output is proportional to the input as well as a row "Non-linear" to denote when the relationship is more complex (e.g., convoluted interplays among features, ambiguous boundary areas, intricate sequences of local, glocal, and global transformations, etc.). Table I is rudimentary since, as noted in Section I, the spawning of "non-monotonic, non-polynomial, and even non-continuous functions" is not infrequent [7]; this greatly complicates matters, and gauges for interpretability are often tied to "measure[s] of model complexity," such as "the Vapnik-Chervonenkis dimension (VC-dim)" [10]; the VC-dim can, by way of example, be indicative of the number of weights, rules, etc., (but does not equate to them).

TABLE II. Exemplar ML model process interpretability

	Monotonic	**Non-monotonic**
Linear		
Non-linear		

 ISBN: 978-1-68558-259-3

To date, STEA Efforts have tended to be on the *post*-side (e.g., Model Agnostic or MA), and those on the *pre*-post-side (e.g., Model Specific or MS) have had varied limitations. Exemplars of MA (e.g., Local Interpretable MA Explanations or LIME, Shapley Additive exPlanations or SHAP, etc.) and MS approaches (e.g., Gradient-weighted Class Activation Mapping or Grad-CAM, which is geared more for Convolutional Neural Networks or CNNs; Deep Learning Important FeaTures or DeepLIFT, which is geared for Keras and TensorFlow implementations; etc.) — the latter being constrained to a more limited set of ML models — are shown in Table II.

TABLE III. ML MODEL TYPES WITH EXEMPLAR I&E TOOLS

Model Specific (MS)	Exemplar I&E Tools	*Model Agnostic (MA)*
Linear Regression (LR)	e.g., InterpretML	e.g., LIME; SHAP
Decision Tree (DT)	e.g., GPTree	
Neural Network (NN)	e.g., Grad-CAM	
Deep Learning (DL)	e.g., DeepLift	

On the MS side, since the LR coefficients "directly represent the influence of each feature on the prediction," LR is construed as green when compared to the yellow of DT (which may have a complicated branching structure), the orange of NN (which may have complex internal workings, as contrasted to the more simplistic rules of DT), and the red of DL (which typically has a far greater number of layers than NN) [11]. On the MA side, LIME is oriented for more localized and individualized instances while SHAP capabilities extend beyond local and can well contribute towards a more global perspicacity across a gamut of instances; SHAP is well-suited to ascertain the more impactful features (i.e., as each feature will have a SHAP value to signify the impact on the posit, the features of import can be ascertained, and feature combinations that are able to maintain posit accuracy can be formulated while also considering the non-dominance principle, wherein no other feature combinations can provide posits without a degradation of efficacy in another facet) at the local, glocal, and global levels.

D. Optimal Shapley-Owen-Nondominated Solution (OSONS) for Enhanced STEA and Construct Validity

The Optimal Shapley-Nondominated Solution (OSNS) paradigm of Section IIC was explored as shown in Table III.

TABLE IV. EXEMPLAR DIGITAL OBJECT IDENTIFIERS (DOI) FOR VARIOUS FACETS OF OSNS

OSNS context	*Facet*	*DOI*
STEA	In general:	• 10.1109/AIIoT61789.2024.10579033 • 10.1109/OETIC57156.2022.10176215
	HON	• 10.1109/AIIoT61789.2024.10579029 • 10.1109/IBDAP62940.2024.10689701
	DAWS	• 10.1109/ICPEA56918.2023.10093212 • 10.1109/ICSGTEIS60500.2023.10424230
	LAHU	• 10.1109/GEM61861.2024.10585580
STEA-related SOS boundary areas	C2 of C2	• 10.1109/IEMCON.2019.8936241 • 10.1109/IAICT62357.2024.10617473
	ML of ML	**This paper**

