Journal of System Safety

Assessing the Safety of Intelligent Systems

Volume 58 No. 2
Summer 2023

A publication of the International System Safety Society -
Professionals dedicated to the safety of systems, products, and services

e-ISSN 2832-305X

Journal of System Safety

Editorial Team

For more information please visit our Editorial Team page

Editor-in-Chief

Dr. Charles Muniak
Syracuse Safety Research, USA
Email: editor@jsystemsafety.com

Associate Editors

Dr. Rod Simmons
Independent Consultant, USA

Stephen Thomas
NVIDIA Corporation, USA

Dr. Rami Debouk
General Motors, USA

Editorial Board

International System Safety Society

For more information please visit the Society Home page

Officers

Pam Alte, *ISSS President*
Dave West, *ISSS Executive Vice President*
Pam Kniess, *ISSS Treasurer*
Dr. Rami Debouk, *ISSS Executive Secretary*
Russ Mitchell, *ISSS Immediate Past President*

Directors

Dr. Rodney Simmons, *Education & Professional Dev.*
Donna DiFiglia, *Chapters & International Outreach*
Don Swallom, *Publicity and Media*
Rita Turner, *Member Services*
Yawa Adonsou, *Government & Inter-society Services*
Mike McKelvey, *Conferences*

Connect on Social Media

Society Corporate Partners

A-P-T RESEARCH, INC. AN EMPLOYEE-OWNED COMPANY

base2 solutions a Belcan company

DSTA Defence Science & Technology Agency

FIRSTMERIT Solutions

Parvati Consulting

An official publication of the International System Safety Society, Inc., a non-profit corporation incorporated in the District of Columbia.

Journal of System Safety is published three times a year by the International System Safety Society for the transmission of technical material and news of topical interest to those associated with the practice of system and product safety. Information, recommendations, statements and opinions expressed herein are those of the individual authors and advertisers and do not necessarily represent those of the International System Safety Society. Certain material is published for the purpose of stimulating independent thought on controversial matters or on problems of vital concern to safety professionals. Although caution is taken to ensure accuracy, the publishers or editors cannot accept responsibility for correctness or accuracy of the information presented.

ARTICLE SUBMISSION
Journal of System Safety welcomes article submissions from its readers. Technical manuscripts and news items of interest should be submitted at jsystemsafety.com or sent to Dr. Chuck Muniak, JSS Technical Editor, at journaleditor@system-safety.org. Authors should include the following: (1) one printed copy of the manuscript, double spaced; (2) electronic file in Microsoft® Word™, Adobe® InDesign® or ASCII format; (3) a statement of copyright ownership; (4) a short (one paragraph) author profile; (5) the author's name, address, daytime phone and fax number, email address, affiliation and professional status. For more information on submissions, please see our author guidelines or email journaleditor@system-safety.org. All submissions are subject to peer review.
If authors wish to have their materials returned, they should send a specific request along with a self-addressed, stamped envelope.

ADVERTISING POLICY
Journal of System Safety welcomes advertising compatible with the objectives of the International System Safety Society, subject to the approval of the Technical Editor. The acceptance of advertising does not imply endorsement by the Society or Journal of System Safety.
For more information on advertising, call 651-265-7856 or contact systemsafety@system-safety.org.

MEMBERSHIP INFORMATION
For information on subscription rates and membership, contact the International System Safety Society, 1000 Westgate Dr. Suite 252 Saint Paul, MN, 55114, USA. Tel: 651-265-7856; email: systemsafety@system-safety.org; Web site: www.system-safety.org.

EDITORIAL DISCLAIMER
The views expressed in the editorials and columns in Journal of System Safety are those of the individual writers and do not necessarily reflect the views of the International System Safety Society.
(e-ISSN 2832-305X)

Table of Contents

In The Spotlight

Features

From the Editor's Desk

JSS Technical Editor
C. G. Muniak P.hD.

Cognition

When accidents occur the question "What were they thinking?" is often asked by those of us who are investigating the situation. When we do system design, especially the system safety aspects, we often consider the cognitive process (and errors associated with this process) of the operator. We often also wonder about the thought processes of our colleagues, our management and probably should also wonder about ourselves. The work of Nobel Prize winner Daniel Kahneman [Ref 1] has stimulated some interesting discussion in the last several years. The main finding of Kahneman is that our minds are susceptible to systematic errors. I will enumerate a few examples that I find particularly interesting.

People are very good intuitive grammarians from the time they are children. However, people are not good intuitive statisticians. We have a bias in that we believe results based on inadequate evidence (i.e., a small number of observations). This problem is true even for actual statisticians, yes, the guys who taught us statistics.

Another common problem is that people tend to overestimate their understanding of the world and to underestimate the role of random chance in events as they unfold. Overconfidence is a factor in many accidents.

The first technical paper in this issue is "Human Reliability Analysis using a Human Factors Hazard Model" by Dustin S. Birch, Erika E. Miller and Thomas H. Bradley. This paper proposes a Human Factors Hazard Model (HFHM), which builds an HRA (Human Reliability Analysis) method from the tools of Fault Tree Analysis (FTA), Event Tree Analysis (ETA), and a novel model of considering serial Human Error Probability (HEP) more relevant to psychomotor-intensive industrial and commercial applications such as manufacturing, teleoperation, and vehicle operation.

The second technical paper is "Proposing the Use of Hazard Analysis for Machine Learning Data Sets" by H. Glenn Carter, Alexander Chan, Chris Vinegar and Jason Rupert. To provide information on the impor-

We Want to Hear from You!

Journal of System Safety is seeking papers and articles on many topics where system safety makes a critical contribution, including:

- Explosive Safety
- Nuclear Safety
- Hazardous Material Management
- Chemical Safety
- Biotech Safety
- Safety Management Issues
- Human Error
- Software Safety
- Safety-Critical Processes
- Lessons Learned
- Industry Book Reviews

Please send summaries or abstracts or letters to the editor to Chuck Muniak, Technical Editor, at journal@system-safety.org.

tance of various attributes in the machine learning data sets, this paper proposes a new technique the authors call data hazard analysis. The data hazard analysis provides an approach to qualitatively analyze the training data set to reduce the risk associated with garbage in garbage out.

The third paper in this issue is "Review of the latest Developments in Automotive Safety Standardization for Driving Automation Systems" by Rami Debouk. With the introduction of complex driving automation systems, new standardization efforts to deal with safety of these systems have been initiated to address emerging gaps such as the human/automation roles and responsibilities in the presence/absence of the driver/user, the impact of the technological limitations and the verification and validation needs of automation systems. This paper highlights some of these gaps and introduces some of the latest developments in automotive safety standardization for driving automation systems.

The TBD article by Charlie Hoes describes the concept of "producer" and how it might apply the ISSS.

Regards,
Chuck

References

1. Kahneman, Daniel. Thinking Fast and Slow. Farrar, Straus and Giroux, New York, 2011

TBD

by Charles Hoes

I want to tell you a story about an encounter I had at a hotel bar in Lancaster California. I appreciate that at first it doesn't appear to have anything to do with System Safety. Trust me, I think you will agree that perhaps there is an important lesson for us and the Society.

On the first day of May I started on a long, slow trip along some of the backroads of America – just to see what I might see. On the fifth day after leaving my home near Sacramento I finally made it to Palmdale where I got a room in the Doubletree hotel.

After checking in and resting a bit I went downstairs to get a glass of wine and see about getting dinner. The bar was quite stark and "sterile" feeling. It was all white - white walls, white bar top, white tables. The room was empty except for two guys sitting at opposite ends of the bar eating dinner and staring at their cellphones. A baseball game was playing on the television, but nobody was paying attention. There was no opportunity to start up a conversation or even make a friendly gesture. Rather dejected, and a bit lonely after five days on the road, I realized I was exhausted and it was best to just have a glass of wine, eat a simple dinner, and go to bed early.

Before I could finish my dinner a big, older gentleman (Whitney) came in and took the middle seat at the bar. He ordered a drink and dinner – then sat up straight, leaned back and started singing! He just flat belted out a song about how to care for a woman. I have been in a lot of bars over the years, but this was the first time that I had seen anyone launch into full-throated song.

After finishing the song, Whitney indicated that he wanted us to sing a few of the backup notes while "we" try it again. Even though I am usually the shy, quiet type I took up the invitation and joined him (not well, but enthusiastically). He paused his singing to give directions to the three of us, indicating who should take the bass lines, who would sing the middle ones and who would take the high notes. I was assigned the middle register because I would normally take the bass and he said that wouldn't be any fun to do what is normal. He moved us all to places in our voices that were uncomfortable for us – and then we all sang! None of this was "normal." It wasn't pretty, but it was fun.

After our singing finally died down, he quizzed the guy at the right end of the bar about his background. It turns out that this guy was from Ireland, is a part time music producer and full-time engineer working in the aerospace industry. The two of them talked to each other about the position of "producer" in the music industry. I was having a hard time understanding what they were talking about so asked them to explain what a producer is and what they do. That really ignited an interesting discussion.

As Whitney described the role, the producer of a song is the one that guides all aspects of the production, from the initial concept, creation of the lyrics, selection of music, the performers, the details of the performance (pace, style, clothing, lighting, room details, microphone selection – everything), financing, obtaining copyright protection – everything. Whitney explained that Michael Jackson was an example of a performer

that was successful because he was also the producer. He was in total control over all aspects of his productions, getting everyone to do exactly what he needed them to do to achieve his vision.

As an example of "producing" the bar experience that we were in, Whitney "redid" the bar/lounge image. He "put" imaginary poles on a table in the middle of the room and populated them with imaginary pole dancers, he added an imaginary lap dancer to spice things up a bit, and he changed the color scheme and blocked the windows so kiddies wouldn't see in. Basically, he completely changed the vision of the place just by playing with a few ideas. Perhaps that vision isn't exactly what the hotel management would like in their family-friendly hotel – but it was an interesting game in that moment and he did a good job of illustrating the possibilities.

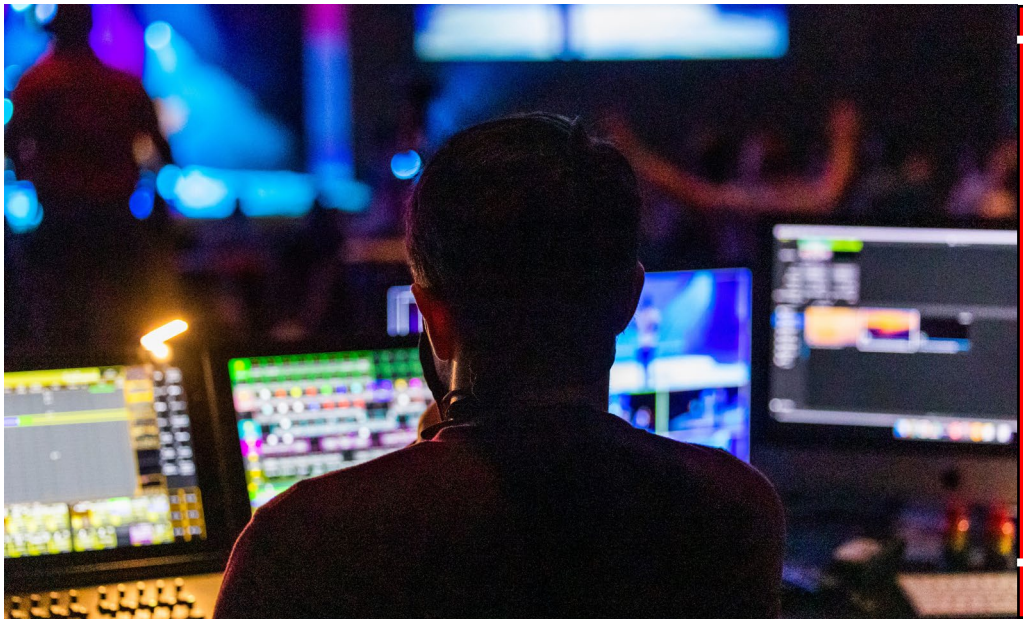
Whitney said that he could "produce" me to sing. I scoffed at the possibility of this. I am not known for my musical prowess. He then asked me to sing a note. I just picked one and sang a note. He said to go higher, then a little lower, and a little longer – and finally said, "There, you got to my vision and you now can do that part. I can now concentrate on other parts of the production."

It dawned on me that he was describing what I do when I am "teaching" newly hired engineers in my consulting firm how to be system safety engineers. I tell them what is expected, let them try, and then come back to adjust their efforts until they "get it" – at which point I can turn my attention to other concerns. However, in order to do that I first have to have a clear vision of the entire business, and their roles and responsibilities within it. I need to know the whole thing, and then I can "produce" an effective business.

In running my engineering consulting firm I am being the "producer" of our services. I had been thinking in terms of being a manager – but it is much more than that. A manager would be tasked with implementing certain parts of "the vision," but not everything. That is my job as the "producer/owner." All of the parts of the "show" (the business, finances, staffing, scheduling, services and relationships with our customers) are important and need to be guided for best results. I also realized that is what is missing in the System Safety Society's management. We don't have a "producer" that can communicate an appropriate vision, who knows how to help others help him/her achieve that vision, who can ensure that all of the necessary parts are in place and working properly.

It became clear to me that if we are to achieve our shared vision of System Safety in the world we need to find an effective "producer" to guide our individual actions. We need a visionary who has the vision, knowledge, funding and authority to take us forward into the "production" of the System Safety Society.

By the end of the day I was totally blown away. I had started an evening of absolutely nothing, and then something happened… something that has changed my understanding of my life, something that will stay with me far into the future. Lessons come from mysterious places if you just relax and let them happen.

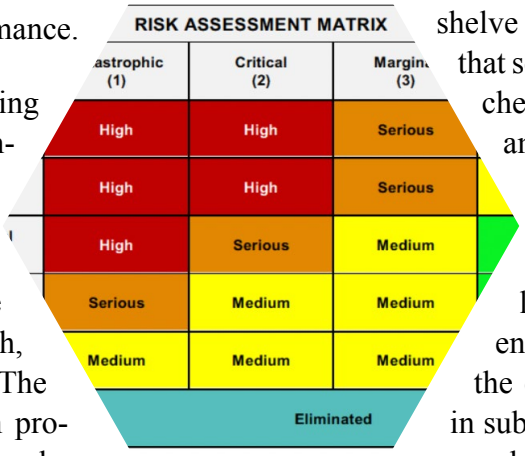


"It became clear to me that if we are to achieve our shared vision of System Safety in the world we need to find an effective "producer" to guide our individual actions."

Photo: Pexels

From Our Readers

In the Charles Hoes article "TBD regarding Risk Assessment" the sample risk assessment matrix and his explanation of how the chart can be used to assign "risk levels" is on par with the basics of the risk management process generally used on programs. Hoes is correct in pointing out that the use of the matrix or a similar risk chart for program risk prioritization would be a mistake. Within the risk management process, once a risk is assessed, program risks should be prioritized based upon program priorities and not just their risk level. Program priorities are usually characterized in terms of cost, schedule, and technical performance.



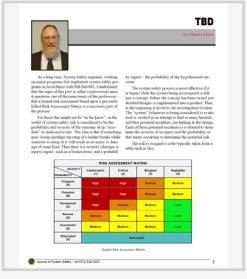
A risk assessment matrix is nothing more than a map containing a complete set of risk values. Its purpose is to show there is a relationship between risk probability and risk severity so that each risk can receive an assessed ranking (i.e., low to high, green to red, or a numerical value). The assessed level of concern can then provide visibility for both the decision makers and the other stakeholders. Using a risk value for each risk gives all stakeholders an equal understanding of the potential threat level per risk. More importantly the measure can assist senior leaders during their review of risk items to implement control actions in a timely fashion.

Hoes's idea of consilience, "the agreement from different disciplines" in forming an opinion makes good sense when assessing risks. However, his recommendation to drop the use of the matrix entirely and in its place provide a "well thought out rationale statement and studies" would likely bring misconceptions and confusion. A risk statement in place of a conventional risk value would likely be perceived differently by the stakeholder community. Some individuals, interpreting the risk to be important while others interpreting the risk as low consequence. Conventional risk rankings ensure all stakeholders interpret threats equally by use of an agreed to risk assessment matrix.

With respect to system safety, engineers must be cautious not to identify or spend time managing the occurrence of "extremely unlikely risks". Risks that fall into this bin are those having an extremely low level of probability, but which could theoretically occur. Studying and managing risks is a time-consuming task and prudent engineers must have a grasp of when to formally pursue them as well as when to shelve them. It has been my observation that some system safety programs left unchecked, spend an inappropriately large amount of effort managing "extremely low probability" system safety risks. This is costly to the program in terms of dollars, resources and loss of reputation among the other engineering disciplines. Moreover, the end-user will unknowingly operate in suboptimal conditions given that higher probability threats may not have been given their due attention.

- Claudio Pantaleo
Software Engineer – Retired

Read the Original Column



TBD
Volume 57 No. 3
Fall 2022

Original Article
https://doi.org/10.56094/jss.v58i2.251

International System Safety Society

www.jsystemsafety.com

Journal of System Safety

Established 1965 Vol. 58 No. 2 (2023)

Human Reliability Analysis using a Human Factors Hazard Model

Dustin S. Birch[ab], Erika E. Miller[c], Thomas H. Bradley[c]

[a] *Corresponding author email: dustinbirch@weber.edu*
[b] *Weber State University, Ogden, UT*
[c] *Colorado State University, Fort Collins, CO*

Keywords
human reliability analysis, human error probability, hazard analysis techniques, fault tree analysis, event tree analysis, system safety, human factors engineering, risk analysis, reliability engineering

Peer-Reviewed
Gold Open Access
Zero APC Fees

Online: 22-Jun-2023

Cite As:
Birch DS. et al, Human Reliability Analysis using a Human Factors Hazard Model. Journal of System Safety. 2023;58(2):7-29. https://doi.org/10.56094/jss.v58i2.251

ABSTRACT

Human Reliability Analysis (HRA) has found application within a diverse set of engineering domains, but the methods used to apply HRA are often complicated, time-consuming, costly to apply, specific to particular (i.e., nuclear) applications, and are not suitable for direct comparison amongst themselves.

This paper proposes a Human Factors Hazard Model (HFHM), which builds an HRA method from the tools of Fault Tree Analysis (FTA), Event Tree Analysis (ETA), and a novel model of considering serial Human Error Probability (HEP) more relevant to psychomotor-intensive industrial and commercial applications such as manufacturing, teleoperation, and vehicle operation. The HEP approach uses Performance Shaping Factors (PSFs) relevant to human behavior, as well as specific characteristics unique to a system architecture and its corresponding operational behavior. The HFHM tool is intended to establish a common analysis approach, to simplify and automate the modeling of the likelihood of a mishap due to a human-system interaction during a hazard event.

The HFHM is executed commercial software tools (MS Excel and SysML) such that trade and sensitivity studies can be conducted and iterated automatically. The results generated by the HFHM can be used to guide risk assessment, safety requirements generation and management, design options, and safety controls within the system design architecting process. Verification and evaluation of the HFHM through simulation and subject matter expert evaluation illustrate the value of the HFHM as a tool for HRA and system safety analysis in a set of key industrial applications.

INTRODUCTION

An engineered system is comprised of numerous human, electrical, mechanical, and software components and subsystems. These system building blocks are combined together into a larger, more complex system, that is used to perform a function per a specified design intent. Human beings (human actors), along with all other components in the design, can interact with the system to respond to off- design behavior to avoid a hazardous situation that may evolve into an accident [2]. These human- system interactions play a significant role in determining the reliability and safety of a system throughout its lifecycle [3]. The combined functionalities and associated interactions of all system elements, including human elements, must be modeled, analyzed, and documented as a matter of Systems Engineering (SE) best practice. System Safety analysis asserts that the reliability and hazard characteristics of the system design must be evaluated and analyzed, with all of the identified potential hazards eliminated or minimized, such that a failure will not result in a catastrophic outcome. To be considered complete, this engineering analysis must consider the interactions and risks posed by all human actors within the system context. A consistent and uniform approach to analyzing the human contribution to safety throughout the system lifecycle management process is preferred.

