11th Annual Environmental Health Summit

Artificial Intelligence in Environmental Health Science and Decision Making

October 18-19, 2018
North Carolina Biotechnology Center, RTP, NC

Introduction

Environmental health sciences (EHS) research and practice have ushered in countless advances in public health over the last century; however, unsolved EHS challenges from the last century remain, and emerging challenges continue to arise as the 21st-century unfolds. Solving both persistent and novel challenges will require considerable changes in EHS paradigms and approaches. Innovations within the field of artificial intelligence (AI) and its subset of machine learning (ML) offer possible catalysts for such changes.

AI techniques have been well established since the 1950s. The exponential increase in data production and advances in computing power and storage in the 2000s–2010s have created a recent surge in AI research and implementation. Although AI and associated terms have been commonly used by the public for over a decade, the field remains somewhat impenetrable for those outside AI practice and research. Much of the progress in the field of AI has been restricted to computer science, data science, finance, and business analytics. The significant potential of AI has not yet been realized in EHS.

The field of EHS is well poised to benefit from AI and ML. From the influx of data from consumer-level genome testing to collaborative efforts on high-throughput screening, conditions are optimal to utilize large and complex datasets in innovative ways. Such innovation has the potential to enable dramatic advances in EHS that will more efficiently and effectively protect human health.

To support the integration of AI/ML into EHS, the Research Triangle Environmental Health Collaborative hosted its 11th Annual Summit (Summit) on Artificial Intelligence in Environmental Health Science and Decision Making on October 18-19 in Research Triangle Park, North Carolina (http://environmentalhealthcollaborative.org/2018-summit/). The Summit, held at the North Carolina Biotechnology Center, brought together over 100 individuals currently working or studying EHS, AI/ML, and data science to identify and prioritize EHS research questions that might be investigated with AI/ML techniques.

The Summit hosted a half-day workshop on October 18 to introduce attendees to basic concepts in AI/ML and to provide an overview of a software platform researchers can use to employ AI/ML techniques. On October 19, the Summit held three plenary presentations, five case-study presentations, three breakout discussions and report backs, and concluding remarks. Descriptions of certain AI/ML concepts, summaries of presentations, and recommendations developed during the Summit are described herein.

Summit Overview

BAYESIAN NETWORKS WORKSHOP

As part of the Summit, Stefan Conrady, managing partner at Bayesia USA, presented a half-day workshop to approximately 40 attendees. The workshop introduced the fundamentals of Bayesian Networks (BNs) and incorporated AI/ML terminology and analytic concepts (e.g., unsupervised v. supervised learning, algorithmic v. parametric model sources, predictive v. explanatory model purposes. etc.). The workshop also employed use-cases for BayesiaLab software, including knowledge modeling, knowledge elicitation to build an expert group-derived BN, a diagnostic decision support tool, structure discovery through exploratory analysis and visualizations, and a policy analysis.

BNs are a type of probabilistic graphical model. Variables are represented as nodes and relationships are represented by arrows; nodes without arrows (parent node) are considered statistically independent, while nodes with arrows (child node) are dependent. Each node has an associated conditional probability table. A simple example of a BN structure is presented in Figure 1[1]. In this example, Variable X1 (Season) is a statistically independent parent node, while variables X2 (Rain) and X3 (Sprinkler) are child nodes dependent on the season.

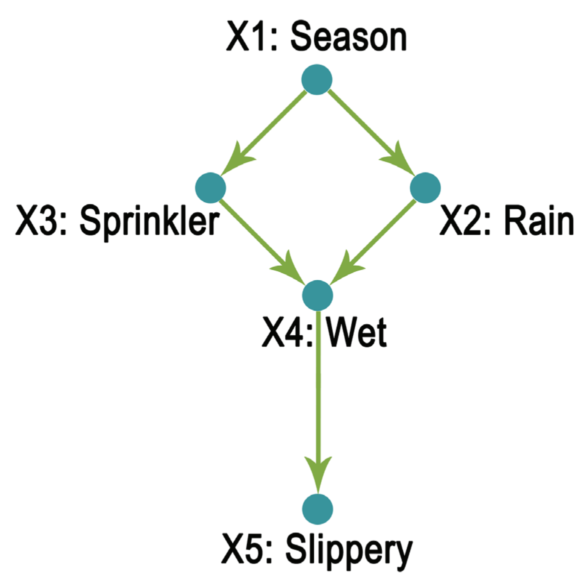

Figure 1.
Source: Adapted from Conrady & Jouffey, 2015

The benefits of BNs are numerous and include:

- Ability to team humans and machines (e.g., domain experts can 'encode' their knowledge and build network structures through theory, or the structures can be machine-learned and further interpreted by domain experts)
- Inclusion of both independent/dependent variables and categorical/continuous variables
- Omni-directional inference
- Flexibility in both approach and use
- Reduction in dimensionality and more compact joint probability distributions

MODERN MACHINE LEARNING: PROBABILISTIC MODELING AND FUNCTIONAL PREDICTION — TOM DIETTERICH

Tom Dietterich, Ph.D., Distinguished Professor Emeritus at Oregon State University's School of Electrical Engineering and Computer Science, delivered the opening plenary. Dietterich began with an introduction into the original motivations behind ML and the evolution of the field as it broadened from pure computer science into the fields of statistics, engineering, and health. Two primary paradigms have developed within the field of ML: probabilistic modeling and functional prediction methods. Probabilistic modeling broadly involves defining and fitting probabilistic models to data, while the functional learning paradigm involves fitting a highly-accurate prediction function from a non-parametric class of functions. Dietterich noted that both paradigms have a well-developed set of software tools for various ML approaches. He demonstrated an example of such software (Stan) using a multi-level, probabilistic modeling approach. Multi-level modeling can account for individual-level and group-level variation when estimating group level coefficients (i.e. incorporating individual-level estimates into group-level estimates). Multi-level modeling can also stabilize subgroups with small sample sizes by borrowing strength from subgroups with larger sample sizes.

To illustrate examples within the functional prediction field, Dietterich described the premises, advantages, and disadvantages of two popular functional prediction approaches: random forests and support vector machines. The random forest approach is a collection of many randomized decision trees that can overcome the statistical instability of individual decision trees and yield excellent predictive accuracy; however, they are generally considered to be "black boxes" because their internal processes are not easily inspected. Support vectors can enhance approaches to classification questions by maximizing boundary margins, for example, by allowing for greater tolerance in measurement error without changing the classification boundary, and can also generate non-linear classification boundaries. This approach does not readily scale to large datasets. Dietterich's final example involved deep learning methods, which have a promising set of applications including computer vision and speech recognition. Unfortunately, deep neural networks must be constructed for each problem or research question and are generally data-hungry and computationally intensive.

Dietterich sees a role for both probabilistic modeling and functional learning in the field of EHS, although different subfields will be better served by some techniques over others. He concluded that probabilistic modeling is likely the most useful ML tool for policy development and decision making in environmental health, as it can allow for causal inference under some conditions and can generate highly interpretable models. Functional learning approaches are better suited for extracting data from unstructured datasets, such as analyzing medical images or electronic health records. Dietterich advised the audience on the inherent presence of bias in modeling and ML and advised that all data analyses must be conducted with caution and care.

