Risk, Trust, and Bias: Causal Regulators of Biometric-Enabled Decision Support

KENNETH LAI[1], **HELDER C. R. OLIVEIRA**[1], **MING HOU**[2], **(Senior Member, IEEE),**
SVETLANA N. YANUSHKEVICH[1], **(Senior Member, IEEE), AND VLAD P. SHMERKO**[1]

[1]Biometric Technologies Laboratory, Department of Electrical and Computer Engineering, University of Calgary, Calgary, AB T2N 1N4, Canada
[2]Defence Research and Development Canada (DRDC), Ottawa, ON K1N 1J8, Canada

Corresponding author: Kenneth Lai (kelai@ucalgary.ca)

ABSTRACT Biometrics and biometric-enabled decision support systems (DSS) have become a mandatory part of complex dynamic systems such as security checkpoints, personal health monitoring systems, autonomous robots, and epidemiological surveillance. Risk, trust, and bias (R-T-B) are emerging measures of performance of such systems. The existing studies on the R-T-B impact on system performance mostly ignore the complementary nature of R-T-B and their causal relationships, for instance, risk of trust, risk of bias, and risk of trust over biases. This paper offers a complete taxonomy of the R-T-B causal performance regulators for the biometric-enabled DSS. The proposed novel taxonomy links the R-T-B assessment to the causal inference mechanism for reasoning in decision making. **Practical details** of the R-T-B assessment in the DSS are demonstrated using the experiments of assessing the trust in synthetic biometric and the risk of bias in face biometrics. The paper also outlines the **emerging applications** of the proposed approach beyond biometrics, including decision support for epidemiological surveillance such as for COVID-19 pandemics.

INDEX TERMS Risk, trust, bias, biometrics, intelligent decision support, Bayesian causal inference, machine reasoning, epidemiological surveillance.

I. INTRODUCTION

Biometric-enabled decision support is a mandatory mechanism of various complex systems, such as:

— security checkpoints (identity management) [1],
— personal health monitoring systems [2],
— e-coaching for health [3],
— driver assistant (e.g., fatigue and stress detection) [4],
— multi-factor authentication systems [5],
— spoken conversational agents [6],
— epidemiological surveillance [7] and,
— preparedness systems for emerging health service [8].

To be integrated into a complex system, a biometric-enabled computational intelligence (CI) must satisfy various requirements for compatibility and standards. In other words, it must adhere to the concept of Decision Support System (DSS). Generally, the goal of any DSS is to support human

The associate editor coordinating the review of this manuscript and approving it for publication was Marina Gavrilova.

experts, operators, or users in their decision-making in real-time, under multiple and constantly evolving factors.

A well-identified trend in DSS is to augment **intelligent features** (learning, training, adaptation, possibility to choose among available decisions) beyond simple information retrieving, e.g., predicting the evolution of the current state situation, known as situational awareness [9]. In intelligent DSS, human cognition can impact the CI, and vice versa. That is, the performance of the DSS depends on complicated factors such as cognitive biases of humans and intelligent machines. This is the area of our interest.

Typically, the DSS performance is evaluated in various dimensions:

— Technical, e.g., false acceptance rate (FAR), false rejection rate (FRR), accuracy rate, and throughput [5], [10],
— Social, e.g., privacy, public acceptance [11],
— Psychological, e.g., efficiency of human-machine interactions (known as teaming and trustworthy intelligent systems) [12]–[14],
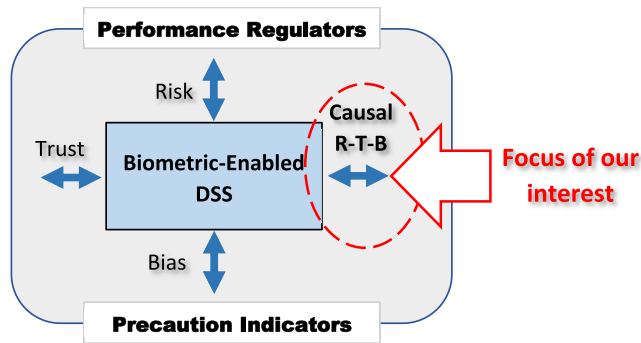
FIGURE 1. The R-T-B impact on the DSS performance and can be used as precaution indicators. The complementary nature of the R-T-B is the focus of our study.

FIGURE 2. Typical integrated performance measures of the biometric-enabled DSS are privacy and public acceptance.

— Security, e.g., vulnerability and sensitivity of personal data [11], [14], [15], and

— Efficiency of teamwork and group decision, e.g., trust, risk, reliance, satisfaction, stress [16].

Risk, Trust, and *Bias* (R-T-B) are essential indicators of the performance evaluation of complex dynamic systems such as intelligent DSS (Fig. 1). The R-T-B measures belong to the class of high-level performance measures. For example, trust in the intelligent interview assistant addresses the intelligent (cognitive) biases [15]. Risk and trust in the DSS are linked to various kinds of biases, for example, racial biases in face-based human identification [17], [18] and attribute biases in social profiles [11].

As follows from Fig. 1, the role of the R-T-B measures is twofold: *performance regulators* and *precaution indicators*. Contemporary approaches often consider the R-T-B impact on the DSS performance **independently, ignoring their causal relationships** such as the risk of bias, the risk of trust, the risk of trust over biases, etc. Adding to the evaluation level, the R-T-B measures become precaution indicators or signs, e.g., high risk, low trust, large bias, and low risk of bias.

The complementary nature of the R-T-B measures in biometric-enabled DSS is the focus of our study (Fig. 1). Specifically, R-T-B can manifest themselves as a *causal ensemble* and convey additional useful information for DSS performance regulation. For example, once a decision regarding a subject's identity is made, a "risk of decision trust" is calculated that 1) assesses the risk of acceptance or rejection of the decision based on the operator's trust in the DSS, and 2) acts as a measure of precaution on the over-trust. The **working hypothesis** of our research is that the R-T-B landscape that includes the causality between the R-T-B measures is evaluated using causal networks. In this paper, we provide the results of such study.

The remainder of this paper is organized as follows. Section II provides a survey of the most important related works. Contributions of this paper are provided in Section III. Our approach to the R-T-B taxonomy is explained in Sections IV, V, including the view on standardization of the R-T-B measures using the Admiralty Code is proposed in
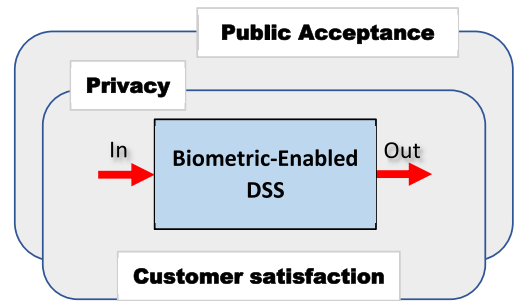
Section VII. The core mechanism of the R-T-B assessment on the causal networks is explained in Section VI and demonstrated through experiments in Section VIII. The forecast of emerging applications and overall summary are provided in Sections IX and X.

II. RELATED WORK

Performance measurement is commonly understood as a regular measurement of a system's outcomes that captures the efficiency of said system. Measures such as privacy, customer satisfaction, and public acceptance belong to a class of *integrated, or high-level* system performance measures (Fig. 2). Each of these measures includes a specific set of quantitative and qualitative indicators, often complementary. For example,

— *privacy* includes indicators such as personal data (collection, storing, sharing, etc.);

— *public acceptance* includes security, privacy, user satisfaction, etc.

— *common indicators* of privacy and public acceptance measures include psychological predictors, social profile factors, and demographic indicators.

In order to achieve the compatibility between systems placed on a unified computational platform, it is reasonable to extend these measures using the R-T-B indicators like "risk of storing personal data", "demographic bias", and "trust of social profiling". Some of the R-T-B projections have been studied in the last decade in a wide spectrum of applications, for example, disclosure risks [11] (privacy losses in social computing networks), biases in facial recognition [17], trust, risk and optimism bias in e-government [19], risk of crime and social trust in the presence of endogeneity bias [20]. In [21], the notion of 'biased trust' addresses the phenomenon of a small set of trusted network users. The adversary can use this bias (prior trust relationships) for the development of an attack strategy in onion-routing networks. In [22], risks of decision-making biases and biases of the trustworthiness are studied for various consumer scenarios. In [23], the trust of reduced risk has been used for mobile shopping analysis.

The risk of bias in CI judgments is a key interest in all CI applications, e.g., in medicine [24] (trust of machine decision) and security [18] (risk of mis-identification due

to "demographic" bias in facial recognition algorithms). Assessing the trust of these phenomena for a given and novel CI algorithm is of critical importance. Inappropriate calibration of trust in human-machine and machine-machine interactions is a serious problem, and when conjoined with bias, risks of various kind of unwanted effects greatly intensifies. Among various approaches, intelligent DSSs are of the greatest demand, e.g., risk-adaptive trust model [25].

Paper [26] is an introduction to the trust management engine using pattern recognition techniques. The key notion is the *trust feature*: "the desired feature to be taken into consideration for a trust assessment". Example of established trust features include knowledge, reputation, and experience. *Trust assessment classes* include "untrustworthy", "neutrally trusted", and "trustworthy". Next, the regulators (measures) of trust discipline are the *trust levels*. The inputs of the machine learning algorithms are the trust features in a certain context, $\mathbf{X} = \texttt{Context}$, some labels $y^{(i)}, i = 1, 2, \ldots, m$, are assigned to each training set, $\{X, y^{(i)}\}$. This is a formulation of a trust assessment problem in terms of pattern recognition. In particular, various opportunities for choosing an appropriate machine learning algorithm are provided, e.g., multi-class classifiers such as a deep learning network, or a support vector machine that trains the model in order to identify the best margin to separate the trustworthy interactions from the other interactions. More details on trust computing using machine learning algorithms are provided in [27]. In our opinion, the approach proposed in [26], [27] can be extended to the broader R-T-B spectrum.

