Editors: Jens Grossklags, jens.grossklags@in.tum.de | Jelena Mirkovic, mirkovic@isi.edu | Bill Newhouse, newhouse@nist.gov

Integrating Cybersecurity and Artificial Intelligence Research in Engineering and Computer Science Education

Fariborz Farahmand | Georgia Institute of Technology

We summarize integrating cybersecurity and artificial intelligence (AI) research in cybersecurity education and implementing a module in an existing noncybersecurity undergraduate engineering course. This initiative will drive the broader community to focus on the convergence of cybersecurity and AI education.

The Joint Task Force on Cybersecurity Education[1] defines cybersecurity as a "*computing-based* discipline involving technology, people, information, and processes to enable assured operations. It involves the creation, operation, analysis, and testing of secure computer systems. It is an *interdisciplinary* course of study, including aspects of law, policy, human factors, ethics, and risk management in the context of adversaries." Our society is undergoing a tremendous exploration and adoption of artificial intelligence (AI) to drive the growth of the economy, enhance economic and national security, and improve quality of life. It is thus vital to expose engineering and computer science students to AI and how it works along with how AI can enable assured operations and how it might be exploited by adversaries to undermine public trust and confidence.

Currently, university *research* offerings in AI and cybersecurity are limited to occasional offerings of special-topic courses. An exploration of AI within engineering and computer science education requires a comprehensive integration of cybersecurity research across the curriculum. This article describes the initial results of our research on fostering new, previously unexplored, collaborations among the fields of cybersecurity, AI, and education, supported by the Secure and Trustworthy Cyberspace Program (SaTC) of the National Science Foundation. Our approach teaches engineering and computer science students that cybersecurity and AI are by and for everyone, while also developing their security mindset. This is our first step to convene a broad community conversation focused on the convergence of cybersecurity and AI education.

Purpose and Scope

This article describes the initial results of a recent and ongoing research effort to create and implement curriculum modules that are focused on AI in cybersecurity and infused with real-world scenarios. We describe the first module, which was developed for use in a course that covers introductory probability and fundamental concepts in discrete mathematics and their efficient realization via algorithms, data structures, computer programs, and hardware. We describe our experience using the module in an advanced undergraduate course comprising university students without a cybersecurity background.

We have obtained an institutional review board approval to allow data gathering from volunteer students who completed extra-credit homework after the first module was completed. The data gathered were used to investigate if the students are able to apply computational applications of AI to develop realistic computational decision making after exploring causal (versus correlative) models in cybersecurity and privacy

Digital Object Identifier 10.1109/MSEC.2021.3103460
Date of current version: 28 October 2021

situations. Based on the students' performance observed in this effort, the integration of cybersecurity and AI research in education can be successful by using hands-on examples and explaining cybersecurity issues in a *computational* form—a language familiar to the engineering and computer science students.

Module Development and Implementation

The first module was integrated in a 40-min lecture titled Probabilistic and Causal Reasoning: A Cybersecurity Application. The materials included in this module were based on cybersecurity research, human-level AI research, and an adaptation of the work of Judea Pearl on tools of causal inference with reflections on machine learning and causal representation learning.[2,3] Examples of the computational tools of AI and causal interference that were used in the first module were 1) graphical causal models, in which causal stories behind the data can be rigorously conveyed, and 2) *do*-calculus, a calculus of causation, composed of simple logical operations for identifying causal effects. These were infused with real-world scenarios, for instance, assessing the probability of cyberattacks, probability of cause of the breach, and probability of necessity in cyberattribution and liability quantification, as well as investment decisions on cyberinsurance.

The first module was implemented in an existing undergraduate engineering course toward the end of the semester. This course had already covered introductory probability and some fundamental concepts in discrete mathematics and their efficient realization via algorithms, graphs, data structures, and computer programs—all background materials that students needed to understand causal reasoning.

The focus on causal reasoning stems from this value statement by Schölkopf et al.: "Despite its success, statistical learning provides a rather superficial description of reality that only holds when the experimental conditions are fixed. Instead, the field of causal learning seeks to model the effect of interventions and distribution changes with a combination of data-driven learning and assumptions not already included in the statistical description of a system."[4]

The module consisted of three parts. Part 1 included a quick review of the basic statistical and probabilistic concepts that students needed to understand the rest of the module. It also included examples on how standard Bayesian inference can be used in the assessment of suspects in a cybercrime investigation.