SYSTEMATIC REVIEWS, MACHINE LEARNING, AND THE LIBERATION OF KNOWLEDGE FROM INFORMATION IN ENVIRONMENTAL HEALTH RESEARCH — PAUL WHALEY

Paul Whaley, Ph.D., a Research Fellow at Lancaster University, UK, presented a plenary on the importance of systematic reviews (SR) in EHS and their potential when combined with AI/ML. He suggested that SRs, if properly proceduralized, can minimize bias in reviewing evidence and help attenuate the problem of data reproducibility in research. The past decade has seen a significant uptake of SR methods in chemical risk assessment and EHS research. SR protocols are now available from major health organizations (e.g. World Health Organization, U.S. Environmental Protection Agency).

Despite the benefits of SRs, major drawbacks remain. Notably, they are time-intensive and limited in scope (e.g. restricted to single exposure/outcome pair). These drawbacks limit their usefulness when thousands of chemicals have yet to be tested and new studies expanding the knowledge base are published at a fast rate. The confluence of these factors yields a significant data volume problem. Semantic databases of subject-predicate-object triples (e.g., Rat Group | Is Dosed with | BPA; Rat Group | is investigated for | Liver Tumors) are better suited to solve such problems than typical relational databases, which summarize studies as tables—a poor reflection of the real complexity in data. Semantic databases can generate graphical databases that mirror adverse outcome pathways. Constructing such databases is not currently possible using human labor but presents an opportunity for AI/ML. Whaley proposed that ML will necessarily have to be integrated into SRs to overcome these challenges and produce better chemical risk assessments.

To achieve progress in this space of automated and accurate SRs, however, humans have a significant amount of front-end work to accomplish. Whaley identified critical areas of work that researchers can contribute to now, including producing SR training sets (detailed annotation of the literature by domain experts), publishing documents that are more readily machine-readable, and improving reporting practices to enable better understanding of the quality of input data. These are all modifiable behaviors that can be undertaken by scientific communities.

According to Whaley, human knowledge is currently dispersed, compartmentalized, and inconveniently accessed. Integrating domain knowledge into graphical databases produced by AI/ML can yield an illustrative and interpretable amalgamation of knowledge that is far more informative than single “data points” contained in single journal articles scattered across the internet and left to individuals to assimilate and disperse.

ENVIRONMENTAL HEALTH… IN CONTEXT — SAMUEL ADAMS

Samuel Adams, Ph.D., a Senior Artificial Intelligence Researcher at RTI International, explored the value, power, and importance of placing data within context. Adams noted that one piece of data in isolation is essentially worthless, and posited that value instead comes from connecting singular pieces of data and that the connection between data is often more valuable than the data itself. These connections create context, he said. Unfortunately, much of that context resides within the minds of the researchers who generated the data. Adams reaffirmed Whaley’s suggestion that the process of distilling, writing, publishing, and extracting data causes much of the context to be lost by creating multiple discrete data points contained within innumerable journals, books, and PDFs. Synthesizing this data in a way that can be reintegrated into the minds of researchers would reintroduce context, he asserted, and thus reintroduce meaning and value. Summary tables and databases have struggled to recapture this lost context, but knowledge graphs (i.e. graphical databases) may be one such way to accomplish this.

A knowledge graph is a method of integrating numerous ontologies. Ontologies are a body of knowledge with formal definitions, relationships, categories, concepts, and data types. Adams noted that although EHS may have thousands of separate ontologies and individuals may describe ontologies differently, knowledge graphs blend ontologies in a way that aggregates discrete data and transcends summary tables. Adams noted that he has experienced some resistance to these concepts from those outside the fields of computer and data sciences, but made the case that large sectors of the world economy are successfully built on such knowledge graphs (e.g. Facebook, Amazon, and government intelligence agencies). He suggested that this is a mature and stable technology that should not be feared but rather pursued in other areas, such as EHS. He noted that there are essentially two pathways for producing graphical databases: Property Graphs or Resource Description Framework Triple Graphs, each of which has unique benefits, drawbacks, and userbases. According to Adams, there are numerous open-source methods to produce such graphical databases.

Collecting and integrating the entirety of EHS knowledge could lead to significant breakthroughs and may seem ideal, but Adams noted significant obstacles, primarily data access barriers. Organizational barriers, agency barriers, proprietary barriers, and privacy barriers would all stymie such an endeavor. Such barriers are largely human-driven rather than technology-driven. Adams stated that cooperation must increase the value for all shareholders and suggested that creating a way for everyone to "win" would encourage cooperation between these spheres; he argued for the equivalent of a digital data barter system where researchers identify pieces of information that would add value or insight into others' work and that have previously been inaccessible to them. He noted that although creating EHS knowledge graphs will be a difficult undertaking, the endeavor is worth the effort because the true value of data is revealed by placing it in context.

Case Studies

HARNESSING MACHINE LEARNING TO PREDICT TOXICITIES — NICOLE KLEINSTREUER

Nicole Kleinstreuer, Ph.D., is the Deputy Director of the National Toxicology Program's Interagency Center for the Evaluation of Alternative Toxicological Methods. Kleinstreuer's case study focused on the regulatory and health challenges of the current chemical landscape and opportunities for AI/ML in this sphere. The overwhelming majority of chemicals humans are exposed to are not studied or regulated; of the 140,000 mono-constituent chemicals found in commerce, less than 10% have been studied to evaluate safety. This number does not include mixtures, natural products, or metabolites that may also pose toxicological risk to human health. This gap is largely due to the limitations of traditional toxicological methods of animal studies, which are costly, time-intensive, low-throughput, ethically challenging regarding animal use, and may not provide exact insight into human disease pathways. This presents a clear opportunity for AI/ML to bridge the gap. Kleinstreuer and others have begun identifying endpoints that may be of interest to regulators such as endocrine disruption and acute systemic toxicity endpoints, to which AI/ML techniques may be applied. Global research projects, such as the Collaborative Modeling Project for Androgen Receptor Activity and the Collaborative Acute Toxicity Modeling Suite, build the large, curated datasets that are needed for ML training sets.

Quantitative Structure-Activity Relationship modeling groups from around the world are provided these training sets and asked to participate in collaborative efforts to produce models. Over 100 groups have participated in such collaborative efforts using a variety of ML techniques. Kleinstreuer noted that this type of crowd-sourced ML compensates for any algorithm limitations that any one group might use, and that building consensus models in this fashion allows for a very robust toxicological approach that can enable regulators to prioritize certain chemicals for further testing. Furthermore, these approaches may eventually allow risk assessors to assess risk probabilistically rather than relying on the current methods of reference doses and uncertainty factors, and provide better methods for approximating risks of exposure to mixtures, which are currently difficult to quantify.

