Fugacity Phase Transition and Hyper-Heuristic Convergence for AI-centric Conceptual Estimating

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

***Abstract*—Levels of effort and timetable posits for the development and operationalization of System Transparency, Explainability, and Accountability (STEA)-centric Artificial Intelligence (AI) Systems (AIS) are beset by underestimation in often overlooked areas, such as the "Optimizing" facet of the "Deploying and Optimizing" phase of the AI Development Life Cycle, among others. This is a high derailment factor in conceptual estimating, particularly for those mission-critical AIS that do not well consider biases stemming from the broader Socio-Technical System (STS), which impact Interpretability & Explainability (I&E). In furtherance of bias mitigation and AIS whitening — STS-STEA-I&E (SSI) — an amalgam construct for facilitating/discerning a Fugacity Phase Transition (FPT) and Hyper-Heuristics (HH) convergence, segueing to an enhanced SSI contribution, is delineated.**

Keywords-AI Development Life Cycle; interpretability; explainability; justification logic; decision engineering.

I. INTRODUCTION

The development and deployment of Artificial Intelligence (AI) Systems (AIS) is on the rise, and the rapid growth in market size for AIS and related supply chains have been abundantly memorialized; Compound Annual Growth Rates (CAGR), such as 36.6% from 2024 to 2030, have been reported [1]. Cisco's 2024 AI Readiness Index asserts that "nearly all companies (98%) report that the urgency to deploy AI has increased in the last year" [2]. Rough Order of Magnitude (ROM) cost estimates for Levels of Effort (LOEs) and the associated timetables for the design/development and operationalization of these AIS are being requested in a torrential fashion to keep pace with the escalating demand/adoption rate [3][4]. This is buttressed by Stanford University's AI Index Report 2024, which notes a dramatic increase of interest in GitHub AI projects (more than doubling between 2022 to 2023) [5]. Simply, AIS are in high demand.

Despite the $184 billion market size for AI as of November 2024, the anticipated $826 billion market size by 2030, and the rising price tags for AIS deployments, conceptual estimating (e.g., positing ROMs prior to the substantial completion of the involved architecture/design) has not yet become sufficiently mature and/or robust; these ROMs are often far off target with a plethora of cost/schedule overruns and project failures populating the landscape [6][7][8][9]. Generally speaking, cost estimates are typically predicated upon the historical costs of successfully completed projects, and since the corpus of historical data is still quite limited in this arena, a myriad of conceptual estimating and cost estimator issues have arisen; ROMs are often erroneous.

To aggravate matters, not all AIS are equal. By way of example, a number of the earlier AIS had been withdrawn from the market due to their problematic "black box" architectures and prospective biases (e.g., algorithmic), which had not been well accounted for during their architectural/design phases [10]. Since that time, the AI ecosystem has progressively moved toward a paradigm of System Transparency, Explainability, and Accountability (STEA) for the prototypical stages/phases of the AI Development Life Cycle (ADLC) (as pertains to the development and operationalization of an AIS). The number of phases varies depending upon organizational preference and model selection — e.g., 3, 5, 6, 8, etc.; for simplicity, 3 phases will be considered herein; of the 3 basic phases — (1) Planning & Collection, (2) Designing & Training, and (3) Deploying & Optimizing — the "Optimizing" facet (a substantive contributor towards the success of the AIS) of (3) constitutes a formidable STEA challenge. *Without careful consideration, the STEA treatment for "Optimizing" can dramatically increase the required LOEs and potentially derail any posited ADLC timetable for the STEA-centric AIS.*

Yet, without even considering the STEA complexities and requisite mitigations against biases stemming from the larger Socio-Technical System (STS) rubric, which includes the ecosystem of "humans, technology, and the environs," there are a variety of staggering statistics to consider: (1) the Project Management Institute has reported that "almost half of business projects fall behind schedule, and up to a third are not completed at all," (2) a Boston Consulting Group (BCG) survey reports that "nearly half of all respondents said that more than 30% of their organization's technology development projects were over budget and late," (3) McKinsey & Company (McK), in collaboration with the [BT Group plc, formerly British Telecom] BT Centre for Major Programme Management at the University of Oxford, reports that "on average, large [Information Technology] IT projects run 45 percent over budget and 7 percent over time, while delivering 56 percent less value than predicted" while McKinsey further reports that "software projects run the highest risk of cost and schedule overruns," (4) [Research & Development] RAND Corporation notes that, "by some estimates, more than 80 percent of AI projects fail — twice the rate of failure for information technology projects," and

(5), The Computing Technology Industry Association (CompTIA) notes that “nearly 80% of the AI projects typically don’t scale beyond a [Proof of Concept] PoC or lab environment” [11]-[16]. Against this backdrop, when the complexities of IT/AI projects are conjoined with the cited STEA and STS complexities, it becomes clear that the devising of a robust STS/STEA-centric AIS architecture is non-trivial. Accordingly, four central aspects, among others, need to be well considered for an STEA-centric AIS architecture prior to providing a ROM.

The first is the desired level of transparency. The literature describes the principal variations in AIS architecture — “black-box,” “gray-box,” and “white-box — as being distinguished by gradations in transparency (most opaque to most transparent). The second is the desired level of interpretability, which centers upon the AIS’s Decision Engineering/Decision-Making (DE/DM) processes. The third is the desired level of explainability, which centers upon the rationale/underlying logic employed to arrive at the, hopefully, non-biased and reasonable outcomes [17]; the University of Toronto’s Schwartz Reisman Institute for Technology & Society and others further distinguish between Explainable AI (which centers upon “fact”) and Justifiable AI (a.k.a., justifiability) (which centers upon “judgment”) [18]. The fourth is the degree of accuracy (ACC) desired. For the second and third aspects, *interpretability* describes *how* the AIS formulates certain posits (e.g., the DE/DM processes), and *explainability* describes *why* the AIS made certain posits (e.g., the justification logic). These (i.e., Interpretability and Explainability) are often referred to as I&E, and along with the fourth aspect, there is an ongoing dialectic in the literature regarding the trade-off between *ACC* and *I&E*. Some argue that reduced ACC AIS are more readily interpreted; along this vein, some argue that enhanced ACC AIS are less able to be interpreted in an intuitive fashion [19][20]. A similar argument has been made regarding explainability [21][22]. Amidst this backdrop, researchers have endeavored to achieve high-performance AIS that still have high I&E [23]. Suffice it to say, this arena constitutes a challenging study space.