Part 2 of the module introduced intervening (versus conditioning) and causal reasoning, that is, reasoning for situations where one intervenes in the world, thereby interfering in the natural course of events. Key to this part was the fundamental distinction between regression coefficients and structural parameters and how students can use both to predict causal effects in linear models and work with Pearl's *do*-calculus, a general calculus for identifying causal effects. For example, one uses $do(X = x)$ to force the variable X to take the value x, having no other immediate effect. Part 2 explained that a causal model can be interpreted as a Bayesian network, which, in addition to answering probability queries, can also answer intervention queries, and that the answer to an intervention query $P(Y|do(z), X = x)$ is not generally the same as its corresponding probability query $P(Y|Z = z, X = x)$.

Part 3 of the module introduced the concept of counterfactuals; that is, what would have happened had we chosen differently at a point in the past. Discussions followed on how to compute counterfactuals, estimate their probabilities (such as the probability of necessity that captures the legal criterion of "but for"), and how to use counterfactuals to answer practical questions in cybersecurity (for example, cyberattribution).

All three parts of the first module included computational examples of the applications of causal inference in either tangible, real-life situations or real-world cybersecurity situations. Two examples from Part 2 of the module are summarized in the following sections.

Example 1: Understanding Causal Hierarchy and Conditioning Versus Intervening (Level 2 in the Hierarchy) With a "Fun Example"

This example summarizes Pearl's causal hierarchy (see Table 1) to explain the difference among the commonly used levels of association, intervention, and counterfactuals. To show the difference between the causal hierarchy levels, it was explained that an increase in ice cream sales is correlated with an increase in crime, not because ice cream causes crime, but because an increase in both ice cream sales and crime is more common in

Table 1. Pearl's causal hierarchy.

Level	Typical Activity	Typical Questions
1. Association: $P(y\|x)$	Seeing (observing a certain phenomenon unfold)	What is? How would seeing *X* change my belief in *Y*?
2. Intervention: $P(y\|do(x), z)$	Doing (acting in the world to bring about some state of affairs)	What if? What if I do *X*?
3. Counterfactuals: $P(y_x\|x', z')$	Imagining (thinking about alternative ways the world could be)	Why? Was it *X* that caused *Y*? What if I had acted differently?

hot weather—a confounding variable. Students called this a "fun example."

Randomized controlled experiments are considered the gold standard of statistics. But, in cases where randomized controlled experiments are not practical, engineers and computer scientists tend to perform observational studies, in which they purely record data rather than controlling it. The problem of such studies is that it is difficult to untangle the causal from the purely correlative relationships.

We introduced intervening and causal surgery in the example depicted in Figure 1(a), using endogenous variables X as ice cream sales, Y as crime rates, and Z as temperature; see Figure 1(b) and (c). In these graphs, the exogenous (versus endogenous) variables U_X, U_Y, and U_Z stand in for any unknown or random effects that may alter the relationship between the endogenous variables. That is, endogenous variables are those that we choose to include in the model, and the exogenous variables are unmodeled, latent variables. Following this example, where students learned the concept of intervention, we introduced *do*-calculus and its mathematical tools.

Example 2: Applying *do*-Calculus in a Cybersecurity Investment Decision

This example introduces students to the "machinery of causal calculus."[3] It shows how to apply the rules of *do*-calculus and *do*-operators to untangle causation in a cybersecurity investment decision to answer a "what if" question (level 2 of the causal hierarchy; see Table 1).

In this example, X, Y, Z, and W are arbitrary disjoint sets of nodes in a causal directed acyclic graph G, as depicted in Figure 2(a). Here, an arrow from one variable to another indicates that the first variable causes the second—that is, the value of the first variable is part of the function that determines the value of the second. Therefore, the second variable depends on the first for its value. $G_{\overline{X}}$ denotes the graph obtained by deleting from G all arrows pointing to nodes in X, and $G_{\underline{X}}$ denotes the graph obtained by deleting from G all arrows emerging from nodes in X. $G_{\overline{X}\underline{Z}}$ represents the deletion of both incoming and outgoing arrows. Figure 2(b) explains the three rules of *do*-calculus to help with eliminating the *do*-operators from the query expression, working with the observational data. For example, Rule 3 provides conditions for introducing (or deleting) an external intervention $do(Z = z)$ without affecting the probability of $Y = y$.