USING BAYESIAN NETWORKS TO DISCOVER RELATIONS BETWEEN GENES, ENVIRONMENT, AND DISEASE — MARK BORSUK

Building on the foundation of the BayesiaLab workshop, Mark Borsuk, Ph.D., Associate Professor of Civil and Environmental Engineering at Duke University, continued the exploration of Bayesian Networks (BNs) in his case study. Borsuk explained structure learning, or structure 'discovery,' in which domain knowledge is not encoded in the structure of machine-learned BNs. Structure learning presents vast opportunities for genetic epidemiology, particularly in discovering complex relationships from substantial numbers of single nucleotide polymorphisms (SNPs), including in both gene-gene associations and gene-environment interactions. Borsuk illustrated such a scenario in his case study of a bladder cancer dataset comprised of 1,477 SNPs in cancer-related genes, as well as demographic/epidemiologic data including smoking status, age, gender, and arsenic exposure. The large number of SNP variables in the dataset precluded encoding prior knowledge into a BN structure and thus produced a non-causal machine-learned structure. Such a structure not only provides insight into relationships that would be difficult to identify using traditional methods, but also into probabilities, and, thus, odds ratios. Borsuk was careful to emphasize that in cases where the number of variables is close to the sample size, the single best-fitting model is based on the given observations (i.e., it is arbitrary) and should not be interpreted as a generally applicable model or mechanism. Relative confidence in this single best-fitting model out of numerous possible models can be evaluated using numerical methods. While incorporating prior knowledge can potentially improve model accuracy and structure learning performance under some conditions (e.g., noisy or sparse data), incorrect prior knowledge can potentially decrease accuracy and performance. Borsuk tested the sensitivity of his approach by adding intentionally incorrect prior knowledge to a network structure he knew to be correct. The bladder cancer structure in this case study allowed for up to half of the priors to be incorrect while still converging on the correct structure.

MACHINE LEARNING IN DOSE-RESPONSE ASSESSMENT: TRANSLATING SCIENCE TO DECISIONS — JACKIE MACDONALD GIBSON

All major proposed environmental regulations in the United States are required to undergo cost-benefit analysis. In such analyses, regulators must quantify the change in measured outcome (e.g., death or disease) and assign a dollar amount to each case in order to establish the cost-benefit of a proposed intervention. Dose-response functions form the critical foundation of the cost-benefit quantification under proposed scenarios. In her case study, Jackie MacDonald Gibson, Ph.D., Associate Professor at UNC-Chapel Hill and RTI University Scholar, proposed integrating ML into dose-response functions to better estimate critical components of the environmental regulatory process, thereby enabling the creation of better environmental policies. She identified the limitations in current dose-response methods employed by the risk assessment community, including disparities in the evaluation and regulation of cancer and non-cancer outcomes. She asserted that such methods are “20th Century” approaches, and that newer technologies and techniques are readily available but underutilized. To illustrate the potential benefits of employing ML in EHS, MacDonald Gibson presented a case study of arsenic exposure and diabetes outcomes in a population in Mexico. She evaluated the benefits of a hypothetical policy using traditional statistical methods (reference dose, slope-factor, and logistic regression) and a machine-learned BN. She found that not only did the BN have the greatest predictive accuracy, but the network also showed variable influences that may not have been identified otherwise. Her results demonstrate that BNs can better quantify the health benefits of interventions and help justify proposed policies to decision makers. Using BayesiaLab software, MacDonald Gibson also created an interactive website based on her findings that allows users to generate dynamic diabetes risk outcomes based on user inputs for a number of variables.

BAYESIAN INFERENCE FOR SUBSTANCE AND CHEMICAL TOXICITY (BISCT) — LYLE BURGOON

Lyle Burgoon, Ph.D., is the leader of the U.S. Army Engineer Research and Development Center’s Bioinformatics and Computational Toxicology Group. In his case study, Burgoon explained that urban warfare and first response scenarios can create exposures to toxicants and associated health risks. Although the military is initially concerned about acute toxicity in such scenarios, chronic toxicity can also negatively impact military readiness. Management of chronic toxicity outcomes poses policy and financial challenges for the U.S. Departments of Defense and Veterans Affairs. Burgoon’s research group uses numerous approaches and techniques within AI/ML to assess chronic toxicity, but his case study focused on his work with BNs.

His product, Bayesian Inference for Substance and Chemical Toxicity (BISCT), can integrate manifold bioassay data within a Bayesian approach. BISCT is a simple way for the end-user to generate probabilities of hormone production, enzyme activation, or molecular initiation for a given chemical; for example, these outputs can be used by a risk assessor to evaluate the steroidogenesis capacity of a chemical of interest. According to Burgoon, the software is extremely user-friendly and requires few steps by the end-user, which he has found encourages use and acceptance within the risk-management community.

Burgoon suggested there are opportunities in the EHS community to combine Adverse Outcome Pathways (AOPs) with BNs. AOP-BNs can suggest causal networks, which play an important role in regulatory and legal considerations. In a trial study comparing results from data fed into an AOP-BN versus more traditional methods of evaluating various compounds’ potential for steatosis, the AOP-BN yielded the same results as the traditional method with extremely high certainty. Burgoon asserted that this type of study could serve as a proof-of-concept model to achieve buy-in from decision makers and stakeholders. The incorporation of BNs into AOPs also means that risk-assessors would not have to test every node in a BN to make conclusions about outcome probability, only those in the so-called Markov Blanket—the set of nodes in a BN that renders the node of interest conditionally independent from the rest of the network. By limiting the need for certain evaluations—as there is no extra statistical value gained in evaluating intermediate nodes—the Markov Blanket makes assessing probabilities more efficient. These methods could also reduce reliance on animal testing and reduce testing costs if sufficiently demonstrated by mathematics. Despite the benefits of the AOP-BN and Markov Blanket condition, Burgoon has found resistance to their implementation among the risk-assessment community, whose members are more likely to believe that every node must be tested. He noted that the Markov Blanket is not an intuitive concept, particularly because it may initially seem contrary to traditional training in this field and because it also relies on mathematical demonstration. This type of resistance is a hurdle to more widespread use and suggests an area for more professional training, particularly for risk managers.

RARE DISEASES AND AI ANALYSIS WITH POTENTIAL USE OF THE ENVIRONMENTAL GENOME — MICHAEL KOWOLENKO AND MICHAEL OVERCASH

While other case studies and plenaries considered how to integrate data, Michael Kowolenko, Ph.D., CEO of NoviSystems and Michael Overcash, Ph.D., Executive Director of the Environmental Genome Project, discussed ways to obtain a complex dataset.

Overcash introduced attendees to his team's efforts to map what they call the "Environmental Genome," which comprises the approximately 100,000 anthropogenic chemicals that make up the global economy. The team, which aims to describe, model, and map this collection of chemicals just as the human genome was mapped, foresees both health and economic benefits from better understanding these chemicals. Overcash noted that work is now underway to utilize AI to investigate chronic diseases, in particular rare diseases, using the comprehensive database of chemicals within the Environmental Genome. He proposed the development of an Environmental Genome Explorer, which would consist of a user-friendly online interface to explore the lifecycle of chemicals and their interaction with biologic systems, which he argued would have to be built from a "smart system" that can filter data with very little human input.

Kowolenko detailed how the development of such a smart system could occur. He emphasized the promise of fusing Natural Language Processing (NLP) with structured data. NLP is an approach to overcoming the problems that 1) algorithms do not easily accommodate the type of ambiguity and nuance found in human text and language, and 2) structuring data from language/text into matrices in a way that still has meaning is a challenge. Kowolenko remarked that working with the right subject matter experts is critical in designing algorithms with appropriate feature-extraction and feature-matching. A text extraction model primed by subject matter experts would initially involve developing a dictionary of expert terms, as well as an expert-derived rules-based system (e.g., if the program sees "X", then extract). A series of filters would be emplaced until a corpus of words and relationships is achieved. This corpus can then be returned to a subject matter expert who can decide if the components are relevant, beginning an iterative process of query and term refinement between computers and humans. The program could search relevant databases daily to harvest data defined by the algorithm, thereby continuously updating databases, as well as fortifying the rules (with feedback from a subject matter expert). The teaming of humans and machines in this sort of self-populating system ensures the data in the databases have meaning to those in that domain. Kowolenko and Overcash emphasized that this process is agnostic and can work in any domain, and proposed it as an effective way to blend a rules-based system with ML—the combination of the two takes advantage of both systems' strengths to achieve specificity and context. Specificity and context, they assert, enrich the data, thereby creating value to those using the systems.