In the interim, research forays have trended towards more transparent white-box (a.k.a., glass-box) architectures, which reputedly have better I&E-by-design [24]. However, the performance tends to, as reported by some, lag behind the more translucent/opaque black-box architectures [25]. Accordingly, researchers have actively investigated the feasibility of middle-ground gray-box architectures. Along the vein of the previously discussed AIS project cost/schedule overruns, the initial *development time* for a high-performance **S**TS/**S**TEA/**I**&E (SSI)-centric AIS architecture can vary greatly (e.g., from months to years), and Gartner notes that, generally, “organizations” take about “7 months to develop AI initiatives, with 47% of the surveyed companies taking between 6 to 24 months from prototype to production” [12][26][27]; some AIS implementers assert that SSI-centric AIS architectures can take several years to devise and realize. The *testing times* can also vary greatly. Generally speaking, black-box testing can require less time than white-box testing since the latter would require additional LOEs (i.e., an increased amount of time) to comprehend the DE/DM pathways and logic employed. The AIS *model training time*/cost is also highly variable, as the training data needs to be refreshed in an ongoing fashion, particularly for Real World Scenario (RWS) AIS applications. With regards to the “Optimizing” facet of (3) of the ADLC, the degree of ACC (versus I&E) needs to be specified, and the various involved optimizations (e.g., pertaining to the involved computational resources, quantity/quality of the training data, heuristics/algorithms employed, tuning/fine-tuning efficacy for [e.g., Deep Neural Network or DNN] weights/hyperparameters, complexity of the AIS model and AIS architecture/design, etc.) is central. Stanford University and Epoch AI (a multidisciplinary research institute that investigates the arc of AI) reviewed AI model training cloud compute times/costs, and MIT Technology Review noted that “the process used to build most of the… [AI] models we use today can’t tell if they will work in” RWS, “and that’s a problem” [28][29][30]. Some AIS implementers argue that the greater the desired level of SSI, the “more time-consuming and resource-intensive” the processes can be — with an ensuing increase to Capital Expenditures (CAPEX). Over time, the seeming CAPEX advantage of “black-box” over “white-box” architectures may potentially be offset by ever-escalating Operational Expenditures (OPEX) related to brittleness and obsolescence issues (e.g., undetected issues, such as data drift may result in dramatic performance degradation) that often beset black-box architectures; in other words, the downstream OPEX-related disadvantages may offset the initial CAPEX advantages of the earlier developmental and testing phases. Gray-box architectures seem to constitute a middle-ground.

Certain SSI challenges that beset the ADLC are illuminated within this paper, such as at the “Optimization” facet of (3) of the ADLC. To assist the reader, a table of acronyms is provided in Table I below.

TABLE I. TABLE OF ACRONYMS

Acronym	*Full Form*
ACC	Accuracy
ACM	Association for Computing Machinery
AdapHH	Adaptive selection Hyper-Heuristics
ADLC	AI Development Life Cycle
AI	Artificial Intelligence
AIS	Artificial Intelligence System
ALGB-WG	Algorithmic Bias Working Group
BCG	Boston Consulting Group
CAGR	Compound Annual Growth Rate
CAPEX	Capital Expenditure
CompTIA	Computing Technology Industry Association
CRITIC	CRiteria Importance through Intercriteria Correlation
CWA	Connection Weights Algorithm
DE	Decision Engineering
DM	Decision-Making
DNN	Deep Neural Network
EO	Expert Opinion
FPT	Fugacity Phase Transition
GA	Garson’s Algorithm
GI	Gini Importance
HH	Hyper-Heuristic
HH-CF	Choice-Function-based Hyper-Heuristic
HH-R	Reward-based Hyper-Heuristic
HH-SF	Statistical Frequency-based Hyper-Heuristic

I&E	Interpretability & Explainability
IEEE	Institute of Electrical and Electronics Engineers
LOE	Level of Effort
MADM	Multi-Attribute Decision-Making
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MCDM	Multi-Criteria Decision-Making
McK	McKinsey & Company
MDA	Mean Decrease in Accuracy
MDI	Mean Decrease in Impurity
MLR	Multiple Linear Regression
MODM	Multi-Objective Decision-Making
NIST	National Institute of Standards and Technology
OA	Olden's Algorithm
OM	Objective Measure
OPEX	Operational Expenditure
OPH	Operator/Procedure/Heuristic
PLSR	Partial Least Squares Regression
POC	Proof of Concept
PROMETHEE	Preference Ranking Organization Method for Enrichment Evaluation
PSO	Particle Swarm Optimization
QoS	Quality of Service
QR	Quantile Regression
RAND Corp.	Research & Development Corporation
RMSE	Root Mean Square Error
ROM	Rough Order of Magnitude
RR	Ridge Regression
RWS	Real World Scenario
SM	Subjective Measure
SSHH	Sequence-based Selection Hyper-Heuristic
SSI	Socio-Technical System-System Transparency, Explainability, and Accountability-Interpretability & Explainability
STEA	System Transparency, Explainability, and Accountability
STS	Socio-Technical System
TransE	Translating Embeddings
WIP	Work-in-Progress
XAI	Explainability in AI

Section I delineates the impetus of the paper — *the illumination and consideration of certain SSI-related derailment facets that may dramatically increase the required LOEs and potentially derail posited ADLC timetables for the SSI-centric AIS.* Section II provides pertinent background information regarding: (1) **S**TS/**S**TEA (in general) and **I**&E (in particular) (collectively, "SSI") for certain facets of the ADLC for an AIS, (2) the "Fugacity Phase Transition" (FPT) (e.g., the series of deviations between the "ideal" training data and the "actual" observed data), and (3) certain other key considerations (e.g., the SSI aspects of the utilized Hyper-Heuristics or HH) that are critical to consider prior to putting forth conceptual estimating ROMs for an SSI-centric AIS. Section III delineates the presets & theoretical foundations as well as benchmarking & insights related to the involved HH/FPT experimentation. Section IV concludes and presents some prospective future work.