The above review leads to the conclusion that neither the R-T-B measures nor their causal relationships have been systematically addressed so far. Our study aims at overcoming these gaps and introduce the state-at-the-art R-T-B taxonomy and related cause-and-effects.

III. CONTRIBUTIONS

The foundation for the proposed taxonomy was laid in [28] and [29]. In [28], risks of biases for facial recognition were investigated, and in [29], risk and trust indicators of synthetic data in cognitive security checkpoints were studied. The quintessence of the experimental results from these sources is analyzed in Section VIII. This paper takes these results one step further by introducing a **systematic approach** to the R-T-B causal performance evaluation of complex biometric-enabled systems.

Our contribution and goal are achieved in conjunction with the following results:

- **Framework of intelligent DSS**; we adopted *Haykin's* fundamental results on cognitive systems [30], in particular, in modeling the DSS for multi-state intelligent checkpoint;
- **Taxonomical view** for causal R-T-B inference; we adopted, for this purpose, *Pearl's* layered causal inference hierarchy [31], as well as fundamentals of causal (Bayesian) networks [32];
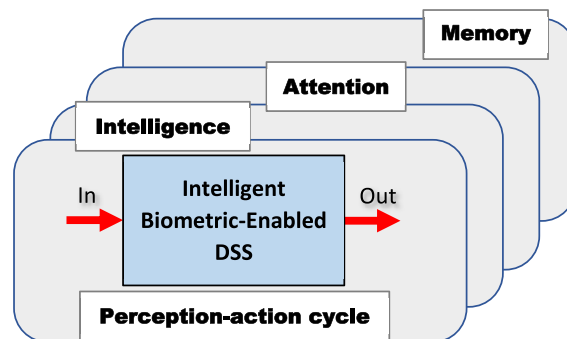

FIGURE 3. Principle elements of the intelligent DSS.

- **Standardization the R-T-B measures**; we referred to a widely used in practice the Admiralty Code [33]–[35]; and
- **Systematic view** of advanced causal networks; we extended a recent review [36] that covers most of the causal networks.

IV. BACKGROUND

This Section provides the basic knowledge of the intelligent DSS over the R-T-B performance regulators.

A. FRAMEWORK OF INTELLIGENT DSS

Intelligent biometric-enabled DSS for identity management is a complex dynamic system [15], [37] with the following elements of a cognitive system [30] (Fig. 3):

Perception-action cycle that enables information gain regarding the state of an identified person;

Memory distributed across the entire system (personal data are collected in the physical and virtual world);

Attention is driven by memory to prioritize the allocation of available resources; and

Intelligence is driven by perception, memories, and attention; its function is to enable the control and decision-making mechanism to help identify intelligent choices.

The R-T-B measures are an integrated part of the high-level measurements used in intelligent systems, e.g., risk of person identification in the perception-action cycle, privacy trust of distributed memory, and attention bias. Intelligent DSS is a semi-automated system, which deploys CI to process the data sources and to assess R-T-B; this assessment is submitted to a human operator for the final decision.

B. THE R-T-B DEFINITIONS

Definition 1: **Risk** is a "measure of the extent to which an entity is threatened by a potential circumstance or event, and typically is a function of: (i) the adverse **impact**, also called **cost** or magnitude of harm, that would arise if the circumstance or event occurs, and (ii) the **likelihood** of event occurrence" [38]. ∎

Formally, $\texttt{risk}$ in this paper is defined as a function F of impact (or consequences) of a circumstance or event and its

occurrence probability:

$$Risk = F(Impact, Probability)$$

For example, in automated decision making, the `Risk` is expressed as an `Impact` of accepting the DSS decision (which can be correct or incorrect) magnified by the likelihood of its correctness or incorrectness.

For computational purpose of this paper, we adapt the following definition of trust.

Definition 2: **Trust** as defined in [39], [40] is "the *subjective probability* by which one entity (the trustor) expects that another entity (the trustee) to perform a specific action of which its goal is dependent on." ∎

Useful details are provided in [41]: "trust is the attitude that an agent (DSS in our case) will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability."

From these definitions follows the key property of trust: the level of trust is *believed probability*.

The framework of trust assessment includes the following issues [12], [25], [41], [42]:
- The concept of trust is subjective by nature, its definition depends on a particular area.
- In order to assess trust, it is reasonable to derive trust as a function of risk.
- Trust is not necessarily proportional to the inverse of risk because risk may exist even under a situation with high trust.
- The balance between trust and risk can be achieved by optimizing gains in decisions.
- In the presence of uncertainty, trust can provide a "credit" to decisions made under uncertainty.

In [43], trust is associated with the expected utility of the decision, expressed in terms of *cost of verification* and *cost of the determined action*. Trust in the currently preferred decision on the action *a* is the probability $P(a)$ that *a* is successful. If trust is high enough, the currently preferred action can be accepted without verification; however, the operator will take risks for not performing verification. For example, face matching (as binary decision) by an operator aided with the DSS has only two options: Acceptance of the DSS recommendation, or its Rejection. If the option Accept is chosen, then the Trust in the DSS is simply the probability that the Accept decision is correct

$$Trust = P(Accept)$$

and the user should accept the DSS recommendation without verification if `Trust` $> 1 - C_v/C_a$ where C_v is the cost of decision verification action and C_a is the cost of incorrectly accepting the DSS solution.

Definition 3: **Trustworthiness** as defined in [39], [40] is "the *objective probability* by which the trustee performs a given action on which the goal of trustor depends". ∎

Useful details are given in [38]: "trustworthiness is the degree to which a system can be expected to preserve the confidentiality, integrity, and availability of the information being processed, stored, or transmitted by the system across the full range of threats."

Hence, in contrast to trust, trustworthiness refers to the actual probability by which the trusted party will perform as expected.

Note that trust is a belief that does not necessarily require observed behavior in the past, that is distinct from trustworthiness, which is a verified objective of trust through observed evidence. For example, a trusted biometric sample acquisition system should satisfy a set of requirements such as resistance to: 1) fake biometric target presentation, 2) communication attack, and 3) acquisition system tampering. In [5], trustworthy values are calculated using the accuracy rates (like FAR and FRR) of different authentication modalities, that is, the higher accuracy rate makes the modality more reliable.

Definition 4: **Bias** in the cognitive DSS refers to the tendency of an assessment process to systematically over- or under-estimate the value of a population parameter. ∎

For example, the bias of trust refers to a phenomenon that is well identified in psychology. In our study, we consider the bias of trust as the difference between the baseline trust (`Trust`) and the trust given some prior knowledge on a specific parameter of the system (`Trust'`):

$$Bias_{Trust} = Trust - Trust'$$

The result can be positive (decreased trust) or negative (increased trust).

Identifying and mitigating bias is essential for building trust and estimating the risk of trust in human-machine and machine-machine interactions. For example, the phenomenon of *own-race bias* is well-known in psychology. It is a tendency for systems to better recognize faces from one's racial in-group rather than for racial out-groups [13]. This tendency was recently shown in face recognition experiments [17], [18], [44]. CI biases were introduced by intelligent support of human-machine interactions [15]. Identity management biases were analyzed in multiple social profiles [11]. In [5], fuzzy reasoning and intelligent adaptive selection of the Multi-Factor Authentication (MFA) credentials have been used. Trustworthiness value functions are used in different authentication modalities such as biometrics, non-biometrics, and cognitive behavior metric. Fuzzy DSS for MFA is a cognitive system where a decision regarding user authentication is made iteratively and adaptively.

V. TAXONOMICAL VIEW ON THE CAUSAL R-T-B INFERENCE

This section represents a crucial part of our approach. The complementary nature of the R-T-B triad is a well-known fact for researchers. Some of the R-T-B causal relationships are successfully used in various fields (see our review of related work in Section II). Gaps between the R-T-B complementary nature and computational methodologies are periodically reviewed [42], [45], standardized [46] and created into guidelines [38]. The focus of our interest is motivated
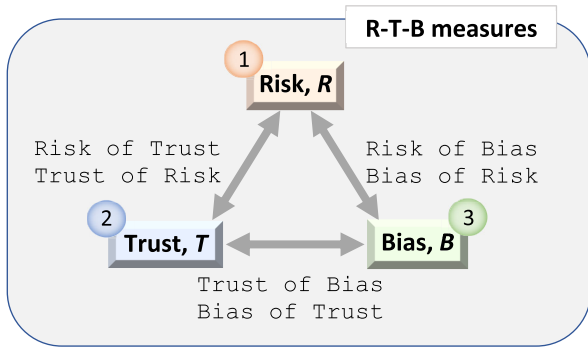
FIGURE 4. The R-T-B causal landscape.

TABLE 1. The three-layer causal inference hierarchy based on [31].

Level	Activity	Typical question
I. Association $P(y\|x)$	Seeing	What is? How would seeing X change my belief in Y?
II. Intervention $P(y\|\mathrm{do}(x), z)$	Doing Intervening	What if? What if I do X?
III. Counterfactuals $P(y_x\|x', y')$	Imagining, Retrospection	Why? Was it X that caused Y? What if I had acted differently?

by the improvement of the DSS performance using machine learning techniques.