Breakout Sessions

Breakout sessions were held to encourage small group discussions regarding the use of three AI/ML areas in EHS identified by the Summit Organizing Committee:

1. Applications and Data,
2. Aspects of Decision Making, and
3. Education and Training.

Summit attendees self-selected their preferred topic of interest and were approximately evenly split between the breakout sessions. All groups had individuals from public, private, and education sectors and broadly included professional experience in EHS research or academia, EHS policy, computer science, data science, and medical practice. The two-hour sessions were facilitated by three Summit organizers with experience in the field of EHS and AI/ML. A list of discussion questions was generated by the Summit Organizing Committee based on identified needs, obstacles, knowledge gaps, and priorities in this emerging field. Conversations were guided by the topics and question sets, but free-form discussion and brainstorming were highly encouraged and readily occurred. Due to the interconnected nature of these topics, group discussions converged around a common set of themes; therefore, the individual groups' discussions are briefly detailed below and themes are fully explored in Discussion Points.

APPLICATIONS AND DATA

The Applications and Data breakout discussion was facilitated by Michelle Angrish, Ph.D., a toxicologist at the EPA. This group was posed the following questions to facilitate discussion:

- What type(s) of questions does your organization or project use AI to solve?
- What AI components are used to solve those problems?
- How are those AI components made consistent and interoperable?

Discussions began by identifying that less than half of group participants knew of any AI being used in their workplace and most participants had not used AI themselves. Conversations generally focused on obstacles to implementing AI/ML in their workplace and more broadly, such as data-sharing and accessibility, data privacy concerns, connecting domain knowledge experts with AI experts, lack of training opportunities, and confidence in AI-generated outputs. These topics will be further discussed in Discussion Points.

ASPECTS OF DECISION MAKING

The Aspects of Decision Making breakout discussion was facilitated by Michele Taylor, Ph.D., a neurotoxicologist at the EPA. This group was posed the following questions to facilitate discussion:

- Who (if anyone) is thinking about how AI can be used to improve decision-making?
- What are the challenges faced by organizations?
- How might AI enhance chemical risk assessment?
- Why are there institutional barriers to wider use by decision-makers and risk analysts?

The Aspects of Decision Making group tended to focus on interpersonal, cultural, or institutional issues that may impede or facilitate incorporating AI into decision making processes. Discussions began by identifying the tendency of decision makers to continue with well-established techniques and the difficulty in challenging customs and norms. This group suggested areas that AI may enhance chemical risk assessment and decision making. These topics will be further discussed in Discussion Points.

EDUCATION AND TRAINING

The Education and Training breakout discussion was facilitated by Jackie MacDonald Gibson, Ph.D., Associate Professor at UNC-Chapel Hill and RTI University Scholar. This group was posed the following questions to facilitate discussion:

- Who needs to be trained and connected to use artificial intelligence in environmental health research and decision-making?
- What curricula are needed to train college students and risk assessors with next generation AI?
- What institutional changes are needed to promote this training?

This group recognized the distinction between constructing models and using models and proposed that education should reflect this distinction. The group also discussed that development of training in this field should consider the diversity of AI/ML usage between students, faculty, practitioners, and consumers. Furthermore, there is a need to educate the public to some extent for AI/ML in EHS to gain stakeholder acceptance; public education may mean marketing the products of AI/ML in EHS or establishing professional standards to achieve legitimacy and establish best practices. Many of the topics discussed in the Education and Training group touched on cultural issues and collaboration. These topics will be further discussed in Discussion Points.

CLOSING REMARKS

Following the breakout sessions, the group reconvened to review the discussions and any findings. It was determined that there was significant overlap in the discussion of the themes and issues identified across the breakout sessions. These intersecting topics also mapped to many of the subjects presented in the plenaries and case studies.

Discussion Points

The plenaries, case studies, and breakout sessions revealed many shared themes. The subthemes proved to be expansive, covering both tangible issues in the field and more philosophical matters broadly applicable to research and emerging technologies. The discussions, insights, questions, and recommendations generated from the various portions of the Summit have been divided into five categories, which are summarized in the sections that follow.

1 Education and Training

2 Data, Research, and Practice

3 Culture and Values

4 Early Opportunities for AI/ML in EHS

5 Identified Action Items

EDUCATION AND TRAINING

Training in both EHS and AI/ML will be critical in connecting the two fields. Training will need to be multi-faceted and must account for usage differences in students, faculty, practitioners, industry, and the public. This will require different approaches to educational curricula, adjustments in professional organizations and practice, and a focus on perceptions and accessibility of AI/ML.

▶ General, introductory seminars could be key to introducing AI/ML into more classrooms. These seminars would serve as exploratory exercises to promote thinking on the subject and could potentially lead to the integration of AI/ML concepts into other fields as well. These types of cross-cutting seminars would also combine many types of students in the same setting, encouraging the interdisciplinary foundation necessary for the benefits of AI/ML to enter more fields.

▶ Beyond simply generating interest in AI/ML, students will need specific curricula changes to support the integration of AI/ML into EHS. For example, scientific communication class offerings need to be expanded, particularly regarding communicating decision making under uncertainty. Efficient scientific communicators will be essential for the acceptance of these technologies by both decision-makers and the public. Scientific communicators who can bridge the gap between EHS professionals and computer/data science professionals will be especially needed.

▶ Deficiencies in modeling and statistics curricula will need to be addressed. Although students do not need to be computer scientists or programmers to engage in this field, it is still critical to understand underlying statistical theories. Additionally, students should be competent in at least one programming language—although preferred languages fall into and out of favor rapidly, understanding just one language makes it easier to understand and acquire additional languages and other programming processes. Tool-based skills are helpful in the short-term, but long-term focus on developing creativity and critical thinking in students should be emphasized. Faculty need to develop students who are ready to adapt and evolve in shifting professional landscapes.

▶ The rapid rate of change within AI/ML means curricula will also need to be adaptable and will require significant cooperation between expert faculty. Some of these curricula changes and approaches in disseminating relevant knowledge may benefit from faculty having more interdisciplinary joint appointments.

▶ It may also be beneficial to reorganize course catalogs. Rather than listing course offerings by department, listing courses by subject area may encourage more interdisciplinary education and broaden students' horizons to previously unknown opportunities.

▶ Education and training will need to extend beyond the classroom and into places of work. Professionals need to be ready and willing to constantly evolve in rapidly-changing fields of practice. Just as students should be specifically instructed to develop creativity and critical thinking, so too should professionals. Places of work should encourage this in their employees and should additionally host training sessions to introduce AI/ML concepts to their staffs. The "Train the Trainer" model is one such opportunity for expanding AI/ML in practice.