II. Background

A. The Import of SSI for AIS

In these contemporary times, there is a heightened expectation for SSI-centric AIS, particularly with regards to I&E. Winfield and others have remarked on various STEA-centric Work-in-Progress (WIP) standards, as well as actual standards that have buttressed the Explainability in AI (XAI) movement; these WIPs/standards include, among others, the U.S. National Institute of Standards and Technology (NIST) Special Publication 1270 "Towards a Standard for Identifying and Managing Bias in Artificial Intelligence," the Association for Computing Machinery (ACM) "Principles for Algorithmic Transparency and Accountability," and the Institute of Electrical and Electronics Engineers (IEEE) Standard for Transparency of Autonomous Systems (P7001), among others. There are also a range of engaged working groups, such as the IEEE Algorithmic Bias Working Group (ALGB-WG) (P7003). On the topic of bias, NIST has opined that certain AI biases (e.g., "human biases and systemic, institutional biases as well") may stem from the larger STS rubric [31]. This includes the involved corpus of data, which may, potentially, derive from problematic "facts," "assessment surveys," and other bias-related problems from the "Collection of Data" facet of (1) of the ADLC [32][33].

Traditionally, it has been opined that, for the ADLC, approximately "80% time" is spent on (1) [34]. For the "Collection of Data" facet of (1) of the ADLC, Westland has noted that the "bias and informativeness" of Subjective Measures (SMs) (e.g., Likert-type measurements) "have been the center of recent" dialectic [35][36]. From an SSI perspective, STS-related biases, such as from a variety of assessment data utilized as input to the AIS (e.g., from surveys) has recently been illuminated as a prospective Achilles heel for AIS. For example, McLeod informs us that "prior research has shown that using Likert scales can be problematic," via a variety of biases (e.g., "social desirability bias, acquiescence bias," central tendency bias, etc.) [37]. Taherdoost affirms this by noting that Likert "scale validity may be difficult to demonstrate[,] and there is a lack of reproducibility" [38]. To further underscore the aforementioned, Louangrath's experimentation reports on the higher reliability levels of non-Likert scales (e.g., "92%") over Likert-type scales (e.g., "90, 89, and 88% reliability") as well as higher validity levels of non-Likert scales (e.g., "93%") over Likert-type scales (e.g., "89, 61, and 57%") [39]. Hence, *the formulation/implementation of enhanced assessments (e.g., STS-related surveys) for the "Collection of Data" facet of (1) of the ADLC, which is a key part of the STS rubric, will likely increase the time needed for formulating and instantiating SSI-centric AIS architectures.*

B. The Fugacity Phase Transition (FPT) between Phases (2) and (3) of the ADLC

With regards to the AIS architecture's DE/DM apparatus, Fattoruso depicts Multi-Criteria Decision-Making (MCDM) as being comprised of Multi-Attribute Decision-Making (MADM) and Multi-Objective Decision-Making (MODM) [40]. Generally speaking, while MODM concurrently addresses a range of objectives ("and endeavors to determine an optimal solution set among "undetermined continuous alternatives"), MADM addresses a single objective and "organizes/sorts/ranks" (in the endeavor to ascertain the optimal solution among "a finite set of discrete alternatives") [41]. For the "Collection of Data" facet of (1) of the ADLC, a more robustly counterpoised MADM/MODM

SM/Objective Measures (OM) construct is crucial for facilitating SSI robustness, as it can better contend with the issue of AIS model drift (a.k.a., model decay) (i.e., shifts in the involved data/relationships that can result in AIS model performance degradation, wherein the posits become increasingly less effective), particularly in situations for which the RWS data encountered is far different "from the data it was trained to recognize or handle" [42]. Generally speaking, it can be easier to discern this drift within a higher SSI-centric than a lower SSI-centric AIS architecture. To assist in contextualizing/delineating this paradigm, the term "fugacity" (an apropos term utilized by Dreyfus-Schmidt-DuPhan-Desfontaines that nicely references the "tendency… to escape from one phase to another") is utilized; "'fugacity' measures the difference between the expected 'ideal' data… [that the AIS] model was trained on and the observed 'real' data" that the AIS model encounters (i.e., the distinction between the "reference distribution" and the "prediction distribution") [43][44][45]. The indicators of low drift and low fugacity can be utilized in ascertaining when a transitioning from phase (2) to (3) of the ADLC (i.e., FPT) is prudent. It should be noted that the FPT is not a singular punctuating event/milestone; rather, it denotes a fairly steady-state paradigm, wherein the fugacities for the successive states of dynamically updated AIS heuristics (acting in conjunction with the involved AIS algorithms) are low enough to be of satisfactory utility for the involved RWS AIS application. *The monitoring of the involved AIS model (and encompassing AIS architecture) will require a sufficient temporal span given the SSI-centric AIS architectural requirement.*

C. *Potential ADLC pitfalls (e.g., HH) and I&E Robustness for the "Optimization" Facet of the ADLC*

As alluded to in Section I and Section IIA, the requisite time to develop a sufficiently robust performance SSI-centric AIS architecture can vary greatly. It consists of the (1), (2), and (3) phases referenced in Section I, as well as various facets, such as that of "Optimization." Within the phases of a 3-phase ADLC, the *"Collection of Data"* and "*Training/Inferencing"* (e.g., which might be subject to the prospective inversion of the classical training:inferencing ratios) facets, as discussed in [46], are noteworthy, for they need to be well considered prior to positing LOEs and their associated timetables for an SSI-centric AIS (i.e., conceptual estimating).