The contemporary inductive perspective of System Safety analysis tends to emphasize scrutiny of the non-human elements (electrical, mechanical, software) that are combined into the larger system architecture [6][20]. Typically, the probabilistic failure rates of these various elements are determined, and then accounted for in the larger system arrangement using established Hazard Analysis Techniques (HAT's). The prospective failure modes and safety related concerns of a system are evaluated based on the results of these HAT activities and documented for future abatement during subsequent design and testing activities [6]. In addition to the electrical, mechanical, and software elements that are commonly recognized as the core building blocks of a system design, human actors and their respective influence on system operations can be of equal or even greater importance, to the performance, reliability, and safety within the system lifecycle [3]. Accident rates attributable to human activity in system operations range from 10%, to as high as 80% depending on the industry and application [7][13]. Also, of note, the National Highway Traffic Safety Administration (NHTSA) reports that human error is the cause of up to 94% of all ground transportation accidents [21]. Although sometimes overlooked or minimized during system analysis and design, the various human interactions within the system context, and their possible impact on safety, should be properly scrutinized, with potential hazard probabilities being quantified explicitly [13].

There is no universal or general technique to evaluate the hazards associated with human-system interaction [7][11]. Several Human Reliability Analysis (HRA) approaches have been developed, but they are typically complicated and time consuming to implement and are not designed to be applied across engineering disciplines or applications [14]. Instead, HRA approaches generally have specific application within certain industries, environments, or operational activities [7]. For example, HRA techniques such as the Technique for Human Error Rate Prediction (THERP) and Success Likelihood Index Method – Multiattribute Utility Decomposition (SLIM-MAUD) have their origins and primary usage in the nuclear power industries, with an emphasis on procedural control room activities. A technique such as Maintenance Personnel Performance Simulation (MAPPS) focus primarily on human hazard analysis as it relates to maintenance activities, and Aeronautical Decision-Making (ADM) is an analysis technique specific to pilot-flight control interface analysis [14][15][16].

Based on this understanding of the state of the field, the proposed Human Factors Hazard Model (HFHM) seeks to provide a novel, commonly applied, and efficient approach to assessing system risk associated with human interactions.

HAZARD ANALYSIS TECHNIQUES AND THEIR APPLICATION IN SYSTEM SAFETY ANALYSIS

System safety analysis as an activity within Systems Engineering (SE) has its origins in the early 1960's, with the earliest contributor being the Department of Defense (DOD) under MIL-STD-38130 (Safety Engineering of Systems & Associated Subsystems) which was later superseded by the current MIL-STD- 882 (Standard Practice – System Safety) [6][17]. Following the development of these guidelines, other agencies were quick to adopt these system safety philosophies including the Nuclear Regulatory Commission (NRC), as well as the

National Aeronautics and Space Administration (NASA). These techniques have gained widespread acceptance and use across government and commercial industries.

It is common to perform detailed safety analysis using one or more of the various analysis techniques that have been developed [6].

Over 100 different HAT approaches are listed in The System Safety Analysis Handbook published by the International System Safety Society (ISSS) [6]. However, only 10-20 different HATs are regularly used by system safety experts [6]. Among the most common HAT approaches utilized in safety analysis include Fault Tree Analysis (FTA) and Event Tree Analysis (ETA). Both FTA and ETA have direct application in Probabilistic Risk Assessment (PRA) and are used extensively to evaluate the likelihood of failure related to system design. Correspondingly, these two HATs are utilized as a significant building block of the analytical basis for the proposed Human Factors Hazard Model (HFHM) described in this work.

As an overview, FTA is a technique used to compile the failure probability of individual events into larger logic networks, accounting for the interdependency and combined probability of failure [6][18][19]. All FTAs are composed of basic events that are combined using AND/OR logic gates into intermediate events. These intermediate events are then combined using the same logic gate structure to determine the probability of the top-level event. The FTA approach is very useful for evaluating the overall likelihood of a particular failure with a quantified probability. The individual FTA results can then be used in subsequent safety analysis activities to assess the hazard event severity and possible negative consequences of the failure.

Unlike an FTA, an ETA is used to evaluate a sequence of independent, but related events, and their cumulative probability of concluding in a desired or undesired outcome [6]. Hence, the primary purpose of an ETA is to determine the probability that a series of sequential pivotal events will culminate in success or failure relative to specific scenario. For the events identified and analyzed using ETA, the probabilities of all possible outcomes (success or failure) are evaluated and documented.

The FTA / ETA combination forms the computational basis of this proposed Human Factors Hazard Model (HFHM). The HFHM requires the development of Human Error Probabilities (HEP's) that can be combined to determine the joint likelihood of failure for a top-level hazard event using an FTA. The four FTA analyses (corresponding to the four pivotal events of a human response model) are then evaluated in an ETA to determine the top-level probability of success (and failure) for the specified human / system interaction.

HUMAN RELIABILITY ANALYSIS AND HUMAN ERROR PROBABILITY

As a field of study, Human Reliability Analysis (HRA) is conducted with the intent of describing human interactions with related system elements and documenting the associated risks and potential failure modes [4]. HRA is also intended to help develop corrective actions and other possible countermeasures intended to reduce or eliminate the possibility of human caused failures. A recent literature review indicates that there are approximately 38 documented and commonly used HRA methodologies in the public domain [7]. Along with the method proposed in this work, there continues to be ongoing proposals for HRA predictive tools using various qualitative and quantitative approaches [26][27].

Among the most commonly utilized and cited HRA techniques are the Technique for Human Error-Rate Prediction (THERP) and Expert Estimation [7].

THERP was developed for application in safety analyses related primarily to nuclear power plant operations [8]. THERP includes well defined procedural steps to hazard analysis, as well as a comprehensive library of Human Error Probabilities (HEP) associated with common human-system interactions. These documented HEP values include considerations of design characteristics including training efficacy, instrumentation interpretation, control system actuation, as well as other common human factors considerations such as fatigue, distraction, and stress effects and their influence on HEP.

Expert Estimation (also known as Expert Judgement) is a general HRA approach with several different basic techniques used to assess HEP values associated with specific human-system interaction. Four basic approaches used for Expert Estimation have been documented, and they include: (1) paired comparison, (2) ranking / rating, (3) indirect numerical estimation, and (4) direct numerical estimation. Paired comparison and ranking / rating

approaches produce equivocal results. Indirect numerical estimation will establish a HEP by relative comparison based on the probabilities of failure determined for other events. The direct numerical estimation technique produces a specific HEP based on an expert or group of expert's estimations of the likelihood of a specific error due to the relevant human factors as well as system characteristics [7][9].

Performance Shaping Factors and Human Error Probability Prediction

The likelihood that a human actor will fail to perform or incorrectly perform a required task, possibly resulting in a mishap, is referred to as Human Error Probability (HEP). The development of an analytical model used to predict HEP is primarily dependent on consideration of human factors and system characteristics. These two elements are referred to as Performance Shaping Factors (PSFs). PSFs are used to calculate HEP relevant to specific operational scenarios. For example, the complexity of a system design, the human actor's knowledge of system operation, the actor's distraction and stress levels, and the nature of the off-design behavior of the system, will all contribute to the probability that the actor will react correctly to the system behavior, and successfully avoid a mishap. A non- comprehensive list of elements that represent human factors and system factors in PSFs are presented in Table 1 [7][11][12]. Typically, the characteristics of PSFs are drawn from established and widely cited Human Reliability Analysis (HRA) and human factors engineering sources [8][10]. The proposed Human Factors Hazard Model (HFHM) uses PSFs and their associated literature in calculating HEP values used to calculate the overall failure probability.

These sets of PSFs are used to develop the conditional and combined probabilities of failure that can be used to determine the HEP values for use in the HFHM. These HEP values are bookkept in the model and are used to produce a prediction of failure related to a given hazard scenario. In general, HEP calculations can be used in their baseline state, or can be modified based on other PSF characteristics. For example, the baseline probability of failure (HEP) due to an actor's intellectual capacity can be modified by their stress level, fatigue level, impairment characteristics, or other relevant PSF values.

When a modified HEP is to be considered, factors attributable to specific PSFs are multiplied to change the baseline probability of failure for the characteristics of the scenario. This calculation is of the form:

$$P_f' = (P_f)\prod_{n=1}^{i} M_n \quad (1)$$

Where:

P_f' = *Modified Event Probability of Failure*
P_f = *Initial Event Probability of Failure*
M_n = *Probability Modifier, or PSF, n*
i = *Total Number of Probability Modifiers Applied*

Table 1: Examples of the set of Human and System Factors Used to Determine PSFs for HFHM's Industrial and Commercial Application Set

HUMAN FACTORS	SYSTEM FACTORS
Training	System Complexity
Practice	Hazard Event Timing
Experience	Hazard Event Duration
Mental Acuity	Observability of System Behavior
Intellectual Capacity	Annunciation of System Behavior / Alarms
Gross and Fine Motor Skills	Instrumentation Availability to Monitor System Behavior
Sensory Acuity (Smell, Vision, Hearing, Touch)	Input Control Capabilities
Fatigue, Vigilance, and Impairment Level	Input Control Accessibility to Actor
Stress and Emotional Stability	System Behavior Feedback Characteristics
Reaction Time	Environmental Conditions (Temperature, Illumination, etc.)
Location and Orientation of Actor withing System Context	System Fail-Safes
Negligence and Malevolent Intent	System Safeguards

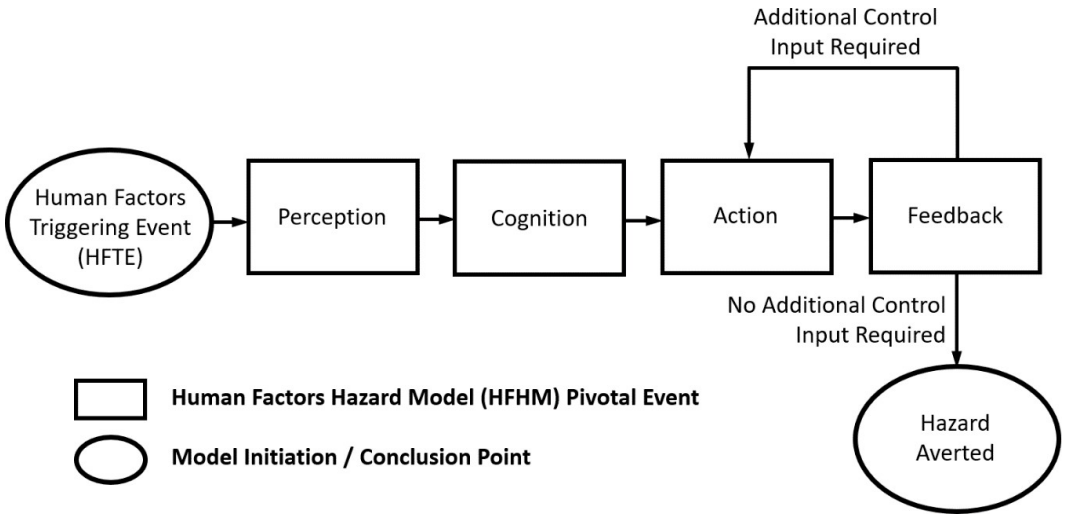


Figure 1: Hazard Event Human Response Model

When considering the chronology associated with the event from hazard initiation through a mishap or successful resolution, human reaction time is adjusted using a multiplier similar to the probability adjustments noted above. In this case, multipliers are not compounded, but applied individually, then summed to adjust the baseline human actor reaction time. If multipliers are used to modify the baseline reaction time, the calculation is of the form:

$$T_r' = T_r + \sum_{n=1}^{i} T_r(R_n - 1) \quad (2)$$

Where:

T_r' = *Modified Reaction Time*
T_r = *Baseline Reaction Time*
R_n = *Reaction Time Modifier or PSF, n*
i = *Total Number of Reaction Time Modifiers Applied*

Using these modifications of the relevant baseline HEP values due to the unique PSFs of a given hazard scenario, the basic event probabilities are established for subsequent processing in individual Fault Tree Analyses (FTA's).

SUMMARY

Based on this understanding of the state of the field of HRA, we can identify the opportunities for development of a new model and process for human factors safety modeling. First, THERP (and its antecedents [8][7][15][28]) use a 1- or 2-stage model of human behavior that does not consider the multi-event feedback-inclusive nature of skilled human operation. The proposed HFHM embeds a computational architecture that implements a formal specification of the psychological theories of cognition, perception, and action [22][23][25][29], that are more complete for consideration of serial and psychomotor tasks. Second, many models of HRA are complicated to use and maintain. The classical methods are largely not computerized and are therefore inaccessible and costly for adaptation to minor commercial or industrial applications. HFHM provides both a MS Excel-based and SysML-based implementation of a relatively comprehensive HRA and extant PSF database, enabling modern document-based and model-based systems engineering application and scalability from small to large HRA problems [7][8][10].

THE HUMAN FACTORS HAZARD MODEL (HFHM)

The proposed Human Factors Hazard Model (HFHM) seeks to predict the likelihood of failure due to an actor's response to a Human Factors Triggering Event (HFTE), where the HFTE is defined as any interaction between a human being and the system which may result in a mishap [1]. The conceptual model of the steps involved in predicting the human response to an HFTE is a serial processing approach as illustrated in Figure 1. First, the event must be perceived and recognized as a hazard. Second, the actor will cognitively process the available observed information, and then establish a corrective action

plan. Third, a planned remedial action by the actor is then communicated to the system via control inputs, and the subsequent system behavior response is then observed. Fourth, based on the system feedback behavior due to control input, the actor must decide whether to terminate control input because the hazard has been resolved, or continue to provide additional control corrections in an effort to eliminate the hazardous behavior completely. Each stage of the process detailed above in Figure 1 (Perception, Cognition, Action, and Feedback), indicates a point in the hazard sequence where a possible human failure could result in a mishap. This sequential approach to human information processing is a widely accepted model used to map a response in discrete, identifiable steps [22][23]. As an example, if the actor perceives the hazard event, but subsequently does not cognitively process it correctly, concluding that corrective action is necessary, the series of events will not progress to the action step, and thus, the HFTE will end in a mishap.

COMPONENTS OF THE HFHM

Under the proposed Human Factors Hazard Model (HFHM) technique, each of the individual pivotal events of Figure 1 are modeled using an embedded Fault Tree Analysis (FTA). The probability of success (or failure) for each of the four pivotal events are predicted via the FTA logic networks composed of basic events determined from the human factors and system characteristics (PSFs) unique to the problem being analyzed. The individual FTAs are each based on an evaluation of probability of failure associated with combinations of the various contributing events due to human interaction with the system. The set of failures considered and modeled in each FTA are derived from the set of HRA-derived failures [7][8][10] that the authors have considered relevant to a broad set of industrial and commercial applications. While this set is not comprehensive, it includes a broad set of human failure events that are relevant to HFHM's set of industrial and commercial applications, referencing the broad literature on HFA [8][9][10].

Each basic event probability of failure is evaluated using Boolean logic, through intermediate events, to subsequently arrive at a combined top-level probability of failure. The relevant symbols used in the HFHM Fault Tree Analysis logic networks are presented in Figure 2.

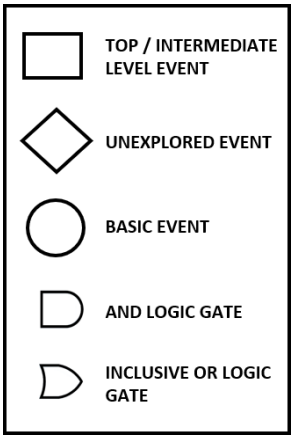


Figure 2: FTA Event Symbol Key

The assignment of AND/OR logic within the FTA model is dependent upon the interrelation of the events being considered. For example, if a visual hazard signal is generated both by observable system behavior as well as instrumentation communication to a control panel indicator, both events must fail for the actor to not receive information communicating the unfolding hazard event. Thus, an "AND" gate to model this scenario would be an appropriate approach to combined probability. Conversely, if no inherent redundancy exits within the relationship of events, an Inclusive "OR" would be appropriate, indicating that any individual or combination of failures would signal a failure at the next highest level within the FTA.

As noted, the HFHM utilizes four different FTAs to model human actor response to a Human Factors Triggering Event (HFTE). Each of these four FTAs represent the pivotal events (Perception, Cognition, Action, and Feedback) associated with human response to a hazard, as noted in Figure 1. All of the basic events introducing failure probabilities into their corresponding FTAs determine their respective values from the PSF information used to modify baseline HEP values using equations (1) and (2) as defined above.

The FTA corresponding to the **Perception** pivotal event is presented in Figure 3 and Table 2. The FTA corresponding to the **Cognition** pivotal event is presented in Figure 4 and Table 3. The FTA corresponding to the **Action** pivotal event is presented in Figure 5 and Table 4. The FTA corresponding to the **Feedback** pivotal event is presented in Figure 6 and Table 5.