▶ Widespread usage and acceptance of AI/ML may be considered a goal by many, but it should be approached with caution. As detailed later in the Culture and Values section, trust in AI/ML methods and outputs will be critical to gain acceptance and usage. Branding and marketing will need to occur not just for the products that come out of this field but for the field itself. Professionals already in this emerging field should aim to better market products and outcomes of their work and to ensure AI/ML is more broadly understood by the public and by decision makers. This evokes the importance of instruction in scientific communication needed at the university level, which is also needed in the workplace.

▶ Many participants were concerned that AI/ML in EHS would be labeled a "pseudo-science" if implementation is not done carefully. This may necessitate criteria, standards, and professional best-practices. A certification in AI/ML in EHS may need to be developed if an official community of practice is established and the standards to achieve certification will need to be codified.

▶ Counter-intuitive concepts (such as not gaining extra value in evaluating all nodes in an AOP-BN) or mathematics-heavy concepts are an obstacle to more widespread use and acceptance of AI/ML. In particular, risk managers and regulators need training to better understand such concepts.

▶ Perceptions among students and professionals that they cannot be involved with AI/ML if they are not computer scientists or cannot code could limit important progress, as expert EHS domain knowledge is a critical component for AI/ML in EHS. Domain knowledge will remain fundamental to both constructing models/structures and interpreting machine learned models. The teaming of human and machine will be more influential than either applied separately, and this necessarily requires EHS experts that may not be trained in AI/ML.

▶ Fortunately, AI/ML technology is becoming more accessible. Software in many areas of AI/ML is becoming extremely user-friendly and trending toward more "point and click," rather than coding-intensive processes. This was evidenced by the BayesiaLab workshop, in which the user interface was extremely intuitive and user-friendly. MacDonald Gibson further highlighted movement in this direction by demonstrating her publicly available and user-friendly website, where her machine-learned research has been translated into intuitive sliders and buttons to dynamically evaluate risk in real time. Burgoon's BISCT software also emphasizes these aspects of ease and accessibility.

▶ Accessibility also necessitates connecting various experts in EHS and AI/ML. Such connections will likely arise from developing a community of practice and from more interdisciplinary contact between faculty in each field.

Quantifiable obstacles related to AI/ML acceptance and implementation largely involved issues related to data volume, data access, and data quality.

- The volume of data in most fields has increased exponentially in recent decades. This has left most professionals struggling to synthesize immense volumes of information from their field's literature. Also, the substantial quantity of chemicals in existence—the majority of which are unregulated—poses issues for evaluation using traditional methods. These two aspects present a significant challenge, but this "deluge of data" has also generated opportunities for humans and machines to work cooperatively using ML. The revolution of Big Data in marketing and business analytics paved the way for AI/ML use in those fields. Although the volume of data in EHS is less than in those fields and the type of data is different, it is sufficient to start using it in AI/ML techniques to potentially revolutionize the field of EHS.

- Most databases do not currently communicate well with one another and many databases are largely incompatible. There were questions concerning whether there are ways to "retrofit" and standardize data to be more efficiently integrated in the future.

- A data repository or a master list of publicly available toxicological/EHS data sources would be a good first step, and eventually linking such databases would be extremely valuable. Some approached this idea with caution and reaffirmed Adams' premise that context and specificity are lost when data is uploaded to a repository, and that significant information is left behind in the mind of the researcher or in documents due to inbuilt limitations of AI/ML.

- Questions of data quality and uniformity present looming issues. Integrating disparate data sources requires evaluation of research methodology and protocols in some way. Improved reporting practices will allow groups to better understand input data and to better validate methods used in AI/ML, both of which are important for building trust in the community and the practice as a whole.

- AI/ML may provide uses for previously difficult data (e.g., small sample sizes with large numbers of variables, or no prior knowledge of underlying mechanisms). New statistical techniques from AI/ML can allow us to extract new insight and information from older datasets. As such, perceptions and long-held paradigms of data quality may have to expand and shift.

- Strong support exists for adoption of FAIR Guiding Principles[2] for scientific data, affirming the idea that scholarly data should adhere to the principles of Findability, Accessibility, Interoperability, and Reusability. Participants largely agreed that formally codifying these principles would go a long way in addressing many data issues discussed at the Summit.

CULTURE AND VALUES

To make significant, groundbreaking progress in the use of AI/ML in EHS, culture and values will be among the most important but difficult aspects to address. Such topics were touched on continuously throughout the Summit but discussion yielded no easy answers. Many of the perceived obstacles for more widespread use of AI/ML in EHS relate to cultural issues in research more broadly, including issues of cooperation, bias, privacy, and trust. Developing a formal community of practice will be critical to tackling these issues within this emerging field and ensuring that the field continues to progress.

Cooperation in research and practice, particularly with respect to data-sharing, is a persistent issue. Organizational barriers exist throughout the field of EHS (as in other fields) and are pervasive in government agencies, where many agencies may work on similar problems but use different approaches. Even within an agency, conflict and competition occur within different offices and research groups. The private sector can be reluctant to share research data and results and tend not to do so unless mandated by laws and regulations.

Breakout session groups characterized cooperation issues as largely stemming from human nature and as a broader cultural problem not restricted to just EHS or AI/ML; however, some cooperation and data-sharing issues are more easily addressed. Many participants identified protected health data and patient privacy concerns as a major obstacle to cooperation and data sharing, despite the importance of privacy. HIPPA modifications or advances in data encryption could make headway in this area. Sharing information from vast datasets in conjunction with advances in AI/ML could result in some of the most substantial medical breakthroughs in more than a century. Requiring data accessibility as a condition of government funding could encourage cooperation in EHS. Adam's plenary also addressed the idea of a data "bartering" system to facilitate data sharing.

Interdisciplinary cooperation will be just as critical as intradisciplinary cooperation for advancing the role of AI/ML in EHS. As discussed in the Education and Training section, most EHS professionals currently are not well-versed in the potential for AI/ML, methodologies, or best practices. Collaborating with professionals from computer science, data science, biostatistics, and engineering will broaden the scope of what is possible in EHS. Bridging the gap between those who have AI/ML expertise and those who have EHS domain knowledge will need to be a deliberate effort. Building a community of practice focused on the development of this field will be critical to breaking down domain silos and forging collaboration.

Participants working for government agencies noted that any software developed would have to be open-source and companies would need to work as contractors to federal agencies, both of which may pose organizational challenges or may even disincentivize AI/ML experts from working on EHS issues.

Bias and systematic error in research can influence results, interventions, and policies and should be evaluated seriously—particularly in emerging fields not well understood by the public. Summit participants observed two key areas of potential bias in AI/ML in EHS that should be addressed:

Deterrence of researchers from publishing inconclusive studies or meager results can create bias in the literature and shape the field more broadly; inconclusive or poor results still have a role in AI/ML.

Implicit and explicit biases of researchers may inherently propagate bias in construction of AI/ML methods and models. Diversity within research teams and thoroughly engaging stakeholders may be an important part of addressing bias in AI/ML and its applications in EHS.

The values and priorities of researchers, practitioners, and society will play an influential role in the way AI/ML progresses within the field of EHS. It was suggested that for AI/ML to gain widespread adoption, techniques must not only produce results more efficiently, but must also yield results that are objectively better than traditional methods.