The counterpoising of SM and OM (for MADM and MODM), such as for the *"Collection of Data"* facet of (1) of the ADLC is non-trivial. This is further complicated with the need to appropriately weight and "organize/sort/rank," which may be accomplished via the utilization of various OM combinatorials; this includes the leveraging of OM methods, such as the CRiteria Importance through Intercriteria Correlation (CRITIC) OM for the ascertainment of apropos weights and the Preference Ranking Organization Method for Enrichment Evaluation (PROMETHEE) OM for the ensuing ranking. *The apropos selection and testing of more SSI-oriented OM (as well as SM) combinatorials will also likely increase the time needed for SSI-centric AIS architectures.*

Moreover, there is a fundamental distinction between the paradigm of static weights and that of dynamically updated weights. The literature is abundant with regards to the criticality of a dynamic weighting strategy for the "*Training/Inferencing"* facet [47]. Along this vein, oftentimes, heuristic approaches are leveraged to complement algorithmic approaches, particularly for RWS AIS applications. After all, the amalgam of heuristics and algorithms lend to numerical methods implementations of higher efficacy, and a dynamically updated heuristic model lends to more optimal convergence for a "better-fit" or "best-fit" approximation, etc. (e.g., the robust convex relaxation discussed in [48]). *The SSI-related issue is that while algorithms have received increasing SSI attention, the myriad of static/brittle heuristics populating the AIS landscape has not received comparable SSI attention; this is an area that can increase the ADLC time needed.*

Beyond the *"Collection of Data"* and the "*Training/Inferencing"* facets, the *"Optimization"* facet of the ADLC (e.g., optimizing the involved AIS model) is critical, for it facilitates more accurate and efficient predictions, which segues to enhanced performance, decreased OPEX, and higher practicality/applicability for RWS. In particular, optimization (e.g., such as with regards to AIS model size, complexity, etc.) can facilitate more rapid inferencing with less computational resources (e.g., energy consumption) and lend toward scalability (e.g., optimized AIS models are more readily deployed). By way of context, the heuristic problem-solving approach is geared for ascertaining a "good enough" solution within a bounded period of time, but there is no certainty that it will provide an optimal solution; in contrast, certain algorithmic approaches are favored for ascertaining an optimal solution, but the "runtimes" may vary greatly. To date, "research in the explainability of optimisation techniques has largely focused on meta-heuristics" (which "directly search the solution space of a problem") [49][50]. There has been far less research on HH (higher-level Operator/Procedure/Heuristic (OPH) methods that "operate on a search space of low[er]-level heuristics…rather than solutions directly"), which can pose herculean SSI challenges due to the use of a plethora of lower-level OPHs, which complicates matters [51]. There have been some notable SSI-related explorations that have shown promise with regards to SSI, such as Misir's Adaptive selection HH (AdapHH) and Drake's Sequence-based Selection HH (SSHH) (which leverages probability matrices to facilitate I&E) [49][51]. By leveraging these lessons learned as well as the presets delineated in Section IIIA, a more SSI-centric HH paradigm can be leveraged.

 ISBN: 978-1-68558-259-3

III. Experimentation

A. Presets & Theoretical Foundations

Ali, Piccialli, and others have noted that a substantive portion of AI researchers opine that "a deeper network is better for decision-making than a shallow network" [24]. Yet, the prototypical DNNs are increasingly more difficult to examine at the deeper layers given the increasingly complex patterns/abstractness (as contrasted to the more straightforward patterns residing at the more shallow/earlier layers). From an SSI perspective, this makes certain reportage of DNN usage for mission-critical HH of even greater import [52]. For the experimentation herein, a preset (i.e., a precursor experimental construct) leveraged was in the form of an RWS-oriented Particle Swarm Optimization (PSO)-based Meta-Heuristic approach, as depicted in [53]. Another preset centered upon the selection of MODM OMs, such as CRITIC and PROMETHEE as well as those delineated in [54]. These presets are reflected in Figure 1 in bright red; the critical counterpoisings shown in lavender are supported by these presets. The focus of the experimentation is at the "Optimization" nexus of FPT/HH (denoted in brick red).

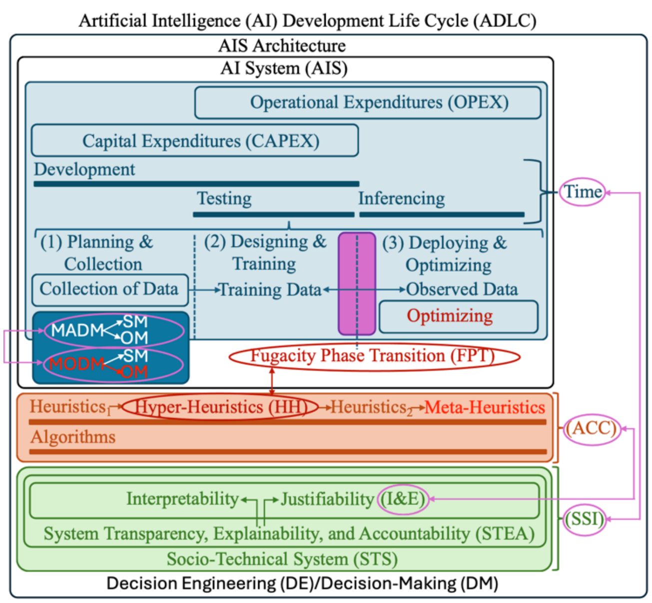

Figure 1. ADLC with the experimental focus, as indicated in brick red.

