AI/ML and Cybersecurity: The Emperor has no Clothes

Walter Willinger
NIKSUN, Inc.
wwillinger@niksun.com

May 24, 2022

Ongoing Collaborative Effort …

- Arthur Selle Jacobs (UFRGS)
- Roman Beltiukov (UCSB)
- Ronaldo Alves Ferreira (UFMS)
- Arpit Gupta (UCSB)
- Lisandro Zambenedetti Granville (UFRGS)

A Cautionary Tale …

- "The emperor has no clothes" ["The emperor's new clothes"] by Hans Christian Anderson (1837)
- Something widely accepted as true or professed as being useful, due to a general unwillingness to criticize it or be seen as going against popular opinion ….
- "Something" = "AI/ML for cybersecurity"

AI/ML & Cybersecurity - Research

- Has been a "hot topic" for a while (Google Scholar search, May 2022)
  - Before 2000: ~1,000 publications
  - 2001 – 2010: ~6,000 publications
  - Since 2011: ~16,000 publications
- Large number of new workshops and conferences
- Large number of journals and books

AI/ML & Cybersecurity – Industry

- The AI/ML for cybersecurity market was estimated to be worth about $10.18B in 2021 (about $50B in 2028/29)
- Some 4,000 cybersecurity companies
- Very active merger and acquisition (M&A) market
- Plenty of startups
- Venture capitalists are shifting some of their focus from AI to concepts like Web3 and decentralized finance (blockchain) …

AI/ML & Cybersecurity – M&A

- Google acquires Mandiant (2022, $5.4B)
- Private equity group purchases McAfee for $14B (2022)
- SentinelOne plans to add Attivo Networks (2022, $600M)
- Private equity firm (Thoma Bravo) deal for Proofpoint (2021, $12.3B)

AI/ML & Cybersecurity - Startups

- CrowdStrike https://www.crowdstrike.com
  - "…enhanced endpoint machine learning that advances and augments CrowdStrike's behavioral-based machine learning prevention in the cloud for complete and effective protection for all endpoints."
- Darktrace https://www.darktrace.com/en/
  - "… recognized globally for its immune system approach to cyber security, Darktrace's Self-Learning AI is today capable of making decisions and taking proportionate, autonomous actions to thwart in-progress attacks."

- Other examples include Fortinet, Cynet, Check Point, Sophos, Cylance, Vectra and many others …

A Naïve Question …

- With so much effort from ...
  - Academia (research publications)
  - Industry (commercial products)
  - VCs/private equity (funding sources)

- … why is cybersecurity not a solved problem?
  - Have we made any progress?
  - If so, how much?
  - Will we ever be able to "solve" the problem?

A "Vision" or "Wishful Thinking" …?

- Today, a typical enterprise network …
  - Uses some 30-50 different security solutions
  - Excessive false alert rates
  - Understaffed IT department
  - One incident can doom the network

- Vision for tomorrow: Autonomous defensive systems
  - Networks that can "defend themselves"
  - Intuitive but "nebulous" concept
  - What counts as "success"?

Hope Springs Eternal …!

- Critical requirements
  - Detection & mitigation of (nefarious) network events in real time, in high-throughput scenarios, and largely without humans in-the-loop
  - High accuracy (detection) and user-defined control over collateral damage (mitigation)

- New capabilities
  - Fully programmable, protocol-independent data planes (with languages such as P4 for programming them)
  - Renewed allure of Artificial Intelligence (AI) and Machine Learning (ML) technique in support of cybersecurity

Towards Realizing this Vision …

- Sensing/monitoring (at line rate)
  - Market pull: Decide on the "right" data …
  - Technology push: Programmable devices
- Automatic (AI/ML-based) inference (in real time)
  - Market pull: Decide on the the "right" learning model …
  - Technology push: AI/ML-based inference in the data plane
- Actuating (in the data plane)
  - Market pull: Decide on the "right" action …
  - Technology push: Programmable devices

1st Challenge: The Data Problem

- AI/ML applications have excelled in areas that have an abundance of (labeled) data
  - Computer vision – ImageNet database
  - Autonomous vehicles – Argoverse, nuScenes

- The cybersecurity domain is not such an area!
  - Publicly available data are by and large of no use in this context
  - Proprietary data (e.g., large providers like Google, Microsoft, Facebook) are in general inaccessible to third parties

