



Protect Your Machine Learning Applications from SolarWinds' Attacks

Mark Sherman
Director, Cybersecurity Foundations, CERT
Oct 19, 2020
AI World Government Conference

Software Engineering Institute
Carnegie Mellon University
Pittsburgh, PA 15213

Copyright 2021 Carnegie Mellon University.

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

The view, opinions, and/or findings contained in this material are those of the author(s) and should not be construed as an official Government position, policy, or decision, unless designated by other documentation.

References herein to any specific commercial product, process, or service by trade name, trade mark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by Carnegie Mellon University or its Software Engineering Institute.

NO WARRANTY. THIS CARNEGIE MELLON UNIVERSITY AND SOFTWARE ENGINEERING INSTITUTE MATERIAL IS FURNISHED ON AN "AS-IS" BASIS. CARNEGIE MELLON UNIVERSITY MAKES NO WARRANTIES OF ANY KIND, EITHER EXPRESSED OR IMPLIED, AS TO ANY MATTER INCLUDING, BUT NOT LIMITED TO, WARRANTY OF FITNESS FOR PURPOSE OR MERCHANTABILITY, EXCLUSIVITY, OR RESULTS OBTAINED FROM USE OF THE MATERIAL. CARNEGIE MELLON UNIVERSITY DOES NOT MAKE ANY WARRANTY OF ANY KIND WITH RESPECT TO FREEDOM FROM PATENT, TRADEMARK, OR COPYRIGHT INFRINGEMENT.

[DISTRIBUTION STATEMENT A] This material has been approved for public release and unlimited distribution. Please see Copyright notice for non-US Government use and distribution.

This material may be reproduced in its entirety, without modification, and freely distributed in written or electronic form without requesting formal permission. Permission is required for any other use. Requests for permission should be directed to the Software Engineering Institute at permission@sei.cmu.edu.

Carnegie Mellon® and CERT® are registered in the U.S. Patent and Trademark Office by Carnegie Mellon University.

DM21-0870

Outline

Anatomy of a Supply Chain Attack: SolarWinds

Understanding the ML Attack Surface

Understanding Risks of Transfer Learning

Remedies and Limitations

Additional Attacks on Machine Learning Applications

Supply Chains are How Most Goods are Created



Supply Chain

In commerce, a supply chain is a system of organizations, people, activities, information, and resources involved in supplying a product or service to a consumer. Supply chain activities involve the transformation of natural resources, raw materials, and components into a finished product and deliver to the end customer

Source: Wikipedia, https://en.wikipedia.org/wiki/Cold_chain

Supply Chains are How Most Goods are Created – With Properties



Cold Chain

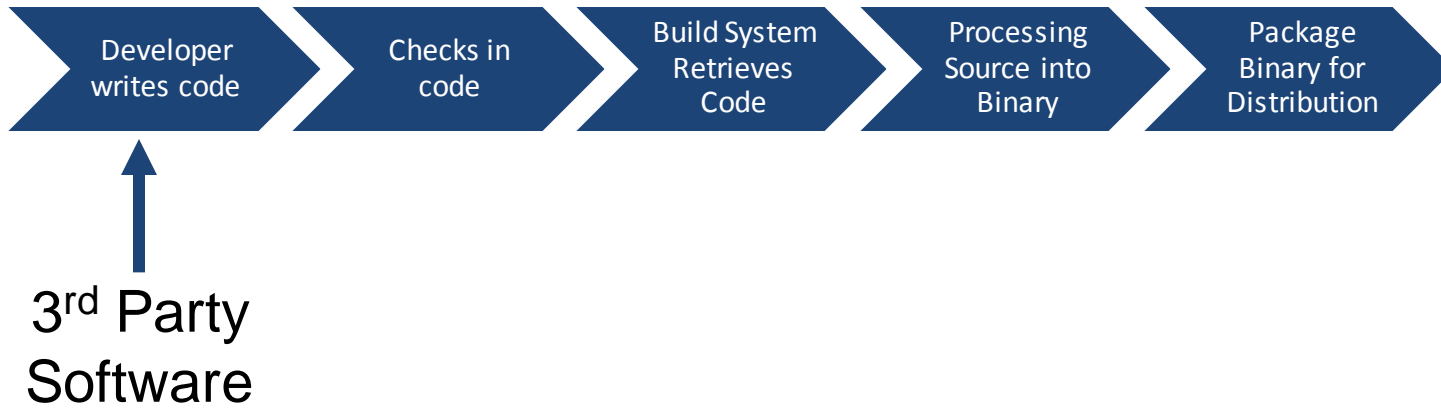
A cold chain is a temperature-controlled supply chain. An unbroken cold chain is an uninterrupted series of storage and distribution activities which maintain a given temperature range.