Drake asserts that there are two types of HHs: (1) selection HH that select a sequence of Low[er]-Level Heuristics (LLHs), and (2) generation HHs that spawn LLHs [51]. From an SSI perspective, the degree of I&E depends upon the involved mechanism; for example, Maashi's Choice-Function-based HH (HH-CF) facilitates the examination of LLHs, as LLHs are designated with a score/normalized score based upon prior performance and chosen accordingly. Qu's Statistical Frequency-based HH (HH-SF) can reveal LLH sequences that relate to the more optimal solutions (thereby making I&E more self-evident). Kheiri's Reward-based HH (HH-R) leverages LLH usage and transitions among LLH to yield transition probabilities for enhanced I&E [49].

B. Benchmarking & Insights

For the purposes herein, operators will be construed as: (1) diversification, (2) intensification, and (3) perturbation, Typically, (1) will leverage randomness to induce a substantive variation (e.g., to avoid stagnation at local optima) to expand the search space (e.g., progress to unexplored areas), (2) will spawn solution variations in high potential areas of the search space, and (3) will induce minute variations (e.g., to facilitate the gauging of LLH performance). In some cases, the sequencing of (1), (2), and (3) is effective; in other cases, (3), (1), and (2) may have efficacy. Our experimentation finds that the (1), (2), (3), (2) sequence has high efficacy; our findings are consistent with Drake's reportage that LLHs/LLH sequences "which are ineffective at the start of the search process prove to be highly effective at the end, and vice versa" [51]. In essence, the efficacy of LLHs/LLH sequences and their concomitant HHs need to be gauged *over time*. For this temporal consideration, the assessment of the LLHs/HHs also needs to include consideration of the long-tail (part of the [statistical distribution], which is far afield from the head and centroid) phenomena prevalent in RWS; Samuel reports that "strongly unbalanced data with a long-tail is ubiquitous in numerous domains and problems" and "learning [*over time*] with unbalanced data causes models to favor head classes" [55][56]. Various techniques (e.g., based upon Wang's Translating Embeddings or TransE) for better balancing across both head and tail classes are discussed in [57]. There is also the matter of AIS model drift *over time*. Along this vein, HHs can be leveraged to avoid a high drift paradigm (i.e., a drift score closer to 1), such as for the case where the features underlying the AIS model drift are of low significance; HHs can also be leveraged to lower the drift paradigm (e.g., moving the drift score closer to 0) by recognizing features of high significance, whose removal would dramatically degrade the AIS model performance. Interestingly, the challenge of feature significance determination centers upon the fact that features are not independent; actually, a substantive portion of features are highly correlated (a.k.a., collinear features). Spearman's and Pearson's correlation [coefficient] (R) can be used to gauge collinearity (e.g., a high R indicates collinearity), and given the plethora of collinear features, the notion of feature families becomes quite useful. Given a high R^2 (a value closer to 1, which implies a perfect fit), wherein R^2 = 1 – Sum Squares of Error or SSE/Total Sum of Squares or SST, the removal of a high dependency feature will likely not have a significant impact upon ACC for the feature family; on the other hand, a lower Root Mean Square Error (RMSE) (square root of the average squared differences between the measured values and actual values) and Mean Absolute Error (MAE) (average of the absolute differences) implies a

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

better fit. As Matel notes, "the larger the drop in R^2 when a variable [/feature] is removed…, the more important it is assumed to be" [58]. This is affirmed by Gini Importance (GI), Mean Decrease in Impurity (MDI), and Mean Decrease in Accuracy (MDA) (a higher GI, MDI, and MDA indicates higher variable/feature significance). In essence, the involved RWS AIS evaluation was conducted *over time* (i.e., the FPT).

Matel's experimentation was utilized for benchmarking purposes, as Matel had reported that his conceptual estimating model exhibited "a 14.5% improvement in the accuracy" over Hyari's model when considering Mean Absolute Percentage Error (MAPE) [58]. Matel's findings are as follows: (1) for the Connection Weights Algorithm (CWA), "the lowest MAPE with all 16 variables was 50.36%," but the MAPE dropped "to 27.41%" "when only the top 5 variables were used," (2) for Multiple Linear Regression (MLR), "when [only] the top 5 to 7 variables" were used, the MAPE was "42.47%," and (3) for Expert Opinion (EO), when only "the top 5 variables" were used, the "MAPE was 93.25%" [58]. Hence, in terms of efficacy, CWA >> MLR >>> EO; this should be no surprise, for while CWA can accommodate non-linear relationships, MLR is not able to. For the case herein, Matel's experimentation was reiterated with HH utilized for determining the top variable/features, and the results were somewhat comparable. The results are shown in Figure 2, which also incorporates Garson's Algorithm (GA) and Olden's Algorithm (OA) as alternatives to CWA as well as Partial Least Squares Regression (PLSR), Quantile Regression (QR), and Ridge Regression (RR) as alternatives to MLR.

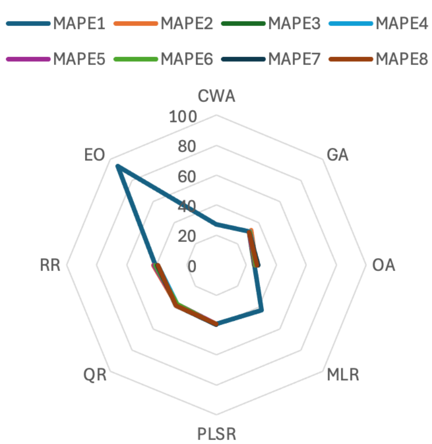

Figure 2. Benchmarking of Section III with Matel's Experimentation

CWA tends to outperform GA, and OA (as an implementation of CWA) is more nuanced than the plain vanilla CWA. PLSR is better suited for multi-collinearity than MLR, and QR can better handle outliers than MLR. Apart from that, the principal distinction was that of a steady-state convergence that was obtained with the amalgam of: (1) low drift, (2) low RMSE and MAE reflecting low fugacities/a more narrow FPT, (3) high GI, MDI, and MDA affirming variable/feature significance, (4) high R (reflecting collinearity) and a high R^2 (wherein the removal of high dependency features did not have a substantive ACC impact), and (5) high efficacy HH ascertainment at 8 variables/features. This logical progression through the amalgam composition and FPT/HH convergence should make clear the FPT/HH SSI contribution.