There are two questions we address: 1) the reflection of the complementary nature of the R-T-B and 2) the need to close the aforementioned gaps. The principal solution to the first question is illustrated in Fig. 4. This R-T-B ensemble can be ordered according to their appropriate discipline. For example, first-order complement quantifies only single factor of the R-T-B domain and is introduced by a single node (variable), e.g., risk R of event $X = x$. There are also second and third order R-T-B complements.

The next step is more complicated and addresses the nature of causal models [32]. Recent work by *Pearl* [31] is a structured view on such model types.

Table 1 is *Pearl's* original causal hierarchy table, provided to explain our approach to the R-T-B causal taxonomy. *Pearl's* hierarchy is based on the rule: questions at level i, $i \in \{1, 2, 3\}$ can only be answered if information from level $j \in \{j \geq i\}$ is available. There are three levels:

1) *Association* (low level); it invokes purely statistical relationships and require no causal information; formally, it is the conditional probability of event $Y = y$ given event $X = x$, i.e., $P(y|x) = p$;
2) *Intervention* (intermediate level); it involves not just seeing what is, but changing what we see; formally, it is the probability of event $Y = y$ given that we intervene and set the value X to x and subsequently observe event $Z = z$, i.e., $P(y|\mathrm{do}(x))$;
3) *Counterfactuals* (top level); a mode of necessitating retrospective reasoning; formally, the probability of event $Y = y$ would be observed had X been x, given that actually observed X to be x' and Y to be y'.

Table 2 together with Fig. 4 provide the R-T-B taxonomical view on the intelligent DSS. Three kinds of ensembles are distinguished in the R-T-B space:

1st order: The R-T-B ensemble represents the simplest (idealized) scenarios of performance regulators, e.g., risk assessment, R, ignoring trust and bias factors.

2nd order: The R-T-B ensemble reflects the simplest causal relationships, that is, knowledge about the 1st order ensemble, e.g., risk caused by trust $R(T)$.

3rd order: The R-T-B ensemble contains knowledge about the 2nd order ensemble, e.g., risk of trust in the presence of bias, $R(T|B)$. For example, in [20], the risk of crime has been studied over social trust in the presence of expected biased parameter estimates. Note that the 3rd order R-T-B include $T(B|R)$, $T(R|B)$, $B(T|R)$, and $B(R|T)$.

In [47], trust is derived from experience while assuming multiple "sources" of trust in a system. The fusion of trusted sources represented by Beta-distributions is performed using *copula* technique resulting in a joint probability distribution of the overall trust. Paper [47] then proposes a method for risk forecasting based on a trust model using a mix of Beta-distributions. It takes positive or negative security incident indications as an input and compiles a numerical value within [0,1] that models the probability of system failure, conditional on the recorded experience. The trust model is updated in a Bayesian fashion, making the trust measure a conditional expectation of a security indicator, based on prior experience.

Summarizing, Table 2 outlines a causal R-T-B relationships. These can be inferred using various types of causal graph models, a brief guide to which is provided in the next Section.

VI. SYSTEMATIC VIEW OF ADVANCED CAUSAL NETWORKS

The crucial requirement for the R-T-B formalization is the ability to reason about the R-T-B state as well as the R-T-B prediction. Specifically, the following conditional dependencies can be derived from the R-T-B causal landscape in Fig. 4:

(a) Given the bias B, assess/predict the risk of trust $R(T|B)$ and trust of risk $T(R|B)$;
(b) Given the trust T, assess/predict the risk of bias $R(B|T)$ and bias of risk $B(R|T)$;
(c) Given the risk R, assess/predict the trust of bias $T(B|R)$ and bias of trust $B(T|R)$.

A causal network is a directed acyclic graph where each node denotes a unique random variable. A directed edge from node n_1 to node n_2 indicates that the value attained by n_1 has a direct causal influence on the value attained by n_2. Uncertainty inference requires a specific type of data structures referred to as *Conditional Uncertainty Tables* (CUTs).

TABLE 2. The R-T-B taxonomy based on *Pearl's* causal inference hierarchy [31].

The R-T-B causal inference	Graph
I. First order R-T-B	
Risk R: Risk of event $X = x$, `Risk= F(Impact, Probability` $X = x$`)` **Trust** T: Trust for event $X = x$ **Bias** B: Bias of event $X = x$	
II. Second order R-T-B	
Risk←Trust: — *Association* $R(x\|t)$: Risk of event $X = x$ given that trust $T = t$ is observed; — *Intervention* $R(x\|\mathtt{do}(t,x))$: Risk of event $X = x$ given that we intervene and set the value of trust $T = t$ and subsequently observe event $Z = z$; — *Counterfactuals* $R(x_t\|t',x')$: Risk of event $X = x$ would be observed had X been x, given that we actually observed X to be x' and trust T to be t'	
Risk←Bias: *Association* $R(x\|b)$; *Intervention* $R(x\|\mathtt{do}(b),x)$; *Counterfactuals* $R(x_b\|b',x')$	
Trust←Bias: *Association* $T(x\|b)$; *Intervention* $T(x\|\mathtt{do}(b),x)$; *Counterfactuals* $T(x_b\|b',x')$	
Trust←Risk: *Association* $T(x\|r)$; *Intervention* $T(x\|\mathtt{do}(r),x)$; *Counterfactuals* $T(x_r\|r',x')$	
Bias←Risk: *Association* $B(x\|r)$; *Intervention* $B(x\|\mathtt{do}(r),x)$; *Counterfactuals* $B(x_r\|r',x')$	
Bias←Trust: *Association* $B(x\|t)$; *Intervention* $B(x\|\mathtt{do}(t),x)$; *Counterfactuals* $B(x_t\|t',x')$	
III. Third order R-T-B	
Bias→Risk←Trust: *Association* $R(x\|b,t)$; *Intervention* $R(x\|\mathtt{do}(b,t),x)$; *Counterfactuals* $R(x_{b,t}\|b',t',x')$	
Bias→Trust←Risk: *Association* $T(x\|b,r)$; *Intervention* $T(x\|\mathtt{do}(b,r),x)$; *Counterfactuals* $T(x_{b,r}\|b',r',x')$	
Trust→Bias←Risk: *Association* $B(x\|t,r)$; *Intervention* $B(x\|\mathtt{do}(t,r),x)$; *Counterfactuals* $B(x_{t,r}\|t',r',x')$	

A CUT is assigned to each node in the causal network. Given a node *n*, the CUT assigned to *n* is a table that is indexed by all possible value assignments according to the parent nodes of *n*. Each entry of the table is a conditional "uncertainty model" that varies according to the choice of the uncertainty metric.

Analysis of a causal network is out of the scope of this paper. However, we introduce in this paper the *systematic criteria* for choosing the appropriate computational tools. In addition, some details are clarified in our experimental study.

The following types of causal networks are deployed in contemporary machine reasoning based on the CUT criterion:

$$
\mathtt{Causal\ network} \equiv
\begin{cases}
\mathtt{Bayesian} & \mathrm{CUT} \equiv \mathrm{CPT} & [32]; \\
\mathtt{Imprecise} & \mathrm{CUT} \equiv \mathrm{CImT} & [50]; \\
\mathtt{Interval} & \mathrm{CUT} \equiv \mathrm{CInT} & [51]; \\
\mathtt{Credal} & \mathrm{CUT} \equiv \mathrm{CCT} & [52]; \\
\mathtt{DS} & \mathrm{CUT} \equiv \mathrm{CDST} & [53]; \\
\mathtt{Fuzzy} & \mathrm{CUT} \equiv \mathrm{CFT} & [54]; \\
\mathtt{Subjective} & \mathrm{CUT} \equiv \mathrm{CST} & [55];
\end{cases}
\tag{1}
$$

In the list (1), the following abbreviations are used:
CPT – Conditional Probability Table;
CImT – Conditional Imprecise Table;

CInT   –   Conditional Interval Table;
CCT   –   Conditional Credal Table;
CDST   –   Conditional Dempster-Shafer (DS) Table;
CST   –   Conditional Subjective Table.

The distinguishing feature of these CUTs is that the uncertainty is interpreted in different ways. For example, uncertainty in risk assessment can be "filled" by weighted compositions of costs of losses, or by a set of alternative decisions. The type of a causal network shall be chosen given the DSS model and a specific scenario. The choice depends on the CUT as a carrier of *primary* knowledge and as appropriate to the scenario:

Bayesian network is defined as a causal network with the CUT being CPT using point probability measures [31]. The key limiting factor is the assumption that modeled events are independent.

Imprecise causal network is defined by using the CUT type such as the CImT, using lower and upper probabilities $\overline{p}(A)$ and $\underline{p}(A)$, respectively [50].

Interval causal network is defined by the specification of the CUT as the CInT and probability interval using a "radius of uncertainty" for each point probability [51].

Credal causal network is defined by specifying the CUT as the CCT using closed intervals of the possible range of probability values [52]; this model can be viewed as a set of Bayesian networks that share the same graphical structure but are associated with different conditional probability parameters.

Dempster-Shafer (DS) causal network is defined by using the CDST that utilizes the formalization of *imprecise* probabilities for evaluating the quality of results, producing optimistic and pessimistic estimations of vulnerability via *plausibility* and *belief* measures [53].

Fuzzy causal network is defined by specifying the CFT using fuzzy measures [54]. The CFTs are similar to CInT, but the lower and upper bounds may be "soft".