Source: Wikipedia, https://en.wikipedia.org/wiki/Cold_chain

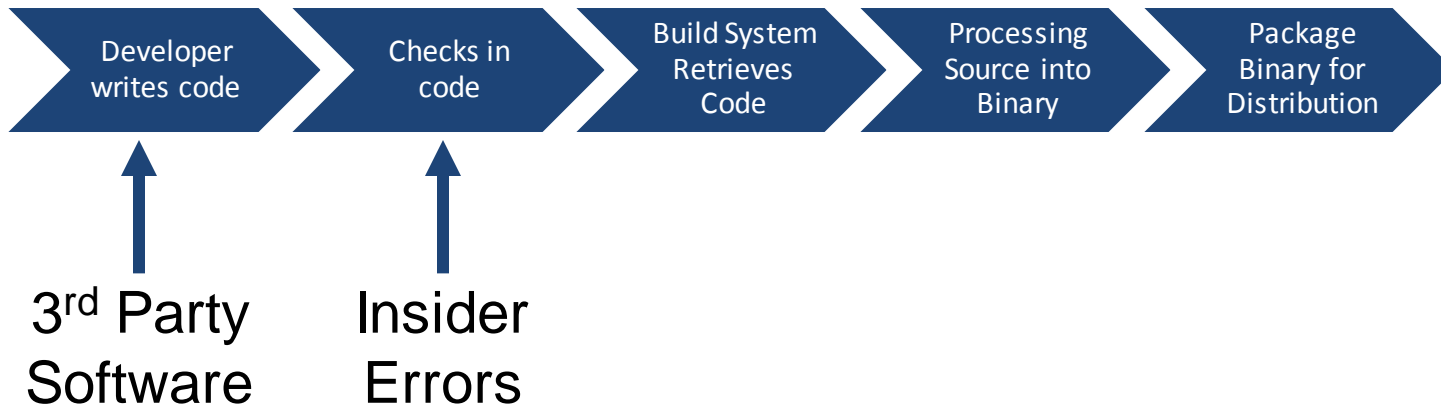
Development Process and Supply Chains



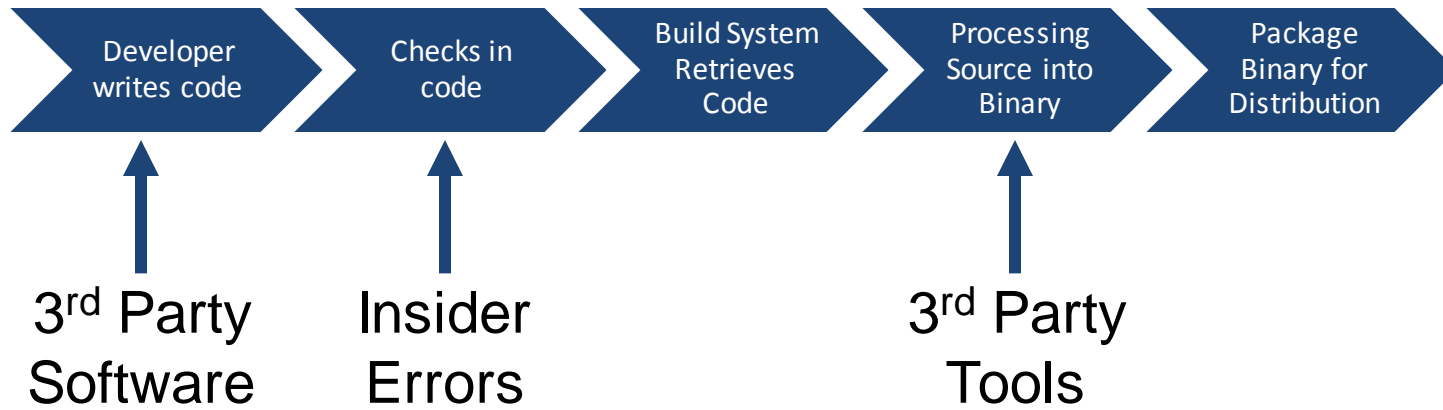
Development Process and Supply Chains



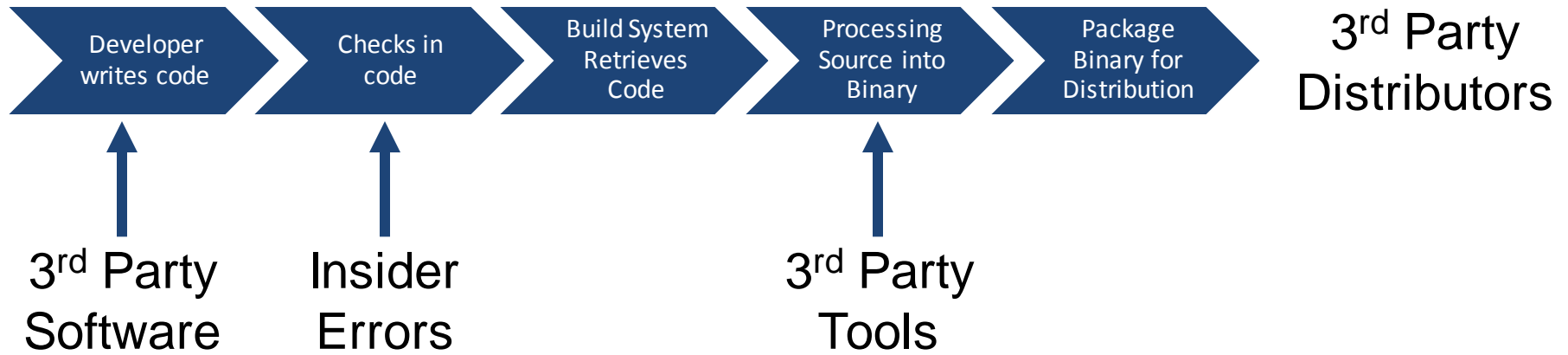
Development Process and Supply Chains



Development Process and Supply Chains



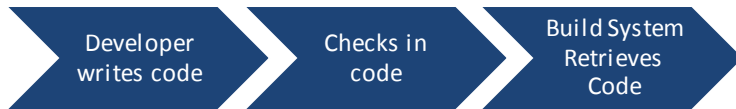
Development Process and Supply Chains



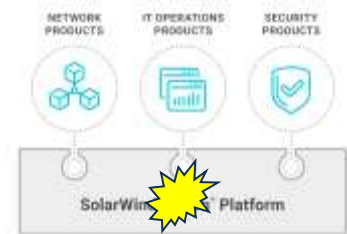
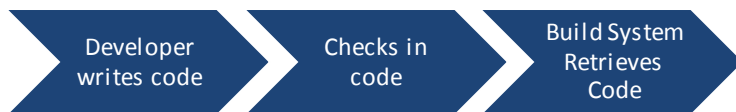
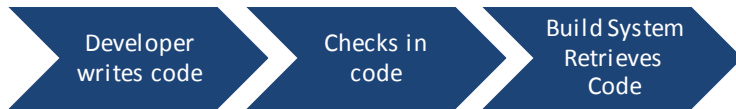
Corrupting the Development Process – Reliving SolarWinds



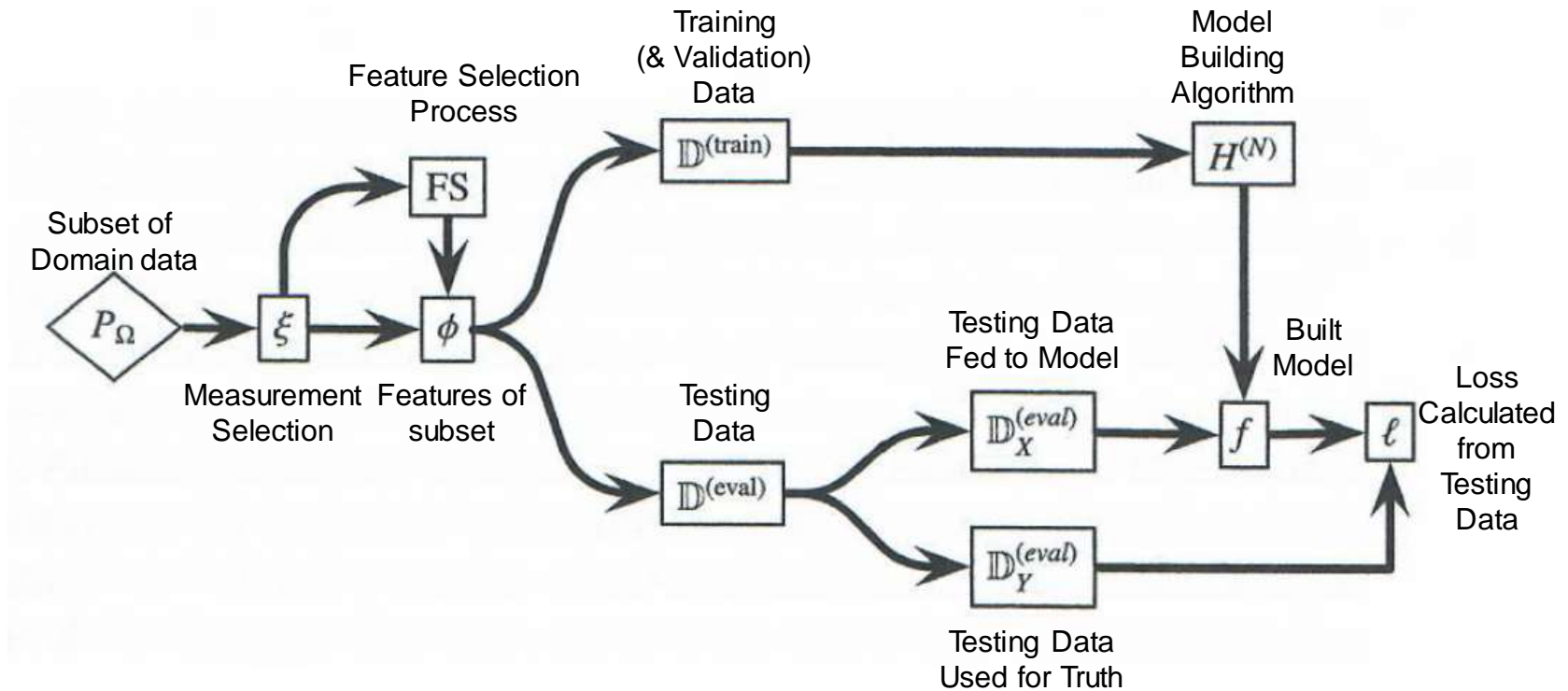
Corrupting the Development Process – Reliving SolarWinds