In essence, it delineates prior work in the context of: (1) enhanced STEA, which facilitates a better understanding of the influence of HON-related drivers, a finer-tuned and more robust DAWS, and a more readily interpretable/leverageable repertoire of experience for a LAHU situation, as well as (2) STEA-related SoS boundary areas, such as those related to Command and Control (C2) of C2 (i.e., now ML of ML). For this paper, the notion of OSNS is expounded upon, as varied SHAP approaches differ in their local and global efficacies. By way of background, Borgonovo had referred to this hybridized efficacy as "glocal" (a portmanteau of "global" and "local"). Among other contributions, as a gauge of feature import (a key tasking of construct validity), SHAP values can be invaluable; Lundberg had advocated for SHAP to "explain various machine learning [ML] algorithms" [12]. With regards to the previously discussed (1) of this Section IID, Balog affirms the import of STEA-related HON-related drivers, and Sundararajan reinforces this perspective [13][14]. Kwon addresses the import of STEA-related DAWS, introduces "WeightedSHAP," and distinguishes it from the standard SHAP, which "uses the same weight for all marginal contributions;" Kwon also "demonstrates that the influential features identified by WeightedSHAP are better able to recapitulate the model's predictions compared to the features identified by the [classical] Shapley value" [15]. Addressing the matter from a different vantage point, Kotthoff raises the significance of utilizing the temporal-sensitive/temporal-centric (as contrasted with the classical) Shapley value, and the temporal-centric LAHU notion is delineated by the associated DOI shown in Table III [16]. With regards to the previously discussed (2) of this Section IID, Guidotti affirms the importance of ML model inspection at the margins (e.g., STEA-related SoS boundary areas) [17]. These SoS boundary areas refer to, among others, regions between/among individual/amalgam constituent components as well as local/glocal/global interstices. With regards to the former, Dhamdhere affirms the notion of "Shapley-Owen values" "for the quantification of joint contributions" [18]. With regards to the latter, Borgonovo advocates the use of *Finite-Change Shapley-Owen or FCSO values*, such as articulated by Dhamdhere), which are well suited for the *testing* facet (e.g., the discussed aspect of Section IB is more focused upon local/hyper-local scrutinization of the ML model) [18]; in conjunction with this, the Shapley-Owen values (generally, the *Generic Shapley-Owen or GSO valu*es, such as articulated by Grabisch, and more granularly, Borgonovo's suggested *Variance-Based Shapley-Owen or VBSO values*) can well serve in a generalized fashion — globally — across the model in its entirety [19][20]. Specifically, Borgonovo underscores the fact that *FCSO values* have equivalence to what Mase deemed to be the Baseline Shapley (i.e., the average of the *FCSO values* function under uncertainty) [20][21]; this Baseline Shapley also relates to the *VBSO*, since the upstream local finite-changes for the *FCSO values* segues to the Glocal Partial Dependence Function (which segues to the Conditional

Regression Function and what Mase deemed to be the "Squared Cohorts" value function) [20][21]. Borgonovo notes that by averaging the *"Squared Cohorts" Shapley-Owen or SCSO values*, the VBSO values can be obtained [20]. This reflects one of the many interplays among local, glocal, and global, and is also indicative of how "additional insights into the [ML] model behavior" are possible [20]; these supplemental insights segue to enhanced construct validity, which provides the basis for more robust ML of ML.

III. Experimentation

ML of ML is a central tenet of this paper. To improve upon the ML model and the involved SoS, the need for interpretability (and STEA) is paramount. After all, constituent component and interstitial analyses is vital for determining whether the prospective ML learnings are of potential benefit; in some cases, RCA will be needed to discern and mitigate against problematic areas affecting performance. Borgonovo's glocal notion can help bridge the gap, and the significance of the OSNS segueing to an Optimal Shapley-Owen-Nondominated Solution (OSONS) paradigm is well articulated by Casajus, Lopez, Beal, and others [22][23][24]. In essence, the Owen value (which well captures the nuanced interactions between/among the members of the feature set) extends the Shapley value (which well captures the individual feature contributions) in a consistent fashion. However, OSONS is also just a precursor, and the utilization of the b-SHAP amalgam (e.g., temporal-centric FCSO values, SCSO values, and GSO values/VBSP values) is central. In turn, the b-SHAP amalgam needs to be leveraged in conjunction with a well-counterpoised MADM/MODM SM/OM paradigm. Wu, Wang and others have advocated for the use of the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) OM in conjunction with SHAP [25][26]. Meanwhile, Hua and others have advocated for the use of the PROMETHEE OM with SHAP (there is a dearth of research for SHAP with other OMs, such as ÉLimination Et Choix Traduisant la REalité or ELECTRE) [27]. The experimentation evaluated both of the former cases, and a finding, among others, is that of utilizing an OM (e.g., CRITIC) to first, derive the criteria weights and second, use a complementary pairing for the ensuing ranking (e.g., TOPSIS, PROMETHEE).