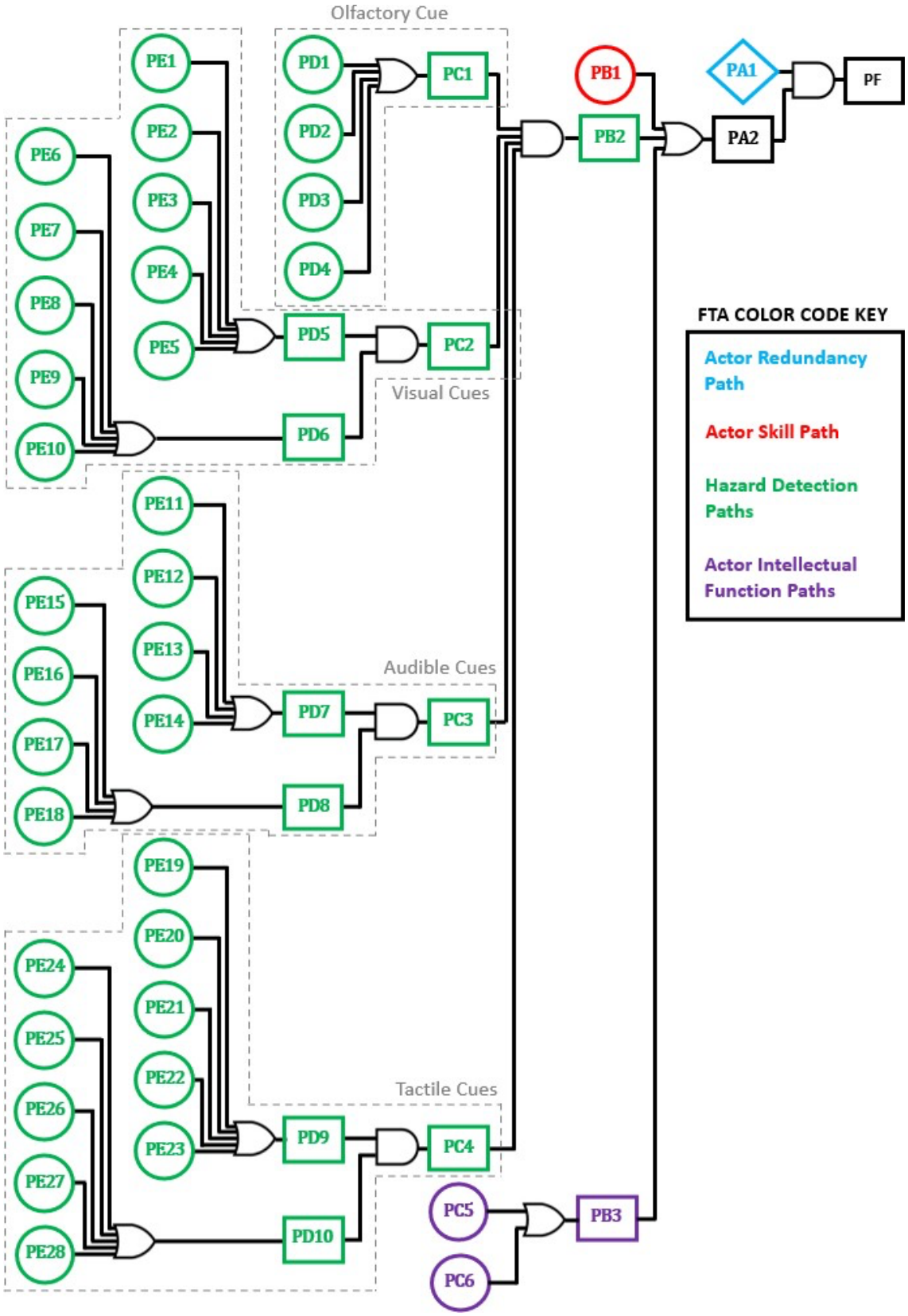


Figure 3: HFHM Perception FTA Used to Model the Probability of Fault for the Operator to be Unable to Perceive the Hazard

Table 2: HFHM Perception FTA Label, Descriptions, and Logic Gate Types

PERCEPTION PIVOTAL EVENT FTA NETWORK									
LABEL, EVENT DESCIPTION, AND EVENT LOGIC NETWORK GATE TYPE									
PF	Top Level Event								AND
	PA1	Redundant Actors							N/A
	PA2	Single Actor							OR
		PB1	Actor Skill						N/A
		PB2	Hazard Detection Cue						AND
			PC1	Olfactory Hazard Cue					OR
				PD1	Timing				N/A
				PD2	Location				
				PD3	Sensory				
				PD4	Olfactory Cue				
			PC2	Visual Hazard Cue					AND
				PD5	System Behavior Cue				OR
					PE1	Timing			N/A
					PE2	Location			
					PE3	Orientation			
					PE4	Sensory			
					PE5	Signal			
				PD6	Instrumentation Cue				OR
					PE6	Timing			N/A
					PE7	Location			
					PE8	Orientation			
					PE9	Sensory			
					PE10	Signal			
			PC3	Audible Hazard Cue					AND
				PD7	System Behavior Cue				OR
					PE11	Timing			N/A
					PE12	Location			
					PE13	Sensory			
					PE14	Signal			
				PD8	Alarm Cue				OR
					PE15	Timing			N/A
					PE16	Location			
					PE17	Sensory			
					PE18	Signal			
			PC4	Tactile Hazard Cue					AND
				PD9	System Behavior Cue				OR
					PE19	Timing			N/A
					PE20	Location			
					PE21	Orientation			
					PE22	Sensory			
					PE23	Signal			
				PD10	Control System Cue				OR
					PE24	Timing			N/A
					PE25	Location			
					PE26	Orientation			
					PE27	Sensory			
					PE28	Signal			
		PB3	Actor Intellectual Function						OR
			PC5	Actor Mental Acuity					N/A
			PC6	Actor Attention					

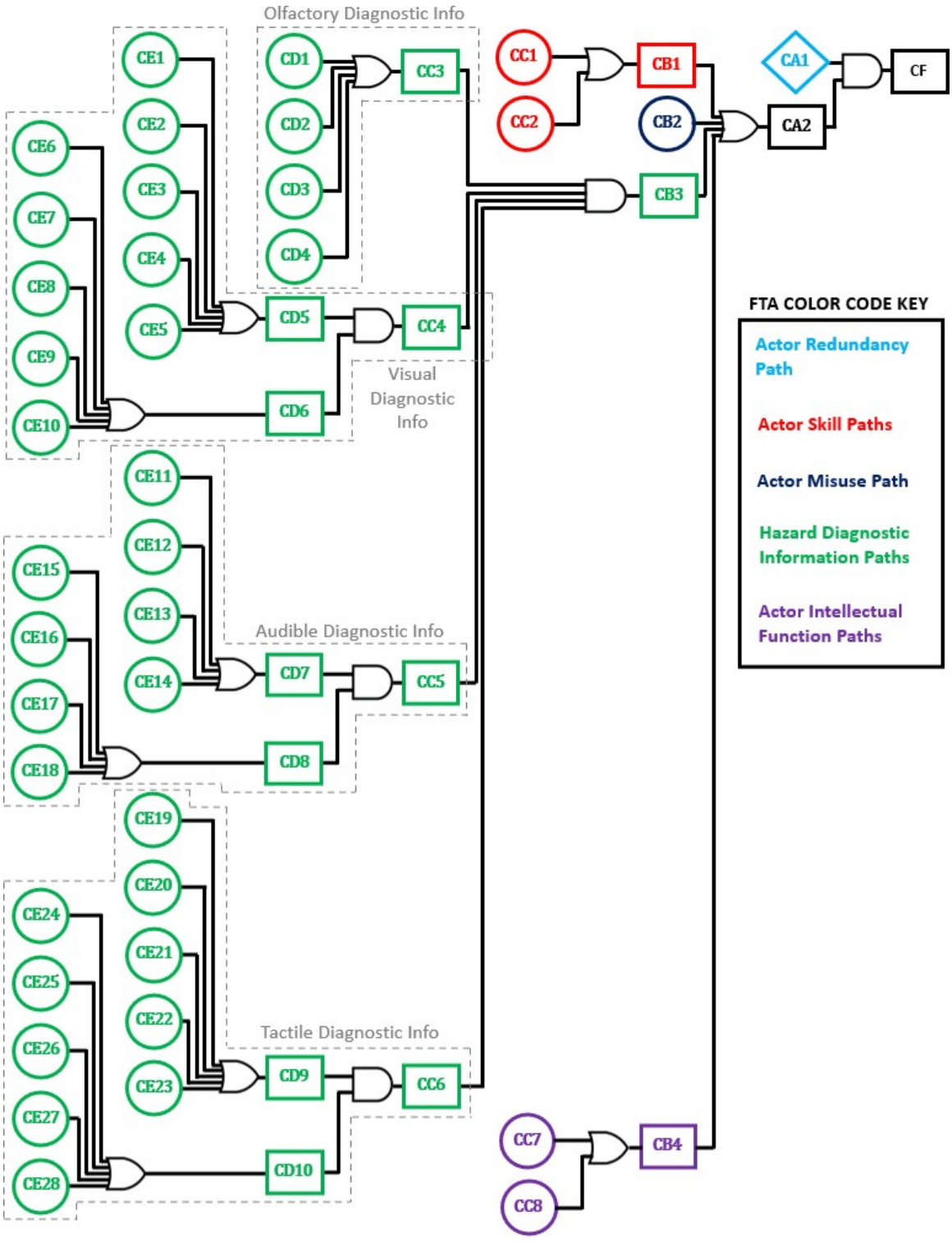


Figure 4: HFHM Cognition FTA Logic Network Used to Model the Probability of Fault for the Operator to be Unable to Cognitively Process the Hazard

Table 3: HFHM Cognition FTA Label, Descriptions, and Logic Gate Types

COGNITION PIVOTAL EVENT FTA NETWORK														
LABEL, EVENT DESCIPTION, AND LOGIC NETWORK GATE TYPE														
CF	Top Level Event													AND
	CA1	Redunant Actors												N/A
	CA2	Single Actor												OR
		CB1	Actor Skill											OR
			CC1	Timing										N/A
			CC2	Diagnostic Approach										
		CB2	Misuse											N/A
		CB3	Hazard Diagnostic Information											AND
			CC3	Olfactory Diagnostic Information										OR
				CD1	Timing									N/A
				CD2	Location									
				CD3	Sensory									
				CD4	Signal									
			CC4	Visual Diagnostic Information										AND
				CD5	System Behavior Information									OR
					CE1	Timing								N/A
					CE2	Location								
					CE3	Orientation								
					CE4	Sensory								
					CE5	Signal								
				CD6	Instrumentation Information									OR
					CE6	Timing								N/A
					CE7	Location								
					CE8	Orientation								
					CE9	Sensory								
					CE10	Signal								
			CC5	Audible Diagnostic Information										AND
				CD7	System Behavior Information									OR
					CE11	Timing								N/A
					CE12	Location								
					CE13	Sensory								
					CE14	Signal								
				CD8	Alarm Information									OR
					CE15	Timing								N/A
					CE16	Location								
					CE17	Sensory								
					CE18	Signal								
			CC6	Tactile Diagnostic Information										AND
				CD9	System Behavior Information									OR
					CE19	Timing								N/A
					CE20	Location								
					CE21	Orientation								
					CE22	Sensory								
					CE23	Signal								
				CD10	Control System Information									OR
					CE24	Timing								N/A
					CE25	Location								
					CE26	Orientation								
					CE27	Sensory								
					CE28	Signal								
		CB4	Actor Intellectual Function											OR
			CC7	Actor Mental Acuity										N/A
			CC8	Actor Attention										

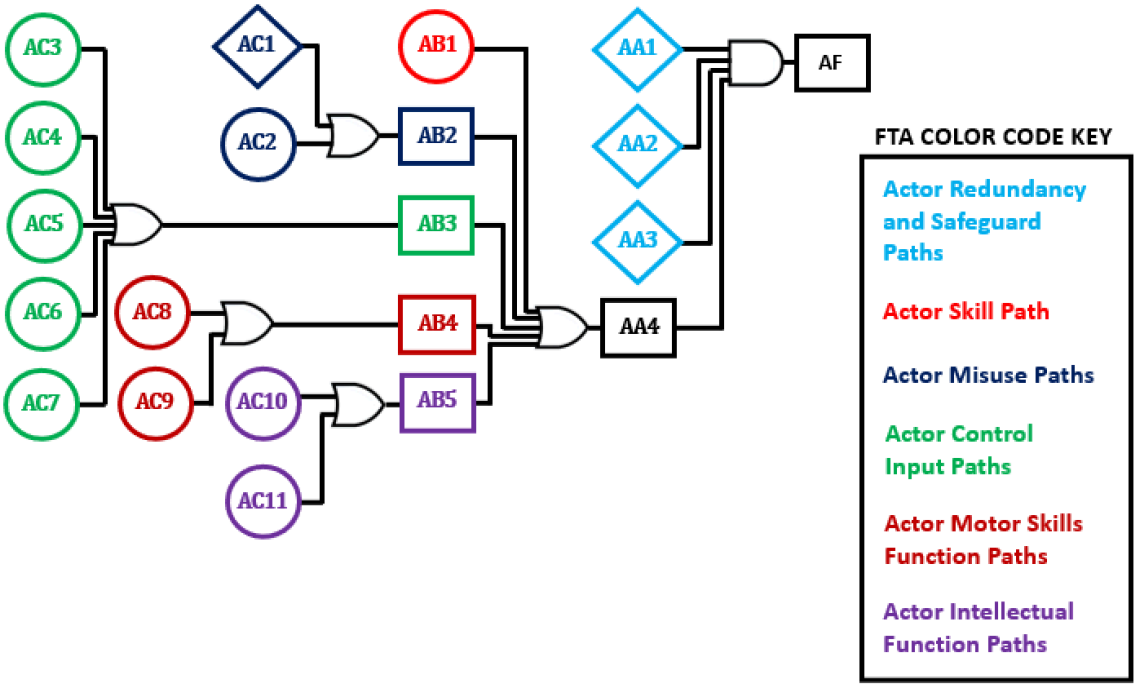


Figure 5: HFHM Action FTA Logic Network Used to Model the Probability of Fault for the Operator to be Unable to Correctly Apply Input Control to Correct the Hazard Behavior

Table 4: HFHM Action FTA Label, Descriptions, and Logic Gate Types

ACTION PIVOTAL EVENT FTA NETWORK LABEL, EVENT DESCIPTION, AND LOGIC NETWORK GATE TYPE					
AF	Top Level Event				AND
	AA1	Redundant Actors			N/A
	AA2	Software Safeguards			
	AA3	Hardware Safeguards			
	AA4	Single Actor			OR
		AB1	Actor Skill		N/A
		AB2	Misuse		OR
			AC1	Intentional Misuse	N/A
			AC2	Malevolence	
		AB3	System Control Input		OR
			AC3	Timing	N/A
			AC4	Location	
			AC5	Orientation	
			AC6	Sensory	
			AC7	Control Interface	
		AB4	Actor Motor Skills Function		OR
			AC8	Gross Motor Skills	N/A
			AC9	Fine Motor Skills	
		AB5	Actor Intellectual Function		OR
			AC10	Actor Mental Acuity	N/A
			AC11	Actor Attention	

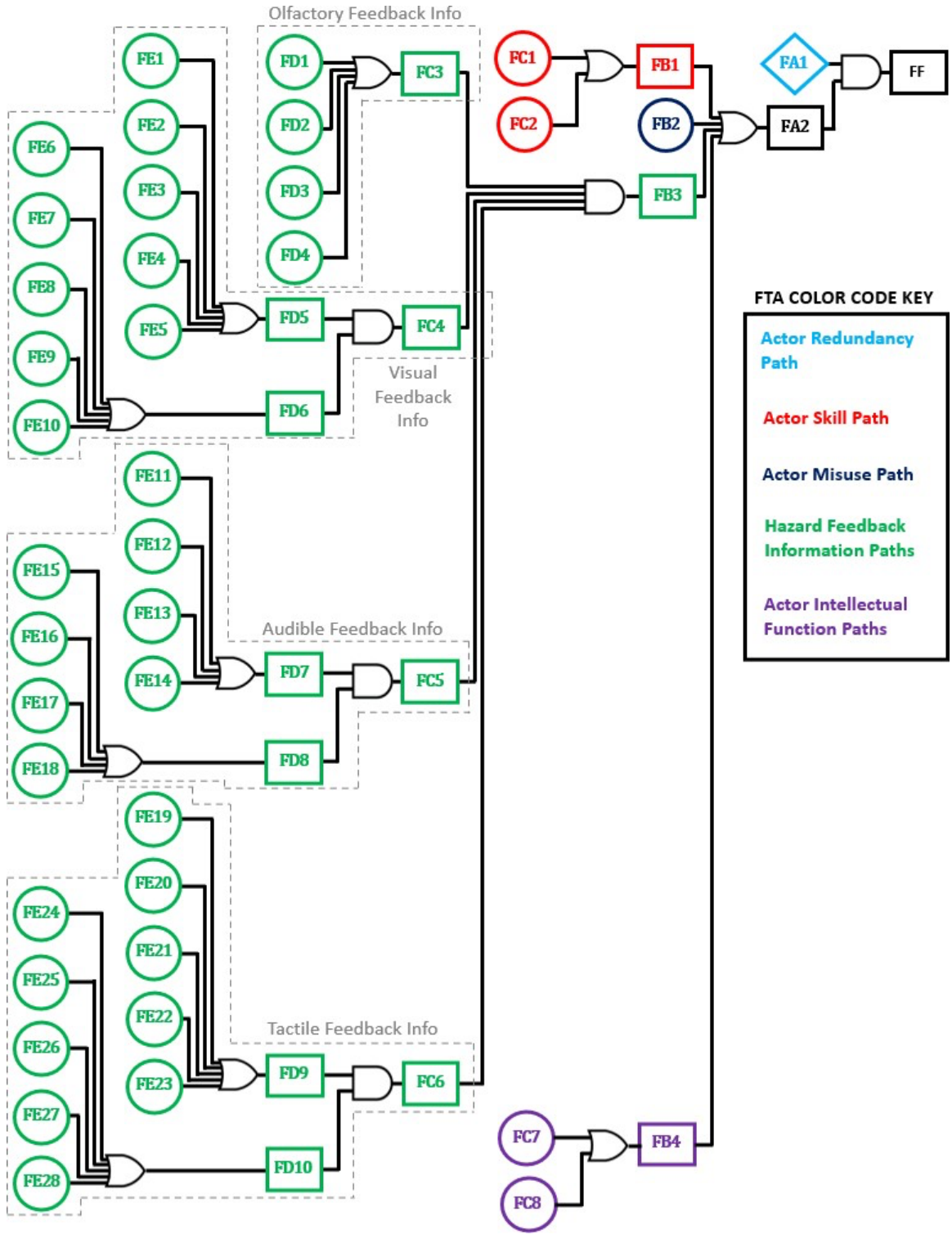


Figure 6: HFHM Feedback FTA Logic Network Used to Model the Probability of Fault for the Operator to be Unable to Receive and React Correctly to System Feedback Generated by Prior Input Control Action

Table 5: HFHM Feedback FTA Label, Descriptions, and Logic Gate Types

FEEDBACK PIVOTAL EVENT FTA NETWORK										
LABEL, EVENT DESCIPTION, AND LOGIC NETWORK GATE TYPE										
FF	Top Level Event									AND
	FA1	Redunant Actors								N/A
	FA2	Single Actor								OR
		FB1	Actor Skill							OR
			FC1	Timing						N/A
			FC2	Feedback Interpretation						
		FB2	Misuse							N/A
		FB3	Hazard Feedback Information							AND
			FC3	Olfactory Feedback Information						OR
				FD1	Timing					N/A
				FD2	Location					
				FD3	Sensory					
				FD4	Signal					
			FC4	Visual Feedback Information						AND
				FD5	System Behavior Information					OR
					FE1	Timing				N/A
					FE2	Location				
					FE3	Orientation				
					FE4	Sensory				
					FE5	Signal				
				FD6	Instrumentation Information					OR
					FE6	Timing				N/A
					FE7	Location				
					FE8	Orientation				
					FE9	Sensory				
					FE10	Signal				
			FC5	Audible Feedback Information						AND
				FD7	System Behavior Information					OR
					FE11	Timing				N/A
					FE12	Location				
					FE13	Sensory				
					FE14	Signal				
				FD8	Alarm Information					OR
					FE15	Timing				N/A
					FE16	Location				
					FE17	Sensory				
					FE18	Signal				
			FC6	Tactile Feedback Information						AND
				FD9	System Behavior Information					OR
					FE19	Timing				N/A
					FE20	Location				
					FE21	Orientation				
					FE22	Sensory				
					FE23	Signal				
				FD10	Control System Information					OR
					FE24	Timing				N/A
					FE25	Location				
					FE26	Orientation				
					FE27	Sensory				
					FE28	Signal				
		FB4	Actor Intellectual Function							OR
			FC7	Actor Mental Acuity						N/A
			FC8	Actor Attention						

The four FTA logic networks are designed to calculate the associated probability of failure for the top- level event based on the modified HEP values and all intermediate probabilities calculated in the logic network. The corresponding probability of success for each top-level FTA is:

$$S = 1 - F \quad (3)$$

Where:

S = Event Probability of being Successful
F = Event Probability of being Unsuccessful (Failure)

Using the values for probability of success, as calculated using equation (3). The Event Tree Analysis (ETA) then calculates the probability of success and failure for each sequential event. The logical basis of the ETA assumes that each pivotal event must occur in order, without failure, for an ultimate successful outcome. In the case of the HFHM, all four events must successfully occur sequentially for the HFTE to be resolved. If any individual pivotal event experiences a failure, then all subsequent events are null, and the HFTE has resulted in a mishap. Per this logic network, each individual pivotal event is considered to be mutually exclusive in that any individual failure precludes success for all subsequent events. The logic network and associated mathematical basis of the ETA is presented in Figure 7. Where PF, CF, AF, and FF are failure probability inputs from each respective FTAs.