Subjective causal network is defined by specifying the CUT as the CST using subjective opinions, a belief-and-uncertainty representation of an unknown probability distribution of a random variable [55].

The choice of a specific causal network model is heavily dependent on the data that is available for creating the CUTs, as well as the information that is expected to be given by the posterior uncertainty model. For instance, if statistical data is in abundance, probability theory will be the most suitable choice of uncertainty model and will provide the most informative results. If statistical data is lacking for certain variables, probability intervals can account for uncertainty in those probabilities for which there is insufficient data. If statistical data is almost completely lacking, DS theory may be appropriate and the expert can provide the DS weights to populate the CDSTs.

Note that specific biases can be observed in reasoning using causal networks, such as *endogeneity bias*. Endogeneity occurs when an omitted variable or a variable's value confounds the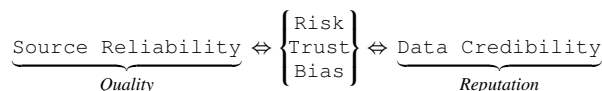 relationship between cause and effect, thereby introducing bias into the estimate of the causal effect and reasoning mechanism. In statistical terms, the endogeneity of a given variable manifests itself as an association between the variable and the error term. For example, in [20], the societal R-T-B assessments have been considered with respect to the endogeneity bias.

There were several attempts to provide researchers with the *"Guidelines"* for choosing the best causal network platform based on the CUT. A recent review [36] covers most of the network types in list (1). Comparison of causal computational platforms for modeling various systems is a useful strategy, such as Dempster-Shafer vs. credal networks [56], Bayesian vs. interval vs. Dempster-Shafer vs. fuzzy networks [37], [48].

R-T-B reasoning is the ability to form an intelligent judgment using the R-T-B data. It is a judgment under uncertainty based on a causal network. For example, in [57], the notion of trust is closely connected to the notion of belief change. Trust is defined in terms of a trust partition over a set of belief states, and the belief is updated based on the trust-sensitive belief revision operators.

VII. STANDARDIZATION OF THE R-T-B MEASURES USING ADMIRALTY CODE

R-T-B and their causal relationships manifest themselves in intelligent DSS in different ways, such as the reliability of information (data) sources and the credibility of the information (data):

$$\underbrace{\text{Source Reliability}}_{Quality} \Leftrightarrow \begin{Bmatrix} \text{Risk} \\ \text{Trust} \\ \text{Bias} \end{Bmatrix} \Leftrightarrow \underbrace{\text{Data Credibility}}_{Reputation}$$

The relationship can be represented as follows:

(*a*) *Source reliability* as the quality of being reliable, or trustworthy, is related to 1) risk as a function of potential adverse impact and the likelihood of occurrence, 2) trust as the confidence in quality, as well as their causal relationships, and 3) bias as systematic over- or under-assessment of the parameter of interest.

(*b*) *Information (data) credibility* as the reputation impacting one's ability to be believed.

In [48], available resources and information for traveler profiling are rated accordingly to the Admiralty Code (Fig. 6). NATO Standardization Agreements such as STANAG 2022 and STANAG 2511 [33]–[35] use the Admiralty Code to resolve conflicting scenarios in human-human, human-machine, and machine-machine interactions. In [35], information trust is defined based on well-formalized reliability and credibility attributes.

The reliability of the decision support provided by the DSS can be increased by using more reliable sources and credible information or can be diminished due to lowered reliability of the source and/or credibility of the information. In this context, trust can be expressed in terms of the reliability of data sources and/or the credibility of prior information. For
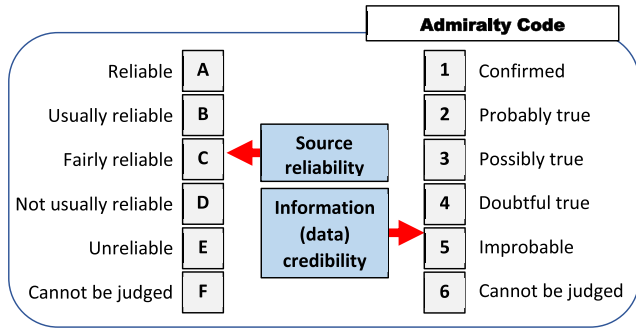
FIGURE 5. Manifestation of the R-T-B via assessments of reliability of source and credibility of information using the Admiralty Code.
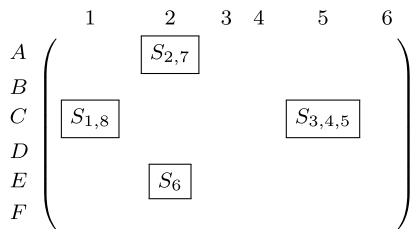

FIGURE 6. Example of the R-T-B reasoning. Given a set of a system states $S = S_i$, $i = 1, 2, \ldots, 8$. The primary task of the DSS is the R-T-B reasoning about these states using available resources such as the Admiralty code. The result is a set system states provided to a human analyst/expert to support their decision-making.

example, scenario $F6$ is composed of the source reliability `F ≡ <Cannot be judged>` and information credibility `6 ≡ <Cannot be judged>`.

There are various ways to use the Admiralty Code standard. For example, notion "credibility" is equivalent to "trustworthiness" over "expertise" where "trustworthiness" represents a fused reliability of source and credibility of data.

Fig. 6 explains the decision support mechanism for assessing different scenarios in terms of *system states*. For example, given the states $\{S_1, S_2, \ldots, S_8\}$ of the epidemiological surveillance, the DSS analyzes the states according to the Admiralty Code resulting in the following decision-making landscape:

— States S_1 and S_8; S_2 and S_7 can be used for decision-making;
— Decision-making based on states S_3, S_4, and S_5; S_6 is very risky.

For security checkpoints, the source reliability and information credibility are represented by the probabilistic variables such as false ID, multiple ID of the same person, and features of intentional data alteration in the chip (e.g., biometric traits and text data), as well as a false life-cycle history [11], [49].

VIII. DEMONSTRATIVE EXPERIMENTS

The goal of this section is to demonstrate how the R-T-B concept works in real-world large scale tasks, and through this method, empirically prove that the R-T-B triad is a system performance regulator. For this, a typi-
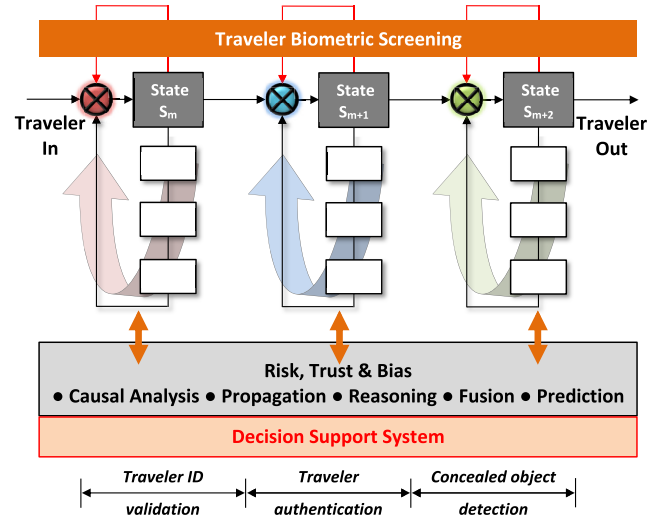

FIGURE 7. The taxonomical view of the multi-state intelligent identity management process. The R-T-B of synthetic data is assessed considering their causal relationships with authentic data. Each state is represented by a perception-action cycle of sub-states.

cal biometric-enabled complex dynamic system was chosen. Two research questions was prioritized for investigation in the R-T-B dimensions:

— Impact of synthetic data on system performance; this problem is motivated, in particular, by research [58], [59];
— Impact of demographic factors on system performance; this problem is critical in facial recognition [17], [18].

A. THE R-T-B OF SYNTHETIC DATA
Experimental study of synthetic data impact on performance of a cognitive checkpoint is reported in [29]. Below, we briefly introduce the quintessence of this report and provide new projections.

1) PROBLEM
Synthetic data often replaces authentic data or is used together with the latter. They are an essential part of modeling and training various components of a checkpoint. Synthetic biometric traits are a class of algorithmically generated biometric, non-biometric, and cognitive behavior authentication credentials (e.g., face and facial expressions, fingerprints, voice, gait, user name, password,) used as a source for constructing a human profile for identity management [60].

2) MULTI-STATE SCREENING MODEL
We consider a multi-state screening in the dynamic cognitive system that (Fig. 7):

1) Monitors the traveler data throughout the process of e-ID checking, face recognition, and continuously assess the R-T-B using various sources such as behavioral biometrics, watchlist, and CI decision assistant results;

2) Updates its states based on the intelligence gathered via human-machine interactions (CI decision assistant), the results of the biometric traits recognition based on machine learning, the results of the concealed object detection (by adjusting radar illumination), and others.

In Fig. 7, the traveler's identity management process is implemented in three states, S_1 (ID validation), S_2 (Traveler authentication), and S_3 (Concealed object detection). Each state S_i and sub-state is a part of the 'Layered Security Strategy', a contemporary security doctrine [61]. Each state S_i and sub-state generates the R-T-B assessments for further processing. Inference using operations such as propagation, causal analysis, and reasoning can also be applied to the R-T-B assessments.