2nd Challenge: The AI/ML Problem

- Most modern learning models are "black-box"models
  - Provide no insight in why the model makes certain decisions (and not some other decisions)
  - Provided no understanding about the decision-making process that gives rise to its decisions
- Network operators don't like "black-box" models
  - Operators look for insights/understanding
  - Operators don't trust "black-box" models
  - Operators are reluctant to deploy products in their production network that they don't trust

3rd Challenge: The P4 Problem

- Programmable data plane technologies
  - Highly resource-constrained (e.g., memory)
  - Limited functionalities (e.g., only basic arithmetic operations)
  - Only primitive runtime programmability capabilities
- This problem may go away in the future …
  - Highly dynamic landscape
  - Constant innovations on hardware- and software side
  - New architectures for per-packet AI/ML (e.g., Taurus, 2022)

Rest of the Talk: The AI/ML Problem

An Illuminating Exercise (Step 1)

- Pick 1,000 research papers from the existing literature on AI/ML for cybersecurity
- Each paper has to
  - Consider a specific problem formulation
  - Develop one or more learning models
  - Report relevant findings (e.g., model accuracy)

- Everyone can do this step, but it's time-consuming …
- We checked out a few hundreds of papers …

An Illuminating Exercise (Step 2)

- Check how many of the 1,000 papers are reproducible
- To be reproducible, a paper has to
  - Contain a detailed description of the proposed learning model(s)
  - Provide access to and description of dataset(s) used
  - Make code base available
- Fact: Only about 10(!) of the 1,000 papers can be fully reproduced

An Illuminating Exercise (Step 2, cont.)

- AI/ML for cybersecurity has a "reproducibility" problem
- "AI/ML causing science crises" (AAAS 2/16/2019)
  - https://www.bbc.com/news/science-environment-47267081
  - *The "reproducibility crisis" in science refers to the alarming number of research results that can not repeated when another group of scientists tries the same experiment. It can mean that the initial results were wrong. One analysis suggested that up to 85% of all biomedical research carried out in the world is wasted effort.*

An Illuminating Exercise (Step 3)

- Check how many of the 10 or so papers that are reproducible are correct in the sense that
  - the developed learning model performs well in real-world deployment settings
  - The developed model generalizes as expected in deployment scenarios

- Fact: At most 1(!) of the 10 reproducible works is correct

An Illuminating Exercise (Step 3, cont.)

- AI/ML for cybersecurity has a ”trust” problem
- Lipton (2018): “A *network operator has trust in an AI/ML model if the operator is comfortable with relinquishing control to the model.*”
- How to ensure that a network operator can trust a given learning model?

Our Research Focus

- How to decide when an operator can trust a given AI/ML Model?

  in the sense of

- When would an operator be comfortable to relinquish control to the model?

Our Research Program (Synopsis)

- Explainable/Interpretable AI/ML
  - How to construct an appropriately-chosen "white-box" model (e.g., decision tree model) that can "explain/interpret" the decision-making process of a given "black-box" model?