We used a cybersecurity example in which the following three board members in a high-tech company are discussing purchasing/not purchasing cyberinsurance:

1. Member 1 presents data indicating that companies that have bought cyberinsurance have actually experienced more data breaches.
2. Member 2 argues that Member 1 has ignored the company type and size in the assessment of the data. However, Member 2 does not present any data to support his argument.
3. Member 3 presents data indicating that purchasing cyberinsurance by the company has contributed to the *anomalous* behavior of the company staff in the past. For example, the staff of those companies that had purchased cyberinsurance felt that they were being protected by cyberinsurance, and, as such, they engaged more in risky behaviors.

The module concludes by demonstrating the application of *do*-calculus to work with the available observational data and assess the probability of experiencing a cyberbreach (C) if the company buys cyberinsurance (B), using the query $P(C|do(B))$. Figure 3 summarizes applying the *do*-calculus rules, step by step, until all of the *do*-operators are eliminated (shown in red).

Homework

Following completion of the module, volunteer students were tasked with extra-credit homework, which included six conceptual and computational questions related to all three

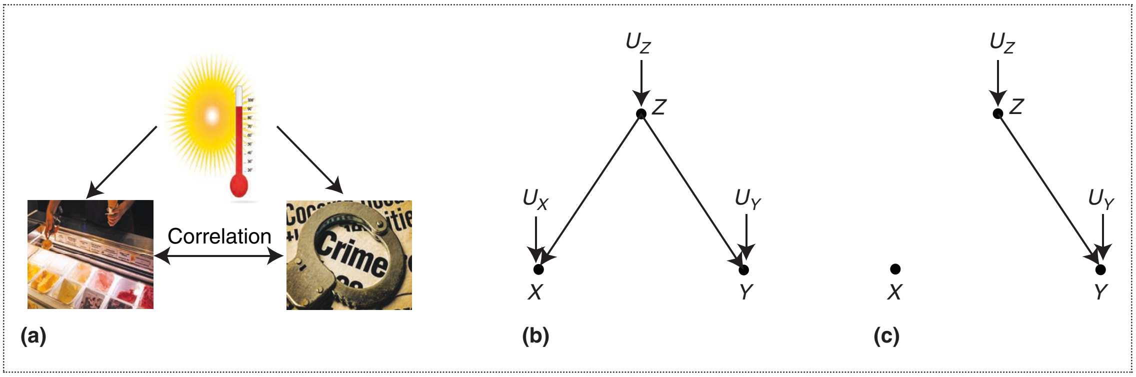
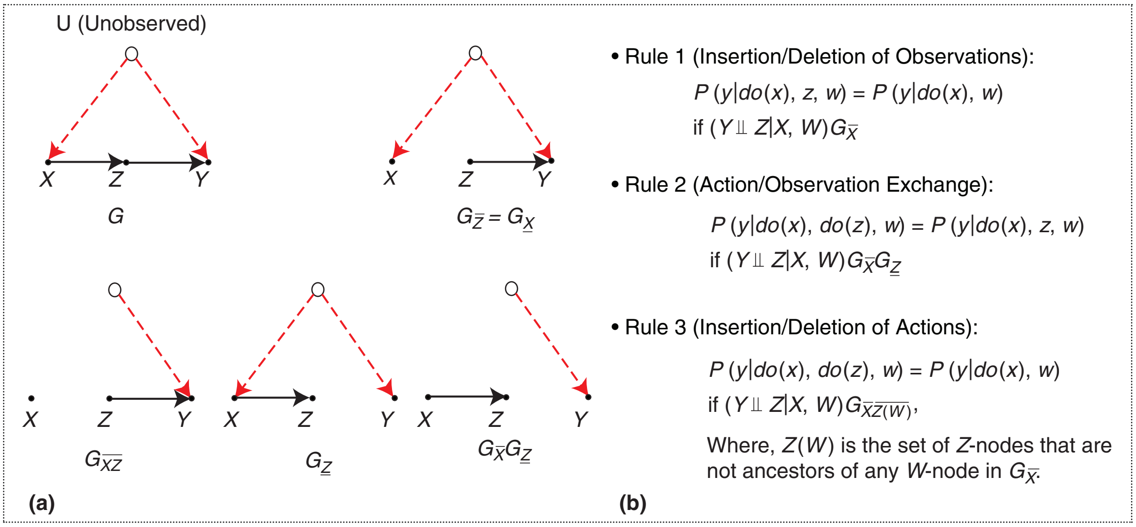


Figure 1. (a) Correlation between ice cream sales and crime rates. (b) Relationship among temperature (Z), ice cream sales (X), and crime rates (Y). (c) An intervention on the model in (b) that lowers ice cream sales.