IV. Conclusion

The use of heuristics, to assist with algorithmic convergence for RWS AIS applications, is on the rise. These applications are likely to have specific stringent RWS timing requirements (e.g., pursuant to the involved Quality of Service or QoS). The adherence to these stringent RWS timing requirements constitutes a key facet of why the dynamically updated heuristic model (e.g., via HH) tangibly contributes towards the utility/practicality expected for RWS applications. *Hence, HHs become critical to the equation, and their SSI orientation becomes central;* it should be noted that HH has gained traction "in addressing NP-hard optimisation problems because it generalises well across problem domains" [59]. This paper presented an FPT/HH convergence approach (i.e., low drift, narrow FPT, and high efficacy HH) that would lend to a more SSI-centric optimization facet of the ADLC; accordingly, conceptual estimating and cost estimator ROMs can be made more robust. To conclude, this paper explores the development and implementation of improved assessment methods, such as STS-oriented surveys, for the "Data Collection" process within ADLC, a key component of the STS framework. The study highlights how these enhancements may impact the creation and deployment timelines of AIS architectures focused on SSI. By refining evaluation approaches, the research aims to improve the efficiency and effectiveness of data-driven decision-making within STS-based systems. Future work will involve more quantitative experimentation and benchmarking.

References

[1] "Global Artificial Intelligence Market Size & Outlook," *Grand View Research*. [Online]. Accessed: Mar. 1, 2025. Available: https://www.grandviewresearch.com/horizon/outlook/artificial-intelligence-market-size/global.

[2] "Cisco's 2024 AI Readiness Index: Urgency Rises, Readiness Falls," *Cisco*, Nov. 2024. [Online]. Accessed: Mar. 1, 2025. Available: https://newsroom.cisco.com/c/r/newsroom/en/us/a/y2024/m11/cisco-2024-ai-readiness-index-urgency-rises-readiness-falls.html#:~:text=Nearly%20all%20companies%20(98%25),from%2014%25%20a%20year%20ago.

[3] "Artificial Intelligence's Use and Rapid Growth Highlight Its Possibilities and Perils," *U.S. Gov. Accountability Office*, Sep. 2023. [Online]. Accessed: Mar. 1, 2025. Available: https://www.gao.gov/blog/artificial-intelligences-use-and-rapid-growth-highlight-its-possibilities-and-perils.

[4] "Groundbreaking Framework for the Safe and Secure Deployment of AI in Critical Infrastructure Unveiled by Department of Homeland Security," *U.S. Department of*

Homeland Security, Nov. 2024. [Online]. Accessed: Mar. 1, 2025. Available from: https://www.dhs.gov/news/2024/11/14/groundbreaking-framework-safe-and-secure-deployment-ai-critical-infrastructure#:~:text=These%20sectors%20are%20increasingly%20deploying,distribute%20mail%20to%20American%20households.

[5] "Artificial Intelligence Index Report 2024," *Stanford University*. [Online]. Accessed: Mar. 1, 2025. Available: https://aiindex.stanford.edu/wp-content/uploads/ 2024/05/HAI_AI-Index-Report-2024.pdf.

[6] "Artificial Intelligence – Worldwide," *Statista*, Mar. 2024. [Online]. Accessed: Mar. 1, 2025. Available from: https://www.statista.com/outlook/tmo/artificial-intelligence/worldwide.

[7] "Artificial Intelligence (AI) market size worldwide from 2020 to 2030," *Statista*, Mar. 2024. [Online]. Accessed: Mar. 1, 2025. Available: https://www.statista.com/outlook/tmo/artificial-intelligence/worldwide# market-size.

[8] B. Cottier, R. Rahman, L. Fattorini, N. Maslej, and D. Owen, "The rising costs of training frontier AI models," *arXiv*, May 2024. [Online]. Accessed: Mar. 1, 2025. Available: https://arxiv.org/abs/2405.21015.

[9] "AI exposes challenges to cost estimation," Acquisition Talk, Jul. 2019. [Online]. Accessed: Mar. 1, 2025. Available from: https://acquisitiontalk.com/2019/07/ ai-exposes-challenges-to-cost-estimation/.

[10] K. Hao, "The two-year fight to stop Amazon from selling face recognition to the police," *MIT Technology Review*, Jun. 2020. [Online]. Accessed: Mar. 1, 2025. Available: https://www.technologyreview.com/2020/06/12/ 1003482/amazon-stopped-selling-police-face-recognition-fight/.

[11] B. Velden, "Explainable AI: Current Status and Future Potential," *Eur. Radiol.,* vol. 34, pp. 1187-1189, Aug. 2023.

[12] M. Lorko, M. Servatka, and L. Zhang, "A Better Way to Avoid Project Delays," *MIT Sloan Manage. Rev.*, Nov. 2024. [Online]. Accessed: Mar. 1, 2025. Available: https://sloanreview.mit.edu/article/a-better-way-to-avoid-project-delays/.

[13] S. Palumbo, B. Rehberg, and H. Li, "Software Projects Don't Have to Be Late, Costly, and Irrelevant," *Boston Consulting Group*, Apr. 2024. [Online]. Accessed: Mar. 1, 2025. Available: https://www.bcg.com/publications/2024/software-projects-dont-have-to-be-late-costly-and-irrelevant.