In summary, the HFHM model allows for users to model human error in considering overall system performance. The characteristics of the system design and human actor are used to determine Performance Shaping Factors (PSFs) and the related modified HEPs. The modified HEP values are then used as the basic event probabilities that are utilized at the entry levels of the four FTA networks. In cases where the human factors and system characteristics are considered to be standard and universally applicable, the baseline HEP values can be derived from existing literature [8][10]. When certain PSF values are more specialized and standard values are not universally established or published, the HEP values can be based on an Expert Estimation / Expert Judgement approach [9]. For unique cases, where empirical data for specific operational scenarios have been derived, the HEP values can be directly specified in the HFHM. Once the individual HEPs are determined, and the respective pivotal event FTAs are calculated, ETA is then used to determine the cumulative probabilities of the individual pivotal events and calculate the overall

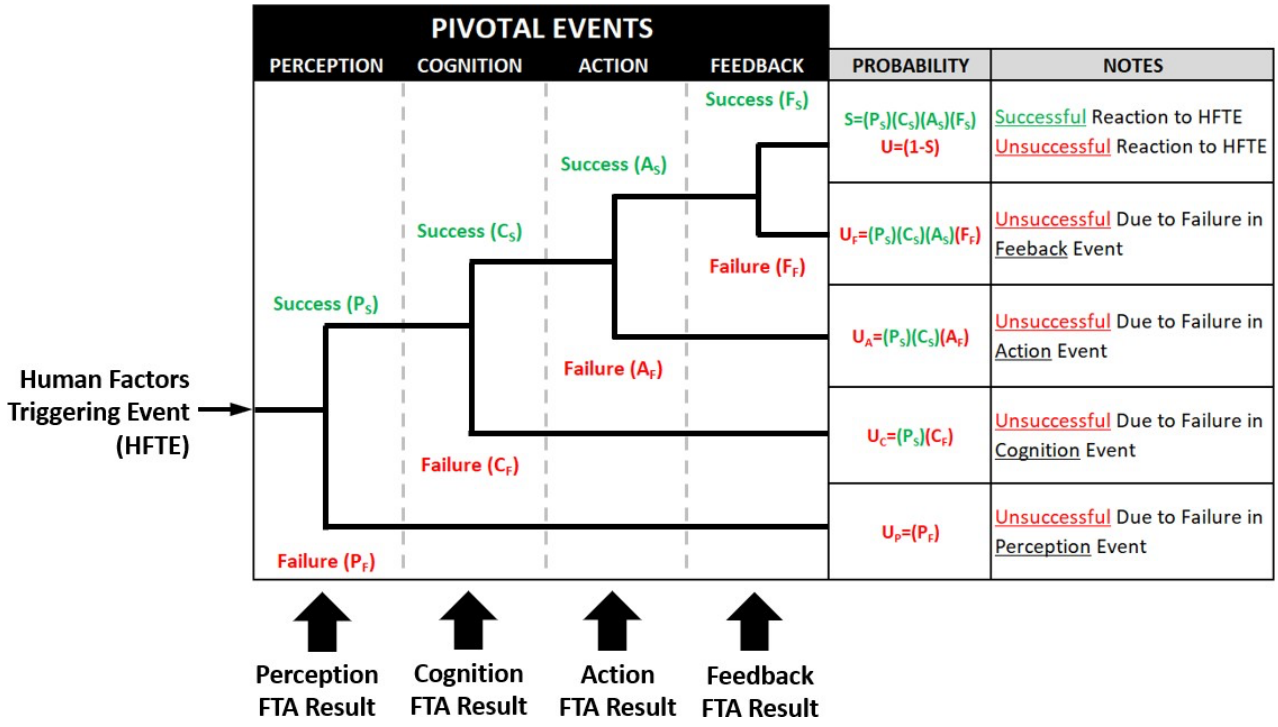


Figure 7: HFTE Sequential Processing Model ETA

probability of human error for the HFTE under consideration.

THE HUMAN FACTORS HAZARD MODEL (HFHM) AUTOMATED SOFTWARE INTERFACE

A large number of calculations are required to establish the Performance Shaping Factors (PSFs) and associated Human Error Probability (HEP) values that feed into the individual Fault Tree Analysis (FTA) models. Additionally, the associated quantity of calculations required to establish all intermediate and top-level probabilities in the FTA and ETA networks are also voluminous. Several thousand individual calculations are required to complete any single design iteration of the HFHM. Performing these calculations manually would require a large amount of time and would likely be prone to errors. The HFHM must therefore rely on a computational platform to efficiently produce results. Microsoft Excel, the spreadsheet software that is included as part of the standard MS Office software suite, was selected to be used as the analytical foundation of the HFHM model. MS Excel is commonly available, and many users are familiar with the software. The structure and functional flow of the HFHM within the spreadsheet software is presented in Figure 8.

As illustrated in Figure 8, Step (1) involves the primary user interface where information specific to the human factors being analyzed as well as characteristics of the system design are entered into the program. The information specified at this step is typically derived from three possible sources. These include:

- Source material (literature derived values from established HRA methods or documented human behavior databases).
- Expert Estimation values based on standardized value of HEP.
- User defined values as determined by the specific hazard scenario circumstances, experimentally derived data, or custom determined human error probabilities.

For programs that pull HEP data from published sources, Step (2) executes the algorithms that utilize the human factor (HF) and system factor (SF) data to define the relevant PSF's used to modify the various HEP's that are then passed to the FTAs of the four pivotal events. As previously discussed, the PSF modifying factors used to adjust baseline HEP values are defined in equations (1) and (2) above. If Expert Estimation or user specified probabilities are specified in Step (1), then the HEP data flows directly into Step (3) without modification. Step (3) includes all four FTAs used to predict the likelihood of failure due to the corresponding pivotal events, namely: Perception, Cognition, Action, and Feedback. The probabilities calculated in the FTAs of Step (3), are

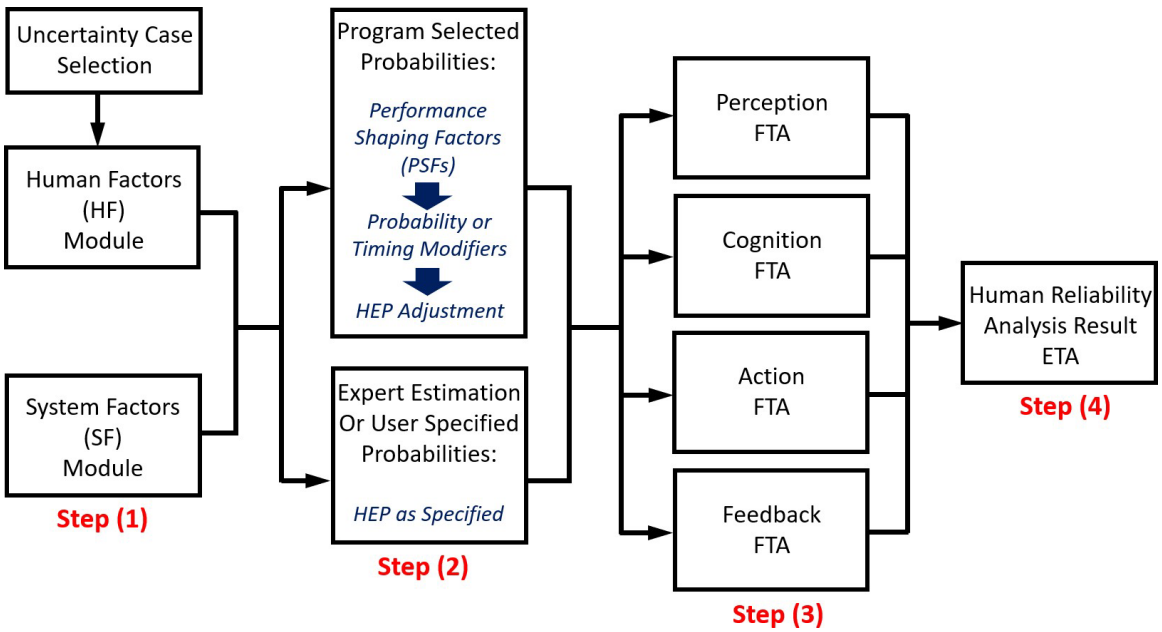


Figure 8: HFHM Software Functional Diagram

then passed to the Event Tree Analysis (ETA) in Step (4) to calculate the overall probability of success (and failure) attributable to the human actor's response the hazard event.

As with all probabilistic analyses, statistical uncertainty is present in all HEP determinations. Uncertainty within the Human Factors Hazard Model (HFHM) can be represented to model a maximum possible (worst case), minimum possible (best case), and most likely (nominal) probability of failure for Human Error Probability (HEP) calculations. Any of these three cases can be specified by the HFHM user in their initial analysis specification. As illustrated in Figure 8, based on user selection of best case, worst case, or most likely case, the entire series of HEP will be calculated and the probabilities will be reported accordingly in the HFHM. As recommended in the Technique for Human Error-Rate Prediction (THERP) [8], uncertainty in the HEP calculations is accomplished by using an Error Factor (EF), that is applied to the nominal (most likely) HEP value. The maximum possible probability of failure is calculated using:

$$P_{max} = P_{nom}(EF) \quad (4)$$

Where:

P_{max} = *Maximum Event Probability*
P_{nom} = *Nominal Event Probability*
EF = *Contributing Event Probability of Failure*

Using an identical Error Factor, the minimum possible probability of failure is calculated using:

$$P_{min} = \frac{P_{nom}}{(EF)} \quad (5)$$

Where:

P_{min} = *Minimum Event Probability*
P_{nom} = *Nominal Event Probability*
EF = *Contributing Event Probability of Failure*

The Error Factors used to establish uncertainty in the model are specified by the user in one of three ways: first, when published HEP data is utilized by the program the associated Error Factor is also selected from that source data. If a probability is selected from the standard Expert Estimation values, a corresponding standard Error Factor is automatically selected for the HEP value used. For user specified HEP entries, the analyst is also required to provide an associated Error Factor to use in the uncertainty calculation. The HFHM analyst selects which extreme case is desired to be calculated (best, or worst) at Step (1), and equations (4) and (5) are used to establish HEP values throughout the model, otherwise the original HEP value is utilized in the model for the most likely case.

VERIFICATION AND EVALUATION OF THE HUMAN FACTORS HAZARD MODEL (HFHM)

Verification is an important aspect of ensuring that a given simulation of a model is accurate and applicable for its intended uses. In this section, we document the verification of HFHM through its quantitative comparison to a baseline conceptual model of HEP developed using the Technique for Human Error-Rate Prediction (THERP). As one of the additional important aspects of HFHM is its usability, this research also executed a survey of systems engineers who would be expected to execute HFHM in industrial, and commercial settings. These can be analyzed numerically and narratively for evidence of the usability of the HFHM tool. Finally, HFHM is evaluated through demonstration in an industrial manufacturing test case.

VERIFICATION OF HFHM BY COMPARISON TO THERP

Quantitative verification of the HFHM analysis is supported through the direct comparison of HFHM results against results derived using THERP. A validation case was established incorporating elements typical of a Human Factors Triggering Event (HFTE). The HFTE included an assumed hypothetical hazard event, communication of system behavior to a human actor via visual and audible signals, cognitive processing by the actor to establish a corrective action response, control system input, and feedback resulting from control system input. A typical HRA event tree was established for the probability analysis per THERP methodology [5][8]. The validation case event tree is presented in Figure 9.

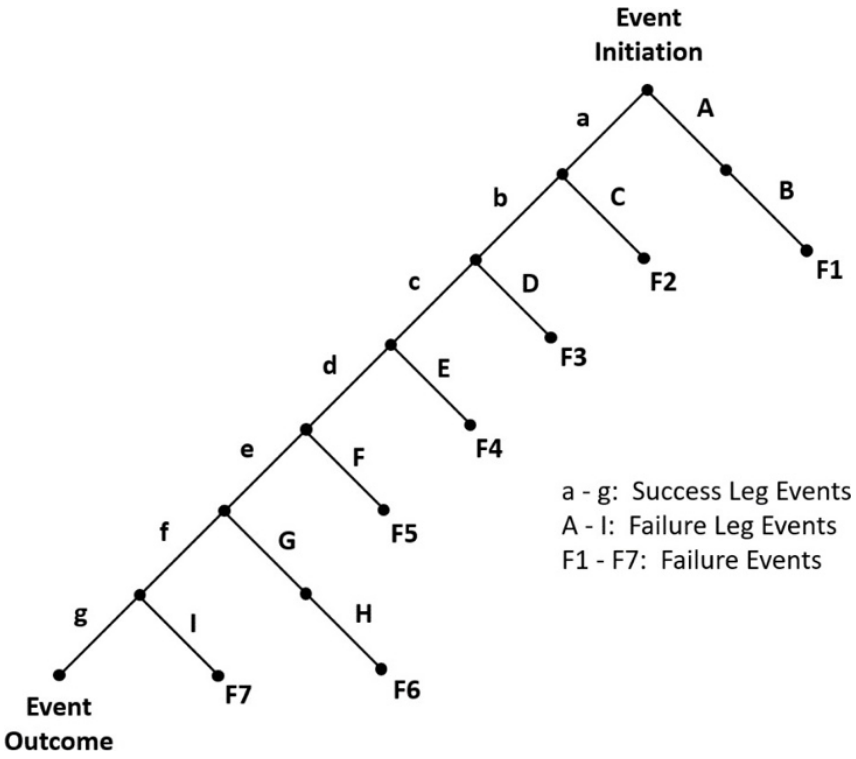


Figure 9: THERP Verification Model Event Tree

For the event specified in Figure 9, several opportunities for a failure related to human interaction with the system are detailed. The possible failure legs are labeled in the HRA event tree as event A-I. Each failure leg represents an opportunity for the human actor interacting with the system to correctly or incorrectly respond to system behavior. For each failure leg, the actor is required to either successfully receive a signal from the system, process that signal, provide appropriate input action to the system, or interpret system feedback relevant to the control input rendered. The various human-system interactions that correspond to these possible failure legs are presented in Table 6.

Several permutations of the baseline case were then established by modifying various human and system factors, thus altering the Performance Shaping Factors (PSFs). These factors include various actor stress levels, training and practice parameters related to the human actor, and instrumentation and control interface organization and ergonomics. Each new human and system factor noted establish revised PSFs that are then used to modify Human Error Probability (HEP) for the various permutations of the baseline

Table 6: THERP Verification Model Failure Leg Event Descriptions

PROBABILITY TREE - FAILURE LEG	DESCRIPTION
A	Recognize Alarm
B	Recognize Indicator Lamp
C	Read Pressure Gage (Analog Meter)
D	Diagnose Hazard
E	Actuate Control (Push Button)
F	Actuate Control (Rotary Dial)
G	Recognize Alarm Shut-Off
H	Recognize Indicator Shut-Off
I	Cease Control Input

Table 7: Human and System Factors Used to Establish PSFs

PERFORMANCE SHAPING FACTOR (PSF)	DESCRIPTION
1	System Training w/ Hazard Practice
2	System Training w/o Hazard Practice
3	Instrumentation & Controls are Organized or Stereotyped
4	Instrumentation & Controls are not Organized or Stereotyped
5	Optimum Stress
6	Extremely High Stress

analysis. As the human and system factors are adjusted in both analyses (THERP and HFHM), the resulting top-level probabilities of success and failure will adjust accordingly. A list of the various human and system factors used to establish the PSFs that are present in the comparative study are presented in Table 7.

As a result of the variations applied to the baseline case, a total of eight operational scenarios are evaluated using THERP and compared with corresponding HFHM analyses. The analysis results for each permutation of the baseline model, utilizing the baseline and updated values are presented in Table 8.

Good agreement between the failure probabilities as calculated by THERP and the HFHM are demonstrated in this verification study. The average variability between the THERP and HFHM probability of success results, over the eight different trial cases, is 4.8%. The ranges of variability between the THERP and HFHM solutions are between a minimum of 0.1% and a maximum 11.5% depending on the exact combinations of PSF employed in the analysis.

These results provide quantitative evidence of the applicability of HFHM to THERP's application domain, and a quantitative estimate of the verification error of HFHM relative to baseline tools in the field.

SUBJECT MATTER EXPERT EVALUATION OF THE HFHM

Evaluation of the Human Factors Hazard Model (HFHM) was accomplished via testing, assessment, and feedback provided by a total of six engineering Subject Matter Experts (SME) with professional positions. The assessment team consisted of personnel representing Systems Engineers and Systems Engineering managers of a large, publicly traded defense and aerospace corporation, production and plant design managers of a mid-sized, privately owned aerospace products corporation, and a former facilities operations manager of a large public university and current faculty member of the Construction Management department at a public 4-year university. A sample size of six evaluators is considered adequate to achieve meaningful feedback in eliciting qualitative input from highly qualified SMEs [24].

Table 8: THERP and HFHM Analysis Results Comparison for All Design Study Cases

	PROBABILITY OF SUCCESS							
	PERFORMANCE SHAPING FACTOR (PSF) COMBINATIONS							
PROBABILITY TREE - SUCCESS LEG	1-3-5	1-3-6	1-4-5	1-4-6	2-3-5	2-3-6	2-4-5	2-4-6
a	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
b	0.997	0.997	0.970	0.970	0.997	0.997	0.970	0.970
c	0.999	0.990	0.999	0.990	0.990	0.900	0.990	0.900
d	1.000	1.000	0.995	0.995	1.000	0.995	0.995	0.995
e	0.999	0.999	0.990	0.990	0.999	0.999	0.990	0.990
f	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
g	0.999	0.990	0.999	0.990	0.990	0.900	0.990	0.900
THERP Probability of Success=	0.994	0.976	0.954	0.936	0.976	0.803	0.936	0.774
HFHM Probability of Success=	0.986	0.908	0.973	0.895	0.950	0.878	0.938	0.863
Agreement Between THERP & HFHM=	0.7%	7.5%	2.0%	4.6%	2.7%	9.4%	0.1%	11.5%
							Average Agreement Between Approaches=	4.8%

None of the Systems Engineering SMEs are experts in HRA, and therefore are making comparisons and evaluations relative to their needs to design and build human-machine systems in the aerospace, construction, and manufacturing applications. The methods and results of their evaluations are summarized below.

The HFHM software was presented and demonstrated to the SMEs. The SMEs were then requested to test the software's functionality in their applications of interest, using their self-designed case studies representing the application of the model to their specific industries and job responsibilities. Following their evaluation of the HFHM, the participants responded to a survey to identify their opinions on the function and fit for purpose of the HFHM model and software. Each respondent answered an online survey utilizing standard 5-point Likert scale responses. All of the questions were worded such that the most desirable answers were in the "Strongly Agree" and "Somewhat Agree" categories. To quantify the survey responses for analytical comparison, a point system was established corresponding to each possible user response. Integer values were assigned to each response ranging from zero (least desirable response) to four (most desirable response), with two indicating a neutral opinion. A composite score for each survey question was then

Table 9: SME Survey Responses on a Likert Scale with Zero (0) Corresponding to Strongly Disagree, Two (2) Indicating Neutral, and Four (4) Corresponding to Strongly Agree

SURVEY QUESTION	COMPOSITE LIKERT SCORE (OUT OF 4.0)	COMMENTS
The Human Factors Hazard Model (HFHM) has an intuitive interface, it is well organized, and can be used efficiently with minimal training and practice.	3.3	Survey participants indicated general satisfaction with the software user interface and its overall usability.
The Human Factors Hazard Model (HFHM) can be utilized in a timely manner, and is able to generate a result in a timeframe useful to a safety analyst or engineer.	3.3	Survey participants generally found the HFHM approach can be utilized in an expedient manner, resulting in timely results.
The Human Factors Hazard Model (HFHM) is an effective tool for use in analysis related to human error probability within a system design, and would be useful in driving a standardized, uniform, and comparable approach to safety analysis.	3.8	Survey participants indicated a high level of confidence that the HFHM approach would be effective in standardizing human hazard response by making results comparable and uniform.
The Human Factors Hazard Model (HFHM) has a high degree of flexibility and can be used to analyze systems ranging from those that are very simple through those that are very complex.	3.7	Survey participants indicated a high level of confidence that the HFHM approach is sufficiently flexible that analyses ranging from simple systems to very complex systems can be evaluated.
The Human Factors Hazard Model (HFHM) will facilitate an efficient approach to design and trade studies as it relates to system safety analysis, and particularly, human factors within a system design.	3.3	Survey participants indicated a general level of confidence that the HFHM approach allows for an efficient approach to design and trade studies as they relate to human factors in system safety.
The Human Factors Hazard Model (HFHM) is useful in identifying potential safety oversights, as they relate to human factors, and can be used to help guide design activities to reduce risk associated with human actors being present in a system design.	3.7	Survey participants indicated a high level of confidence that the HFHM approach will act as a guide to system design activities related to risks associated with human factors.
The Human Factors Hazard Model (HFHM) has a high level of utility, and would improve my organization's ability to predict the hazards associated with human activity within a system, and reduce overall safety risk.	3.3	Survey participants indicate a general consensus that the HFHM approach is an improvement over their current approach to Human Reliability Analysis (HRA).

calculated. The survey questions, composite scores, and response commentary are presented in Table 9.