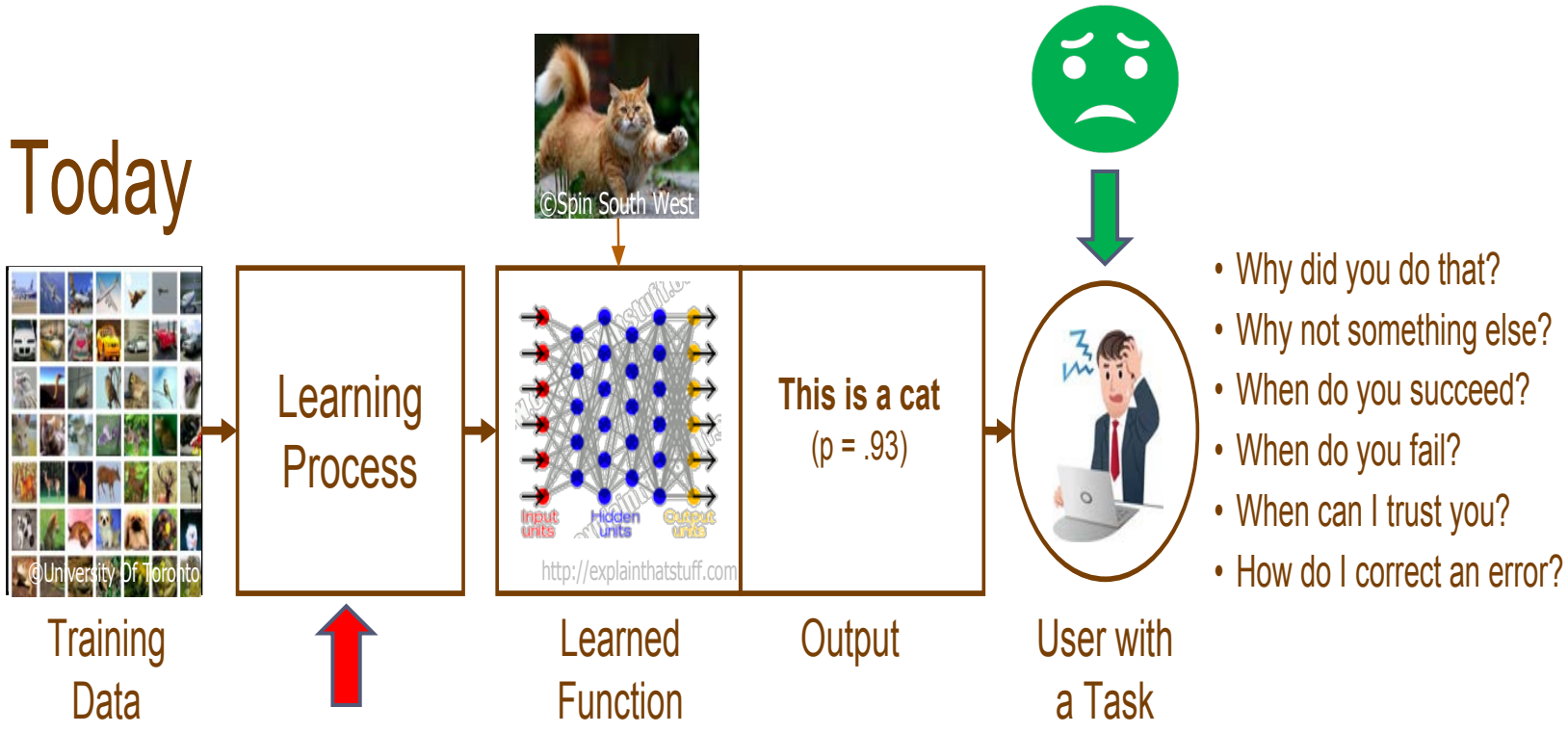
- The problem of underspecification in modern AI/ML
  - The problem refers to the failure to specify an AI/ML model in adequate detail (A. D'Amor et al., 2020)

On Explainable/Interpretable AI/ML

- How to extract high-fidelity "white-box" models such as decision trees (DT) from a given "black-box" model
- New DARPA program (2017): XAI – Explainable AI
  - https://www.darpa.mil/program/explainable-artificial-intelligence