A. Theoretical Foundations

As described in the last paragraph of Section I, the issue of Non-deterministic Polynomial-time Hardness (NP-hard) problem spawning is problematic, such that *Spawn Reduction* becomes critical [8]. The involved optimization problem transformation pathways, such as those shown in Figure 1, strive to effectuate the non-convex to convex transmogrification.

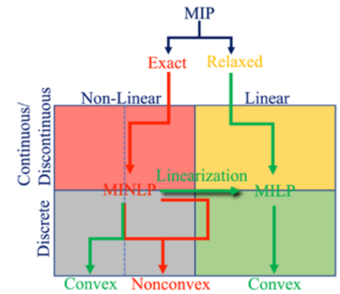

Figure 1. Non-convex to convex Transformation Pathways (e.g., non-convex discontinuous non-linear MINLPs to convex form)

A similar phenomenon is shown in Figure 2; after all, ML algorithms have a propensity to spawn "non-monotonic, non-polynomial, and even non-continuous (i.e., discontinuous) functions" [7]. Of note, the transformation of non-convex to convex can often inadvertently spawn further NP-hard problems. However, once in a convex form, a variety of Semi-Definite Programming (SDP) solvers can be employed to resolve the optimization problems in polynomial time [28].

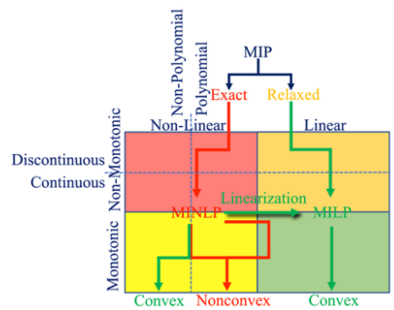

Figure 2. Non-convex to convex Transformation Pathways (e.g., non-convex [non-monotonic, discontinuous] non-polynomial MINLPs to convex form)

B. Experimental Testbed

Taking the case of NN, as depicted in Table II of Section II, the interpretability is in the orange (medium/low interpretability), as NN is more complex that DT and LR. However, for an enhanced STEA/construct validity-centric paradigm, a tasked ML can well learn atop the other MLs, adjust the involved ML model[s], and ascertain ways to mitigate against/lower the inadvertent spawning (i.e., *Spawn Reduction*). For this reason, the *testing* facet (at the constituent component level and interstices) of the *performance metrics* conjoined with *construct validity* considerations become central to the ML of ML task for the reduction of the spawning of further non-convex MINLP (e.g., from the transformation pathways of non-convex MINLP to convex MILP). In this case, the *testing* facet mechanisms and the utilized SDP solvers were implemented aboard GNU's Not Unix (GNU) Octave (a "numerical computation platform" that is "under the GNU [General Public License] (GPL) v3 license" and is generally "compatible with the likes of MATLAB") along with a myriad of Octave Forge packages [28]. As noted in [28], “the source code was modified in the lab environment” so as to implement accelerants for the referenced SDP solvers to quickly address the various involved convex optimization problems described herein. Also, as noted in [28], “GPLv3 avoids the issue of tivoization (the instantiation of a system that incorporates software under the terms of a copyleft software license but leverages hardware restrictions or digital rights management to prevent users from running modified versions of the software on the involved hardware)” [28]. Testing was conducted using a variety of open-source software packages, such as Automatic Differentiation Model Builder (ADMB) (for non-linear statistical modeling) and Interior Point OPTimizer (IPOPT) (for large-scale nonlinear optimization) [28]; other promising software packages, such as LOQO (like IPOPT, it is based upon the interior-point method) and Sparse Nonlinear OPTimizer (SNOPT) (it leverages Sequential Quadratic Programming or SQP for resolving large-scale non-linear optimization problems) were examined, but they were not utilized given their licensing caveats.