Corrupting the Development Process – Reliving SolarWinds

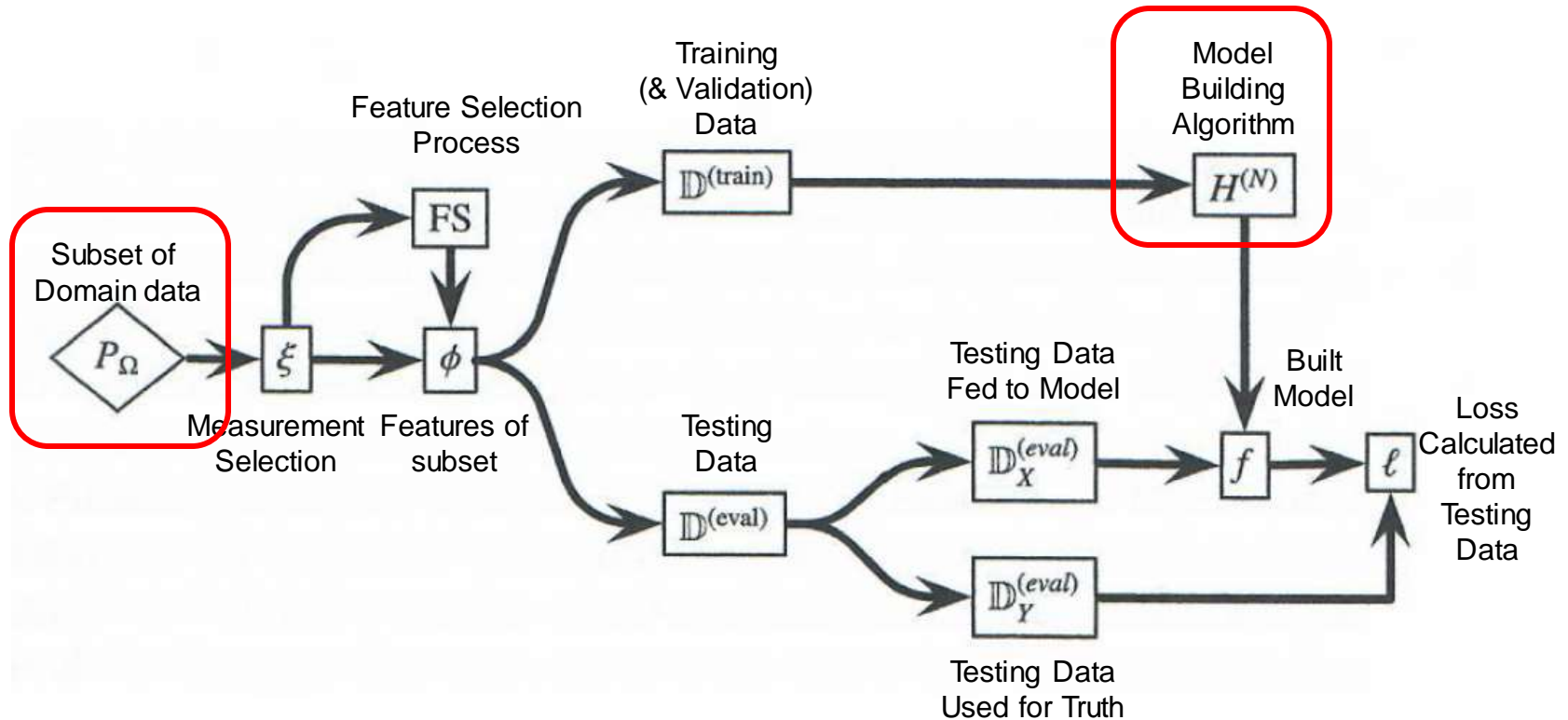


Developing a Machine Learning Application



Adapted from Joseph, Nelson, Rubinstein, Tygar; Adversarial Machine Learning, Cambridge University Press, 2019

Developing a Machine Learning Application



Adapted from Joseph, Nelson, Rubinstein, Tygar; Adversarial Machine Learning, Cambridge University Press, 2019

Software Supply Chain for Machine Learning

Machine learning depends on frameworks and data sets
Relatively less is known about the security of these “supplies”

Data Sources

- Kaggle
- UCI Machine Learning Repository
- Find Datasets
- Data.gov
- xView
- ImageNet
- Google’s Open Images

Machine Learning Frameworks

- Pandas
- Numpy
- Scikit-learn
- TensorFlow
- Keras
- Pytorch & Torch

Lots of Deep Fake Examples are Available



Rich supplies of “deep fakes” are readily accessible

Source: <https://ai.googleblog.com/2019/09/contributing-data-to-deepfake-detection.html>

Poor detection of deep fakes



Cannot reliably verify training data

Fake data easy to introduce into a model

FaceForensics Benchmark

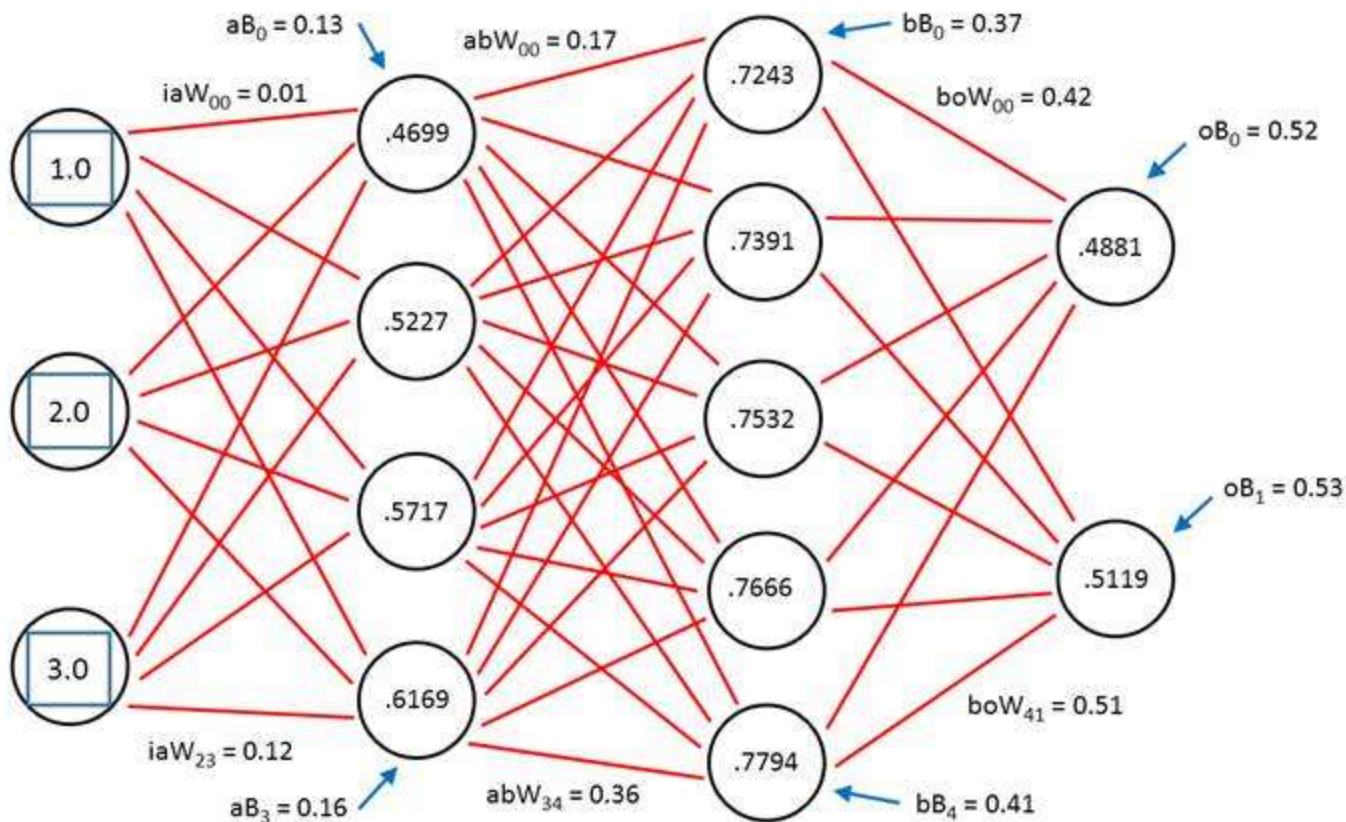
This table lists the benchmark results for the Binary Classification scenario:

Method	Info	Deepfakes	Face2Face	FaceSwap	NeuralTextures	Pristine	Total
Xception		0.964	0.869	0.903	0.807	0.524	0.710
<small>Andreas Rossler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Julia Thies, Matthias Nießner: FaceForensics++: Learning to Detect Manipulated Facial Images, ICCV 2019</small>							
MesoNet		0.873	0.562	0.612	0.407	0.726	0.660
<small>Dariusz Alchar, Vincent Niziol, Junichi Yamagishi, and Iain Echizen: MesoNet: a compact facial video forgery detection network, arXiv</small>							
XceptionNet Full Image		0.745	0.759	0.709	0.733	0.510	0.624
<small>Andreas Rossler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Julia Thies, Matthias Nießner: FaceForensics++: Learning to Detect Manipulated Facial Images, ICCV 2019</small>							
Bayar and Stamm		0.845	0.737	0.825	0.707	0.462	0.616
<small>Behzad Bayar and Matthew C. Stamm: A deep learning approach to universal image manipulation detection using a new convolutional layer, ACM Workshop on Information Hiding and Multimedia Security</small>							
Rahmouni		0.855	0.642	0.563	0.607	0.500	0.581
<small>Nicolas Rahmouni, Vincent Niziol, Junichi Yamagishi, and Iain Echizen: Distinguishing computer graphics from natural images using convolution neural networks, IEEE Workshop on Information Forensics and Security</small>							
Recasting		0.855	0.679	0.738	0.780	0.344	0.552
<small>Davide Cozzolino, Christian Poggi, and Luisa Verdoliva: Recasting residual-based local descriptors as convolutional neural networks: an application to image forgery detection, ACM Workshop on Information Hiding and Multimedia Security</small>							
Steganalysis Features		0.736	0.737	0.689	0.633	0.340	0.518
<small>Jawad French and Jan Kodovský: Rich Models for Steganalysis of Digital Images, IEEE Transactions on Information Forensics and Security</small>							

Preconfigured machine learning (i.e., teacher) systems provide a vehicle to distribute bad training data

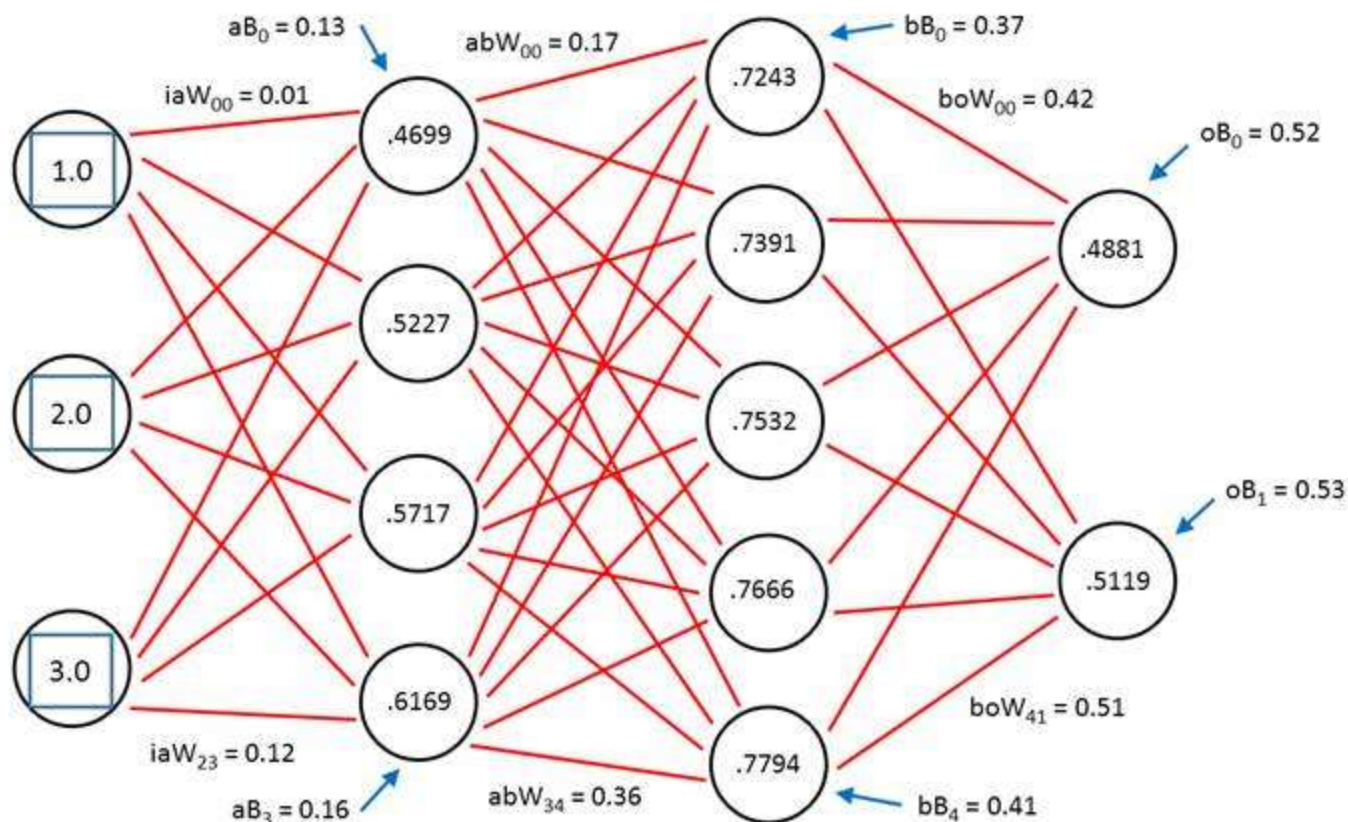
Source: http://kaldir.vc.in.tum.de/faceforensics_benchmark/index.php (as of 9/25/19)

Trained Deep Neural Network



Sources: Sergey Golubev, Deep Neural Networks: A Getting Started Tutorial, Part #1, 30 June 2014, <https://www.mq15.com/en/blogs/post/203> ; Brown, et al, "Language Models are Few-Shot Learners," July 22, 2020, <https://arxiv.org/pdf/2005.14165.pdf> ; Lucian Constantin, "How data poisoning attacks corrupt machine learning models," CSO Online, Apr 12, 2021, <https://www.csoonline.com/article/3613932/how-data-poisoning-attacks-corrupt-machine-learning-models.html>