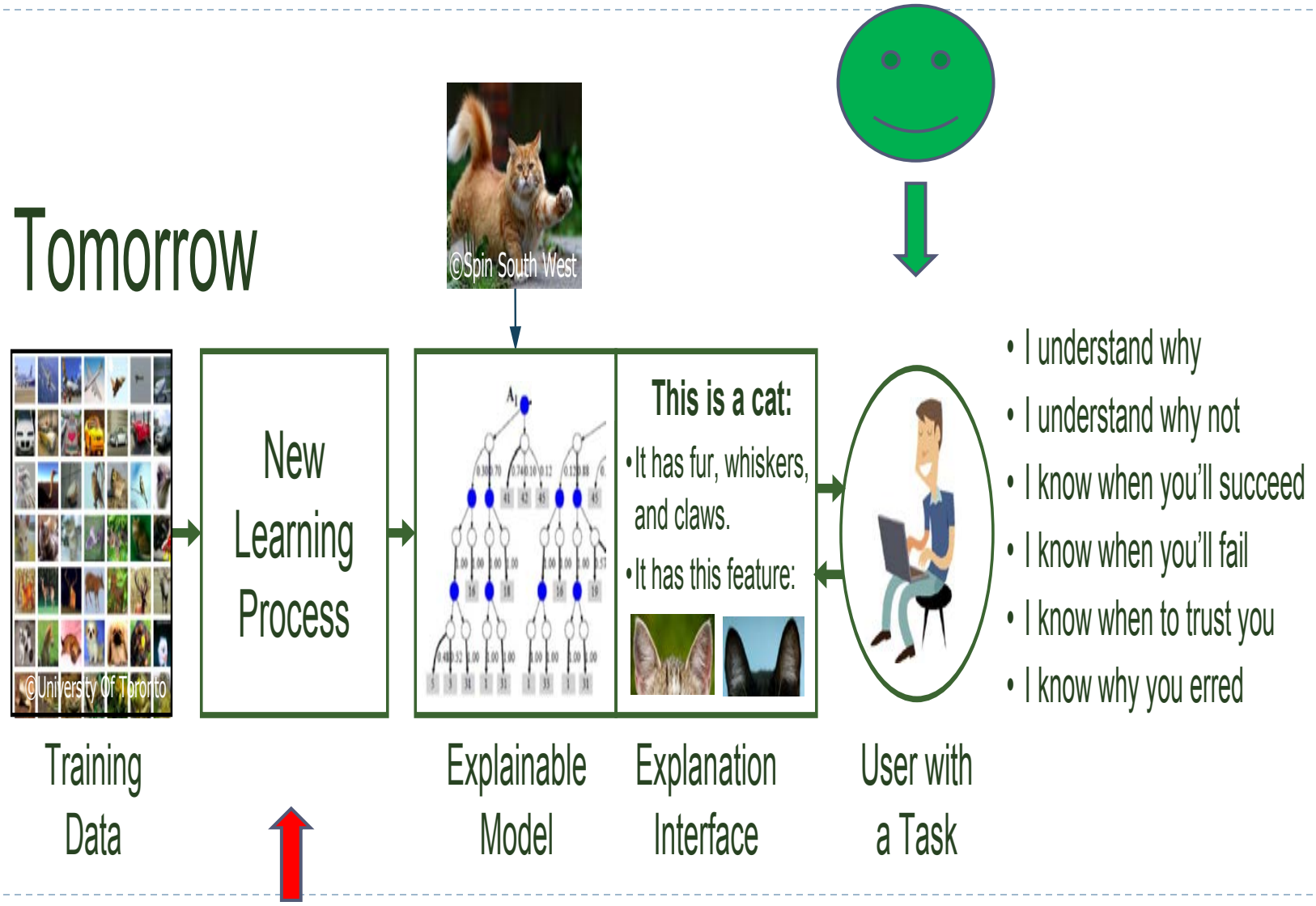
- How to translate heavyweight learning models (e.g., DNN) into lightweight versions that are
  - "explainable" or "white box" and
  - can also run at line rate in the data plane

From "Black Box" Models ….