U (Unobserved)
X
Z
Y
G
$G_{\overline{Z}} = G_{\underline{X}}$
$G_{\overline{XZ}}$
$G_{\underline{Z}}$
$G_{\overline{X}\underline{Z}}$
(a)
(b)

- Rule 1 (Insertion/Deletion of Observations):

  $P(y|do(x), z, w) = P(y|do(x), w)$

  if $(Y \perp\!\!\!\perp Z|X, W)G_{\overline{X}}$

- Rule 2 (Action/Observation Exchange):

  $P(y|do(x), do(z), w) = P(y|do(x), z, w)$

  if $(Y \perp\!\!\!\perp Z|X, W)G_{\overline{X}}G_{\underline{Z}}$

- Rule 3 (Insertion/Deletion of Actions):

  $P(y|do(x), do(z), w) = P(y|do(x), w)$

  if $(Y \perp\!\!\!\perp Z|X, W)G_{\overline{X}\overline{Z(W)}}$,

  Where, $Z(W)$ is the set of Z-nodes that are not ancestors of any W-node in $G_{\overline{X}}$.

Figure 2. (a) Subgraphs of *G* used in the derivation of causal effects. (b) The rules of *do*-calculus.

parts of the module. For instance, students were presented a recommendation by a chief security officer (CSO) and a counterargument from a chief financial officer, who noted that the CSO's recommendation is unnecessary and too costly. Students were asked to assess if the CSO's recommendation is necessary to protect the company from certain malicious attacks, using the observational and experimental data and mathematical tools to identify a probability of necessity, *PN*, as follows:

$$PN = \frac{P(y|x) - P(y|x')}{P(y|x)} + \frac{P(y|x') - P(y|do(x'))}{P(x, y)}.$$

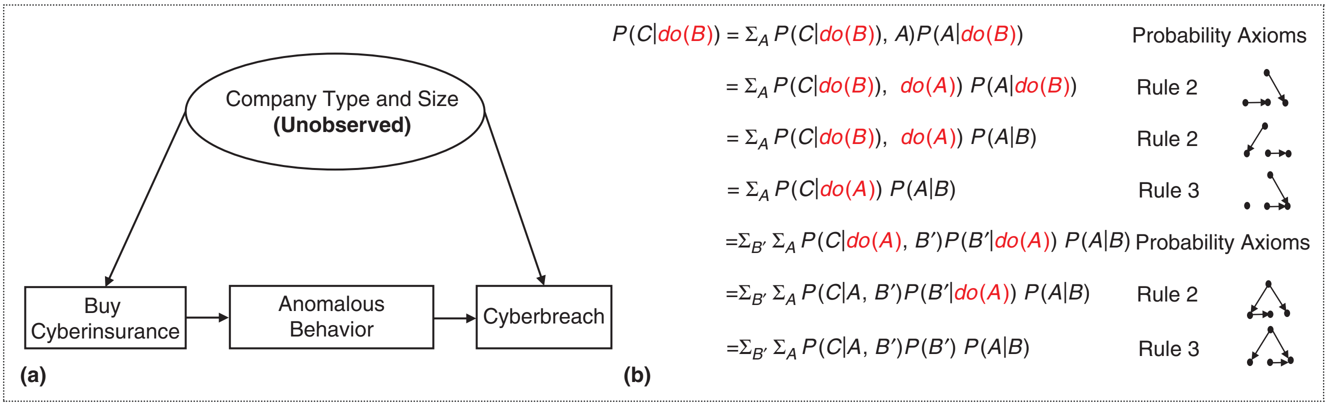
In this equation, the first term on the right-hand side is the familiar excess risk ratio that experts have been using as a surrogate for PN in court cases. The second term (the confounding factor) represents a correction needed to account for confounding bias; that is, $P(y|do(x')) \neq P(y|x')$. Here, events are assumed binary, with $X = x$ and $Y = y$ representing recommendation and outcome, respectively, and $X = x'$ and $Y = y'$ their negations.