[14] M. Bloch, S. Blumberg, and J. Laartz, "Delivering large-scale IT projects on time, on budget, and on value," *McKinsey Digital*, Oct. 2012. [Online]. Accessed: Mar. 1, 2025. Available: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/delivering-large-scale-it-projects-on-time-on-budget-and-on-value.

[15] J. Ryseff, B. Bruhl, and S. Newberry, "The Root Causes of Failure for Artificial Intelligence Projects and How They Can Succeed," *RAND*, Aug. 2024. [Online]. Accessed: Mar. 1, 2025. Available: https://www.rand.org/pubs/research_reports/RRA2680-1.html.

[16] "Artificial Intelligence in Business: Top Considerations Before Implementing AI," *CompTia*. [Online]. Accessed: Mar. 1, 2025. Available: https://connect.comptia.org/content/guides/business-considerations-before-implementing-ai.

[17] "Logic & Reasoning for XAI," *The 2nd World Conf. on Explainable Artif. Intell.* [Online]. Accessed: Mar. 1, 2025. Available from: https://xaiworldconference.com/2024/logic-reasoning-for-xai/

[18] G. Hadfield, "Explanation and justification: AI decision-making, law, and the rights of citizens," *Schwartz Reisman Inst. for Technol. and Soc., Univ. of Toronto*, May 2021. [Online]. Accessed: Mar. 1, 2025. Available from: https://srinstitute.utoronto.ca/news/hadfield-justifiable-ai.

[19] B. Seo and J. Li, "Explainable machine learning by SEE-Net: closing the gap between interpretable models and DNNs," *Sci. Rep.*, vol. 14, pp. 1-16, Nov. 2024.

[20] W. Esser-Skala and N. Fortelny, "Reliable interpretability of biology-inspired deep neural network," *npj Systems Biology and Appl.*, vol. 9, pp. 1-8, Oct. 2023.

[21] A. Nussberger, L. Luo, L. Celis, and M. Crockett, "Public attitudes value interpretability but prioritize accuracy in Artificial Intelligence," *Nature Commun.*, vol. 13, pp. 1-13, Oct. 2022.

[22] V. Hassija, et al., "Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence," *Cognitive Computation*, vol. 16, pp. 45-74, Aug. 2024.

[23] F. Candelon, T. Evgeniou, and D. Martens, "AI Can Be Both Accurate and Transparent," *Harvard Bus. Rev.*, May 2023. [Online]. Accessed: Mar. 1, 2025. Available: https://hbr.org/2023/05/ai-can-be-both-accurate-and-transparent.

[24] S. Ali, et al., "Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence," *Inf. Fusion*, vol. 99, pp. 1-52, Nov. 2023.

[25] O. Loyola-Gonzalez, "Black-Box vs. White Box: Understanding Their Advantages and Weaknesses From a Pracitcal Point of View," *IEEE Access*, vol 7., pp. 154096-154113, Oct 2019.

[26] D. Lambert, "How To Architect and Deliver AI At Scale," *Enterprise Architecture Professional J. (EAPJ)*, Oct. 2023. [Online]. Accessed: Mar. 1, 2025. Available: https://eapj.org/how-to-architect-and-deliver-ai-at-scale/#:~:text=According%20to%20Gartner%20it%20takes,months%20from%20prototype%20to%20production.

[27] "Gartney Survey Finds Generative AI Is Now the Most Frequently Deployed AI Solution in Organizations," *Gartner*. May 2024. [Online]. Accessed: Mar. 1, 2025. Available: https://www.gartner.com/en/newsroom/press-releases/ 2024-05-07-gartner-survey-finds-generative-ai-is-now-the-most-frequently-deployed-ai-solution-in-organizations.

[28] "The Training Costs of AI Models Over Time," *Voronoi*, May 2024. [Online]. Accessed: Mar. 1, 2025. Available: https://www.voronoiapp.com/technology/The-Training-Costs-of-AI-Models-Over-Time-1334.

[29] *"How Much Does It Cost to Train Frontier AI Models," Epoch AI.* [Online]. Accessed: Mar. 1, 2025. Available from: https://epoch.ai/blog/how-much-does-it-cost-to-train-frontier-ai-models.

[30] W. Heaven, "The way we train AI is fundamentally flawed," *MIT Technology Review*, Nov. 2020. [Online]. Accessed: Mar. 1, 2025. Available: https://www.technologyreview.com/2020/11/18/1012234/training-machine-learning-broken-real-world-heath-nlp-computer-vision/.

[31] R. Schwartz, et al., "Towards a Standard for Identifying and Managing Bias in Artificial Intelligence," *NIST Special Publication 1270*, 2022.

[32] M. Gruet, " 'That's Just Common Sense'. USC researchers find bias in up to 38.6% of 'facts' used by AI," *USC Viterbi School of Engineering*, May 2022. [Online]. Accessed: Mar. 1, 2025. Available: https://viterbischool.usc.edu/news/2022/05/thats-just-common-sense-usc-researchers-find-bias-in-up-to-38-6-of-facts-used-by-ai/.

[33] E. Ntoutsi et al., "Bias in data-driven artificial intelligence systems-An introductory survey," *WIREs Data Mining and Knowledge Discovery*, vol. 10, pp. 1-14, May/June 2020.

[34] S. Sagiraju, "Managing the Data For The AI Lifecycle," *Forbes*, Apr. 2022. [Online]. Accessed: Mar. 1, 2025.

Available: https://www.forbes.com/councils/ forbestechcouncil/ 2022/04/20/managing-the-data-for-the-ai-lifecycle/

[35] B. Choi and A. Pak, “A Catalog of Biases in Questionnaires,” *Prev Chronic Dis*., vol. 2, pp. 1-13, Dec 2004.

[36] J. Westland, “Information loss and bias in likert survey responses,” *PLOS One,* vol. 17, pp. 1-17, Jul. 2022.