Models and results still need to be interpreted in the mind of the decision maker, so the community may consider prioritizing models that are as transparent as possible. BNs are likely a good first step in incorporating AI/ML into decision making, as their graphical representation makes them easily interpretable. Also, there are methods to assess confidence in AI/ML results and to evaluate thresholds in data, which may encourage adoption by decision makers

Within the systematic review community, reviewer fatigue is a real and significant phenomenon; involving AI/ML in systematic reviews should help to highlight the most important or relevant documents first, helping to ensure that reviewer fatigue does not contribute to critical pieces of information being overlooked.

AI/ML can potentially provide cheaper results, more easily scalable studies, and studies that do not require animal testing. The importance of such values was also reaffirmed by Summit participants.

As discussed in the section on Data, Research, and Practice, issues surrounding privacy will need to be reconciled at a broader scale than just the EHS community. Privacy laws, such as HIPPA, may need reformation in the future to accommodate trends in technology and privacy. The advent of consumer genetic testing can present challenges to medical privacy but may also benefit health research, particularly when combined with AI/ML.

EARLY OPPORTUNITIES FOR AI/ML IN EHS

Identifying early opportunities can establish initial wins, create buy-in, and build both momentum and interest. Early opportunities for the application of AI/ML techniques in EHS were identified through the plenaries, case studies, and breakout sessions. Major areas of early potential for this field included but were not limited to:

- Systematic review support
- Encoding domain/expert knowledge
- Databases that better capture data complexity and context (e.g., graphical databases, semantic databases)
- Combining Adverse Outcome Pathways (AOPs) with BNs
- Better approaches for evaluating non-linear relationships
- Better approaches for evaluating non-cancer outcomes in risk assessment
- Potential for chemical mixture risk assessments
- Alternatives to animal studies
- Assessing micro-differences within populations

IDENTIFIED ACTION ITEMS AND RECENT DEVELOPMENTS

The early opportunities for AI/ML in EHS will usher in future progress in this field, but attempting early opportunities will only be possible through extremely targeted efforts at this nascent stage. Summit attendees and presenters identified many areas of either necessary or potential progress to further advance the field; these action items are summarized in Table 1 and are roughly grouped into thematic categories.

Given that portions of the 2018 Summit were structured differently than previous years (i.e. guided discussions rather than working groups and subsequent recommendations), the findings presented in this document are general insights rather than targeted recommendations. Should a community of practice develop in this field, similar themes and topics might be more formally considered and detailed recommendations developed.

Following the October 2018 Summit, The Foundations for Evidence-Based Policymaking Act (H.R. 4174) was signed into federal law on January 14, 2019. H.R. 4174 incorporates text from the Open, Public, Electronic, and Necessary [OPEN] Government Data Act (H.R. 1770), which is potentially consequential for the field of AI/ML in EHS. Although the bill was not discussed at the Summit, it affirms many of the topics discussed during the Summit and presents numerous opportunities for a community of practice to consider. Briefly, three important outcomes from this new law include:

- Formalized data inventories and repositories: requires federal agencies to develop comprehensive data inventories that "account for all data assets created by, collected by, under the control or direction of, or maintained by the agency."[3] The law additionally requires agency heads to include extensive "metadata on each data asset of the agency" in their comprehensive data inventories[3].

- Required machine-readability: requires open government data assets to be "published as machine-readable."[3]

- Improved stakeholder engagement and collaboration: requires agency "collaboration with non-Government entities (including businesses), researchers, and the public for the purpose of understanding how data-users value and use government data."[3] The law additionally recommends government agencies host "challenges, competitions, events, or other initiatives designed to create additional value from public data assets of the agency."[3]

The potential implications resulting from this law include publicly-available data that is easier to identify and use, data that is more compatible with AI/ML techniques, and renewed partnerships and dialogue between United States government agencies and their data-users. The law will additionally provide mechanisms for data-users to "request specific data assets to be prioritized for disclosure and to provide suggestions for the development of agency criteria with respect to prioritizing data assets for disclosure,"[3] which is a potentially urgent action item for a community of practice to broach with relevant agencies.

Table 1: Identified Action Items for Advancing the Field and Practice of Artificial Intelligence in Environmental Health Science and Decision Making	
Data, Research, and Practice	Leaders in EHS and AI/ML should immediately review H.R. 4174 and develop cohesive action plans to advocate for certain data prioritizations, leverage opportunities created by this legislation, and understand implications for their fields of practice
	Compile lists of databases and repositories that could provide relevant datasets and potentially link these at a later stage; review and integrate comprehensive data inventories associated with H.R. 4174 as they become available
	Create "seed stock" of annotated document training sets for supervised ML
	Identify best practices for making literature more machine-readable and determine ways to create buy-in from the broader research community to enact these practices; more methodical collaboration between academia, industry, and government arising from H.R. 4174 could create standardized best practices and broader buy-in
	Agencies or research groups with more flexibility can lead the way with integrating AI/ML into low-stake areas and bring proof-of-concept studies to regulators to demonstrate superiority over existing approaches—this will potentially encourage more research/investiture in this area
	Identify what roles and skill sets are needed for a successful cross-domain team in this area; could form the foundation of best practices at a later stage
	Develop a larger pool of AI-EH experts for peer-reviewed publications
Education and Training	Employ the Train-the-Trainer model (e.g., beta-testers learn methods and train others down-stream through demonstrations and workshops) in EHS workplaces that could benefit from AI/ML
	Pilot a general AI/ML seminar for university students
Community and Field Development	Maintain momentum by creating a local community of practice, and potentially foster similar communities across the country; many participants have embraced the "meet-up" culture, where informal, frequent meetings take place with speakers, discussions, and socializing
	Compile lists of cross-domain specialists to seed a community of practice
	Special issue journal editions related to AI/ML in EHS
	Host another workshop and adopt a modular approach: break the problems down into discrete, solvable parts to identify low-hanging fruit in this field. Hosting a workshop that yields a publicly-distributed product at the end would increase awareness and buy-in.
	Utilize the friendly competitiveness within data science, computer science, and biostatistics communities. If an AI/ML in EH community of practice partnered with large conferences in those fields of study (e.g., presented a case study, provided the data, and hosted a competition) this could produce both results and further interest in the field. Example conferences could be the Conference on Neural Information Processing Systems, the Association for the Advancement of Artificial Intelligence, the Conference on Empirical Methods in Natural Language Processing, or the Association for Computational Linguistics, or competitions hosted by federal agencies in response to H.R. 4174.

Conclusion

The 11th Annual Environmental Health Collaborative Summit held one of the first meetings with the specific objective of investigating the potential role of artificial intelligence (AI) within environmental health sciences (EHS). Although the 21st century has seen the integration of AI into many aspects of everyday life, this integration has largely stopped short of the field of EHS research and practice. The 2018 Summit aimed to spark a conversation such that the field of EHS can begin to realize the benefits of AI technologies and methodologies. Attendees of the 2018 Summit were challenged to think about the opportunities that currently exist, the sources of data they may need, and how to prepare current and future generations of researchers and professionals.

Through a series of plenaries, case studies, and breakout sessions, discussions converged around a common set of themes that broadly included but were not limited to: Culture and Values; Data, Research and Practice; and Education and Training. Attendees and presenters also identified specific action items that are needed to grow this interdisciplinary field and brainstormed numerous potential early opportunities for AI in EHS. Some opportunities are already in the early stages of implementation or used in other fields, while other opportunities are still years away. Many of the topics and action items developed at the Summit were affirmed by the passage of The Foundations for Evidence-Based Policymaking Act in January 2019, which supported the ideas of data inventories, machine-readability, and stakeholder engagement, among other principles. This legislation will likely create new action items and may influence the development of this field of practice.