Students were provided information on writing the *PN* formula, as explained in cyberattribution and liability quantification; see Farahmand 2020.[5]

Findings and Implications

A total of 51 students, out of the total of 73 students enrolled in the class, voluntarily completed the homework. The volunteer students included 43 males (out of the total 61 males) and eight females (out of the total 12 females). The module was viewed a total of 3,524 min by the students, according to Canvas Analytics.

In the learning assessment, the six levels of the cognitive domain in the canonical taxonomy of Bloom and Anderson[6] were applied, progressing from the lowest-order processes to the highest: 1-Remember, 2-Understand, 3-Apply, 4-Analyze,



$P(C|do(B)) = \Sigma_A P(C|do(B)), A)P(A|do(B))$ Probability Axioms

$= \Sigma_A P(C|do(B)), do(A)) P(A|do(B))$ Rule 2

$= \Sigma_A P(C|do(B)), do(A)) P(A|B)$ Rule 2

$= \Sigma_A P(C|do(A)) P(A|B)$ Rule 3

$=\Sigma_{B'} \Sigma_A P(C|do(A), B')P(B'|do(A)) P(A|B)$ Probability Axioms

$=\Sigma_{B'} \Sigma_A P(C|A, B')P(B'|do(A)) P(A|B)$ Rule 2

$=\Sigma_{B'} \Sigma_A P(C|A, B')P(B') P(A|B)$ Rule 3

Figure 3. (a) A graphical model representing the relationships among buying cyberinsurance, the anomalous behavior of company staff, a cyberbreach, and an unobserved confounder (company type and size). (b) Applying the rules of *do*-calculus until all *do*-operators are eliminated from the query expression.

5-Evaluate, and 6-Create. All 51 students who participated in this study reached level 4; that is, they were able to remember, understand, apply, and analyze the lecture materials in answering the homework questions. Fifty-three percent of the students who participated in this study reached levels 5 and 6. That is, in answering their homework questions, they were able to work with the *do*-operator to evaluate and justify a decision and put elements together in a creative new way.

Explaining crosscutting cybersecurity concepts in a *computational* form, that is, a language that is understandable to engineering and computer science students, can significantly help them to understand the knowledge area, regardless of the disciplinary lens.

Out of the total possible 100 points, the average and the highest scores for the male students were 85 and 100, and for the female students were 81 and 97, respectively. A Mann–Whitney *U* test was performed to assess if there were any gender differences in the students' performance. The *p*-value was found as 0.25, and the result was not significant at $p < 0.05$. That is, gender did not make a significant difference in the students' performance according to the homework scores.

Considering the novelty of the module and the homework questions, the performance of both male and female students was quite encouraging. This has two important implications for cybersecurity and AI researchers and educators.

First, understanding and addressing cybersecurity issues when they are taught through computational and tangible real-world examples is a realistic expectation from computer science and engineering students. This echoes De Millo et al.'s seminal argument that we can be successful by seeing mathematics as a "social, informal, intuitive, organic, human process, a community project," versus outsiders who see mathematics as a "cold, formal, logical, mechanical, process of sheer intellection."[7]

Second, explaining crosscutting cybersecurity concepts in a *computational* form, that is, a language that is understandable to engineering and computer science students, can significantly help them to understand the knowledge area, regardless of the disciplinary lens. This is a key to the successful implementation of the Cybersecurity Curricular Framework[1] in the engineering and computer science curriculum. Specifically, it helps students to develop

1. *adversarial thinking*: a thinking process that considers the potential actions of the opposing force working against the desired result
2. *systems thinking*: a thinking process that considers the interplay between social and technical constraints to enable assured operations.

"Cybersecurity must be understood as a multifaceted domain".[8] As such, all cybersecurity issues cannot and should not be left only to cybersecurity experts who studied computer science or electrical and computer engineering. As argued by De Millo et al., "mathematicians' errors are corrected, not by formal symbolic logic, but by other mathematicians."[7] Using the *do*-calculus that was used in the first module was just one example of the common languages that can be taught to our students to help them collaborate with the "other" experts. This enables our future workforce to collaborate in eliminating the flaws in the implementation, design, specification, or requirements of AI-based cybersecurity systems.

This article described a recent and ongoing effort toward advancing cybersecurity education through the integration of AI concepts. We have introduced causal analysis versus traditional correlation analysis.