Because R-T-B are measured as probability events, they can be combined using *propagation* and *fusion* techniques. Synthetic data is required at various CI operations and processes. For example, the sub-state $S_m^{(1)}$ of state S_m is defined under *learnt* ID source reliability using authentic data from previous experience, while the potential attack data can be synthesized. This enables assessment of the R-T-B of such rare events (attacks).

3) FORMALIZATION ACCORDINGLY THE ADMIRALTY CODE
Consider a typical real-world scenario of the ID management process: Given an e-ID, assess the ID information credibility. At the descriptive level of the Admiralty Code, this scenario is represented as (Section VII):

`<Credibility>` ≡ `<Trustworthiness>`+`<Expertise>`

where `<Trustworthiness>` manifests itself as

$$\text{Trustworthiness} \equiv \begin{cases} \text{Source Reliability} \\ + \\ \text{Data Credibility} \end{cases}$$

4) CAUSAL NETWORK
Assessment of the ID information credibility is represented in Fig. 8 in the form of a Bayesian network and the corresponding CPTs are as follows:

Node *'ID source reliability'* denotes the three reliability levels of the e-passport/ID authentication, which depends on many risk factors such as the country of issue, the number of defence levels in the document, the life cycle history, the type of chip, the type of biometric modality, the type of encryption, and the type of RFID mechanism.

Node *'Valid ID'*, or *'Trusted ID'* denotes whether the e-passport ID should pass the validation procedure using factors such as watermarks, holograms, ultraviolet threads, micro text, and optical variable ink.

Node *'ID validation'* denotes the outcome of the authentication process of the e-passport.

Node *'ID credibility'* describes the three credibility level of the outcome of the validation process. If the credibility of the validation process is known as a *priori*, it can
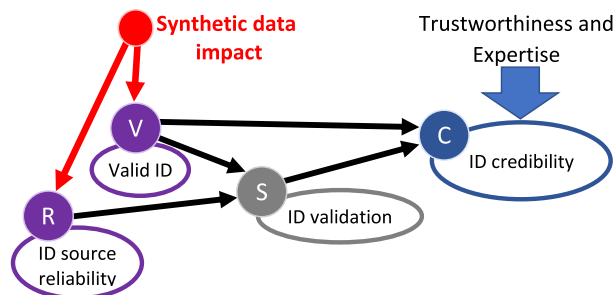

FIGURE 8. Assessment of the ID credibility (trustworthiness and expertise) using an IV-echelon (state) identity management scenario and its implementation. Synthetic data impact is incorporated using the CPTs of the nodes *R* and *V*.

be used to compute the posterior beliefs related to the validity of the individual document (node *V*).

5) SCENARIO AND REASONING EXAMPLE
Consider the following particular scenario: IF the reliability of the ID source is known to be 'low' and the resulting credibility is 'high', THEN what is the posterior probability that the ID is valid? This scenario models a situation of conflict where an unreliable source produces a credible outcome. It is very likely that the ID was valid. That is, the trustworthiness of the statement 'the ID was valid' is coherent with the expert knowledge (incorporated in the algorithms) [29].

6) SYNTHETIC DATA RISKS
Let us assume that to train algorithms for validating ID (node *V*) and identifying ID source reliability (node *R*), synthetic data was used to represent rare events, such as false ID, multiple ID of the same person, and features of intentional data alteration in the chip (e.g., biometric traits and text data) as well as a false life cycle history. Probabilities of these threats are represented by the CPTs for nodes *V* and *R*. There is always a risk that the validation algorithm makes a mistake should the real rare event occur. For example, features of the forged e-ID are not detected, or these features can be mistakenly detected in a valid ID. The goal is to assess these risks caused by the usage of synthetic data.

It is well understood that the frequency of object occurrence in the identity management process follows a long-tailed distribution. For example, people with true IDs and expired IDs are much more common than people with false IDs and multiple IDs. This problem relates the novelty detection, also known as *anomaly detection*, or one-class classification, the task of recognizing that the test data differ in some respect from the data that are available during training [62]. The tailed probability distributions have been used, for example, in the study of cyber-risks such as ID theft [63].

B. BIAS ENSEMBLE IN FACIAL RECOGNITION
There are three phases of bias analysis in the DSS: 1) Bias identification, e.g., what kind of biases are manifested in a system? 2) Bias assessment, e.g., a unified metric for different
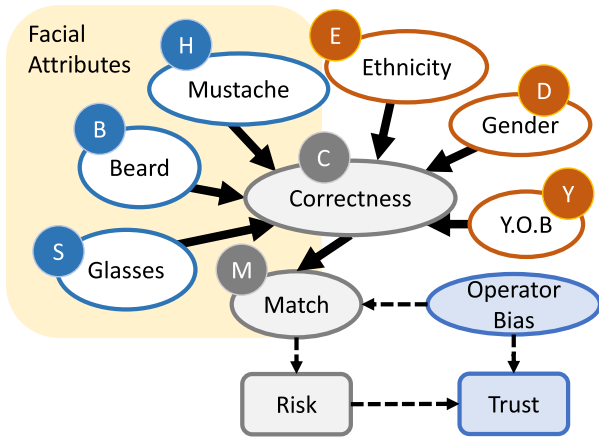
FIGURE 9. A simplified causal network of biases in facial recognition. Risk is derived based on the results of the "Match", and Trust is affected by the Operator's Bias.

kinds of biases, such as the risk of bias; and 3) Bias operation, e.g., the fusion of risks of biases.

1) PROBLEM

The traveler's identity management process is implemented as a process that goes through multiple states [15], [37]. Each state is characterized by a specific bias such as the bias in face recognition. Statistics of these biases are being used for machine learning. These biases are mostly represented by the tails of the probability distributions. A unified metric of bias that we consider in this study is the risk of bias and the related trust in the technology that is biased.

Our experiment addresses multiple biases in a cognitive security checkpoint such as "ID Reliability Bias", "ID Validation Bias", and "Trustworthiness Bias". Among various candidates of biases, we consider specifically "Face Recognition Bias". Few results on demographic bias in facial recognition have been recently reported, in particular, in [17], [18], [44]. The experimental details are described in [28], which aims to highlight the practical details of assessing an ensemble of biases.

2) CAUSAL MODEL

The causal network shown in Fig. 9 describes how the quality of facial recognition can be compromised by the various facial attributes that are "biased" based on the year-of-birth (YOB) Y, gender D, ethnicity E, mustache H, beard B, and glasses S. The parent nodes to the "Correctness" node represents the bias attributed to face recognition. The "Correctness" node presents the probability of the neural network in predicting a positive (genuine subject) or negative (imposter) identity, whereas the "Match" node determines whether the positive or negative prediction matches the ground truth label.

3) FORMALIZATION

Risk of error in the decision due to bias is estimated as `Risk = F(Impact, Probability)` which relates to the error rates of the system, specifically the false non-match rate (FNMR) and false match rate (FMR). In addition, the

risk value is associated with the probability of a random user being genuine given a particular bias `P(Genuine|Bias)`. At a high level of abstraction (e.g., ignoring metric and dependencies), the risk given a particular bias `Risk_Bias` is defined as follows:

$$\text{Risk}_{\text{Bias}} = \underbrace{\text{Impact}_{\text{FMR}}}_{\text{Cost of a FMR}} \times \text{FMR}_{\text{Bias}} \times (1 - \text{P(Genuine|Bias)})$$
$$+ \underbrace{\text{Impact}_{\text{FNMR}}}_{\text{Cost of a FNMR}} \times \text{FNMR}_{\text{Bias}} \times \text{P(Genuine|Bias)}$$

For example, given the scenario of a security checkpoint, the FMR is related to a wrongly granted access, while the FNMR contributes to travelers' inconvenience. The impact of the FMR is a breach of security which, given this scenario, should have a high impact. The impact of the FNMR is a negative user experience, which is of a lower impact. Based on this scenario, we assign a 10:1 impact ratio and a 90% genuine user probability, that is `Impact_FMR = 10`, `Impact_FNMR = 1`, `P(Genuine|Bias) = 0.9`. Given the YOB attribute, the risk for individuals born in the 1930s is computed as follows: $10 \times 0.0208 \times 0.1 + 1 \times 0.0012 \times 0.9 = \boxed{0.02188}$.

The ensemble risk bias in identifying (matching) a particular individual `Risk_Bias(Ensemble)` is assessed as the sum of the risk biases according to his/her attributes:

$$\text{Risk}_{\text{Bias}}(\text{Ensemble}) = \sum_{\text{Bias}=1}^{N} \text{Risk}_{\text{Bias}} \qquad (2)$$

where Bias represents one of the attributes, Y, D, E, H, B, and S in Fig. 9, $N = 6$.

4) EXPERIMENTAL RESULTS

For this experiment, we demonstrate biases using the FERET face database that contains a total of 14,126 images of 1199 subjects [64]. The typical performance measure for face recognition includes accuracy, FNMR, and FMR. The features used for face identification is extracted using a pre-trained Resnet50 convolutional neural network.

The numerical results are presented in detail in [28]. The identified biases include gender, year-of-birth, ethnicity, and facial attributes (glasses, beard, and mustache). The causal relationship between the biases and face recognition accuracy is represented by the causal network shown in Fig. 9. A significant bias in face recognition accuracy was observed with respect to the year-of-birth: **the accuracy decreased by 17.65% between those born in the 1920s and the 1980s**.