The common themes identified during the Summit build off and inform each other. These relationships are summarized in Figure 2. At a structural level, cultural issues and predominant values will influence how professional practice is conducted, and professional practice will also shape what people value and how communities behave. The professional landscape will influence how students and professionals are educated and trained, while novel insights developed by students and employees will continue to advance the practice. All three themes converge to present novel opportunities to implement AI in EHS.

The potential benefits of integrating AI into EHS are numerous. By gaining a greater understanding of the complex interactions between humans and the environment, decision-makers can promulgate interventions, policies, and resources that better protect both health and the environment.

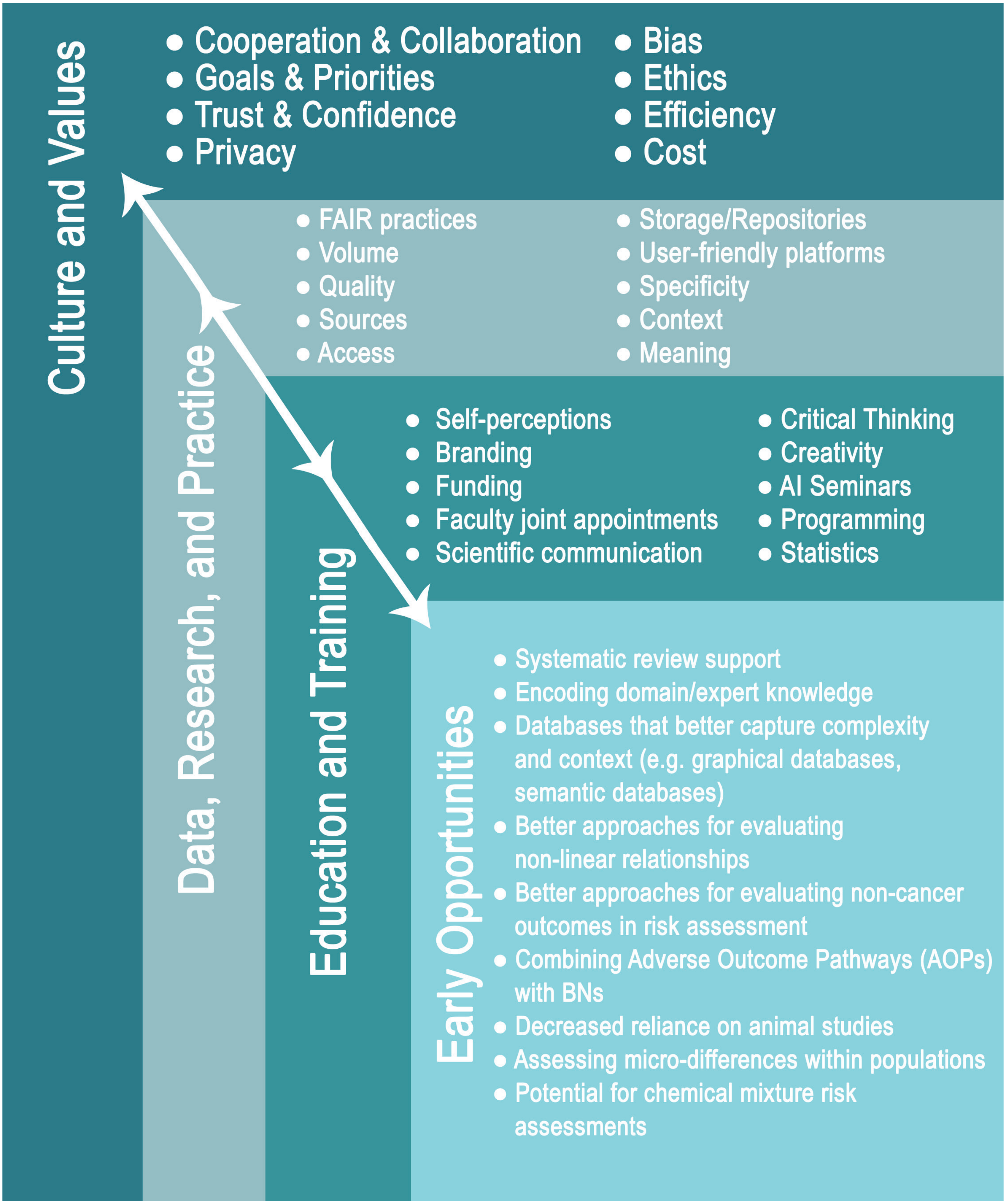

Figure 2: Themes and Topics Considered during the 2018 Summit. Overarching themes and topics were identified through plenaries, case studies, and breakout discussions. Themes build off and inform one another.

Artificial Intelligence Glossary

Artificial Intelligence (AI)

- AI involves teaching computers to make intelligent decisions.

"The goal of AI is to teach computers to do what humans currently do better."

Algorithm

- An algorithm is a set of instructions for solving a specific class of problem.
- The name derives from the Latinization of Muḥammad ibn Mūsā al-Khwārizmī (c. 780 – c. 850), translated as Algorithmi. Algorithimi was the astronomer and head of the library of the House of Wisdom in Baghdad and, during the ninth century AD, developed the basis for the field of algebra.

"An algorithm is a sequence of instructions telling a computer what to do."

Bayesian Belief Network (BN)

- A Bayesian belief network is a type of statistical model representing a set of variables and their relationships. Bayesian belief networks form the basis for one type of machine learning. These models are called "networks" because they can be represented as pictures in which nodes (circles) represent variables and edges (or arrows) connecting nodes represent relationships among the variables. *Bayesian network, Bayes network, belief network, and probabilistic directed acyclic graphical model* are alternative names for Bayesian belief networks.
- Bayesian belief networks were first defined by Judea Pearl, computer science professor at UCLA, in 1985.
- Pearl received the 2011 A.M. Turing Award in Computer Science (the equivalent of a Nobel Prize in computer science) in recognition of his development of the field of Bayesian networks and of algorithms for solving them.

Cross Validation

- Cross validation involves training a machine-learned model on part of a data set and then testing how well it performs in predicting observations in the rest of the data set (the test set).

Machine Learning (ML)

- A subfield of AI, machine learning uses statistical techniques to give computers the ability to "learn"—that is, to make progressively better (more accurate) decisions

"Machine learning takes many different forms and goes by many different names: pattern recognition, statistical modeling, data mining, knowledge discovery, predictive analytics, data science, adaptive systems, self-organizing systems, and more."

Testing Set

- A testing set is a set of observed data used to evaluate how well a machine-learned model is able to make predictions. The test set is not used to train the model but is set aside to check how well the trained model has "learned" patterns in the data.

Training Set

- A training set is a set of observed data used to train a computer model to make future predictions. Machine learning algorithms are designed to determine what kinds of models best match the observed data in the training set.

* All quotations are from The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World, by Pedro Domingos (New York: Basic Books, 2015).

Recommended Reading

Pearl, J. & Mackenzie, D. (2018). The Book of Why: The New Science of Cause and Effect. New York City, New York. Basic Books.