The reality of teaching behavioral learning in cybersecurity and AI education is that computer scientists and engineers follow traditional economic approaches, and humans are considered to be rational agents who always choose the actions that maximize the expected utility.[9] This can safely be equated with the von Neumann–Morgenstern perspective on expected utility theory.[10] However, most cybersecurity and privacy incidents are indeed caused by humans, by either the user or the developer, and could have been prevented. In fact, brain-mapping tools have provided neurobiological evidence, based on the human brain's reaction to privacy risks, that cyberprivacy behaviors cannot be well described by the expected utility theory.[11]

Upcoming modules will be focused on addressing the preceding issues and will contribute to the goal of achieving human-level AI. For example, as an alternative to traditional expected utility theory, the upcoming modules will integrate Kahneman's System 1 and System 2 thinking in the computational learning of cybersecurity and privacy behaviors:[11]

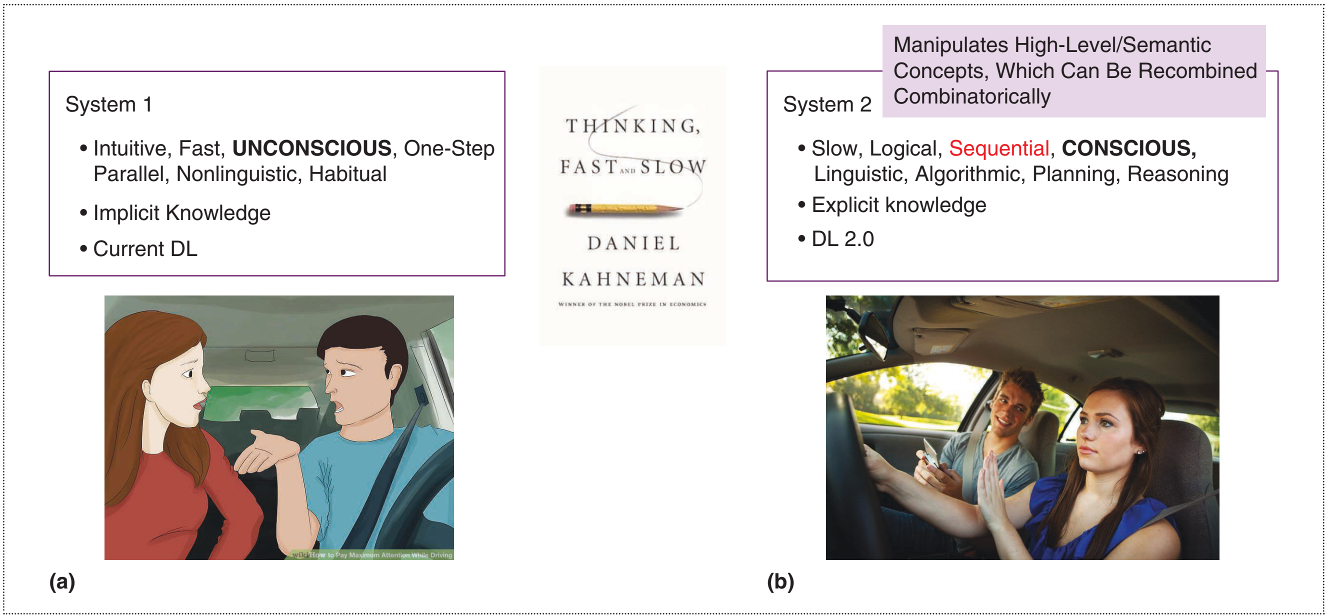

System 1
• Intuitive, Fast, UNCONSCIOUS, One-Step Parallel, Nonlinguistic, Habitual
• Implicit Knowledge
• Current DL
THINKING, FAST AND SLOW
DANIEL KAHNEMAN
WINNER OF THE NOBEL PRIZE IN ECONOMICS
Manipulates High-Level/Semantic Concepts, Which Can Be Recombined Combinatorically
System 2
• Slow, Logical, Sequential, CONSCIOUS, Linguistic, Algorithmic, Planning, Reasoning
• Explicit knowledge
• DL 2.0
How to Pay Maximum Attention While Driving
(a)
(b)

Figure 4. Conscious processing and (a) System 1 versus (b) System 2 thinking. (Source: Y. Bengio[12]; used with permission.)