It had been discussed in [8] that a particular numerical implementation of Continuous Wavelet Transforms (CWTs), aboard a CPRLD architectural paradigm, well contributes to STEA by way of the intrinsic “successive convolutional layers (which contain the cascading of ever smaller ‘CWT-like’ convolutional filters)” [8]. The referenced CPRLD construct handled the various transformation pathways delineated in Figures 1 and 2 (e.g., convex approximations, series of convex relaxations, etc.), and the architectural implementation for this paper was unique in that a ML of ML paradigm was implemented for *Spawn Reduction* (SR_2 on SR_1), such as shown in Figure 3.

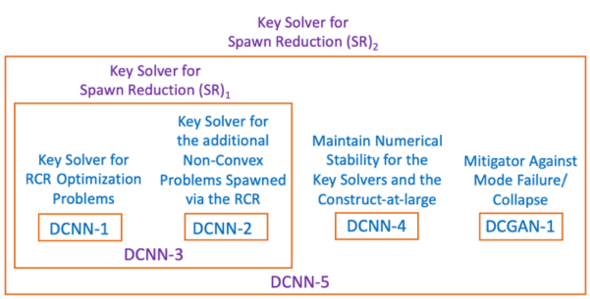

Figure 3. CPRLD Architectural Construct with a ML of ML (SR_2 on SR_1) Spawn Reduction paradigm

In terms of implementation details, a DCNN-centric instantiation was chosen for the requisite sufficient balance of reduced computational complexity along with sufficient robustness to be fit for purpose. The assigned tasks of the various DCNN are labeled accordingly in Figure 3. For example, as DCNN-1 was tasked with being the key solver for the involved convex optimization problems, it required a high degree of numerical stability, and PyTorch version 0.4.1 was selected; DCGAN-1 leveraged a “forward stable” TensorFlow-based DL Convolutional GAN (DCGAN) implementation to be able to well address the potentiality of mode collapse/mode failure (a phenomenon that may occur when adversarial GANs, which are being trained in tandem, are either unable to converge or undergo an anomalous convergence) [8].

C. *Experimental Construct*

With regards to the involved experimental construct, as can be seen in Figure 4, prior experimentation aspects used as presets are reflected in blue font while current experimental elements are shown in purple font. The “t-” elements (e.g., f-FCSO, t-SCSO, t-GSO, t-VBSO) of b-SHAP are extrapolations of Borgonovo’s work (previously discussed in Section IID) that more fully consider Kotthoff’s emphasis on temporal-sensitive/temporal-centric Shapley values [20]. STEA-related experimental forays for various OM were conducted. The OM of CRITIC was utilized as a preset for deriving the criteria weights, and the OMs of PROMETHEE, TOPSIS, and ELECTRE were utilized for the subsequent rankings. Initial selections and avoidances, among others, were based upon the following rationale. For example, PROMETHEE was known to be “easily… understood” and interpretable, so it was selected for testing [29][30]. Along this vein, [fuzzy] VIKOR was not selected, as it was known to be less interpretable and “less explainable than other more intuitive methods” [31].