Conrady, S. & Jouffe, L. (2015). Bayesian Networks & BayesiaLab—A Practical Introduction for Researchers. Available for free download at http://www.bayesia.com/book

Domingos, P. (2015). The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake our World. New York City, New York. Basic Books.

References

1. Conrady, S. & Jouffe, L. Bayesian Networks & BayesiaLab--A Practical Introduction for Researchers. 73, (Bayesia USA, 2015).

2. Wilkinson, M. D. The FAIR Guiding Principles for scientific data management and stewardship. 1–9 (2016). doi:10.1038/sdata.2016.18

3. The 115th Congress of the United States. H.R.4174 - Foundations for Evidence-Based Policymaking Act of 2018. Pub. L. No. 115-435. (2018).

Summit Attendees

Samuel Adams, RTI International
Surafel Adere, Duke University
Michelle Angrish, US EPA
Martin Armes, The Collaborative
Scott Auerbach, NIEHS
Maureen Avakian, MDB Inc.
David Aylor, NC State
Maryam Azimi, Lenovo
Mamta Behl, NIEHS
Shannon Bell, ILS
Alexandre Borrel, NIEHS
Mark Borsuk, Duke
David Brown, The Collaborative
Lyle Burgoon, US Army
Neal Cariello, Integrated Laboratory Systems
Xialquing Chang,
Integrated Laboratory Systems
Rada Chirkova, NC State
Sarah Colley, UNC-Chapel Hill
Gwen Collman, NIEHS
Stefan Conrady, Bayesia USA
Jesse Cushman, NIEHS
Sivanesan Dakshanamurthy, Georgetown
Sally Darney, NIEHS
Deepa Dawadi Demosmita, NC State
Rob DeWoskin, etioLogic
Tom Dietterich, Oregon State
Cedric Dongmo, University of Yaounde
Chris Duncan, NIEHS
Steve Dutton, US EPA
Steve Edwards, RTI International
Neeraja Erraguntla,
American Chemistry Council
Jianing Fan, Duke University
Lydia Feinstein, Social & Scientific Systems
Kenda Freeman, MDB Inc.
Jim French, Live Learn Innovate Foundation
Stavros Garantziotis, NIEHS
Patrick Gray, Duke University
Lauren Gridley, RTI International
Hanbing Guan, Cisco
John Hardin, NC Board of Science,
Technology & Innovation
Linchen He, Duke University
Gina Hilton,
PETA International Science Consortium
Stephanie Holmgren, NIEHS
Kennedy Holt,
UNC-Chapel Hill/NC Public Health
Beibei Hu, Duke University
Sara Imhof, NC Biotech Center
Jaronda Ingram, Zoetis
Kristin Inman, NIEHS
Agnes Janoshazi, NIEHS
Peer Karmaus, NIEHS
Chandana Kasireddy, NIEHS
Channa Keshava, US EPA
Manal Khan, UNC-Chapel Hill
Nicole Kleinstreuer, NIEHS
Les Klimczak, NIEHS
Michael Kowolenko, NoviSystems
Hamid Krim, NC State
Dhirendra Kumar, NIEHS
Archana Lamichhane, NIEHS
Christopher Lavender, NIEHS
Janice Lee, US EPA
Tess Leuthner, Duke University
Jian-Liang Li, NIEHS
Xing Li, Duke University
Yuanyuan Li, NIEHS
Rui Liu, Social & Scientific Systems
Yun Liu, UNC-Chapel Hill
Ming Lu, Health Canada
Jackie MacDonald Gibson, UNC/RTI
Alexandra Maertens, Johns Hopkins
Elizabeth Mannshardt, US EPA
Dwi Sianto Mansjur, IBM
Kamel Mansouri, ILS
Courtney McCortsin, Duke University
Sena McCrory, Duke University
Vanessa Michelou, Novozymes
Erika Munshi, Duke University
Reshma Nargund, Duke University
Emmanuel Obeng-Gyasi, NC A&T
Michael Overcash,
Environmental Genome Initiative
Okan Pala, NC State
Shannon Parker, Duke University
Rajneesh Pathania, NIEHS
John Phillips, Georgia International
Terry Pierson, RTI International
Sunil Rajgopal Prasad, MResult

Asif Rashid, DS Technologies
Caroline Ridley, US EPA
Leon Rosentsvit, Technion IIT
Marianna Rosentsvit, NIEHS
Risa Sayre, US EPA
Kate Scholfield, US EPA
Frederic Seidler, Duke University
Joseph Shaw, Indiana University
Mina Shehee, NC DHHS
Shanshan Shi, Duke University
Susan Simmons, NC State
Linda Smail, Zayed University
Raquel Silva, US EPA
Solomon Tamkabari, Rivers State College
Michele Taylor, US EPA
Shane Thacker, US EPA
Kimberly Thigpen Tart, NIEHS
Jacob Traverse, Triangle Global Health Consortium
Natalia Vinas, US Army Research and Development Center
Jimmy Washington, NIEHS
James Weaver, US EPA
Leah Wehmas, US EPA
Paul Whaley, Lancaster University
Emily Woolard, US EPA
Ya Xue, Infinia ML
Chis Yoo
Hong Zu, NIEHS
Hal Zenick, US EPA
Yongjie Zhou, US FDA

SUMMIT ORGANIZING COMMITTEE

Michelle Angrish, U.S. Environmental Protection Agency
Martin Armes, Research Triangle Environmental Health Collaborative
Scott Auerbach, National Institute of Environmental Health Sciences
Maureen Avakian, MDB, Inc.
Mark Borsuk, Duke University
David Brown, Research Triangle Environmental Health Collaborative
Kenda Freeman, MDB, Inc.
Stephanie Holmgren, National Institute of Environmental Health Sciences
Jackie MacDonald Gibson, University of North Carolina at Chapel Hill, RTI International
Michael Overcash, Environmental Genome Initiative
Terrence Pierson, RTI International
Michele Taylor, U.S. Environmental Protection Agency
Kimberly Thigpen Tart, National Institute of Environmental Health Sciences
Hal Zenick (Retired), U.S. Environmental Protection Agency

RESEARCH TRIANGLE ENVIRONMENTAL HEALTH COLLABORATIVE AND SUMMIT SUPPORTERS

United States Environmental Protection Agency

NIH National Institute of Environmental Health Sciences

RTI International

Nicholas School of the Environment, Duke University — forging a sustainable future

NC STATE UNIVERSITY

UNC Gillings School of Global Public Health

MDB, Inc. — Doing Well by Doing Good

Social & Scientific Systems, Inc. — Improving Public Health Worldwide

visionpoint marketing

North Carolina Biotechnology Center

ILS — Advancing Science, Improving Lives

ECU — CAPTURE YOUR HORIZON

The Research Triangle Environmental Health Collaborative supports a united environmental health resource that connects organizations and institutions; links research and policy; and joins government, academia, industry, and public interest groups to mutually consider, discuss, and debate the future of environmental health on a regional, national, and international level.

www.EnvironmentalHealthCollaborative.org

Editorial Acknowledgments

In addition to the excellent work of summit participants and the team of notetakers, this document benefited from the writing and editorial assistance of Sarah Katherine Colley of UNC Chapel Hill, with special contributions from Jackie MacDonald Gibson of UNC Chapel Hill, as well as the expertise of Holly W. Ross, an independent graphic designer.