IX. FORECAST OF EMERGING APPLICATIONS

Results of our work are common among processes that can be modeled based on the principles of complex dynamic systems, e.g., learning, teaching, observation, conflict resolving, proactive computations, and countermeasures in real world and cyberspace. Below we introduce several emerging DSS over R-T-B applications:

- Epidemiological surveillance [7], [65]–[67];
- DSS for autonomy systems [68], [69];
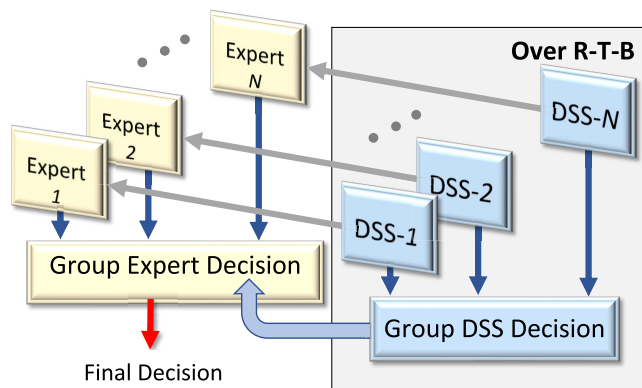- Combat DSS [16], [70], [71];

FIGURE 10. The framework of the DSS emerging applications over the R-T-B measures: human-machine using the DSS. The Final decision is the consensus of each expert's decision supported by the consensus of the DSS.

— Ambient DSS [72];
— E-coaching [3], [73].

These and other potential applications are based on the concept of group decision-making [74], [75]. Given an evidence and N experts, each expert is supported by the DSS to make a group decision (illustrated in Fig. 10).

5) EPIDEMIOLOGICAL SURVEILLANCE

Epidemiological surveillance experts (healthcare, first response, transportation, education, business and communication, media, security, police, etc.) need a near real-time, accurate picture of the extent and patterns of disease transmission at the community-level. Each expert in a specific field needs an intelligent DSS in order to better understand current and evolving healthcare demands, in order to be able to make low-risk, time-sensitive rapid decisions over various kinds of biases related, for example, on how to allocate limited and/or secure additional resources and how to relax mitigation efforts [76]. Part of these tasks uses biometrics, e.g., body vitals monitoring including thermal patterns, as well as personal protective equipment (PPE) detection and PPE-wearing person identification.

A recent survey on Covid-19 pandemic [65] covers decision-making support in many projections such as data-driving modeling, testing, tracing contacts, benchmarks, data hubs, machine learning, and privacy.

In [66], reasoning mechanism using a Bayesian causal network is used for studying *collider bias* of Covid-19 disease risk and severity. In causal networks, a variable is a collider when it is causally influenced by two or more variables; this results in either over-estimation or under-estimation of causal effects. Paper [66] also indicates potential Covid-19 collider biases caused by blood type, demographic factors, and related diseases. In [7], the DSS concept is implemented as causal reasoning on contact tracing to reduce the Covid-19 spread by providing diagnostic-oriented feedback (user symptoms) to citizens with near real-time Covid-19 surveillance, as well as an accurate picture of the extent and patterns of disease

transmission at the community-level for a constantly changing situation. Authors discuss various privacy risk mitigation approaches, public compliance, and trust. Causal reasoning upon the infection prevalence and fatality rates is used in [67].

6) DSS FOR AUTONOMOUS SYSTEMS

According to the taxonomical view in [68], the prevalent types of the **algorithmic biases** in autonomous systems include training data bias, algorithmic focus and processing biases, transfer context bias, and interpretation bias. Responses to the algorithmic bias, in particular, include identifying and intervening problematic biases. A related bias is known as the artificial intelligence bias [77]. Paper [69] addresses the mitigation of the various kinds of biases in autonomous systems using the concept of human-machine *trust repair*, described as a certain act that makes trust more positive. This is the same kind of action as a feedback loop in a cognitive DSS.

7) COMBAT DSS

Contemporary military combat teams include both soldiers and autonomous robots [16]. Situational awareness tasks for each soldier are supported in combat by a biometric-enabled, wearable DSS, e.g., stress and fatigue detector [70], [71]. This addresses the problem of individual effects of stress (cognitive, emotional, behavioral, and physiological) and team effect of stress such as decreased cooperation, ineffective communication, and decreased coordination. Decision-making in such a unit is radically different from a human-only team, since in a human-robot team, a portion of the responsibility is delegated to the intelligent machines. Rapid trust calibration becomes a task of high priority [78]. This problem formulation is known as a *human-CI teaming situational awareness* [79].

8) AMBIENT SYSTEMS

Ambient adaptive systems such as ambient CI assistants or monitors of human occupant vitals/biometrics have to use mechanisms to regulate themselves and change their structure in order to operate efficiently within dynamic ubiquitous computing environments. As a consequence of the increasingly aging population, it is necessary to find solutions to improve the living condition and develop more robust, usable, safe, and low-cost healthcare systems. This leads to a fixed DSS to be incorporated in ambient systems such as smart home, mobility and health assistants [72].

9) E-COACHING

E-coaching systems are aimed at supporting individuals in their self-regulation [73] using various biometrics. E-coaching offers support in the following areas: social ability, credibility, context-awareness, personalization (user tailoring), learning of user behavior, proactiveness, and guidance (coaching planning) [3], [73]. Measures of efficiency in such systems can naturally be expressed in terms of R-T-B.

X. CONCLUSIONS

Biometric-enabled systems are becoming an integral part of more complex intelligent systems. Such system-to-system embedding requires a deep unification of computational platforms and performance regulators. The proposed approach in this study use R-T-B as DSS performance indicators. These indicators shall become a mandatory assessment tool for all stages of the DSS development and deployment for the following reasons:

1) The R-T-B causal-based taxonomy provides efficient resources for deriving the knowledge (from biometrics) required for decision-making in biometric-enabled systems.
2) The DSS core, the reasoning over the R-T-B projection is implemented using causal networks (including Bayesian); each of these network types provides specific interpretation and approximation of uncertainty stemmed from the nature of biometric data.
3) The DSS with R-T-B indicators are most appropriate for forecasting applications, including risk assessment in the biometric-enabled systems that are planned to be implemented given the specific security, privacy, and usability scenarios.

From a *practical standpoint*, the R-T-B indicators are useful performance evaluation tools in any biometric-enabled system. From a *theoretical standpoint*, measuring the R-T-B is an ultimate probabilistic and computational intelligent problem because it aims at the development of a proactive mechanism to detect ill-defined phenomena from observable data. The problem is extremely challenging because the R-T-B are conceptual constructs (often psychological indicators) that are not directly observable and are computed from the multiple sources of factors embedded in a noisy context within a system operation. Among various challenges, we emphasize that consensus methodology for the group DSS is an open problem (Fig. 10).

REFERENCES

[1] R. D. Labati, A. Genovese, E. Munoz, V. Piuri, F. Scotti, and G. Sforza, "Biometric recognition in automated border control: A survey," *ACM Comput. Surv.*, vol. 49, no. 2, pp. A1–A39, Jun. 2016.

[2] Y. Andreu, F. Chiarugi, S. Colantonio, G. Giannakakis, D. Giorgi, P. Henriquez, E. Kazantzaki, D. Manousos, K. Marias, B. J. Matuszewski, and M. A. Pascali, "Wize mirror -a smart, multisensory cardio-metabolic risk monitoring system," *Comput. Vis. Image Understand.*, vol. 148, pp. 3–22, Jul. 2016.

[3] S. F. Ochoa and F. J. Gutierrez, "Architecting E-coaching systems: A first step for dealing with their intrinsic design complexity," *Computer*, vol. 51, no. 3, pp. 16–23, Mar. 2018.

[4] G. Rigas, Y. Goletsis, and D. I. Fotiadis, "Real-time driver's stress event detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 1, pp. 221–234, Mar. 2012.

[5] A. Roy and D. Dasgupta, "A fuzzy decision support system for multifactor authentication," *Soft Comput.*, vol. 22, no. 12, pp. 3959–3981, Jun. 2018.

[6] F. Weng, P. Angkititrakul, E. E. Shriberg, L. Heck, S. Peters, and J. H. L. Hansen, "Conversational in-vehicle dialog systems: The past, present, and future," *IEEE Signal Process. Mag.*, vol. 33, no. 6, pp. 49–60, Nov. 2016.

[7] S. McLachlan, P. Lucas, K. Dube, G. S. McLachlan, G. A. Hitman, M. Osman, and N. E. Fenton, "The fundamental limitations of COVID-19 contact tracing methods and how to resolve them with a Bayesian network approach," Risk Inf. Manage., London, U.K., Tech. Rep., 2020, doi: 10.13140/RG.2.2.27042.66243.

[8] K. Neville, S. O'Riordan, A. Pope, M. Rauner, M. Rochford, M. Madden, J. Sweeney, A. Nussbaumer, N. McCarthy, and C. O'Brien, "Developing a decision support tool and training system for multi-agency decision making during an emergency," in *Proc. Eur. Secur. Res. Conf.* Dublin, Republic of Ireland: National Standards Authority of Ireland Committee, 2015.

[9] A. Poursaberi, S. Yanushkevich, M. L. Gavrilova, V. P. Shmerko, and P. S. P. Wang, "Situational awareness through biometrics," *Computer*, vol. 46, no. 5, pp. 102–104, May 2013.

[10] R. Bolle, J. Connel, S. Pankanti, N. Ratha, and A. W. Senior, *Guide to Biometrics*. Berlin, Germany: Springer, 2004.

[11] A. Andreou, O. Goga, and P. Loiseau, "Identity vs. Attribute disclosure risks for users with multiple social profiles," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining*, Jul. 2017, pp. 163–170.