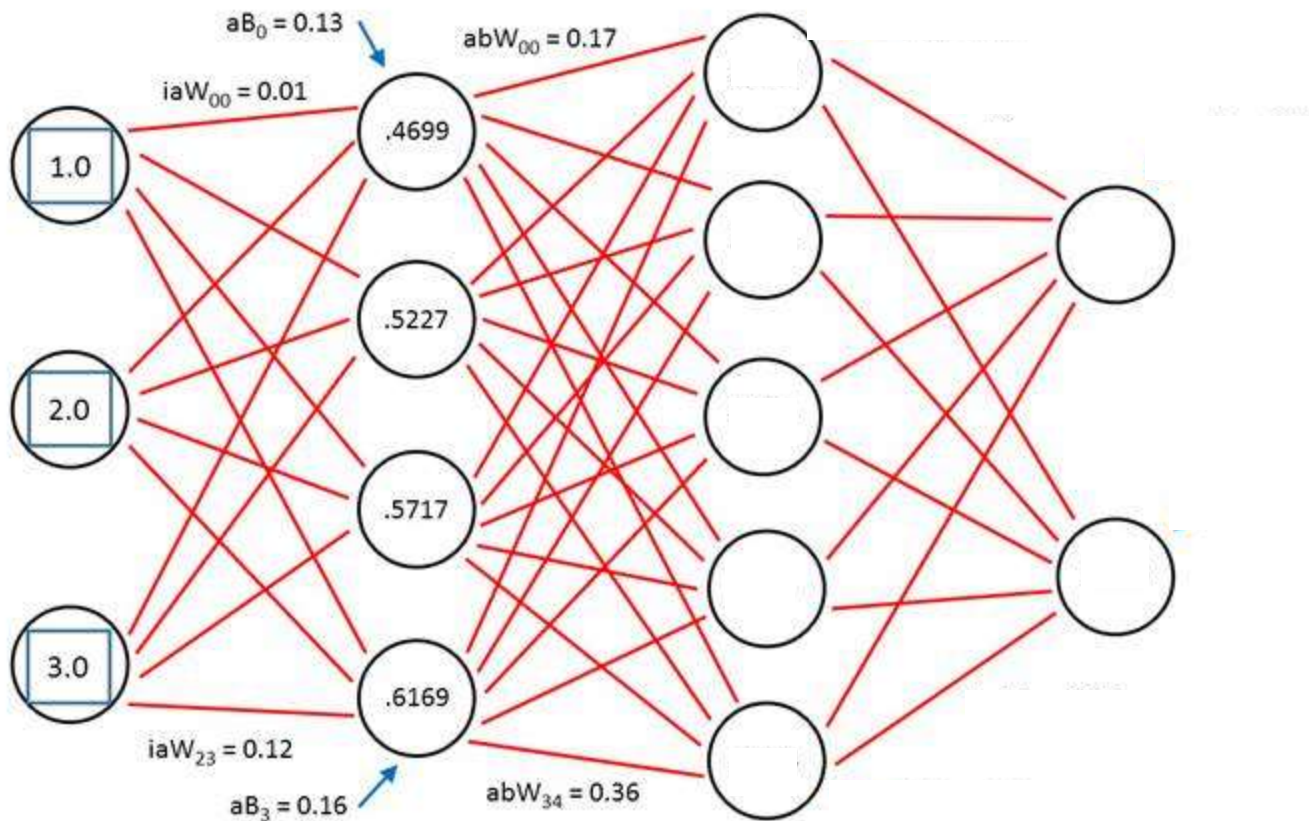
Trained Deep Neural Network



The GPT-3 deep neural net for natural language processing has 175 billion parameters and cost ~\$16M to train once

Sources: Sergey Golubev, Deep Neural Networks: A Getting Started Tutorial, Part #1, 30 June 2014, <https://www.mq15.com/en/blogs/post/203> ; Brown, et al, "Language Models are Few-Shot Learners," July 22, 2020, <https://arxiv.org/pdf/2005.14165.pdf> ; Lucian Constantin, "How data poisoning attacks corrupt machine learning models," CSO Online, Apr 12, 2021, <https://www.csoonline.com/article/3613932/how-data-poisoning-attacks-corrupt-machine-learning-models.html>

Transfer Learning Reuses Trained model



Reusing model layers saves time and resources in training but is a vector for hidden processing

Reducing software supply chain risk factors

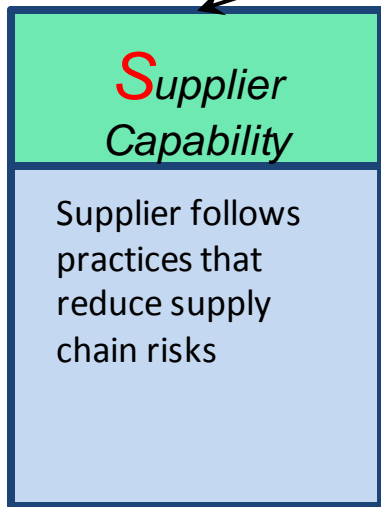
Software supply chain risk for a product needs to be reduced to acceptable level

*Supplier
Capability*

Supplier follows practices that reduce supply chain risks

Reducing software supply chain risk factors

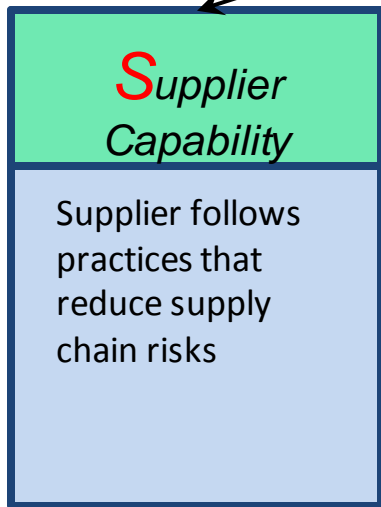
Software supply chain risk for a product needs to be reduced to acceptable level