[37] A. McLeod, S. Pipp, and J. Wong, “Revisiting the Likert scale: can the fast form approach improve survey research?” *Int. J. of Behav. Accounting and Finance*, vol. 2, pp. 310-327, 2011.

[38] H. Taherdoost, “What Is the Best Response Scale for Survey and Questionnaire Design; Review of Different Lengths of Rating Scale/ Attitude Scale/Likert Scale,” *Int J. of Acad. Res. in Manage. (IJARM)*, vol. 8, pp.1-12, Mar. 2019.

[39] P. Louangrath, “Reliability and Validity of Survey Scales,” *Int. J. of Res. & Methodology in Social Sci*., vol. 4, p. 50-62, Mar. 2018.

[40] G. Fattoruso, “Multi-Criteria Decision Making in Production Fields: A Structured Content Analysis and Implications for Practice,” *J. Risk Financial Manag.*, vol. 15, pp. 1-21, Sep. 2022.

[41] S. Chan, “AI-Facilitated Dynamic Threshold-Tuning for a Maritime Domain Awareness Module,” *2024 IEEE Int. Conf. on Ind. 4.0, Artif. Intell., and Commun. Technol. (IAICT)*, Bali, Indonesia, Jul. 2024, pp. 192-198.

[42] S. Bigelow, “How to identify and manage AI model drift,” *TechTarget*, May 2024. [Online]. Accessed: Mar. 1, 2025. Available from: https://www.techtarget.com/ searchenterpriseai/tip/How-to-identify-and-manage-AI-model-drift

[43] “Model Drift Monitoring,” *Dataiku*. [Online]. Accessed: Mar. 1, 2025. Available: https://www.dataiku.com/product/plugins/model-drift/.

[44] M. Khan and M. Islam, “Sustainable Waste Management,” in *The Petroleum Eng. Handbook: Sustainable Oper.*, Gulf Publishing Company, Houston, Texas, 2007, pp. 135-188.

[45] “Dataiku 11.3 Is Hot Off the Presses,” *Dataiku*. [Online]. Accessed: Mar. 1, 2025. Available:https://community.dataiku.com/discussion/31960/new-year-new-features-dataiku-11-3-is-hot-off-the-presses.

[46] S. Chan, “Inference-Optimized Metaheuristic Approach for a Prospective AI Training/Inference Inversion Paradigm in Optimized Energy-Aware Computing,” *2024 IEEE World AI IoT Congr. (AIIoT)*, Seattle, WA, USA, Jul. 2024, pp. 370-379.

[47] L. Liu, B. Zhou, Z. Zhao, and Z. Liu, “Active Dynamic Weighting for multi-domain adaptation,” *Neural Netw.*, vol. 177, Sep. 2024.

[48] S. Chan, “Enhanced Robust Convex Relaxation Framework for Optimal Controllability of Certain Large Complex Networked Systems,” *The 2022 IARIA Annual Congr. on Frontiers in Sci., Technol., Services, and Appl.*, Nice, France, May 2022, pp. 87-96.

[49] W. Yates, E. Keedwell, and A. Kheiri, “Explainable Optimisation through Online and Offline Hyper-heuristics,” *ACM Trans. on Evol. Learn.*, pp. 1-29, Oct. 2024.

[50] M. Blocho, “Heuristics, metaheuristics, and hyperheuristics for rich vehicle routing problems,” *Smart Delivery Syst.*, Jan. 2020, pp. 101-156.

[51] J. H. Drake, E. Özcan, and E. K. Burke, “A modified choice function hyper-heuristic controlling unary and binary operators,” *2015 IEEE Congr. on Evolutionary Comput. (CEC)*, Sendai, Japan, May 2015, pp. 3389-3396.

[52] M. Panzer, B. Bender, and N. Gronau, “A deep reinforcement learning based hyper-heuristic for modular production control,” *Int. J. of Prod. Res.h*, pp. 2747-2768,, Jul. 2023.

[53] S. Chan, “Inference-Optimized Metaheuristic Approach for a Prospective AI Training/Inference Inversion Paradigm in Optimized Energy-Aware Computing,” *2024 IEEE World AI IoT Congr. (AIIoT)*, Seattle, WA, USA, Jul. 2024, pp. 370-379.

[54] S. Chan, “The Triumvirate of an Adaptive Criteria Weighting Methodology, Isomorphic Comparator Similarity Measure, and Apropos High Dimensional Data Cluster Validation Index Measures for the Ascertainment of Bespoke Dynamic Fuzzy Lists,” *The 16th Int. Conf. on Future Comput. Technol. and Appl. (Future Comput. 2024)*, Venice, Italy, Apr. 2024, pp. 8-17.

[55] D. Samuel, Y. Atzmon, G. Chechik, “From generalized zero-shot larning to long-tail with class descriptors,” *2021 IEEE Winter Conf. on Appl. of Comp. Vision (WACV)*, Waikoloa, Hawaii, 2021, pp. 286-295.

[56] M. Buda, A. Maki, and M. Mazurowski, “A systematic study of the class imbalance problem in convolutional neural networks,” *Neural Netw*., vol. 106, pp. 249-259, Oct 2018.

[57] S. Chan, “Cascading Succession of Models for an Enhanced Long-Tail Discernment AI System,” *2024 IEEE World AI IoT Congr. (AIIoT)*, Seattle, WA, USA, 2024, pp. 393-401.

[58] E. Matel, “An artificial neural network apporach for cost estimation of engineering services,” *Int. J. of Construction Manage*., vol. 22, pp. 1274-1287, Nov. 2019.

[59] E. Burke, M. Hyde, G. Kendall, G. Ochoa, E. Ozcan, and J. Woodward, “A Classification of Hyper-heuristic Approaches,” in *Handbook of Metaheuristics*: *Int. Ser. in Oper. Res. & Manage. Sci.*, M. Gendreau and J. Potvin, Eds. Boston: Springer, 2010, pp. 449-46.