As noted in the table, the survey participants were posed with these seven questions eliciting their impressions and assessment of the HFHM and its functionality. The first question in the survey is regarding the software interface and general usability of the model. This question was intended to exclusively solicit user satisfaction (or dissatisfaction) with regards to the user friendliness of the program. This question was used to establish if follow-up inquiries were likely required to guide design of an improved user interface for future software versions. The other survey questions were intended to support the validation of the HFHM's ability to standardize, simplify, be flexible, timely to use, and provide an overall improvement, both in functionally, as well as in overall accuracy, to the current Human Reliability Analysis (HRA) techniques being employed by the user. Feedback from the SME team provides evidence that HFHM has utility in application to system safety analysis, particularly with regards to human factors and risk assessment in the industrial and commercial engineering applications favored by these SMEs.

EVALUATION OF HFHM BY APPLICATION TO A MANUFACTURING ENVIRONMENT HAZARD SCENARIO

Additional evaluation of the HFHM was conducted via a design study of a manufacturing system experiencing a malfunction that depends upon human intervention to recover successfully. For this hypothetical case, a workpiece is assumed to be manufactured using a semi-manual machine tool (lathe). In this type of machining operation, the work piece is turned on a rotational centerline, and material is removed using a shaped cutting insert. During the material removal, a fluid is discharged onto the insert and work piece to remove machining debris as well as lubricate and cool the workpiece and tooling. If the part envisioned in this study is machined too aggressively, or done so without adequate cooling, it risks the generation of an alpha case defect due to surface heating, thus damaging the part beyond salvaging. In this scenario it would be scrapped at a high cost to the company, thus constituting the hazard event.

The design study being used to assist in the HFHM evaluation considers the operator's (human actor's) reaction to an unexpected low coolant flow. The Human Factors Triggering Event (HFTE) is defined as during a normal machining operation, the system experiences a drop in coolant flow, which potentially endangers the component being manufactured. The low coolant flow can be the result of three different possible root causes. The low flow rate root causes include: 1) an obstruction in the flow path restricting the coolant flow, 2) insufficient pump flow (pressure and / or pumping capacity), or 3) a low fluid level in the supply reservoir, thus starving the system of coolant. The manufacturing system design being analyzed includes the machining mechanism, coolant tank and pumping hardware, the control panel / user interface, and a human actor. The human actor via the control panel provides input control to the machining center and coolant management system. The control panel also provides instrumentation feedback to the actor regarding system performance and operational parameters.

A baseline case is established with Performance Shaping Factors (PSFs) based on the human factors as well as system characteristics. Two subsequent updates to the design were then analyzed within the HFHM. The two updates reflected what would be considered improvements to the system safety, which should in turn reduce the hazard probability associated with human intervention in system operations. The HFHM results, including a breakdown of probability for all four pivotal events, were established for the baseline case and the two update analyses. These analysis results are presented in Table 10.

As noted in the table, the baseline analysis indicates a probability of success that the actor will react correctly to the Human Factors Triggering Event (HFTE) of approximately 25.2%. With improvements made to the control panel, as well as improved observability of the system operations, and lowered distraction and stress levels, the probability of success is increased to approximately 96.2%. With the final improvement specified in the second update being hazard simulation and practice related specifically to the undesired system behavior, a final probability of success is determined to be approximately 97.9%.

Table 10: Evaluation of HFHM by Application to a Manufacturing Environment Hazard Scenario Summary

HFHM DEFINITION AND PSF INPUTS			HFHM OUTPUTS				
DESIGN CASE & DESCRIPTION	PERFORMANCE SHAPING FACTORS		PROBABILITIES OF SUCCESS				
Baseline Case			**PERCEPTION**	**COGNITION**	**ACTION**	**FEEDBACK**	**OVERALL**
Machining operation coolant flow failure with a required human operator intervention to avoid a mishap.	Moderately young actor (30 yrs)	→	0.433	0.814	0.853	0.839	0.252
	No Impairment						
	No appreciable fatigue						
	Normal actor visual acuity						
	Typical actor reaction time						
	Actor trained and experienced with system operations						
	Actor has no practice with specific HFTE behavior						
	Event occurs early in shift (1st hour of 8 total hours)						
	Instrumentation not organized or stereotyped						
	Instrumentation not annunciated for hazard alert						
	No audible alarm for hazard alert						
	Input controls not organized or stereotyped						
	No direct observation of machining operation by actor						
	Moderately high stress						
	Moderate distraction level						
	No adverse environmental conditions to inhibit actor response to HFTE						
	System operations are considred to be simple to understand						
Update 1			**PERCEPTION**	**COGNITION**	**ACTION**	**FEEDBACK**	**OVERALL**
All characteristics carried over from the Baseline Case with the noted revsions.	Instrumentation organized and stereotyped	→	0.988	0.997	0.980	0.997	0.962
	Annuciated indicators added for HFTE behavior						
	Audible alarm added for HFTE behavior						
	Input controls organized and stereotyped						
	Viewport added for direct actor observation of machining process						
	System organized for simultaneous viewing of process, instrumentation, and input control						
	Optimal stress level						
	Low distraction level						
Update 2			**PERCEPTION**	**COGNITION**	**ACTION**	**FEEDBACK**	**OVERALL**
All characteristics carried over from the Baseline Case with the noted revsions.	Consistent practice of HFTE response by actor	→	0.997	0.996	0.989	0.997	0.979

These results support a hypothesis that as design improvements are implemented, the HFHM will predict an overall positive trend in hazard reduction related to human-system interaction[1]. HFHM was able to be used to perform this analysis in ~1hr of engineer time, and sensitivity analysis and design revision was performed automatically in minutes of additional effort. This can be contrasted to the time to develop a THERP or similar quantitative probabilistic HRA, which would be measured in 10s of hours.

CONCLUSIONS

The proposed Human Factors Hazard Model (HFHM) is intended to provide a simplified, standardized, broadly applicable, and repeatable approach to assessing human error probabilities and their relationship to mishaps. The model is based on established error probabilities and human performance characteristics that have been experimentally derived over the past several decades but embeds these into a multi-staged and feedback-enabled model of human psychomotor response that is more applicable to common industrial, commercial, and manufacturing conditions. The model makes allowances for Expert Estimation or case specific empirical data to be combined or substituted for Human Error Probability (HEP) data embedded in the base functionality. An Excel and SysML implementation allow for design and sensitivity studies to be quickly and efficiently performed.

Based on evaluation by industry experts, design studies, and quantitative verification relative to existing Human Reliability Analysis (HRA) methods, the HFHM generates results comparable to other established methods in conventional applications, has utility for system engineering activities, and is easy to apply to manufacturing, industrial and commercial applications. HFHM can help guide system design activities to minimize or eliminate those hazards before they are much more hazardous, difficult, or costly to manage.

AUTHORSHIP CONTRIBUTIONS

Dustin Birch: Conceptualization, Methodology, Investigation, Software, Formal Analysis, Writing – Original Draft Preparation. Thomas Bradley: Project Administration, Supervision, Visualization, Writing – Review and Editing. Erika Miller: Supervision, Visualization, Writing – Review and Editing.

COMPETING INTERESTS

All authors declare they have no potential competing interests.

ORCID IDS

Dustin S. Birch https://orcid.org/0000-0003-4066-2802
Erika E. Miller https://orcid.org/0000-0001-5009-9916
Thomas H. Bradley https://orcid.org/0000-0003-3533-293X

REFERENCES

[1] D.S. Birch, T.H. Bradley, Development of a Human Factors Hazard Model Using HEP / FTA / ETA, Wasatch Aerospace & Systems Engineering Conference (AIAA-INCOSE), 2021

[2] N. Siu, Dynamic Accident Sequence Analysis in PRA: A Comment on 'Human Reliability Analysis - Where Shoudst Thou Turn?', Reliability Engineering & System Safety, Volume 29, Issue 3, 1990 https://doi.org/10.1016/0951-8320(90)90019-J

[3] G. W. Hannaman, D.H. Worledge, Some Developments in Human Reliability Analysis - Approaches and Tools, Reliability Engineering & System Safety, Volume 22, Issus 1-4, 1988 https://doi.org/10.1016/0951-8320(88)90076-2

[4] A. Spurgin, Another View of the State of Human Reliability Analysis (HRA), Reliability Engineering & System Safety, Volume 29, Issue 3, 1990 https://doi.org/10.1016/0951-8320(90)90020-N

[5] N.A.A. Aziz, A. Fumoto, K. Suzuki, Assessing Human Error During Collecting a Hydrocarbon Sample of the Chemical Plant Using THERP, Journal of Fundamental and Applied Sciences, ISSN: 1112-9867, 2017

[1] It is important to note that the revisions made to the design and operational procedures to improve system safety will likely incur additional cost and potentially complicate the system, thus introducing other possible reliability concerns, etc. As such, for all system improvements specified, appropriate trade studies should be conducted to verify the net benefit of each revision. Note that the HFHM results only evaluate the issue from the perspective of a human actor's reaction to a hazard event.

[6] C.L. Ericson II, Hazard Analysis Techniques for System Safety, 2nd Edition, Wiley, 2016
[7] D.I. Gertman, H.S. Blackman, Human Reliability and Safety Analysis Data Handbook, 3rd Edition, Wiley-Interscience, 1993
[8] A.D. Swain, H.E. Guttmann, Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications, NUREG/CR-1278, SAND80-0200, 1983 https://doi.org/10.2172/5752058
[9] M.K. Comer, D.A. Seaver, W.G. Stillwell, C.D. Gaddy, "Generating Human Reliability Estimates Using Expert Judgment", NUREG/CR-3688 - SAND84-7115, VOL 1 & 2, 1984 https://doi.org/10.2172/6180932
[10] K.R. Boff, J.E. Lincoln, "Engineering Data Compendium: Human Perception and Performance", Harry G. Armstrong Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, 1988
[11] B.S. Dhillon, Human Reliability with Human Factors, Pergamon Press, 1986 https://doi.org/10.1016/B978-0-08-032774-7.50018-0
[12] S.J. Guastello, Human Factors Engineering and Ergonomics, 2nd Edition, CRC Press, 1986
[13] M.V. Stringfellow, Accident Analysis and Hazard Analysis for Human and Organizational Factors, PhD Dissertation, Massachusetts Institute of Technology, 2010
[14] R.B. Shirley, C. Smidts, M.Li, A. Gupta, Validating THERP: Assessing the Scope of a Full-Scale Validation of the Technique for Human Error Rate Prediction, Annals of Nuclear Energy, Vol. 77, Ohio State University, 2014 https://doi.org/10.1016/j.anucene.2014.10.017
[15] D.E. Embrey, P. Humphreys, E.A. Rosa, B. Kirwan, K. Rea, SLIM-MAUD: An Approach to Assessing Human Error Probabilities Using Structured Expert Judgment, United States Nuclear Regulatory Commission - Human Factors and Safeguards Branch, Office of Nuclear Regulatory Research, Contract No. DE-AC02-76CH00016 Fin. No. A-3219, 1984
[16] Pilot's Handbook of Aeronautical Knowledge, Federal Aviation Administration, U.S. Department of Transportation, FAA-H-8083-25B, 2006
[17] Department of Defense Standard Practice - System Safety, MIL-STD-882E, Revision E, 2012
[18] M. Stamatelos, J. Caraballo, W. Vesely, J. Dugan, J. Fragola, J. Minarick, J. Ralsback, Fault Tree Handbook with Aerospace Applications, NASA Office of Safety and Mission Assurance, V 1.1, 2002
[19] G. Biggs, K. Post, A. Armonas, N. Yakymets, T. Juknevicius, A. Berres, OMG Standard for Integrating Safety and Reliability Analysis into MBSE: Concepts and Applications, INCOSE International Symposium, Volume 29, Issue 1, 2019 https://doi.org/10.1002/j.2334-5837.2019.00595.x
[20] E. Schlosser, Command and Control: Nuclear Weapons, the Damascus Accident, and the Illusion of Safety, Penguin Books, 2013
[21] Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey, National Highway Traffic Safety Administration, U.S. Department of Transportation, DOT HS 812 115, 2015
[22] M. Rauterberg, Perception, Cognition, Action: an Action Theoretical Approach, Systematica, Volume 14, Number 1, 1999
[23] D.W. Carruth, M.D. Thomas, B. Robbins, A. Morais, Integrating Perception, Cognition, and Action for Digital Human Modeling, Digital Human Modeling, Lecture Notes in Computer Science, Volume 4561, 2007 https://doi.org/10.1007/978-3-540-73321-8_39
[24] G.J. Burkholder, K.A. Cox, L.M. Crawford, J.H. Hitchcock, Research Design and Methods - An Applied Guide for the Scholar-Practitioner, Sage Publications, 2020
[25] A.M. Williams, K.A. Ericsson, "Introduction to the Theme Issue: Perception, Cognition, Action, and Skilled Performance", Journal of Motor Behavior, Vol. 39 No. 5, 2007 https://doi.org/10.3200/JMBR.39.5.338-340
[26] S.M.L. Hendrickson, G.W. Parry, J.A. Forester, V.N. Dang, A.M. Whaley, S. Lewis, E. Lois, J. Xing, "Towards an Improved HRA Method", Sandia National Laboratory, SAND2012-1319C, 2012
[27] N.J. Ekanem, A. Mosleh, S. Shen, "Phoenix - A Model-Based Human Reliability Analysis Methodology: Qualitative Analysis Procedure", Reliability Engineering and System Safety, Vol. 145, 2016 https://doi.org/10.1016/j.ress.2015.07.009
[28] D. Gertman, H. Blackman, J. Marble, J. Byers, C. Smith, "The SPAR-H Human Reliability Analysis Method", U.S. Nuclear Regulatory Commission, NUREG/CR-6883, 2005
[29] J.R. Anderson, D. Bothel, M.D. Byrne, S. Douglass, C. Lebiere, Y. Qin, "An Integrated Theory of the Mind", Psychological Review, Vol. 111 No. 4, 2004 https://doi.org/10.1037/0033-295X.111.4.1036
[30] HFHM Project, https://doi.org/10.5281/zenodo.7352422

Original Article
https://doi.org/10.56094/jss.v58i2.253

International System Safety Society

www.jsystemsafety.com

Journal of System Safety

Established 1965 Vol. 58 No. 2 (2023)

Proposing the Use of Hazard Analysis for Machine Learning Data Sets

H. Glenn Carter[b], Alexander Chan[b], Chris Vinegar[b], Jason Rupert[ac]

[a] *Corresponding author email: mailto:jason.rupert@mtsi-va.com*
[b] *U.S. Army Combat Capabilities Development Command Aviation & Missile Center (DEVCOM AvMC); Redstone Arsenal, AL USA*
[c] *Modern Technology Solutions, Inc.; Huntsville, AL USA*

Keywords
machine learning, data assurance, data governance

Peer-Reviewed
Gold Open Access
Zero APC Fees

Online: 22-Jun-2023

Cite As:
Carter, H.G. et al. Proposing the Use of Hazard Analysis for Machine Learning Data Sets. Journal of System Safety. 2023;58(2):30-39. https://doi.org/10.56094/jss.v58i2.253

ABSTRACT

There is no debating the importance of data for artificial intelligence. The behavior of data-driven machine learning models is determined by the data set, or as the old adage states: "garbage in, garbage out (GIGO)." While the machine learning community is still debating which techniques are necessary and sufficient to assess the adequacy of data sets, they agree some techniques are necessary. In general, most of the techniques being considered focus on evaluating the volumes of attributes. Those attributes are evaluated with respect to anticipated counts of attributes without considering the safety concerns associated with those attributes. This paper explores those techniques to identify instances of too little data and incorrect attributes. Those techniques are important; however, for safety critical applications, the assurance analyst also needs to understand the safety impact of not having specific attributes present in the machine learning data sets. To provide that information, this paper proposes a new technique the authors call data hazard analysis. The data hazard analysis provides an approach to qualitatively analyze the training data set to reduce the risk associated with the GIGO.

INTRODUCTION

This paper focuses on a critical building block on the path to certifying machine learning software items - establishing assurance practices for the data set used to train, validate, and test the machine learning models. Key to addressing data assurance concerns associated with certifying machine learning is conducting the hazard analysis of data sets and assuring the adequacy of the data set. Thus, this paper works through what makes up data assurance for machine learning and devotes additional time on establishing hazard assessment artifacts for the data set. This paper also presents some techniques the industry is proposing for conducting data set adequacy, completeness, and representativeness, as well as an example of data hazard analysis.

OUTLINE

An introduction to highlights of traditional software assurance is provided, which includes a comparison of what type of additional assurance is needed for machine learning, where data assurance plays a key role. After that introduction, what is necessary to successfully accomplish data assurance is covered, where data hazard assessment plays a foundational role. Given that foundational role, additional time is spent in this paper proposing what would be necessary for data hazard assessment. This topic is presented to the safety community to generate discussion and engagement. There are certainly additions that should be made to the approach, and we hope the introduction of this concept generates some of that feedback and recommendations. Also, with this introduction we hope to begin to prepare the safety community for the arrival of data assurance techniques, and their role in the certification of machine learning based software items.

BACKGROUND

As indicated by SAE International Aerospace Information Report (AIR) 6988 (Artificial Intelligence in Aeronautical Systems: Statement of Concerns, AIR6988™, 2021) and Aerospace Vehicle Systems Institute (AVSI) AFE-87 (AFE 87 Project Members, 2020), the traditional aviation framework certification guidance is not adequate for the uncertainty added by the probabilistic development techniques used by machine learning:

"Industry standard development assurance processes such as ED-12C/DO-178C, ED-109A/DO-278A, ED-80/DO-254, do not have guidance for AI techniques such as Machine Learning algorithms. For some AI techniques, it may not be possible to meet all ED-12C/DO-178C, ED-109A/DO-278A and ED-80/DO-254 objectives such as those associated with the low-level requirements, implementation, integration, and verification activities. For artificial neural networks, there may be no meaningful representation of the internal structure of Machine Learning algorithm." (Artificial Intelligence in Aeronautical Systems: Statement of Concerns, AIR6988™, 2021)

Moreover, AVSI AFE-87 indicates the "fundamentally different nature of data-based systems", i.e., machine learning. AFE-87 goes on to indicate, "Traditional physical models are explicitly constrained, while data driven models are implicitly constrained by the observed phenomenon in the training data."

Our approach for the development of airworthiness certification guidance for machine learning considers the recommendations laid out in AIR6988 for establishing a framework for AI/ML, and also that of the AVSI AFE-87, SAE International Aeronautical Standard (AS) AS-6983 (SAE G-34, 2022), EASA Level 1 (Soudain, 2021), and SCSC-153B (The SCSC Safety of Autonomous Systems Working Group (SASWG), 2022).