Repeatable Builds

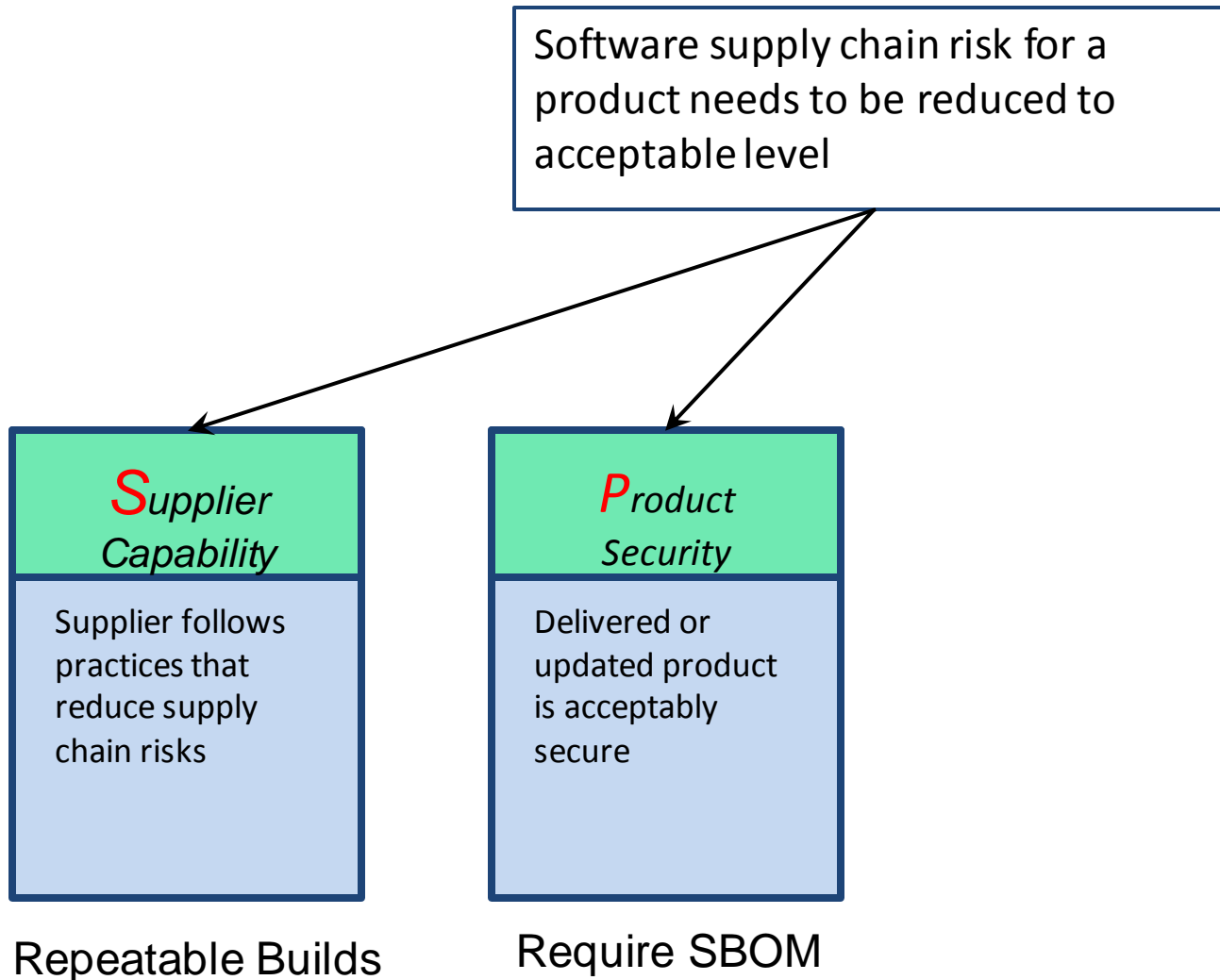
Reducing software supply chain risk factors