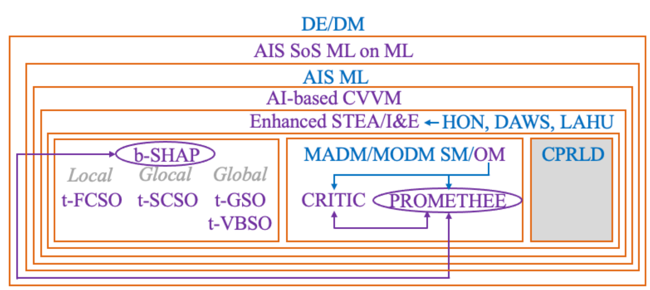

Figure 4. AI-based CVVM (ISA) Experimentation Aspects

Overall, selections were made to improve STEA/I&E. Yet, there were other technical considerations as well. A number of methodologies are subject to a phenomenon known as “Rank Reversal” (RR), wherein ranking results might change when the method changes or when the set of alternatives changes (leading to inconsistent and/or inaccurate results). The select OMs experimented with were known to be the most resistant to RR (yet are still subject to the phenomenon), and preliminary results are shown in Figure 5 below [32]. The key for the chart is as follows. First, the referenced “select OMs” of this Section IIIC are self-evident: ELECTRE, TOPSIS, and PROMETHEE. Second, these “select OMs” were benchmarked by execution time (E), sensitivity (S), performance under uncertainty (U), validity (V), and interpretability (I). Third, the aforementioned were benchmarked against classical SHAP (c-SHAP), as well as the b-SHAP approach described within this paper. Using the CPRLD as a preset,

Copyright (c) IARIA, 2025. ISBN: 978-1-68558-259-3

collectively, this forms the basis of the ISA described herein. The relative values were normalized against a scale of one to ten for ease of comparison.

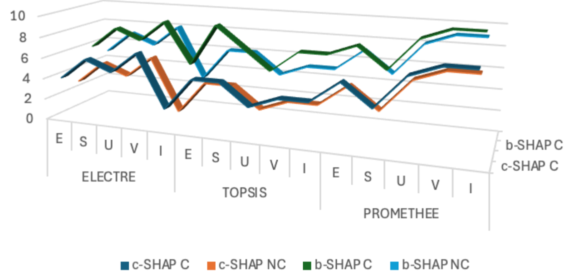

Figure 5. Preliminary Results from b-SHAP/select OM Benchmarking

The V and I were higher for PROMETHEE than for TOPSIS or ELECTRE. The E for TOPSIS was notably higher than that of the others, but the computational complexity is known to be less, and the performance under conditions of U was weaker than that of the others; the performance of PROMETHEE under conditions of U were seemingly better than ELECTRE and TOPSIS, in that order. Overall, the performance of b-SHAP was better than that of c-SHAP across the board for the range of E, S, U, V, I (for all the "select OMs" of ELECTRE, TOPSIS, and PROMETHEE). Hence, the b-SHAP-PROMETHEE amalgam (along with the CRITIC, CPRLD, etc. presets) exhibits promise.

IV. CONCLUSION

In consideration of Abraham Maslow's notion regarding the predilection that follows when there is only one tool to utilize, Section IIIC depicted some of the metrics underpinning the selection of a variety of methods and the comparative performance. For example, with regards to I&E, PROMETHEE was initially chosen over [fuzzy] VIKOR. As another example, PROMETHEE, TOPSIS, and ELECTRE were selected for testing, as they were reported to be more resistant to RR than certain other methods. As yet another example, Figure 5 depicted the relative performance of the methods for E, S, U, V, I; TOPSIS had a comparatively better E when E was considered in isolation, but it did not fare well under U, and along this vein, PROMETHEE did fare reasonably well under conditions of U when compared to ELECTRE and TOPSIS, etc. This brings us to the primary impetus of this paper, which centered upon enhancing robustness of the *testing* facet (with more granularity) at the interstices (e.g., *interstitial areas* between/among individual/amalgam component at the local, glocal, and global levels), better illuminating *I&E*/STEA DE/DM pathways, and operationalizing *AI-based CVVM* for the purposes of achieving higher efficacy AI SoS ML on ML for RWS. The hitherto lack of methodologies in this regard have led to RWS paradigms, wherein AIS adversely impact other AIS with the potentiality of cascading failure for the involved AI SoS (a.k.a., "near misses"). Moreover, the *testing* facet involves *performance metrics* conjoined with *construct validity* considerations. On the performance metrics front, OSONS was found to have greater efficacy than OSNS. Similarly, the b-SHAP (which involves various temporal-centric SHAP instantiations for local, glocal, and global) and PROMETHEE (along with CRITIC) amalgam was found to be more robust than the b-SHAP/TOPSIS or b-SHAP/ELECTRE amalgams on the *interpretability* front. Also on the performance front, *spawn reduction* turns out to be central, for once in the convex form, a myriad of SDF solvers can be leveraged to handle the involved optimization problems in polynomial time; otherwise, NP-hard spawn can congest matters with an indefinite impasse. The advancement of STEA/I&E necessarily involves HONs, DAWS, and LAHU, and these presets were discussed; the enhanced STEA/I&E discernment segues to more robust feature alignment, robust interpretation, etc., which constitutes enhanced *construct validity*. For this reason, it seems apropos to have the "Enhancement of an AI-based Construct Validity Approach" be the overarching descriptor of this paper. Future work will involve more quantitative and qualitative experimentation in the aforementioned areas.