TRADITIONAL SOFTWARE ITEM ASSURANCE

In general, the traditional software item assurance approach can be summarized as shown in Figure 1. Of course, Figure 1 is a bit of an oversimplification for the purposes of this paper. Other critical ML-based system assurance processes are not shown because they are similar to assurance processes for traditional systems. These include planning process, configuration management process, quality assurance process, and certification liaison processes. Processes not shown in Figure 1 but included in requirement assurance are high-level requirements (HLRs) processes, low-level requirements (LLRs) processes and the bi-traceability between HLRs and LLRs. In addition, requirement assurance includes the bi-directional traceability from HLRs to system/sub-

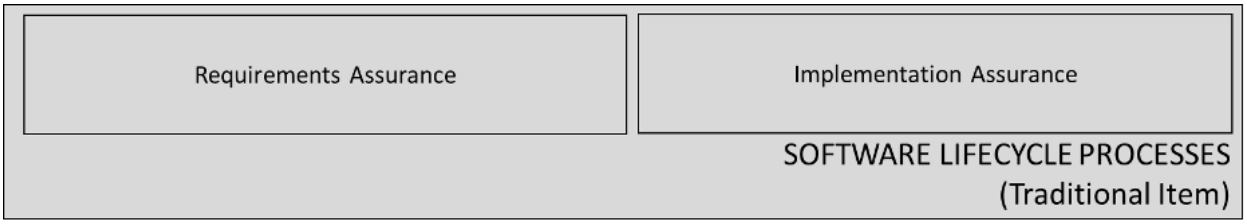


Figure 1: Traditional Software Lifecycle Assurance

Note for Figure 1: Grey fill is used to illustrate existing traditional processes, i.e., processes covered under existing processes and standards. For example, software lifecycle is covered under RTCA DO-178C (SC-205, 2011).

system requirements. Implementation assurance includes design, coding, verification, and implementation, and the appropriate bi-directional traceability between those processes. It is through the execution of the objectives and activities associated with those processes that assurances are provided for the software item to ensure it will "perform [its] intended functions under all foreseeable operating conditions." (14 CFR 25.1309 Equipment, systems, and installations.) Similar quotes are applicable for CFR Parts 23, 27, and 29.

MACHINE LEARNING SOFTWARE ITEM ASSURANCE

As indicated in AIR 6988 and AFE-87, additions are necessary to the traditional software lifecycle process assurance approach to account for the unique aspects of data-driven machine learning software development techniques. Grey fill was used in Figure 1 to indicate traditional software development processes, while in Figure 2 white fill is used to indicate necessary modifications or additions. For machine learning based software item development, data assurance and learning assurance are necessary assurance additions. In addition, enhancements are necessary to the processes that enable requirements and implementation assurance. Enhancements will also be necessary to the planning, configuration management, quality assurance, and certification liaison processes, but those are beyond the scope of this paper.

For machine learning-based software item assurance, the traditional software item assurance of requirement and implementation assurance processes will be augmented by the addition of data and learning assurance. We propose that data assurance consists of ensuring, for example, the training, verification, and test data set correctness, completeness, and representativeness of the operational design domain. Correctness, completeness, and representativeness of the operational design domain are three attributes of the data set that will determine the accuracy and performance of a machine learning model in the operational design domain.

Learning assurance consists of activities to confirm the intended machine learning model generalization performance is reached, e.g., not underfitting or overfitting, not being susceptibility to bias or drift, and appropriate behavior for out of distribution samples. Underfitting occurs when unacceptable error occurs during model validation. This is often a symptom model susceptibility to bias and is an indication of too small of a training data set. Overfitting occurs when validation error is low, but test error is high. Overfitting is an indication that the model has memorized the training and validation set, i.e., is fitting the variance noise in the data, but is not generalizing. Addressing the expectations associated with the machine learning model requirements and learning assurance is out of scope of this paper but will be addressed in follow-on papers.

Such follow-on work will go through the new machine learning development lifecycle (MLDL), which augments the machine learning implementation lifecycle (MLIL). The machine learning development lifecycle includes the processes to ensure the new development assurance expectations for machine learning are met.

Data-driven machine learning software development does not use traditional software development methodologies. That is, data-driven machine learning development does not develop implementation source code and parameter values directly from low-level requirements (LLRs). Instead, machine learning trains a machine learning model, which is a set of hyperparameters, neurons, and layers from a training data set. The data set and machine learning model are based on a set of data and model requirements. The machine learning data requirements are used to drive the data collection process, where the data set must be correct, complete,

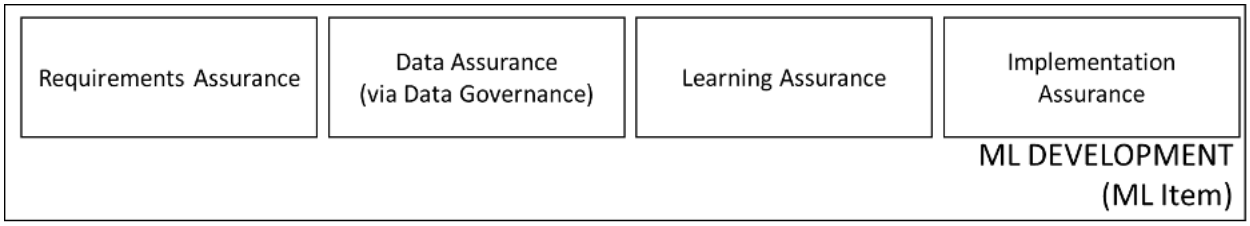


Figure 2: Machine Learning Development Lifecycle Assurance

and representative of the deployment operational design domain. Once collected, processed, and curated, the data set is divided into three independent subsets that are used to train, validate, and test machine learning algorithms. We call the approach to ensure the data set is adequate (i.e., correct, complete, and representative, and collected, processed, and curated correctly, and split appropriately) data assurance, where data governance is the process to ensure data assurance. In this paper, learning assurance is the term used for adequately designing, developing, training, validating, and testing the machine model.

Both new types of assurances (data and learning) are necessary to ensure machine learning is developed in a mature way to consider its use in flight and safety critical applications. Moreover, the addition of these two assurance approaches compensates for the loss of some traditional software development objectives and activities, e.g., loss of low-level requirements, and meaningfulness of traceability and design detail. The loss of the assurance provided by those traditional software development objectives and activities must be accounted for when contemplating the possible use of machine learning in flight and safety critical applications.

DATA GOVERNANCE

Data is one of the bigger technical debts (D. Sculley, 2015) of the machine learning processing: "Data Dependencies Cost More than Code Dependencies" and "Changing Anything Changes Everything (CACE)". These technical debts, i.e., resources and risks, are spread across all the processes associated with data governance, i.e., data planning to allocation. Data collection alone spans various types of data acquisition which could involve discovering existing collected data sets or synthetic generation and augmentation of data sets. After the collection process, the preparation and processing begin where labeling and other improvements are necessary. The labeling can be an intensive and technical process involving manual labeling or a semi-supervised labeling technique. Where necessary, improvements to the data may be necessary, which can also be intensive. Through each of these processes, care must be taken to maintain the data set's validity and authenticity. The data set has a large impact on the performance of the model properly reflecting the required generalization behavior in the operational design domain. Benign and even imperceptible modifications to data can cause unexpected, unanticipated, and undesired behavior of the models when exposed to deployed operational design domain native data sources.

As shown in Figure 3, the data governance process manages the data's sourcing, collection, processing, hazard assessment, and allocation. Data Governance provides the following processes, objectives, and activities to enable data assurance: data integrity, data hazard assessment, data planning, data completeness, data representativeness, data accuracy, correctness, data traceability, data reproducibility (i.e., collection, augmenting, transformation, labelling), dataset independence, data verification, data configuration management (e.g., corruption guards). The Configuration Management Process addresses the data configuration management, and the data verification is addressed by the machine learning verification process.

Notes for Figure 3:

- Note 1: Data set includes the features, attributes, and classes as well as the samples, signals, sources, and collection of those features, attributes, and classes.
- Note 2: Data configuration management, executed in the ML Configuration Management process, will ensure data is only used appropriately, e.g., avoiding data leakage, via methods like blockchain, and data integrity.
- Note 3: Data verification, executed in the ML Verification Process, will ensure the data is adequate, appropriate, representative, and complete as described in the Data Requirements.

Updates to the data set output from ML data governance processes may be driven by the ML Model Development Process or ML Verification Process. Should those updates occur, the ML data requirements, system safety requirements, and operational design domain requirements should be re-examined to determine if updates are necessary to those as well. Because requirements-based testing should be used for flight and safety-critical

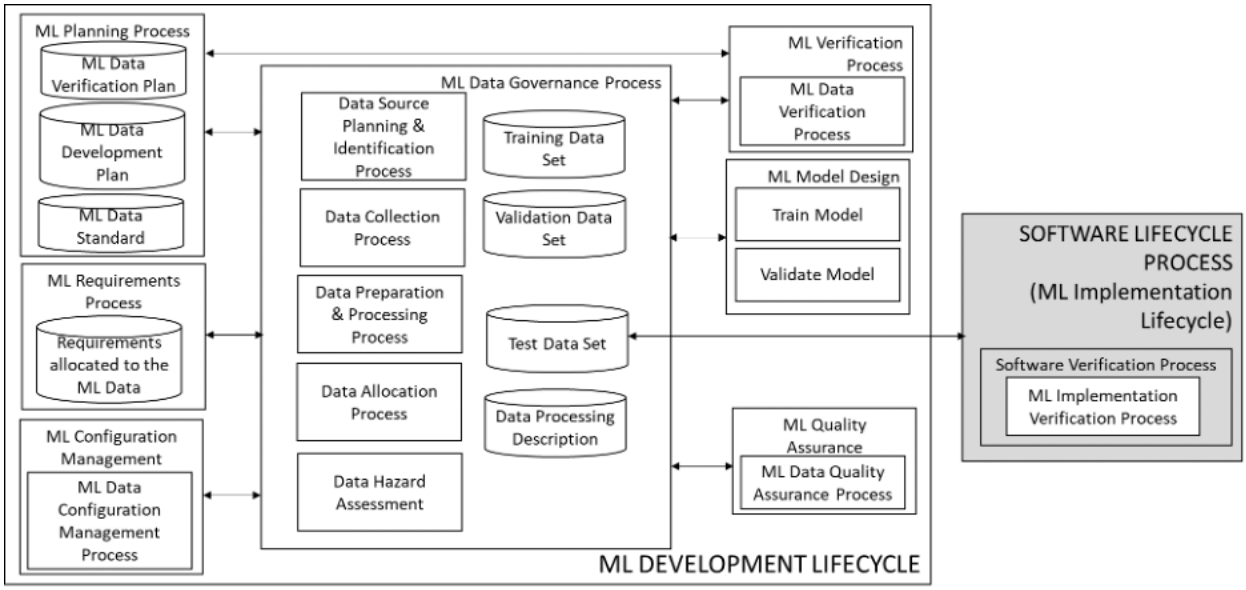


Figure 3: ML Data Governance Process

applications, any updates to those requirements should also be reflected in updates to verification cases. Because data set changes can be costly, all efforts should be made to correctly produce fully representative data set requirements and complete data sets as early in the process as possible.

Except for data hazard assessment, detailed definitions of each of these attributes of data assurance is beyond the scope of this of this paper but will be covered in general. Follow-on work is necessary to fully define the expectation for each of these attributes. The follow-on work will look to luminary guidance like that provided by SCSC-127G Data Safety Guidance (Version 3.4) (Data Safety Initiative Working Group, 2022).

The following sections specifically focus on ensuring data safety and data completeness and representativeness, accuracy, and correctness through the use of data hazard analysis and data verification techniques.

DATA HAZARD ASSESSMENT PROCESS

Data hazard assessment is a hazard assessment process that leverages techniques applied to traditional system and software hazard analysis techniques. In addition, the approach introduces novel techniques to assess the hazard impacts associated with the use of data sets for machine learning model training, validation, and testing.

Figure 4 shows the bi-directional traceability of the data hazard assessment to the traditional system safety hazard assessment process (shown in grey fill), e.g., those associated with and identified in SAE International Aerospace Recommended Practice (ARP) 4754 (S-18, 2010) and ARP 4761 (S-18, 1996). Analysis is on-going to determine if and how traditional hazard assessment processes may need to be augmented for machine learning based systems, e.g., accounting for autonomy level (classification) and methodologies may impact the functional hazard assessment (Copeland, 2019) or development assurance levels.

Notes for Figure 4:

- Note 1: Data includes the features, attributes, and classes present in the data as well as the samples, signals, sources, and collection of those features, attributes, and classes.
- Note 2: Data is most applicable to data-driven ML techniques; however, similar techniques more applicable to reinforcement learning will be specifically covered in the future, e.g., scenario planning, assessment, and verification.

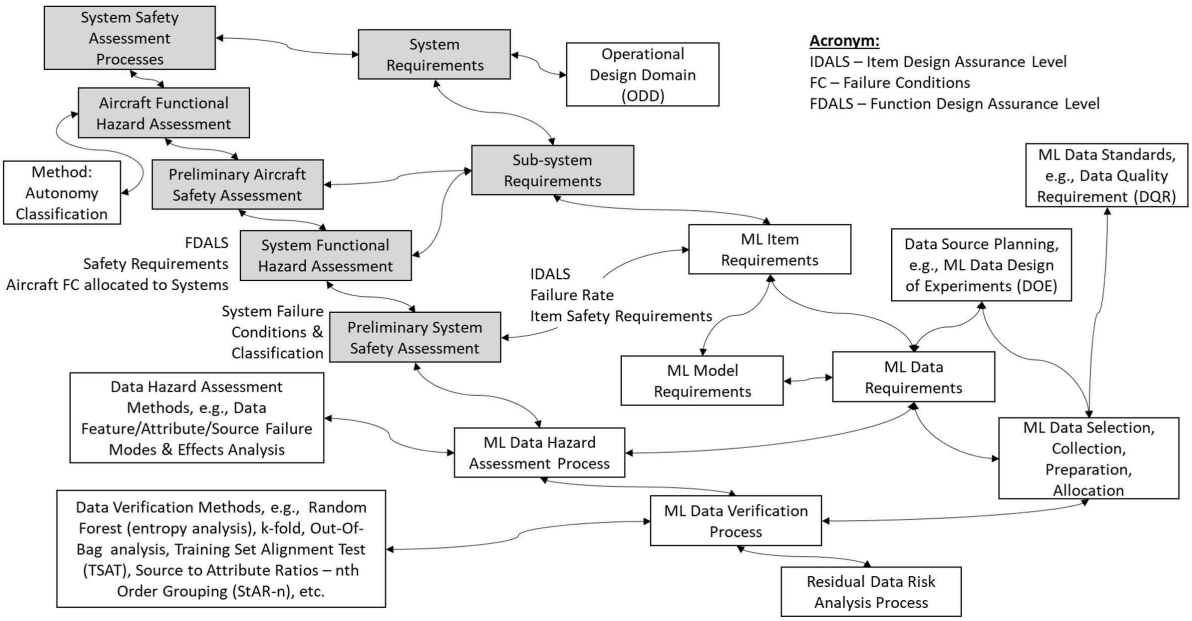


Figure 4: Hazard Assessment Processes

In the discussion that follows, the failure modes and effects analysis (FMEA) is offered up as one type of data hazard analysis technique, but others might be more appropriate, e.g., SCSC-127G proposes the establishment and use of the Data Safety Assurance Level (DSAL).

As indicated in ARP 4761: "An FMEA is a systematic, bottom-up method of identifying the failure modes of a system, function, or item and determining the effects on the next higher level. It may be performed at any level within the system (e.g., piece-part, function, black box). Software can also be analyzed qualitatively using a functional FMEA approach. Typically, an FMEA is used to identify failure effects resulting from single failures." An additional resource for the application and development of the FMEA is SAE International ARP 5580 (Recommended Failure Modes and Effects Analysis (FMEA) Practices for Non-Automobile Applications, ARP 5580, 2020).

As shown in Figure 5, for the discussion in this paper, the machine learning data FMEA is a data hazard analysis technique that would be used to assess the failure effects of the data, e.g., attributes, features, signals and sources, and the hazard effect on the ML model.

As shown in Figure 5, inputs to the ML Data FMEA include the following:

- As shown in the grey boxes, the systems-level hazard assessment artifacts, including the systems functional hazard assessment which should assess the hazard impacts of the functions within the ML-based system.
- From the ML requirements process, the ML data requirements.
- From the ML planning process, the ML data source planning and identification process artifacts, which may include data design of experiments type artifacts indicating what data is necessary and why.

Other artifacts that influence the machine learning data FMEA are the following:

- ML Data Development Plan
- ML Data Standard
- ML Data Requirements Standard

Data FMEA is a safety assessment of data features, attributes and sources, samples, and signals anticipated to drive the AI/ML model generalization in its operational design domain. Each of those is analyzed similarly to the approach used by a Software Interface FMEA described in ARP5580. Ultimately,

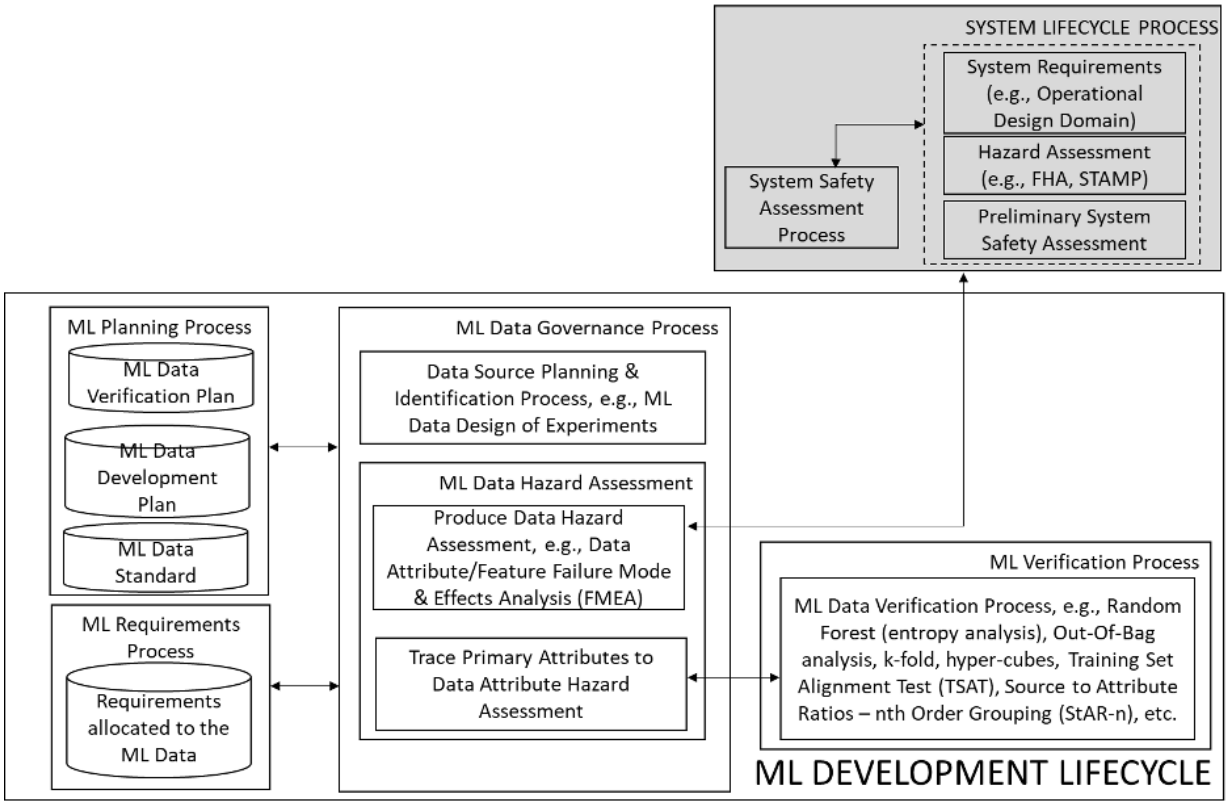


Figure 5: Data Hazard Assessment

the Data FMEAs may be summarized into a Data Failure Modes Effect Summary (FMES) to support the failure modes analysis considerations. Given the sensitivity of the machine learning model to input data sources, developing the data FMEA or its equivalent, is encouraged.