Software supply chain risk for a product needs to be reduced to acceptable level

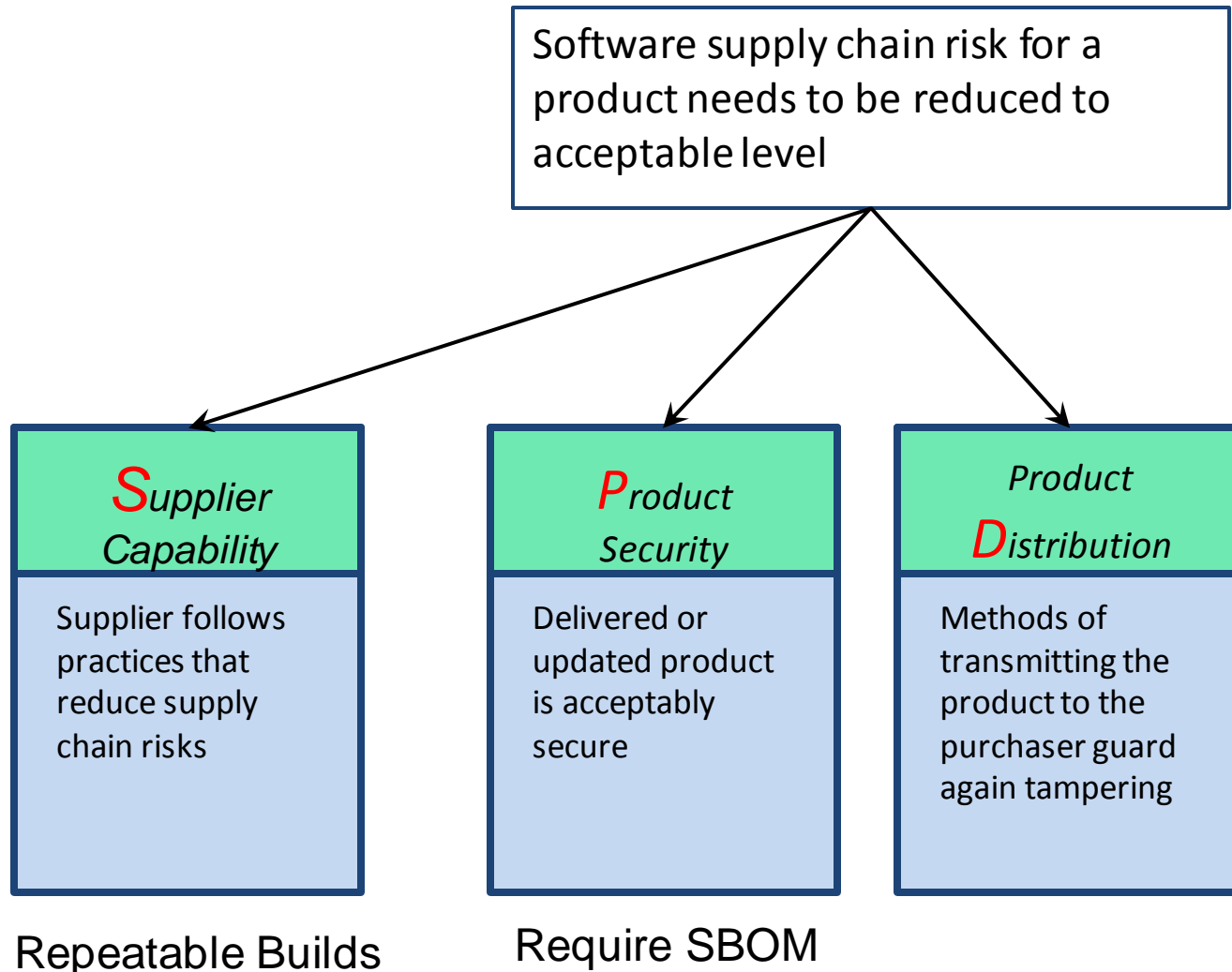


Repeatable Builds

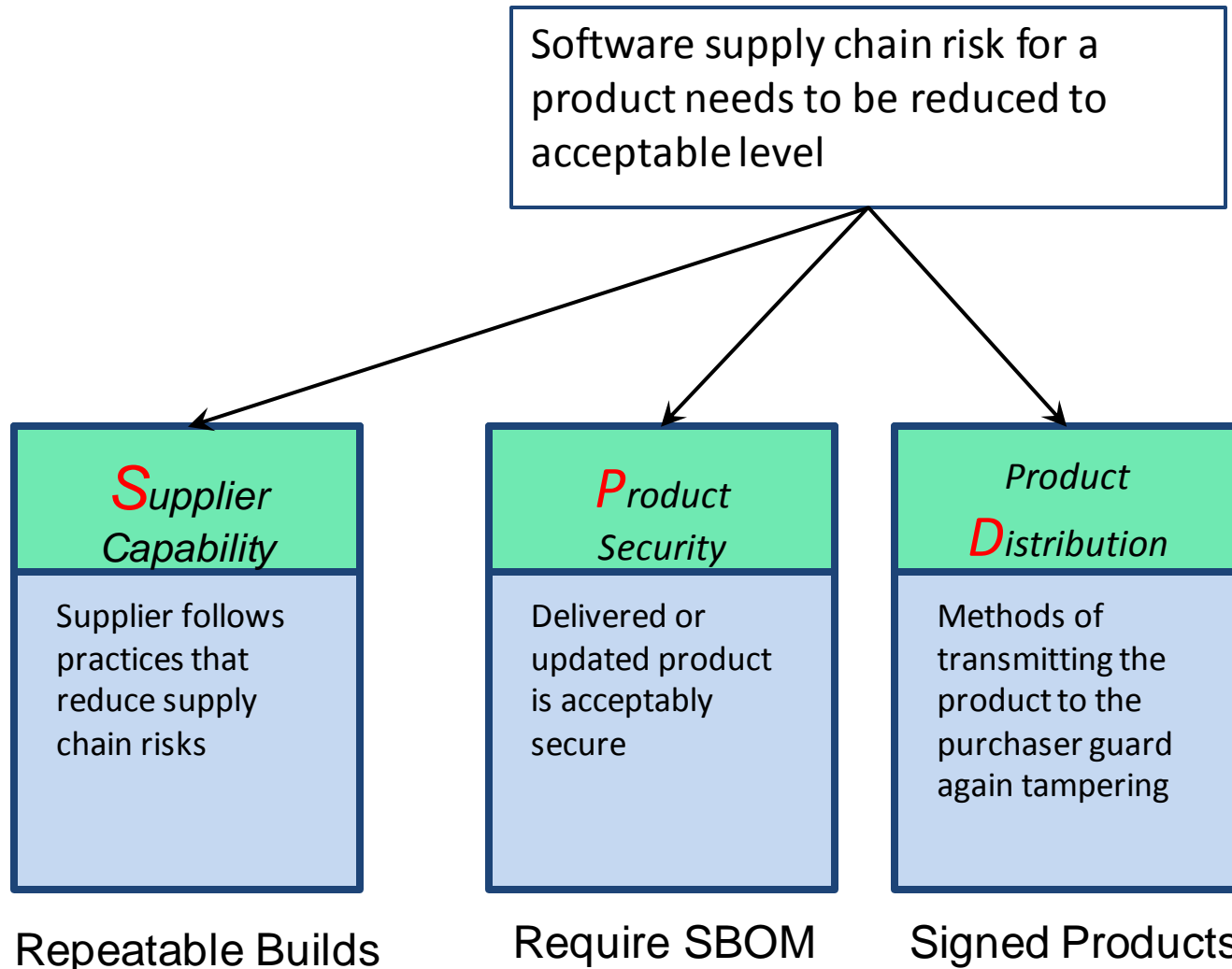
Reducing software supply chain risk factors



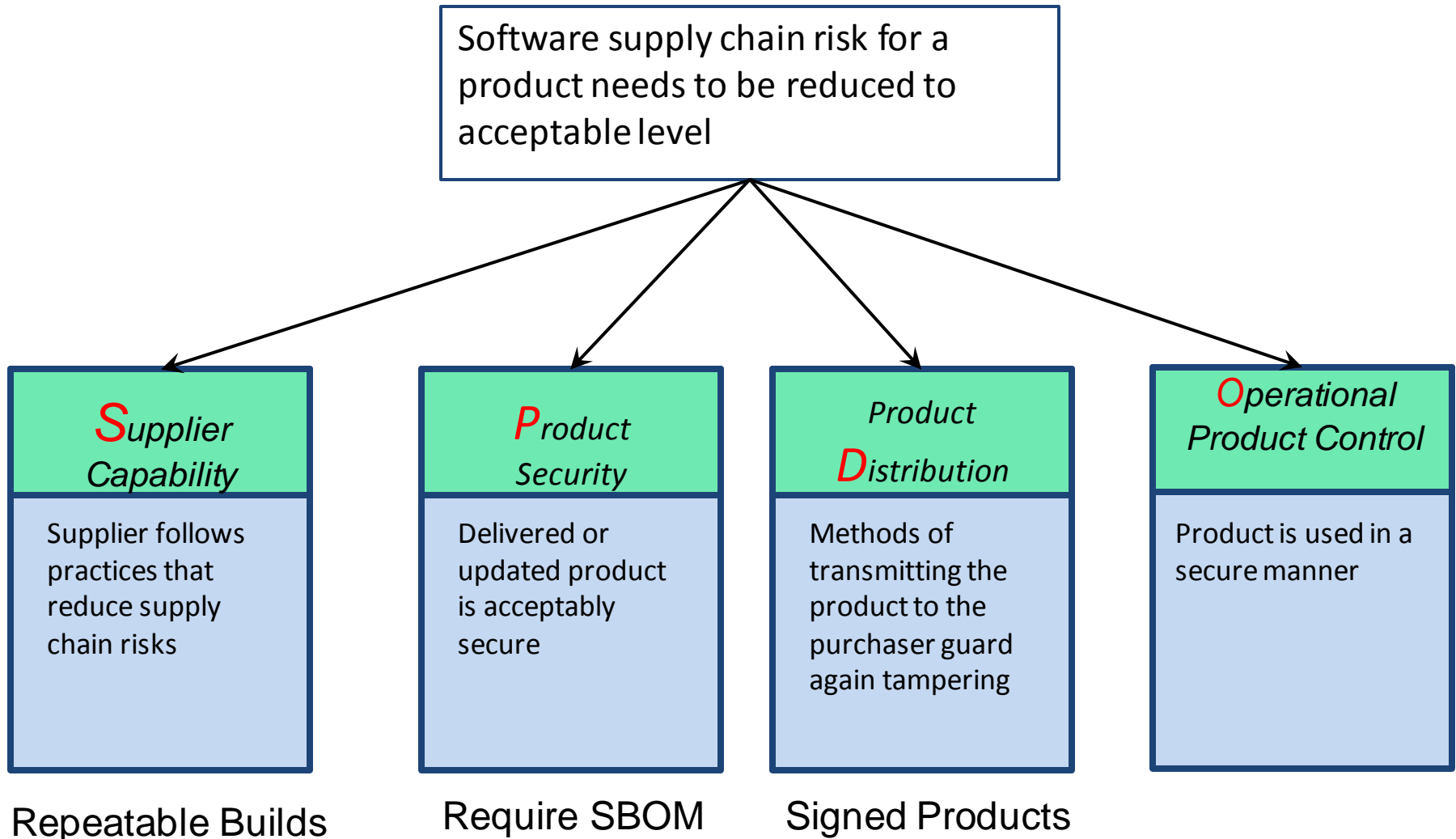
Reducing software supply chain risk factors