REFERENCES

[1] D. Subhadra, et al., "Rise of Artificial Intelligence in Business and Industry," *J. of Inform. Educ. and Res.*, pp. 869-875, May 2024.

[2] J. McKendrick, "The Many, Many Ways AI Spurs Innovation," *Forbes*, Dec 2023. [Online]. Accessed: Mar. 1, 2025. Available: https://www.forbes.com/sites/joemckendrick/2023/12/10/ai-spurs-innovation-heres-how/.

[3] "Industrial AI Insights," *Honeywell*, Jul. 2024. [Online]. Accessed: Mar. 1, 2025. Available: https://www.honeywell.com/us/en/ai/research.

[4] D. Sjoberg and G. Bergersen, "Construct Validity in Software Engineering," *IEEE Trans. on Softw. Eng.*, vol. 49, pp. 1374-1396, Mar 2023.

[5] X. Zhou, Y Jin, H. Zhang, S. Li, and X. Huang, "A Map of Threats to Validity of Systematic Literature Reviews on Software Engineering," *23rd Asia-Pacific Softw. Eng. Conf. (APSEC)*, Apr 2017, pp. 153-160.

[6] S. Deets, C. Baulch, A. Obright, and D. Card, "Content Analysis, Construct Validity, and Artificial Intelligence: Implications for Technical and Professional Communication and Graduate Research Preparation," *J. of Bus. and Tech. Commun.*, vol. 38, pp. 303-315, Mar 2024.

[7] P. Hall, S. Ambati, and W. Phan, "Ideas on interpreting machine learning," *O'Reilly*, Mar. 1, 2017. [Online]. Accessed: Mar. 1, 2025. Available: https://www.oreilly.com/radar/ideas-on-interpreting-machine-learning/.

[8] S. Chan, "AI-Facilitated Dynamic Threshold-Tuning for a Maritime Domain Awareness Module," *IEEE Int. Conf. on Industry 4.0, Artif. Intell., and Commun. Technol. (IAICT)*, Aug. 2024, pp. 192-198.

[9] L. Grennan, A. Kremer, A. Singal, and P. Zipparo, "Why businesses need explainable AI – and how to deliver it," *McKinsey*, Sep. 2022. [Online]. Accessed: Mar. 1, 2025. Available:https://www.mckinsey.com/capabilities/quantumblack/our-insights/why-businesses-need-explainable-ai-and-how-to-deliver-it

[10] F. Scarelli, A. Tsoi, and M. Hagenbuchner, “The Vapnik-Chervonenkis dimension of graph and recursive neural networks,” *Neural Netw.*, vol. 108, pp. 248-259, Dec. 2018.

[11] G. Vilone and L. Longo, “Notions of explainability and evaluation approaches for explainable artificial intelligence,” *Inf. Fusion*, vol. 76, pp. 89-106, May 2021.

[12] S. Lundberg and S. Lee, “A Unified Approach to Interpreting Model Predictions,” *Proc. of the Adv. in Neural Inf. Proc. Syst.*, Dec. 2021, pp. 4765–4774.

[13] D. Balog, T. Bátyi, P. Csóka, and M. Pintér, “Properties and comparison of risk capital allocation methods,” *European J. of Oper. Res.*, vol. 259, pp. 614-625, 2017.

[14] M. Sundararajan and A. Najmi, “The Many Shapley Values for Model Explanation,” *Proc. of the 37th Int. Conf. on Mach. Learn.*, vol. 119, pp. 9269-9278, 2020.