- *System 1 (affective)*: It operates automatically and quickly, with little or no effort and no sense of voluntary control.
- *System 2 (cognitive)*: It allocates attention to the effortful mental activities that demand it, including computations. The operation of System 2 is often associated with the subjective experience and concentration.

Thanks to Bengio[12] for introducing Kahneman's System 1 and 2 thinking and conscious processing as missing parts of human-level AI and encouraging the AI community to move from current deep learning (DL) to DL 2.0 (see Figure 4). It is notable that expected utility theory does *not* recognize the difference between System 1 and 2 thinking and conscious processing. It disregards that humans minimize their cognitive cost and considers them as rational agents with stable, well-defined preferences who always choose the option with the maximum utility in a unitary cognitive process.

Some future modules under consideration will integrate formal methods and model checking. Considering the probabilistic nature of AI and machine learning, they may include probabilistic computation tree logic, which is appropriate for expressing a large class of properties in a rather elegant manner.[13,14] ■

Acknowledgments

The author would like to thank all the students who participated in this study. This research was supported by the National Science Foundation under award number 2041788. Any opinions, findings, conclusions, or recommendations presented in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

References

1. Joint Task Force on Cybersecurity Education, "Curriculum guidelines for post-secondary degree programs in cybersecurity," ACM, IEEE-CS, AIS SIGSEC, IFIP WG 11.8, New York, NY, 2017.
2. J. Pearl, *Causality: Models, Reasoning and Inference*. Cambridge, U.K.: Cambridge Univ. Press, 2009.
3. J. Pearl, "The seven tools of causal inference, with reflections on machine learning," *Commun. ACM*, vol. 62, no. 3, pp. 54–60, 2019. doi: 10.1145/3241036.
4. B. Schölkopf et al., "Towards causal representation learning," *Proc. IEEE*, vol. 109, no. 5, pp. 612–634, 2021. doi: 10.1109/JPROC.2021.3058954.
5. F. Farahmand, "Quantitative issues in cyber insurance lessons from behavioral economics, counterfactuals, and causal inference," *IEEE Security Privacy*, vol. 18, no. 2, pp. 8–15, 2020. doi: 10.1109/MSEC.2019.2930054.
6. L. Anderson et al., *A Taxonomy for Learning, Teaching, and Assessing, A Revision of Bloom's Taxonomy of Educational Objectives*. White Plains: Longman, 2001.
7. R. A. De Millo, R. J. Lipton, and A. J. Perlis, "Social processes and proofs of theorems and programs," *Commun. ACM*, vol. 22, no. 5, pp. 271–280, 1979. doi: 10.1145/359104.359106.
8. "Federal cybersecurity research and development strategic plan," National Science and Technology Council, Washington, D.C., 2019. [Online]. Available: https://www.nitrd.gov/pubs/Federal-Cybersecurity-RD-Strategic-Plan-2019.pdf
9. S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 4th ed. London: Pearson, 2020.

10. J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton Univ. Press, 1953.
11. F. Farahmand and F. Farahmand, “Privacy decision making: The brain approach,” *IEEE Comput.*, vol. 52, no. 4, pp. 50–58, 2019. doi: 10.1109/MC.2018.2885971.
12. Y. Bengio, “From system 1 deep learning to system 2 deep learning,” in *Proc. Conf. Neural Inf. Process. Syst.*, Vancouver, BC, 2019. [Online]. Available: http://www.iro.umontreal.ca/~bengioy/NeurIPS-11dec2019.pdf
13. E. M. Clarke, T. A. Henzinger, H. Veith, and R. Bloem, *Handbook of Model Checking*. New York: Springer-Verlag, 2018.
14. C. Baier, B. R. Haverkort, H. Hermanns, and J.-P. Katoen, “Performance evaluation and model checking join forces,” *Commun. ACM*, vol. 53, no. 9, pp. 76–85, 2010. doi: 10.1145/1810891.1810912.

Fariborz Farahmand is a research faculty member with the School of Electrical and Computer Engineering at the Georgia Institute of Technology (Georgia Tech), Atlanta, Georgia, 30332, USA. His research interests include cybersecurity and privacy, computational decision making, human-level artificial intelligence, causal inference, cyberphysical systems, human factors, and cyberinsurance. Farahmand received a Ph.D. in computer science from Georgia Tech. He is a Senior Member of IEEE. Contact him at fariborz@ece.gatech.edu.

Digital Object Identifier 10.1109/MSEC.2021.3117329