As an example, Table 1, from the Computer Vision- Hazard and Operability Study (CV-HAZOP) (Oliver Zendel, 2017), shows a type of data FMEA establishing traceability from failures (errors) in the data attributes, features, samples, sources, and signals to resultant impact of the ML model. The table identifies the source (location) and feature/attribute (guide word) and the consequences and risk.

Additional steps would be to indicate the expected machine learning effect, and eventually, if the data is available, the probability of each. Currently the CV-HAZOP has 1,469 entries. Such a systematic approach by computer vision experts allows for the machine learning based software item to be appropriately representative and complete data sets to be collected and used for machine learning training, validation, and testing. Data sets used can be assessed against the data hazard assessment to determine if all negative consequence and high risks effects are covered. Where gaps in the data set exist, appropriate mitigations can be determined, e.g., creation of synthetic data, creation of a derived subsystem requirement to mitigate, or other. Any gaps that remain in the data set should be indicated in the ML data processing and MLDL verification output data item and brought to the attention of the certification authority.

With respect to chronology, a data hazard assessment occurs before data source identification and collection, so as to drive the collection of safety critical features/attributes over those less so.

Similar methodology would be employed for reinforcement learning, but instead of data set/signal attributes/features, the reinforcement training scenario attribute/features would be analyzed.

Table 1: Example Data FMEA

Risk Id	Location	Guide Word	Parameter	Meaning	Consequence	Risk
0	Light Sources	No (not none)	Number	No light sources	No light available	Sensor will receive no light, but thermal noise or black current can cause wrong input
1	Light Sources	More (more of, higher)	Number	Many light sources (more light sources than expected)	Too much light	Overexposure (of whole image)
2	Light Sources	More (more of, higher)	Number	Many light sources (more light sources than expected)	Too few shadows	Algorithms using shadows can be confused
3	Light Sources	Less (less of, lower)	Number	Few light sources (fewer light sources than expected)	Too faint light (in parts of the scene)	Sensor will receive too faint light from some scene regions
4	Light Sources	Less (less of, lower)	Number	Few light sources (fewer light sources than expected)	Too many shadows	Algorithms can be confused by shadows
5	Light Sources	Less (less of, lower)	Number	Few light sources (fewer light sources than expected)	Very sharp shadows	
6	Light Sources	As well as	Number	Mirrors fake additional light sources	Light sources can appear at locations other than where they are	Algorithm confuses position of light sources

DATA VERIFICATION ASSESSMENT PROCESS

Statistical data verification assessment of machine learning data sets is an emerging field, so various methods are mentioned where their applicability depends on the situation. For this paper, a few different techniques will be mentioned with appropriate references to guide their application:

- Random Forest, e.g., feature importance
- Clustering, e.g., feature redundance (Tabular Modeling Deep Dive, 2022)
- k-fold cross-validation
- Training Set Alignment Test (TSAT) (Nagy, 2021)
- Source to Attribute Ratios – nth Order Grouping (StAR-n)
- hyper-cubes (focus - data completeness) (Kevin Fuchs, 2016)
- distribution discriminator framework (out-of-distribution)

The machine learning based software item developer may have different techniques they prefer. In such a situation the vendor should indicate their selection. The evaluation and justification of the statistical relevance of the data set should be conducted regardless of the approach for determining such validity. The goal of these approaches is to quantitatively show, through statistical analysis, that the data set selected, e.g., samples, signals, sources, attributes, and features, contains a complete representation of the operational design domain. While the method matters, more important is that the processes is pursued. Approaches to present a valid statistical representation of the data set may involve the following techniques or others unique to the vendor's approach:

- Exploratory Data Analysis (Brillinger, 2011)
- Boxing clever (Rob Ashmore, 2018)
- Datasheets for datasets (Timnit Gebru, 2021)

These machine learning data set verification results should present these statistical examinations of the data sets. The results should explain where the data set does not statistically fulfill the requirements associated with the operational design domain. The goal of these approaches is to ensure the proper data sets were collected, so these processes are complementary to the data hazard assessment processes.

CONCLUSION

For machine learning the data set is critical and ultimately determines how well the machine learning model generalizes on previously unseen data when deployed in complex operational design domain. Data assurance, specifically data hazard assessment and data verification, is a necessary assurance addition to the certification of machine learning based software items. Data assurance provides the necessary confidence that the data set is adequate, complete, and representative of the operational design domain. The data hazard assessment determines the impact of features, attributes and sources, samples, and signals. Through this process, the data hazard assessment provides guidance for the collections of features, attributes and sources, samples, and signals that should be present in the data set. The data hazard assessment process output will be used to guide the data governance collection and processing processes to help ensure data set adequacy, completeness, and representativeness. The complementary data set verification process ensures those features, attributes and sources, samples, and signals were collected. Through the addition of the data assurance process, and others to be addressed more thoroughly in follow-on work, the assurance community can begin to consider the inclusion of data-driven machine learning based software items in flight and safety critical applications.

DISCLAIMER

This paper is for information/education purposes only and does not provide the official position of U.S. Army Combat Capabilities Development Command Aviation & Missile Center (DEVCOM AvMC) with respect to establishing the Airworthiness Assurance argument for Artificial Intelligence and Machine Learning. Review of the document and associated briefing for Public Release was successfully completed (ID 6995).

AUTHORSHIP CONTRIBUTIONS

H. Glenn Carter: Funding acquisition, Conceptualization, Supervision, Writing - Review & Editing; Alexander Chan: Supervision, Writing - Review & Editing; Chris Vinegar: Supervision, Writing - Review & Editing; Jason Rupert: Writing - Original Draft.

COMPETING INTERESTS

This work was funded through the U.S. Army Combat Capabilities Development Command Aviation & Missile Center (DEVCOM AvMC). The authors declare they have no potential competing interests.

ORCID IDS

Jason Rupert https://orcid.org/0009-0004-5778-4747

REFERENCES

[1] AFE 87 Project Members. (2020). Machine Learning, AFE-87. College Station: Aerospace Vehicle Systems Institute. Retrieved June 1, 2022, from https://avsi.aero/projects/current-projects/cert-of-ml-systems/afe-87-machine-learning/

[2] Brillinger, D. R. (2011). Data Analysis, Exploratory. Retrieved June 1, 2022, from https://www.stat.berkeley.edu/~brill/Papers/EDASage.pdf

[3] Copeland, R. (2019). An Analysis and Classification Process towards the Qualification of Autonomous Systems in Army Aviation. Vertical Flight Society's 75th Annual Forum & Technology Display. Philadelphia. Retrieved from https://vtol.org/store/product/an-analysis-and-classification-process-towards-the-qualification-of-autonomous-systems-in-army-aviation-14727.cfm

[4] D. Sculley, G. H.-F. (2015). Hidden Technical Debt in Machine Learning Systems. Advances in Neural Information Processing Systems 28.

[5] Data Safety Initiative Working Group. (2022). Data Safety Guidance (Version 3.4). Safety-Critical Systems Club. Retrieved June 1, 2022, from https://scsc.uk/scsc-127G

[6] Kevin Fuchs, P. A. (2016). INTUITEL and the Hypercube Model - Developing Adaptive Learning. SYSTEMICS, CYBERNETICS AND INFORMATICS, 14(3), 7-11. Retrieved June 1, 2022, from http://iiisci.org/journal/pdv/sci/pdfs/EA039OY16.pdf

[7] Nagy, B. (2021). Increasing Confidence in Machine Learned (ML) Functional Behavior during Artificial Intelligence (AI) Development using Training Data Set Measurements. Acquisition Research Program. Retrieved June 1, 2022, from https://dair.nps.edu/handle/123456789/4393

[8] Oliver Zendel, K. H. (2017). Analyzing Computer Vision Data — The Good, the Bad and the Ugly. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/CVPR.2017.706

[9] Recommended Failure Modes and Effects Analysis (FMEA) Practices for Non-Automobile Applications, ARP 5580. (2020). SAE International.

[10] Rob Ashmore, M. H. (2018). "Boxing Clever": Practical Techniques for Gaining Insights into Training Data and Monitoring Distribution Shift. SAFECOMP 2018 Workshops, LNCS 11094, 393–405. Retrieved June 1, 2022, from https://doi.org/10.1007/978-3-319-99229-7_33

[11] S-18. (1996). Guidelines for Conducting the Safety Assessment Process on Civil Aircraft, Systems, and Equipment, ARP4761. SAE International.

[12] S-18. (2010). Guidelines for Development of Civil Aircraft and Systems, ARP4754A. SAE International.

[13] SAE G-34. (2021). Artificial Intelligence in Aeronautical Systems: Statement of Concerns, AIR6988™. SAE International. Retrieved June 1, 2022, from https://www.sae.org/standards/content/air6988/

[14] SAE G-34. (2022). Process Standard for Development and Certification/Approval of Aeronautical Safety-Related Products Implementing AI, AS6983. SAE International.

[15] Safety of Autonomous Systems Working Group. (2022). Safety Assurance Objectives for Autonomous Systems V3, SCSC-153B. Safety Critical Systems Club. Retrieved June 1, 2022, from https://scsc.uk/SCSC-153B

[16] SC-205. (2011). Software Considerations in Airborne Systems, DO-178C. Washington: RTCA, Inc.

[17] Soudain, G. (2021). First usable guidance for Level 1 machine learning applications. European Union Aviation Safety Agency. Retrieved June 1, 2022, from https://www.easa.europa.eu/newsroom-and-events/news/easa-releases-its-concept-paper-first-usable-guidance-level-1-machine-0

[18] Tabular Modeling Deep Dive. (2022, April). Retrieved June 1, 2022, from https://github.com/fastai/fastbook/blob/master/09_tabular.ipynb

[19] Timnit Gebru, J. M. (2021). Datasheets for Datasets. Communications of the ACM, 64(12), 86-92. https://doi.org/10.1145/3458723

[20] United States Code of Federal Regulations. (n.d.). 14 CFR 25.1309 Equipment, systems, and installations. US Government. Retrieved June 1, 2022, from https://www.ecfr.gov/current/title-14/chapter-I/subchapter-C/part-25/subpart-F/subject-group-ECFR9f24bf451b0d2b1/section-25.1309

Original Article
https://doi.org/10.56094/jss.v58i2.252

International System Safety Society

www.jsystemsafety.com

Journal of System Safety

Established 1965 Vol. 58 No. 2 (2023)

Review of the Latest Developments in Automotive Safety Standardization for Driving Automation Systems

Rami Debouk[ab]

[a] *Corresponding author email: rami.debouk@gm.com*
[b] *General Motors R&D, Warren, MI*

Keywords
functional safety, safety of the intended functionality, driving automation systems

Peer-Reviewed
Gold Open Access
Zero APC Fees

Online: 22-Jun-2023

Cite As:
Debouk, R. Review of the Latest Developments in Automotive Safety Standardization for Driving Automation Systems. Journal of System Safety. 2023;58(2):40-45. https://doi.org/10.56094/jss.v58i2.252

ABSTRACT

The ISO 26262: Functional Safety – Road Vehicles Standard has been the de-facto automotive functional safety standard since it was first released in 2011. With the introduction of complex driving automation systems, new standardization efforts to deal with safety of these systems have been initiated to address emerging gaps such as the human/automation roles and responsibilities in the presence/absence of the driver/user, the impact of the technological limitations and the verification and validation needs of automation systems to name a few. This paper highlights some of these gaps and introduces some of the latest developments in automotive safety standardization for driving automation systems.

INTRODUCTION

Safety-critical systems are systems that have the ability to create potentially hazardous issues in case they do not operate properly or as designed (Ericson-II, 2005), (Leveson, 2001). These systems are in general analyzed using rigorous and systematic safety processes (Bahr, 1997), for instance ISO 26262 (ISO 26262, 2018), Functional Safety – Road Vehicles, in the automotive domain.

The effort of standardization in the area of automotive functional safety accelerated in the last couple of decades as automotive systems became more complex, integrated and software intensive. As a matter of fact, the automotive industry is not as regulated as other industries, hence harmonized guidelines and best practices across the industry may have not been widely available. This definitely helped kick off the automotive functional safety standardization into a higher gear, and it all started

with the adaptation of existing standards to the automotive domain.

ISO 26262 was launched as the adaptation of (IEC 61508, 2010) to comply with needs specific to the application sector of Electrical/Electronic systems within road vehicles. ISO 26262 applies to all activities during the safety lifecycle of system development. At the concept phase, the hazard and risk assessment process focuses on identifying possible hazards caused by malfunctioning behavior of E/E safety-related systems and mitigating them through the identification of safety goals. The design phase includes system, hardware, and software development with requirements derived from the safety goals. ISO 26262 also prescribes the functional safety management activities to be performed during the safety lifecycle and provides requirements on the supporting processes.

However, ISO 26262 application faced some challenges, especially with the introduction and development of automations levels 2 and above driving automation systems (DAS) (SAE J3016, 2021). These systems split the roles and responsibilities of performing the dynamic driving tasks between the driver and the automation system: Levels 2 and 3 still have the driver responsible for some of these tasks while the automation system is fully responsible for these tasks in Levels 4 and 5. Moreover, they may be impacted by some technological limitations in the components they use not to mention that some of these components may not be fully specified, e.g., a Machine Learning (ML) component. Consequently, many standards/documents were drafted and published to address these issues that were not fully addressed by ISO 26262.

This paper is organized as follows. An overview of ISO 26262 and some identified challenges in applying it to DAS are presented first. Next, some of the recently developed automotive safety standards, specifications and guidelines are listed and a brief overview of some of these standards, specifications, and guidelines is provided while focusing on the specific issues they address. Finally, some thoughts on the current state in using the automotive safety standards is provided.

ISO 26262

Overview

ISO 26262 is the de facto standard for functional safety in the automotive electronics domain. It is the adaptation of IEC 61508 to comply with needs specific to the application sector of Electrical/Electronic systems within road vehicles. The adaptation applies to the automotive safety lifecycle of safety-related systems comprised of electrical, electronic, and software elements that provide safety-related functions.

ISO 26262 develops a structured and systematic process for safety analysis to guarantee product integrity and avoid recalls in the field. Requirements cover concept phase to decommissioning alongside safety management and supporting processes. Below is a highlight of the major activities described in the standard. The reader is referred to (Debouk, Overview of the 2nd Edition of ISO 26262: Functional Safety - Road Vehicles, 2019) and (Debouk & Joyce, ISO 26262 Hazard and Risk Assessment Methodology, 2010) for a comprehensive overview of the standard and its hazard analysis and risk assessment process.

At the concept phase of ISO 26262 is the hazard analysis and risk assessment process. This process provides an automotive specific risk-based approach for determining risk classes. Potential hazards caused by malfunctioning behavior are identified and categorized and safety goals related to the prevention or mitigation of these potential hazards are formulated. Each safety goal is assigned an Automotive Safety Integrity Level (ASIL) and the ASIL is determined by a systematic evaluation of hazardous situations. In determining the ASIL one considers the estimation of the following factors: severity, probability of exposure and controllability. It is worth noting here that controllability is defined as the ability to avoid harm by actions of traffic participants. Functional safety requirements needed to avoid an unreasonable risk for each potential hazard are derived from the safety goals which are not expressed in terms of technological solutions, rather in terms of functional objectives. Functional safety requirements inherit the ASIL of the safety goal from which they are derived.

The product development at the system level per ISO 26262 starts with developing the technical safety concept. The technical safety concept specifies the

technical safety requirements and their allocation to system elements (hardware and software). The technical safety requirements inherit the ASIL of the functional safety requirements they refine and specify safety mechanisms to detect faults and mitigate or control failures that may lead to the violation of these functional safety requirements and hence the safety goals. Safety mechanisms are technical solution to detect and mitigate (through avoidance or control) faults/failures in order to maintain intended functionality or achieve or maintain a safe state. The technical safety concept defines the system architectural design as well. The development of the technical safety concept is then detailed at both the hardware and software levels. Once the hardware and software developments are complete, all elements are integrated and tested. Finally, safety validation is completed at the vehicle level, that is evidence is provided that safety goals have been met. Figure 1 below graphically represents this development.

A safety case is published before releasing to production and it is a documentation to communicate a clear, comprehensive, and defensible argument (supported by evidence compiled in work products) that a system is acceptably safe to operate in a particular context.

KEY CHALLENGES

In the context of L2 to L5 DAS, the application of ISO 26262 faces a couple of challenges. A few of these challenges is discussed below:

- *Determination of the controllability parameter:* controllability is defined in ISO 26262 as the ability to avoid harm by actions of traffic participants. Since the role of the automation system in performing the dynamic driving tasks increases as the level of automation increases, the relevance of the controllability parameter becomes somehow questionable in determining the ASIL when performing the hazard analysis and risk assessment.
- *Definition of the safe state:* in the presence of human drivers, many systems relied on them as part of the definition of the safe state making the system fail safe or silent. However, with the reduced responsibility of human drivers in performing the dynamic driving tasks, fail-operational behavior and availability requirements maybe needed to maintain that the automation system achieves or reaches a safe state following the occurrence of a malfunctioning behavior.
- *Addressing hazards due to nominal performance:* ISO 26262 did not analyze hazards of nominal performance such as ones due to incomplete specifications or technology

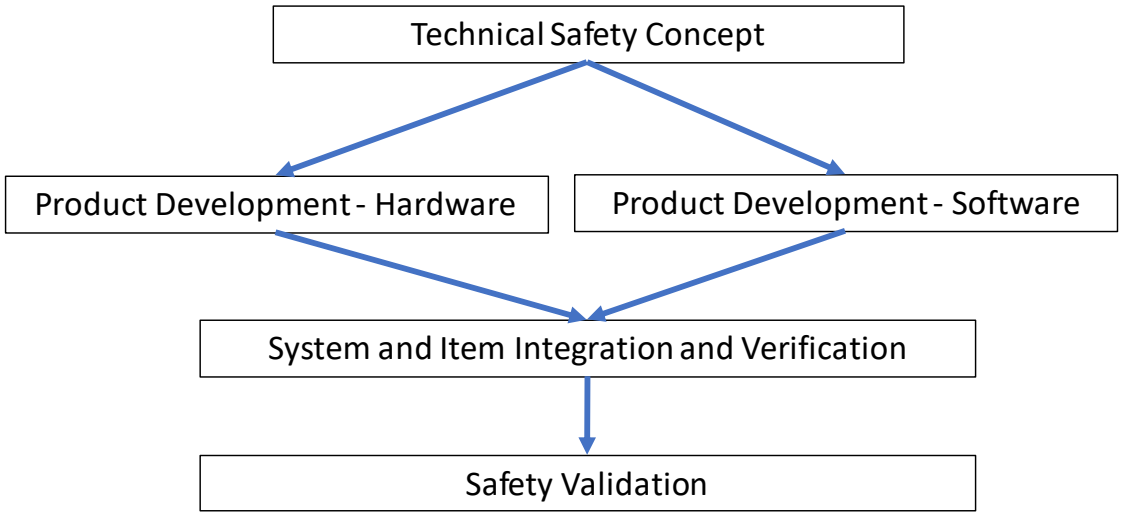


Figure 1: Product Development per ISO 26262

limitations. The latter are referred to as functional insufficiencies and are addressed in the Safety Of The Intended Functionality (SOTIF) standard ISO 21448 (ISO 21448, 2022).