[15] Y. Kwon, and J. Zou, “WeightedSHAP: analyzing and improving Shapley based feature attributions,” *Proc. of the 36th Int. Conf. on Neural Inf. Process. Syst.*, pp. 34363-34376, Oct 2022.

[16] L. Kotthoff, et al., “Quantifying Algorithmic Improvements Over Time,” *Proc. of the Twenty-Seventh Int. J. Conf. on Artif. Intell. (IJCAI-18)*,” pp. 5165-5171, Jul. 2018.

[17] R. Guidotti, et al., “A Survey of Methods for Explaining Black Box Models,” *ACM Comput. Surv. (CSUR)*, vol. 51, pp. 1-42, Aug. 2018.

[18] K. Dhamdhere, A. Agarwal, and M. Sundararajan, “The Shapley Taylor Interaction Index,” *Proc. of the 37th Int. Conf. on Mach. Learn.*, pp. 9259–9268, 2020.

[19] M. Grabisch and M. Roubens, “An axiomatic approach to the concept of interaction among players in cooperative games,” *Int. J. of Game Theor.*, vol. 28, pp. 547-565, Nov. 1999.

[20] E. Borgonovo, E. Plischke, and G. Rabitti, “The many Shapley values for explainable artificial intelligence: A sensitivity analysis perspective,” *European J. of Oper. Res.*, vol. 318, pp. 911-926, Nov. 2024.

[21] M. Mase, A. Owen, and B. Seilver, “Explaining black box decision by Shapley cohort refinement,” *Arxiv.org*, Oct. 2020. [Online]. Accessed: Mar. 1, 2025. Available: https://arxiv.org/pdf/1911.00467.

[22] A. Casjus, “The shapley value, the owen value, and the veil of ignorance,” *Int. Game Theor. Rev*., vol. 11, pp. 453-457, Dec. 2009.

[23] S. Lopez, “On the relationship between Shapley and Owen values,” *Central European J. of Oper. Res*., vol. 17, pp. 415-423, Dec. 2009.

[24] S. Beal, M. Diss, and R. Takeng, “New axiomatisations of the Diversity Owen and Shapley values,” *CRESE Working Paper 2024-09*, Feb. 2024. [Online]. Accessed: Mar. 1, 2025. Available: https://hal.science/hal-04502031v1/document.

[25] Z. Wu, “Evaluation of Provincial Economic Resilience in China Based on the TOPSIS-XGBoost-SHAP Model,” *J. of Math.*, pp. 1-12, Oct. 2023.

[26] X. Wang and Z. Piao, “An interpretable dynamic risk assessment approach in insurance: integrating TOPSIS, GM (1,1) and SHAP,” *Proc. of the 2024 Guangdong-Hong Kong-Macao Greater Bay Area Int. Conf. on Digit. Econ. and Artif. Intell. (DEAI)*, pp. 697-701, Jul. 2024.

[27] Z. Hua and Xiaochuan Jin, “A generalized Shapley index-based interval-valued Pythagorean fuzzy PROMETHEE method for group decision-making,” *Soft Comput.*, vol. 27, pp. 6629-6652, Feb. 023.

[28] S. Chan, “Mitigation Factors for Multi-domain Resilient Networked Distributed Tessellation Communications,” *Fifth Int. Conf. on Cyber-Technol. and Cyber-Syst. (CYBER 2020)*, Aug. 2020, p. 66-73.

[29] M. Brans and B. Mareschal, “Promethee Methods,” *Multiple Criteria Decision Analysis: State of the Art Surveys*, pp. 163-186, Jan. 2005.

[30] J. Brans and P. Vincke, “A Preference Ranking Organisation Method: (The PROMETHEE Method for Multiple Criteria Decision-Making), *Manage. Sci*., vol. 31, pp. 647-656, Jun. 1985.

[31] Y. Lin and T. Chen, “Type-II fuzzy approach with explainable artificial intelligence for nature-based leisure travel destination selection amid the COVID-19 pandemic,” *Digit. Health,* vol. 8., pp. 1-15, Jun. 2022.

[32] M. Aazadfallah, “A New Feature of Rank Reversal in Some of MADM Models,” *J. of Appl. Inf. Sci*., vol. 4, pp. 1-11, 2015.