[12] W.-L. Hu, K. Akash, T. Reid, and N. Jain, "Computational modeling of the dynamics of human trust during human–machine interactions," *IEEE Trans. Human-Mach. Syst.*, vol. 49, no. 6, pp. 485–497, Dec. 2019.

[13] K. Hugenberg, J. P. Wilson, P. E. See, and S. G. Young, "Towards a synthetic model of own group biases in face memory," *Vis. Cognition*, vol. 21, nos. 9–10, pp. 1392–1417, Sep. 2013.

[14] G. Montibeller and D. von Winterfeldt, "Cognitive and motivational biases in decision and risk analysis," *Risk Anal.*, vol. 35, no. 7, pp. 1230–1251, Jul. 2015.

[15] S. N. Yanushkevich, K. W. Sundberg, N. W. Twyman, R. M. Guest, and V. P. Shmerko, "Cognitive checkpoint: Emerging technologies for biometric-enabled watchlist screening," *Comput. Secur.*, vol. 85, pp. 372–385, Aug. 2019.

[16] M. W. Brown, "Developing readiness to trust artificial intelligence within warfighting teams," *Mil. Rev.*, vol. 100, pp. 38–43, Jan./Feb. 2020.

[17] A. Das, A. Dantcheva, and F. Bremond, "Mitigating bias in gender, age and ethnicity classification: A multi-task convolution neural network approach," in *Proc. Eur. Conf. Comp. Vis.* Cham, Switzerland: Springer, 2019, pp. 573–585.

[18] P. Grother, M. Ngan, and K. Hanaoka, "Face recognition vendor test (FRVT) Part 3: Demographic effects," Nat. Inst. Standards Technol., Gaithersburg, MD, USA, Tech. Rep. 8280, 2019.

[19] L. C. Schaupp and L. Carter, "The impact of trust, risk and optimism bias on E-file adoption," *Inf. Syst. Frontiers*, vol. 12, no. 3, pp. 299–309, Jul. 2010.

[20] L. Zanin, R. Radice, and G. Marra, "Estimating the effect of perceived risk of crime on social trust in the presence of endogeneity bias," *Social Indicators Res.*, vol. 114, no. 2, pp. 523–547, Nov. 2013.

[21] P. Zhou, X. Luo, A. Chen, and R. K. Chang, "SGor: Trust graph based onion routing," *Comput. Netw.*, vol. 57, no. 17, pp. 3522–3544, 2013.

[22] Y.-C. Lee, "Impacts of decision-making biases on eWOM retrust and risk-reducing strategies," *Comput. Hum. Behav.*, vol. 40, pp. 101–110, Nov. 2014.

[23] M. Groß, "Impediments to mobile shopping continued usage intention: A trust-risk-relationship," *J. Retailing Consum. Services*, vol. 33, pp. 109–119, Nov. 2016.

[24] A. Gates, B. Vandermeer, and L. Hartling, "Technology-assisted risk of bias assessment in systematic reviews: A prospective cross-sectional evaluation of the RobotReviewer machine learning tool," *J. Clin. Epidemiology*, vol. 96, pp. 54–62, Apr. 2018.

[25] T. Van Hamme, D. Preuveneers, and W. Joosen, "Managing distributed trust relationships for multi-modal authentication," *J. Inf. Secur. Appl.*, vol. 40, pp. 258–270, Jun. 2018.

[26] J. Lopez and S. Maag, "Towards a generic trust management framework using a Machine-Learning-Based trust model," in *Proc. IEEE Trustcom/BigDataSE/ISPA*, Aug. 2015, pp. 1343–1348.

[27] U. Jayasinghe, G. M. Lee, T.-W. Um, and Q. Shi, "Machine learning based trust Computational model for IoT services," *IEEE Trans. Sustain. Comput.*, vol. 4, no. 1, pp. 39–52, Jan./Mar. 2019.

[28] K. Lai, H. C. R. Oliveira, M. Hou, S. N. Yanushkevich, and V. Shmerko, "Assessing risks of biases in cognitive decision support systems," *Proc. 28th Eur. Signal Process. Conf., Special Session Bias Biometrics*, Amsterdam, The Netherlands, Jan. 2021, pp. 1–5.

[29] S. Yanushkevich, A. Stoica, V. Shmerko, G. Howells, K. Crockett, and R. Guest, "Cognitive identity management and synthetic data risk and trust," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Glasgow, U.K., Jul. 2020, pp. 1–8.

[30] S. Haykin, *Cognitive Dynamic Systems (Perception-Action Cycle, Radar, and Radio)*. New York, NY, USA: Cambridge Univ. Press, 2012.

[31] J. Pearl, "The seven tools of causal inference, with reflections on machine learning," *Commun. ACM*, vol. 62, no. 3, pp. 54–60, Feb. 2019.

[32] J. Pearl, *Models, Reasoning, and Inference*. Cambridge, U.K.: Cambridge Univ. Press, 2000.

[33] U.S. Army Training and Doctrine Command (TRADOC), *Joint Warfare Publication 2-00 Intelligence Support to Joint Operations*, Admiralty Code, Joint Doctrine and Concepts Centre: Ministry of Defence, Shrivenham, U.K., 2012.

[34] U.S. Army Training and Doctrine Command (TRADOC), "Intelligence source and information reliability, U.S. Army field manual FM 2-22.3," in *Human Intelligence Collector Operations*, Washington, DC, USA: Dept. Army, U.S. Army Field Manual, 2006.

[35] E. Blasch, K. B. Laskey, A.-L. Jousselme, V. Dragos, P. C. G. Costa, and J. Dezert, "URREF reliability versus credibility in information fusion (STANAG 2511)," in *Proc. 16th Int. Conf. Inf. Fusion*, Jul. 2013, pp. 1600–1607.

[36] J. Rohmer, "Uncertainties in conditional probability tables of discrete Bayesian belief networks: A comprehensive review," *Eng. Appl. Artif. Intell.*, vol. 88, Feb. 2020, Art. no. 103384.

[37] S. N. Yanushkevich, W. G. Howells, K. A. Crockett, J. O'Shea, H. C. R. Oliveira, R. M. Guest, and V. P. Shmerko, "Cognitive identity management: Risks, trust and decisions using heterogeneous sources," in *Proc. IEEE 1st Int. Conf. Cognit. Mach. Intell. (CogMI)*, Dec. 2019, pp. 33–42.

[38] *Security and Privacy Controls for Information Systems and Organizations, NIST Special Publication 800-53, Revision 5*, National Institute of Standards (NIST), Gaithersburg, MD, USA, 2017.

[39] D. Gambetta, "Can we trust trust," in *Trus, Making and Breaking Cooperative Relations*, D. Gambetta, Ed. Oxford, U.K.: Dept. Sociology, Univ. Oxford, 2000, pp. 213–237.

[40] B. Solhaug, D. Elgesem, and K. Stolen, "Why trust is not proportional to risk," in *Proc. 2nd Int. Conf. Availability, Rel. Secur. (ARES)*, Apr. 2007, pp. 11–18.

[41] J. D. Lee and K. A. See, "Trust in automation: Designing for appropriate reliance," *Hum. Factors, J. Hum. Factors Ergonom. Soc.*, vol. 46, no. 1, pp. 50–80, 2004.

[42] J.-H. Cho, K. Chan, and S. Adali, "A survey on trust modeling," *ACM Comput. Surv.*, vol. 48, no. 2, pp. 1–40, Nov. 2015.

[43] M. S. Cohen, R. Parasuraman, and J. T. Freeman, "Trust in decision aids: A model and its training implications," in *Proc. Command Control Res. Technol. Symp.*, 1998, pp. 1–37.

[44] M. Merler, N. Ratha, R. S. Feris, and J. R. Smith, "Diversity in faces," 2019, *arXiv:1901.10436*. [Online]. Available: http://arxiv.org/abs/1901.10436

[45] *European Union Agency for Fundamental Rights, Data quality and Artificial Intelligence-Mitigating Bias and Error to Protect*, Publications Office of the EU, Brussels, Belgium, 2019.

[46] *Risk Management. Principles and Guidelines*, document ISO. BS ISO 31000:2009, 2009.

[47] S. Rass and S. Kurowski, "On Bayesian trust and risk forecasting for compound systems," in *Proc. 7th Int. Conf. IT Secur. Incident Manage. IT Forensics*, Mar. 2013, pp. 69–82.

[48] S. Yanushkevich, S. Eastwood, M. Drahansky, and V. Shmerko, "Understanding and taxonomy of uncertainty in modeling, simulation, and risk profiling for border control automation," *J. Defense Model. Simul., Appl., Methodol., Technol.*, vol. 15, no. 1, pp. 95–109, 2018.

[49] B. Yang, M. Aukrust, C. Busch, J. Bringer, E. Kindt, W. R. Belser, U. Seidel, E. Springmann, U. Rabeler, and A. Wolf, "Towards standardizing trusted evidence of identity," in *Proc. ACM Workshop Digit. Identity Manage. DIM*, 2013, pp. 63–72.

[50] F. P. A. Coolen, M. C. M. Troffaes, and T. Augustin, "Imprecise probability," in *International Encyclopedia of Statistical Science*, M. Lovric, Ed. Berlin, Germany: Springer, 2011.

[51] L. M. De Campos, J. F. Huete, and S. Moral, "Probability intervals: A tool for uncertain reasoning," *Int. J. Uncertainty, Fuzziness Knowl.-Based Syst.*, vol. 2, no. 2, pp. 167–196, Jun. 1994.