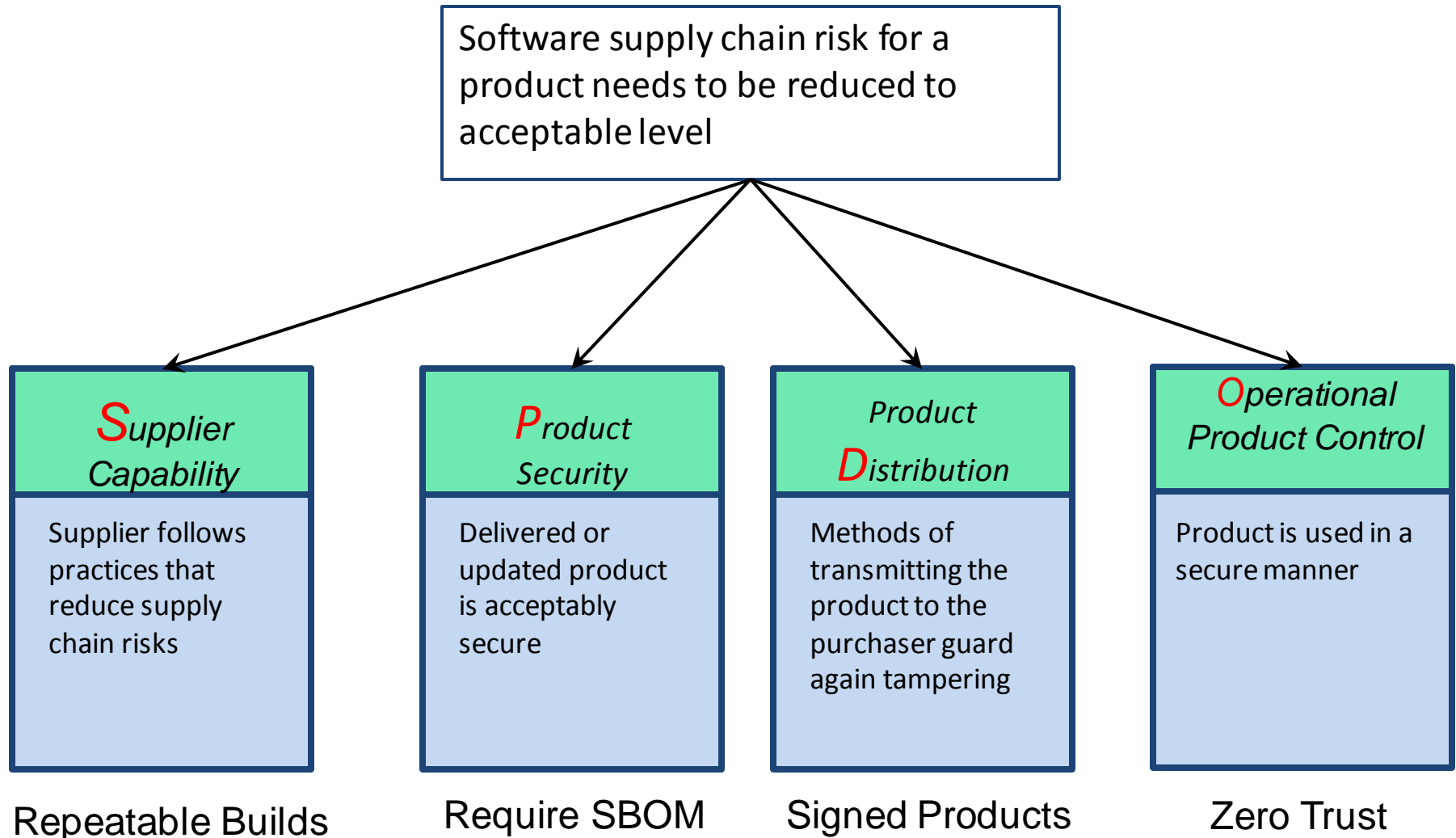
Reducing software supply chain risk factors



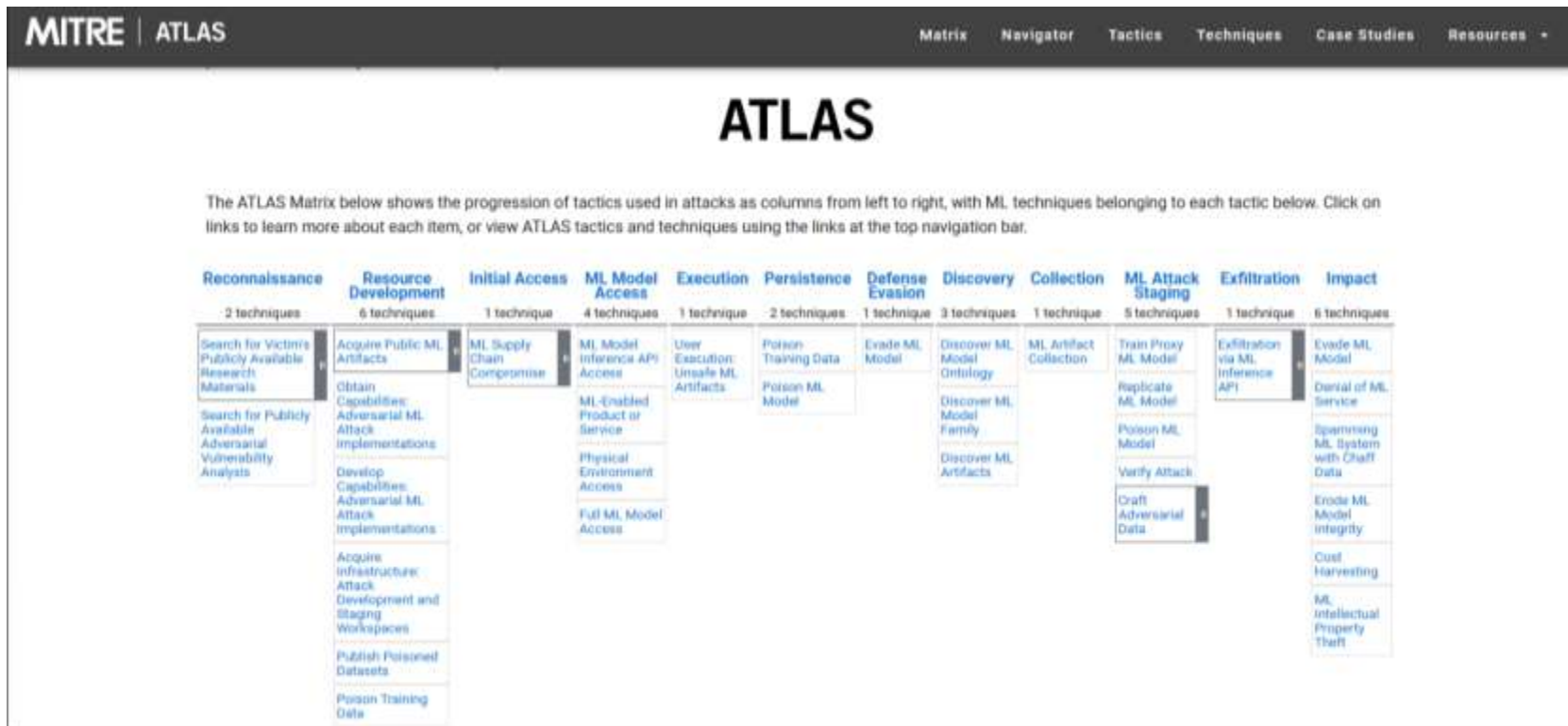
Reducing software supply chain risk factors



Reducing software supply chain risk factors



Additional Attacks on Machine Learning Applications



Summary

Development Processes for Machine Learning Applications are Complex

Each Step in the Process is an Opportunity for Corruption

Extensive Supply Chains for Machine Learning Applications is a Ready Vector for Introducing Corruption

Managing Supply Chain Risk Requires Positive Attention

Ways to Engage with Us



- Download [software and tools](#)
- Explore [research and capabilities](#)
- Participate in [education](#) offerings
- Attend an [event](#)
- Search the [digital library](#)
- Read the [SEI Year in Review](#)
- [Collaborate](#) with the SEI on a new project

Software Engineering Institute

Carnegie Mellon University
4500 Fifth Avenue
Pittsburgh, PA 15213-3890
412-268-5800 - Phone
888-201-4479 - Toll-Free
412-268-5758 - Fax
info@sei.cmu.edu - Email
www.sei.cmu.edu - Web