- *Analyzing cybersecurity threats and their impact on safety:* As hazards may be caused or triggered by security threats, these threats are to be considered and analyzed as part of the hazard analysis and risk assessment. ISO 26262 recognized this issue and required synchronization of analyses between safety and security responsible teams at few instances in the vehicle development process.
- *Considering operational safety:* operational safety considers in general the health of the systems and components of the vehicle and with a less involved human driver such topic requires some planned procedures to monitor these systems and components.
- *Dealing with the use of Artificial Intelligence (AI) and ML models or components:* AI/ML models or components are usually treated as black boxes making them not fully specified and resulting in challenges when analyzing and verifying them.

AUTOMOTIVE SAFETY STANDARDS, SPECIFICATIONS AND GUIDELINES FOR DAS

In order to address the challenges listed above, many standards, specifications and guidelines were drafted and published by many organizations. A non-exhaustive list of these standards/documents is provided in Table 1, and some of these standards/documents are briefly discussed afterwards.

ISO FDIS 21448: ROAD VEHICLES - SAFETY OF THE INTENDED FUNCTIONALITY

SOTIF by definition deals with the absence of unreasonable risk resulting from functional insufficiencies or due to reasonably foreseeable misuses. A functional insufficiency is either an insufficiency of specification or a performance limitation, hence SOTIF complements the scope of ISO 26262 by addressing hazards caused by the intended functionality, i.e., the nominal performance. This is depicted in Figure 2 below.

Table 1: Automotive safety standards, specifications, and guidelines

ISO 21448: Road vehicles - Safety of the intended functionality (https://www.iso.org/standard/77490.html)
UL 4600 Ed. 2-2022: Standard for Evaluation of Autonomous Products (https://www.shopulstandards.com/ProductDetail.aspx?productid=UL4600)
ISO/FDIS 34502: Road vehicles - Engineering framework and process of scenario-based safety evaluation (https://www.iso.org/standard/78951.html)
ISO/TR 4804: Road vehicles - Safety and cybersecurity for automated driving systems - Design, verification and validation methods (https://www.iso.org/standard/80363.html)
ISO AWI TS 5083: Road vehicles - Safety for automated driving systems - Design, verification and validation (https://www.iso.org/standard/81920.html)
SAE J3016: Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles (https://webstore.ansi.org/standards/sae/sae30162021)
SAE J3018: Safety-Relevant Guidance for On-Road Testing of Prototype Automated Driving System (ADS)-Operated Vehicles (https://webstore.ansi.org/standards/sae/sae30182020)
SAE J2980: Considerations for ISO 26262 ASIL Hazard Classification (https://webstore.ansi.org/standards/sae/sae29802018)
SAE J3206: Safety Principles (https://webstore.ansi.org/standards/sae/sae32062021)
BSI PAS 1880: Guidelines for developing and assessing control systems for automated vehicles (https://www.bsigroup.com/en-GB/CAV/pas-1880/)
BSI PAS 1881: Assuring the safety of automated vehicle trials and testing – Specification (https://www.bsigroup.com/en-GB/CAV/pas-1881/)
BSI PAS 1883: Operational design domain (ODD) taxonomy for an automated driving system (ADS) – Specification (https://www.bsigroup.com/en-GB/CAV/pas-1883/)

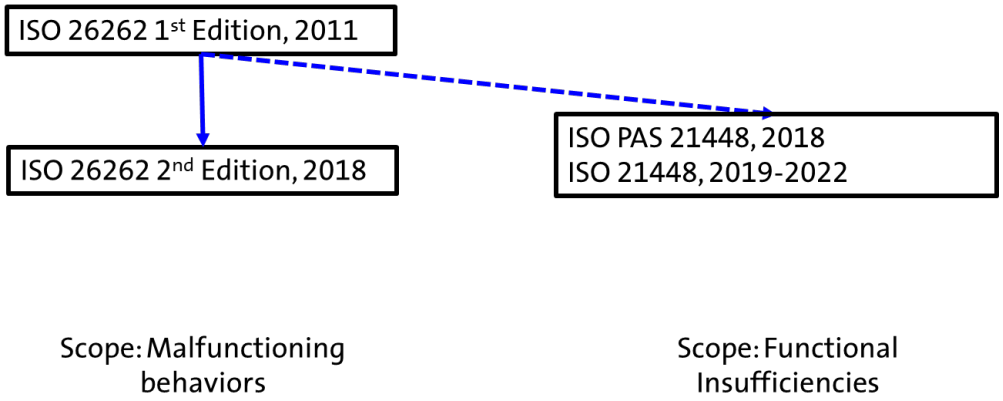


Figure 2: Scope of ISO 26262 vs ISO 21448

ISO 21448 provides guidance on the applicable design, verification and validation measures needed to achieve SOTIF. This includes the system specification, identification and evaluation of hazards caused by the intended functionality, and any modifications needed to reduce the risk due to SOTIF. In addition, the verification and validation strategy and activities are discussed as well as the method to accept the residual risk following the SOTIF activities. ISO 21448 includes an annex to address AI/ML components. The expectation is that ISO 21448 is complementing the safety activities performed while following ISO 26262.

UL 4600 Ed. 2-2022: Standard for Evaluation of Autonomous Products

UL 4600 is intended to work with existing standards to provide the additional elements necessary to assure that safety aspects of fully autonomous item operation have been considered in a comprehensive manner when creating a safety case. It is currently in its second edition with the first edition released in 2020. While use of existing functional safety standards is highly desirable, it is likely that there will be gaps between successful conformance to those standards and the creation of an acceptable safety case for complex autonomous items.

The main goal of UL 4600 is to make sure that the cumulative work products produced as a consequence of following other standards and other best practices do not leave any holes that present an unreasonable risk to autonomous product safety. In particular, compatibility with ISO 26262 and ISO21448 has been considered.

Two areas out of scope for this standard are setting acceptable risk levels and setting forth requirements for ethical product release decisions and any ethical aspects of product behavior.

ISO/FDIS 34502: Road vehicles - Engineering framework and process of scenario-based safety evaluation

ISO 34502 provides guidance and a state-of-the-art engineering framework for automated driving systems test scenarios and scenario-based safety evaluation processes. Therefore, ISO 21448 would benefit from the proposed process in identifying and evaluating scenarios, the latter being integral to the SOTIF safety activities.

ISO/TR 4804: Road vehicles – Safety and cybersecurity for automated driving systems – Design, verification and validation methods

ISO 4804 describes guidelines in developing, verifying, and validating driving automation systems based on basic safety principles. It also considers safety- and cybersecurity-by-design. ISO 4804 is merely a technical report and will be withdrawn once ISO TS 5083 is published.

ISO AWI TS 5083: Road vehicles — Safety for automated driving systems — Design, verification and validation

This document provides an overview and guidance of the steps for developing and validating an automated vehicle equipped with a safe automated

driving system. It considers and details steps for developing a safety concept, designing for safety, verifying, and validating DAS of Levels 3 and 4 as well as post deployment safety activities. In addition, it outlines cybersecurity considerations throughout all described steps. ISO TS 5083 includes an annex to address AI/ML components.

ISO TS 5083 will benefit from both ISO 26262 and ISO 21448 as the "generic" standards to which its application is intended, that is DAS features of Levels 3 and 4.

SAE J3016: Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles

This foundational report defines automation levels, operational design domains (ODD), object and event detection and response, minimal risk conditions among many others, all of which are fundamental in the development of the standards, specifications and guidelines for DAS features of Levels 2 and above.

BSI PAS 1881: Assuring the Safety of Automated Vehicle Trials and Testing – Specification

This publicly available specification specifies the minimum requirements for safety cases for automated vehicle trials and development testing in the United Kingdom to demonstrate activities can be undertaken safely. Even though BSI PAS 1881 deals with development vehicles, its application would benefit vehicle manufacturers in assessing their vehicles ahead of releasing them on public roads.

BSI PAS 1883: ODD Taxonomy for an Automated Driving System – Specification

This publicly available specification provides requirements for the minimum hierarchical taxonomy for specifying an ODD to enable the safe deployment of an automated driving system (Levels 3 and above in J3016). The ODD comprises the static and dynamic attributes within which an automated driving system is designed to function safely. It clearly aligns itself in support of the vehicle manufacturers designing DAS features.

FINAL THOUGHTS

For higher automation level systems (Levels 3 and above in J3016), no direct safety design or assessment guidance is provided in ISO 26262. Therefore, automotive safety engineers performing safety analysis on higher automation level systems need to go beyond what ISO 26262 requires. This can be achieved by interpreting and/or adapting ISO 26262 requirements in the context of the higher automation level systems they are analyzing. ISO TC22/SC32/WG08 that developed ISO 26262 is looking at the gaps and challenges currently in order to provide some guidance until the work on the 3rd Edition of ISO 26262 starts. In the meantime, ISO 21448 and ISO TS5083 (as well as others) are attempting to address some of these issues as well.

COMPETING INTERESTS

The author declares they have no potential competing interests.

ORCID IDS

Rami Debouk https://orcid.org/0009-0000-0542-5356

REFERENCES

[1] Bahr, N. J. (1997). System Safety Engineering and Risk Assessment: A Practical Approach. Taylor and Francis.
[2] Debouk, R. (2019). Overview of the 2nd Edition of ISO 26262: Functional Safety - Road Vehicles. Journal of System Safety, 55(1). https://doi.org/10.56094/jss.v55i1.55
[3] Debouk, R., & Joyce, J. (2010). ISO 26262 Hazard and Risk Assessment Methodology. Proceedings of the International System Safety Conference.
[4] Ericson-II, C. A. (2005). Hazard Analysis Techniques for System Safety. New Jersey: John Wiley & Sons. https://doi.org/10.1002/0471739421
[5] IEC 61508. (2010). IEC 61508, Functional Safety of Electrical/Electronic/Programmable Electronic Safety Related Systems Parts 1-7. Switzerland.
[6] ISO 21448. (2022). Road Vehicles - Safety of the Intended Functionality.
[7] ISO 26262. (2018). ISO 26262 2nd Ed. Road Vehicles - Functional Safety Parts 1-12.
[8] Leveson, N. (2001). Safeware: System Safety and Computers. Addison Wesley.
[9] SAE J3016. (2021). Taxonomy and Definitions for Terms Related to Driving Automation Systems for On Road Motor Vehicles.

From the JSS Archives

Effective April 2022, JSS announced that it was transitioning to a Gold Open Access publishing model, and we launched our new website at jsystemsafety.com.

To date we have published eight years (2014-2022) of our back issues to the new website. We ultimately plan to republish the entire 57 year archive! This page highlights a few of the many articles currently available in our archives.

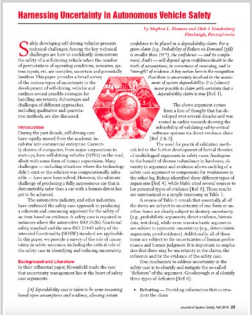

Notes on Society History Note No. 6

Pathfinders of System Safety

Notes on Society History: Pathfinders of System Safety

Rex Gordon

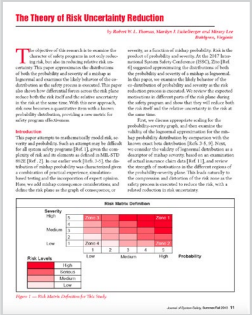
The Society historian honors some of the pioneers in system system


Harnessing Uncertainty in Autonomous Vehicle Safety

Harnessing Uncertainty in Autonomous Vehicle Safety

Stephen L. Thomas
Dirk J. Vandenberg

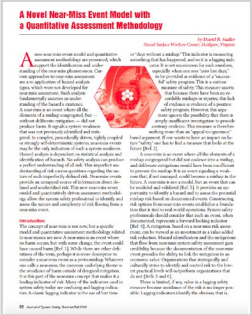
Provides a survey of the types of uncertainty in the development of self-driving vehicles and outlines possible strategies for handling uncertainty.


The Theory of Risk Uncertainty Reduction

The Theory of Risk Uncertainty Reduction

Robert W. L. Thomas
Marilyn J. Eichelberger
Missey Lee

Examines the character of safety programs in not only reducing risk, but also in reducing relative risk uncertainty.


A Novel Near-Miss Event Model with a Quantitative Assessment Methodology

A Novel Near-Miss Event Model with a Quantitative Assessment Methodology

David Sadler

Presents new near-miss event model and quantitative assessment methodology, supporting the understanding of the near-miss phenomenon.

System Safety On Demand

Past Webinars Available On Demand

Challenges and Potential Benefits of AI to Software Safety Assurance

Model-based Systems Engineering or System Safety: An Introduction

Occupational Hazard and Risk Management Techniques

Requirements Analysis using Karnaugh Maps

International System Safety Society Chapter Contacts

ASIA PACIFIC
Singapore Chapter
Eng Ling Onn
011-65-9632-6256
onnel@stengg.com

CANADA
Tony Zenga
514-825-7845
tzenga@cmtigroup.com

UNITED STATES OF AMERICA

ALABAMA/TENNESSEE/MISSISSIPPI
Tennessee Valley Chapter
Tim Browning
tbrowning@apt-research.com

ARIZONA
Saguaro Chapter
Adam Hughes
978-852-8053
ahughes3245@gmail.com

CALIFORNIA
Bay Area Chapter
Graham Murray
408-756-2674
Graham.t.murray@lmco.com

Central California Chapter
Miguel Trujillo
805-606-1533
Miguel_trujillo@yahoo.com

Sierra High Desert Chapter
Glen McCue
760-939-3531
glen.s.mccue.civ@us.navy.mil

Southern California Chapter
Francis McDougall
310-653-1309
Francis.mcdougall@us.af.mil

MAINE/NEW HAMPSHIRE/VERMONT/MASSACHUSETTS/RHODE ISLAND/CONNECTICUT/PENNSYLVANIA/NEW YORK/NEW JERSEY
Northeast Chapter
John Hewitt
203-522-3974
john.e.hewitt@lmco.com

NEW MEXICO
Stacey Durham
riparian77@hotmail.com

TEXAS
North Texas Chapter
Tom Haeussler
505-284-9748
Thomas.a.haeussler@lmco.com

VIRGINA/MARYLAND/DELAWARE
Washington DC
Chapter
John Burchett
301-744-2307
john.burchett@navy.mil

VIRTUAL CHAPTER
Doanna Weissgerber
831-278-0800
Doanna@pacbell.net

RVP ASIA PACIFIC
Eng Ling Onn
(Singapore)
011-65-9632-6256
onnel@stengg.com

RVP EUROPE
Gabriele Schedl
(Austria)
43 (1)811-50-2758
gabriele.schedl@frequentis.com

Mark Your Calendar

33rd European Safety and Reliability Conference (ESREL)
Sep. 3-8, 2023
University of Southampton, UK
https://www.esrel2023.com/

International Air Safety Summit 2023
Nov. 6-8, 2023
Paris Airport Marriott, Paris, FR
https://flightsafety.org/event/international-air-safety-summit-2023/

70th Annual Reliability & Maintainability Symposium (RAMS)
Jan. 22-25, 2024
Albuquerque, NM, USA
https://rams.org/

ANNUAL INTERNATIONAL SYSTEM SAFETY
SUMMIT AND TRAINING
Portland, OR | Aug 28 - Sept 1, 2023

ISSC 2023
KEYNOTE SPEAKERS

We are excited to announce this year's distinguished Keynote Speakers!

Greg Benn

Senior Director, Safety & Airworthiness Functional Chief Engineer, Boeing

Greg Benn is the Boeing Functional Chief Engineer for Safety & Airworthiness Engineering. In this role, he is responsible for the development of capabilities and technical excellence for certification, product safety, investigative, and safety data analytics engineering.

This includes development of technical strategy, enhanced capabilities such as model based certification and safety, knowledge curation, lessons learned/feedback loops, people development in the skill, and continuous improvement. Greg also leads the Enterprise Safety organization, specifically focused on the development of safety engineering capability, addressing enterprise wide product

Todd Zarfos

President and Chair of the Board of Directors, SAE International

Todd Zarfos is the 113th president of SAE International, leading the 138,000 member organization in all its efforts in education, standards, and professional development.

He started his career with Boeing in 1985 and served as Vice President of Engineering Functions at the Washington State Engineering Centers and Senior Chief Engineer of Airplane Systems with Boeing Commercial Airplanes. In this role, he was responsible for driving technical excellence within the airplane systems community while also ensuring the technical integrity and success of development and production programs.

Russ DeLoach

Chief, Office of Safety & Mission Assurance, NASA

Russ DeLoach is NASA's chief of Safety and Mission Assurance (SMA). Appointed to this role in January 2021, DeLoach is responsible for the development, implementation and oversight of SMA policies and procedures for all NASA programs

Prior to this assignment, DeLoach served as the SMA director at NASA's Johnson Space Center, where he led a dedicated team of experts in assuring workforce safety and collaborating on smart solutions to human spaceflight risks since February 2019. His team worked to identify, characterize, mitigate and communicate risk to accomplish safe and successful human space exploration.

Activities

Through its local chapters, committees, executive council, publications and meetings, the Society provides many opportunities for interested members to participate in a va-riety of activities compatible with Society objectives. In addition to the basic operating committees, Society activities include several noteworthy publications and events.

Publications

- Journal of System Safety is the official Society journal. Published three times a year, JSS is a peer-reviewd scholarly journal that keeps members informed of the latest developments in the field of system safety.
- Chapter newsletters are published periodically to disseminate news of chapter activities and items of interest to chapter members.
- Proceedings of Society-sponsored conferences and symposia are made available to members at a special discount.

Meetings — Conferences — Symposia

- International System Safety Conferences are sponsored annually. These conferences have proven to be a very popular and effective means for highlighting the latest techniques, applications and social/legal aspects of system safety.
- Mini-symposia are sponsored by local chapters to provide an in-depth exploration of a specific system safety-related topic.
- Chapter dinner meetings, field trips and panel discussions are held at intervals throughout the year.
- The Society is a co-sponsor of various system safety-related symposia and conferences.

Membership in the Society is open to all persons having an interest in or currently involved in work related to system safety or an allied discipline. Professional member-ship grades are available for those able to demonstrate sufficient qualifications, experi-ence and training. Annual dues are $150 (USD), while student memberships are free. Society members and subscribers are located in all areas of the United States and many countries around the world:

Australia
Austria
Cameroon
Canada
Chile
China
France
Germany
Greece
Israel
Italy
Japan
Netherlands
Nigeria
Norway
Russia
Saudi Arabia
Singapore
South Aftrica
Spain
Sweden
Switzerland
United Kingdom
(England, Northern Ireland, Scotland and Wales)
United Arab Emirates
United States of America

Requests for membership applications, subscription orders, requests for Conference Proceedings and other matters related to membership and services should be addressed to the International System Safety Society, 1000 Westgate Dr., Suite 252, Saint Paul, MN 55114, or contactsystemsafety@system-safety.org. Visit our Website at http://www.system-safety.org.

The **International System Safety Society** is a non-profit organization of professionals dedicated to the safety of systems, products and services through the effective implementation of the system safety concept. Under this concept, appropriate technical and managerial skills are applied so that a systematic, forward-looking hazard identification and control function becomes an integral part of a project, program or activity at the planning phase and continues through the design, production, testing, use and disposal phases.

The Society's Objectives

- To advance the art and science of system safety
- To promote a meaningful management and technological understanding of system safety
- To disseminate advances in knowledge to all interested groups and individuals
- To further the development of the professionals engaged in system safety
- To improve public understanding of the system safety discipline
- To improve the communication of system safety principles to all levels of management, engineering and other professional groups

International System Safety Society

Professionals Dedicated to the Safety of Systems, Products and Services

International System Safety Society
1000 Westgate Dr, Suite 252
Saint Paul, MN 55114

Society Website: https://www.system-safety.org
Journal Website: https://www.jsystemsafety.com