[52] F. G. Cozman, "Credal networks," *Artif. Intell.*, vol. 120, no. 2, pp. 199–223, 2000.

[53] C. Simon, P. Weber, and A. Evsukoff, "Bayesian networks inference algorithm to implement dempster shafer theory in reliability analysis," *Rel. Eng. Syst. Saf.*, vol. 93, no. 7, pp. 950–963, Jul. 2008.

[54] J. F. Baldwin and E. Di Tomaso, "Inference and learning in fuzzy Bayesian networks," in *Proc. 12th IEEE Int. Conf. Fuzzy Syst. FUZZ*, May 2003, pp. 630–635.

[55] M. Ivanovska, A. Jøsang, L. Kaplan, and F. Sambo, "Subjective networks: Perspectives and challenges," in *Proc. 4th Int. Workshop Graph Struct. Knowl. Represent. Reasoning*, M. Croitoru, Ed. Cham, Switzerland: Springer, 2015, pp. 107–124.

[56] A. Misuri, N. Khakzad, G. Reniers, and V. Cozzani, "Tackling uncertainty in security assessment of critical infrastructures: Dempster-shafer theory vs. Credal sets theory," *Saf. Sci.*, vol. 107, pp. 62–76, Aug. 2018.

[57] A. Hunter, "Reasoning about trust and belief change on a social network: A formal approach," in *Proc. 13th Int. Conf. Inf. Secur. Pract. Exper.*, J. K. Liu and P. Samarati, Eds. Cham, Switzerland: Springer, 2017, pp. 783–801.

[58] M. L. Gavrilova and R. Yampolskiy, "Applying biometric principles to avatar recognition," in *Transactions on Computational Science XII*, M. L. Gavrilova, Ed. Berlin, Germany: Springer, 2011, pp. 140–158.

[59] S. M. Bellovin, P. K. Dutta, and N. Reitinger, "Privacy and synthetic datasets," *Stanford Tech. Law Rev.*, vol. 22, no. 1, pp. 1–52, 2019.

[60] Active Authentication. (2016). *U.S. Defense Advanced Research Projects Agency (DARPA)*. [Online]. Available: http://www.darpa.mil/program/active-authentication

[61] Transportation Security Administration. (2013). *Layers of Security*. [Online]. Available: http://www.tsa.gov/about-tsa/layerssecurity

[62] M. A. F. Pimentel, D. A. Clifton, L. Clifton, and L. Tarassenko, "A review of novelty detection," *Signal Process.*, vol. 99, pp. 215–249, Jun. 2014.

[63] T. Maillart and D. Sornette, "Heavy-tailed distribution of cyber-risks," *Eur. Phys. J. B*, vol. 75, no. 3, pp. 357–364, Jun. 2010.

[64] P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss, "The FERET database and evaluation procedure for face-recognition algorithms," *Image Vis. Comput.*, vol. 16, no. 5, pp. 295–306, Apr. 1998.

[65] T. Alamo, D. Reina, M. Mammarella, and A. Abella, "Covid-19: Open-data resources for monitoring, modeling, and forecasting the epidemic," *Electronics*, vol. 9, no. 5, p. 827, May 2020.

[66] N. Fenton, "A note on 'Collider bias undermines our understanding of COVID-19 disease risk and severity' and how causal Bayesian networks both expose and resolve the problem," 2020, *arXiv:2005.08608*. [Online]. Available: http://arxiv.org/abs/2005.08608

[67] M. Neil, N. Fenton, M. Osman, and S. McLachlan, "Bayesian network analysis of Covid-19 data reveals higher infection prevalence rates and lower fatality rates than widely reported," *J. Risk Res.*, pp. 1–14, Jun. 2020.

[68] D. Danks and A. J. London, "Algorithmic bias in autonomous systems," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, Aug. 2017, pp. 4691–4697.

[69] E. J. de Visser, R. Pak, and T. H. Shaw, "From 'automation' to 'autonomy': The importance of trust repair in human–machine interaction," *Ergonomics*, vol. 61, no. 10, pp. 1409–1427, Oct. 2018.

[70] P. Schmidt, A. Reiss, R. Dürichen, and K. V. Laerhoven, "Wearable-based affect recognition—A review," *Sensors*, vol. 19, no. 19, p. 4079, Sep. 2019.

[71] P. Pandey, E. K. Lee, and D. Pompili, "A distributed computing framework for real-time detection of stress and its propagation in a team," *IEEE J. Biomed. Health Inf.*, vol. 20, no. 6, pp. 1502–1412, Nov. 2016.

[72] R. Li, B. Lu, and K. D. McDonald-Maier, "Cognitive assisted living ambient system: A survey," *Digit. Commun. Netw.*, vol. 1, no. 4, pp. 229–252, Nov. 2015.

[73] B. A. Kamphorst, "E-coaching systems What they are, and what they aren't," *Pers. Ubiquitous Comput.*, vol. 21, no. 4, pp. 625–632, 2017.

[74] T. Bedford, "Decision making for group risk Reduction: Dealing with epistemic uncertainty," *Risk Anal.*, vol. 33, no. 10, pp. 1884–1898, 2013.

[75] N. H. Kamis, F. Chiclana, and J. Levesley, "Preference similarity network structural equivalence clustering based consensus group decision making model," *Appl. Soft Comput.*, vol. 67, pp. 706–720, Jun. 2018.

[76] D. Manheim, M. Chamberlin, O. A. Osoba, R. Vardavas, and M. Moore, *Improving Decision Support for Infectious Disease Prevention and Control: Aligning Models and Other Tools with Policymakers' Needs*. Santa Monica, CA, USA: RAND Corporation, 2016.

[77] M. Whittaker. (2018). *AI Now Report, New York University*. [Online]. Available: https://ainowinstitute.org/AI_Now_2018_Report.pdf

[78] R. Tomsett, A. Preece, D. Braines, F. Cerutti, S. Chakraborty, M. Srivastava, G. Pearson, and L. Kaplan, "Rapid trust calibration through interpretable and uncertainty-aware AI," *Patterns*, vol. 1, Jul. 2020, Art. no. 100049, doi: 10.1016/j.patter.2020.100049.

[79] M. Hou, C. M. Burns, and S. Banbury, *Intelligent Adaptive Systems: An Interaction-Centered Design Perspective*. Boca Raton, FL, USA: CRC Press, 2014.

KENNETH LAI received the B.Sc. and M.Sc. degrees from the University of Calgary, Calgary, AB, Canada, in 2012 and 2015, respectively, where he is currently pursuing the Ph.D. degree with the Department of Electrical and Computing Engineering. His areas of interest include biometrics and its application to security and health care systems.

HELDER C. R. OLIVEIRA received the B.Sc. degree in computer science from the São Paulo State University, Brazil, in 2013, and the M.Sc. and Ph.D. degrees in electrical engineering from the University of São Paulo, Brazil, in 2015 and 2019 respectively. He is currently Postdoctoral Fellow with the Biometric Technologies Laboratory, University of Calgary, Canada. His research interests include probabilistic models, pattern recognition, and biometrics.

MING HOU (Senior Member, IEEE) is currently a Senior Defence Scientist with DRDC and the Principal Authority of Human-Technology Interactions with the Department of National Defence (DND), Canada. He is responsible for providing science-based advice at national and international levels to the Canadian Armed Forces (CAF) and coalition partners about the investment in and application of advanced technologies for human–machine systems requirements. He is an Integrator for the Canadian government 16 billion IDEaS program and one of the three Scientific Advisors to the Canadian National Centre of Expertise in Human Systems Performance with responsibilities for guiding national research and development activities in automation, robotics, and telepresence. He also gives advice for the development of National Defence AI Science and Technology Strategy and Roadmap to the CAF and DND. As the Canadian National Leader of the Joint Panel on Human Systems Performance—Air under The Technical Cooperation Program (TTCP), he leads Canadian science and technology activities for military air systems and manages large international collaborative projects (e.g., a $200M Human-Autonomy Teaming technology demonstration project). He was the Canadian National Lead and Scientific Authority during the TTCP Autonomous Warrior 2018 Joint Service Exercise. He is the Co-Chair and the Head of Canadian Delegation to NATO Human Factors Specialist Committee within NATO Joint Capability Group on Unmanned Aircraft Systems (UAS). His book *Intelligent Adaptive Systems: An Interaction-Centered Design Perspective* became a guiding document to the development of NATO Standard Recommendations on UAS Human Systems Integration Guidebook, *UAS Human Factors Experimentation Guidelines* and *UAS Sense and Avoid Guidance*. As one of the four invited lecturers, he delivers NATO Lecture Series on UAVs: Technological Challenges, Concepts of Operations, and Regulatory Issues. He also serves for multiple international associations/programs as a chair and board members.

SVETLANA N. YANUSHKEVICH (Senior Member, IEEE) received the Dr.Tech.Sc. (Dr. Habilitated) degree from the Warsaw University of Technology, in 1999. She is currently a Professor with the Department of Electrical and Computer Engineering, University of Calgary. She is directing the Biometric Technologies Laboratory and conducting research in the area of biometric-based authentication technologies.

VLAD P. SHMERKO is a Developer of authentication machines for security applications. He is currently a Researcher with the Biometric Technologies Laboratory, University of Calgary, Calgary, AB, Canada. He took part in about 30 projects in this area and coauthored over 50 patents. He is a Fellow and a Chartered Engineer with the Institution of Engineering and Technology, U.K.

• • •