https://www.darpa.mil/attachments/XAIProgramUpdate.pdf

… to "White Box" Models

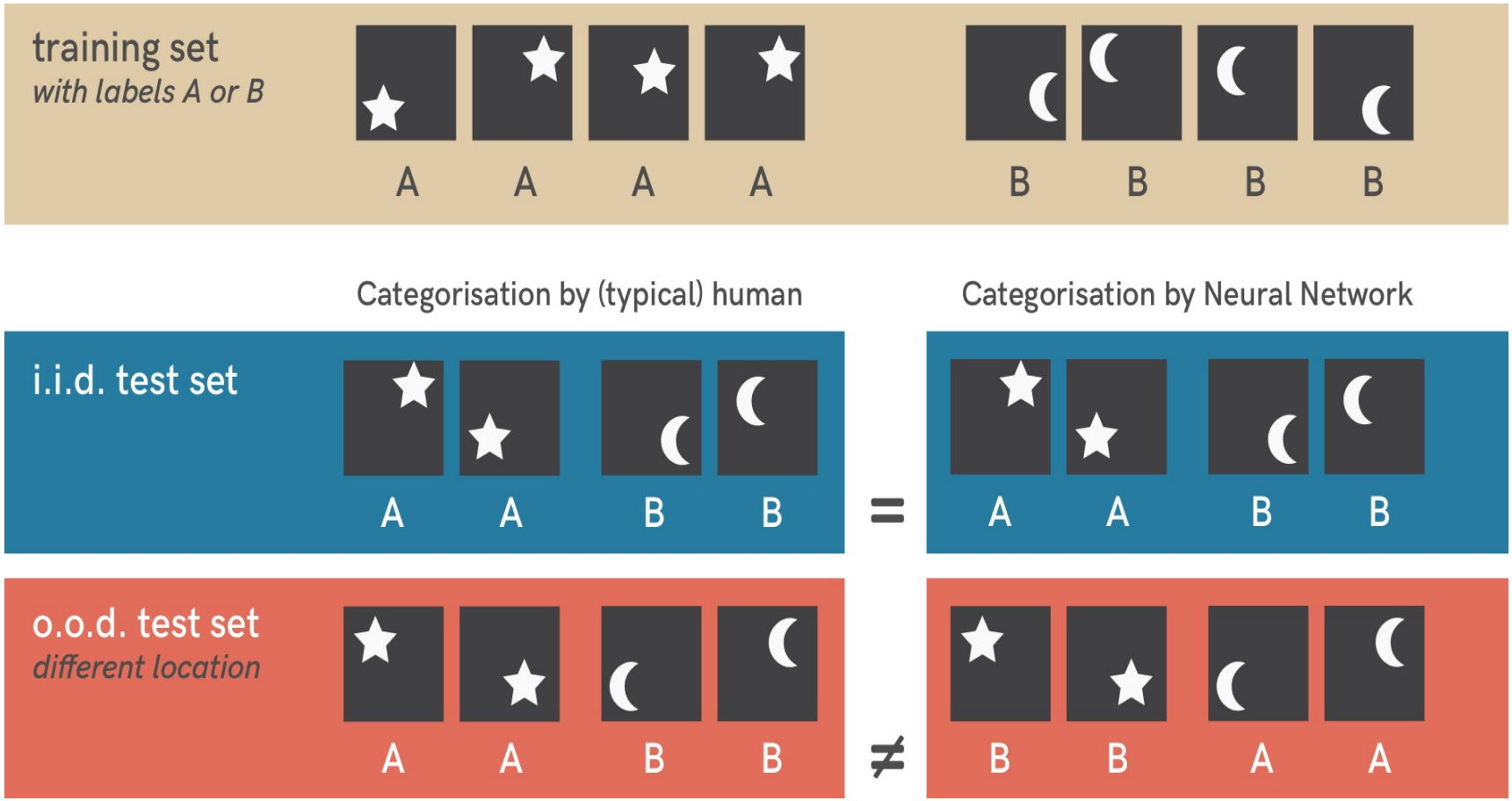


On the Problem of Underspecification

- How to decide if the success of a trained model is
  - due to its innate ability to encode some essential structure of the underlying system or data or
  - simply the result of some inductive biases that the trained model happened to encode?

- **Most common examples of inductive biases**
  - Shortcut learning strategies
  - Spurious correlations
  - Inability to generalize to out-of-distribution samples

Common Example – Inductive Biases



https://arxiv.org/pdf/2004.07780.pdf

Concluding Observations

AI/ML for Cybersecurity: "The Good"

- Lots of efforts
  - Research (academia)
  - Products (industry)
  - Funding (VCs, Private Equity, …)
- Important & promising new developments in AI/ML & networking
  - Explainable/Interpretable AI/ML
  - "Trustworthy AI" (new focus on underspecification problem)
  - The potential of programmable data plane technologies

AI/ML for Cybersecurity: "The Bad"

- AI/ML has a "reproducibility" problem
  - How to avoid that some 90% of all research in the area of AI/ML and cybersecurity will be wasted effort?
  - How to transition from "sharing data" to "sharing learning models/algorithms"?
- There is no "reproducibility" when it comes to commercial AI/ML-based solutions!
  - Proprietary solutions, intellectual property
  - Nothing is open-source

AI/ML for Cybersecurity: "The Ugly"

- AI/ML has a "trust" problem
  - AI/ML-based solutions reported in the existing research literature cannot be trusted …
  - Developing trustworthy AI/ML models for cybersecurity is hard
- Network operators are at the mercy of the vendors
  - Impossible to check if AI/ML-based solutions marketed by the various cybersecurity companies perform as "advertised"
  - Lack of interest in developing a "standard" for evaluating the performance of commercial AI/ML-based cybersecurity solutions

AI/ML for Cybersecurity: The Future

- We can and know how to do better in the future
  - Insist on reproducibility
  - Focus on explainable/interpretable AI/ML
  - Emphasize trustworthy AI/ML

- Doing better in the future requires
  - Broad access to datasets from actual production networks
  - Advances in the area of explainable/interpretable AI/ML
  - Well-designed user studies to evaluate trust in AI/ML models (we can learn from the HCI community …)

THANK YOU!

- Contact me at: wwillinger@niksun.com