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

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.

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

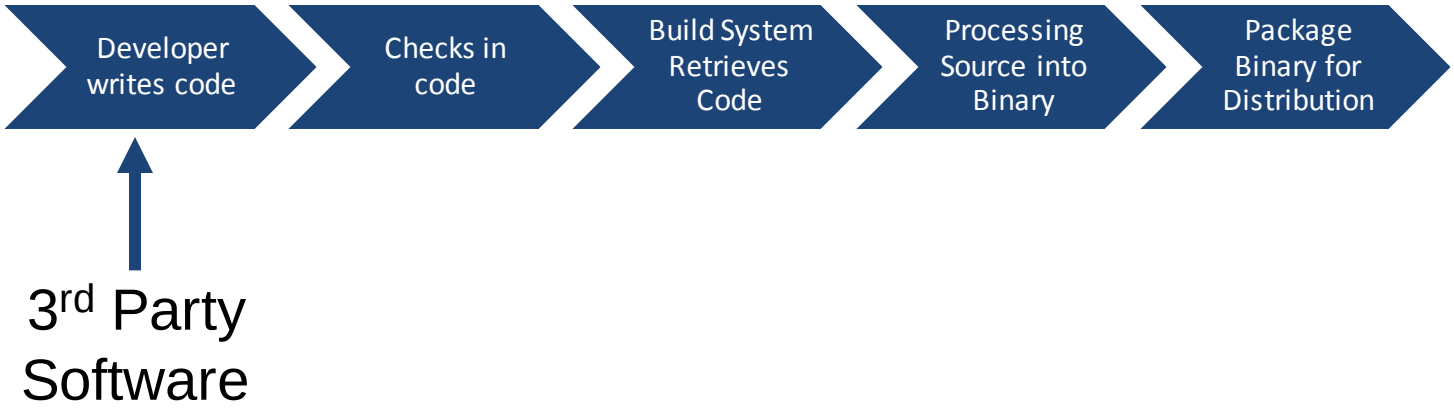
Cold Chain

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.

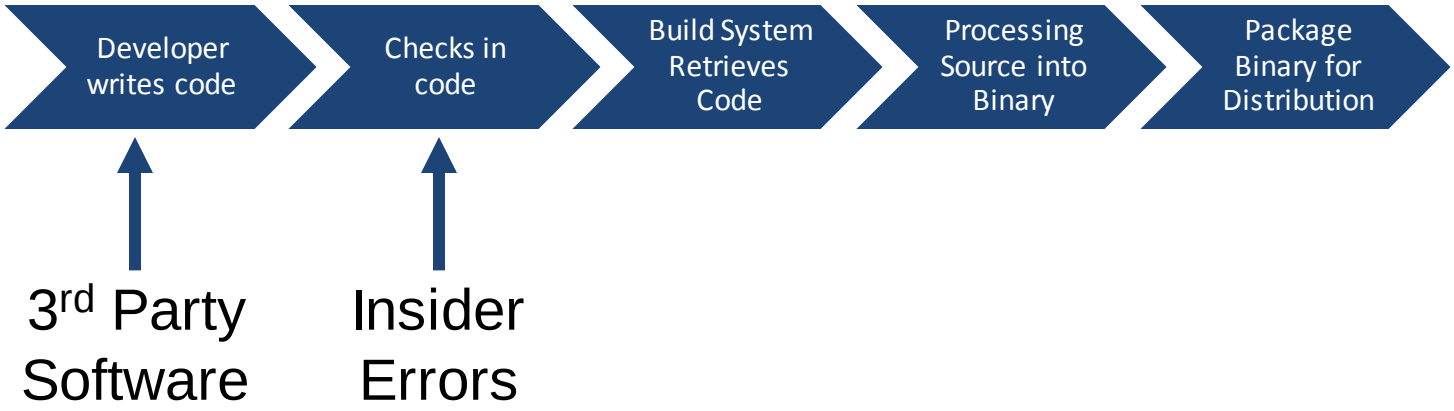
Source: Wikipedia, https://en.wikipedia.org/wiki/Cold_chain

Development Process and Supply Chains

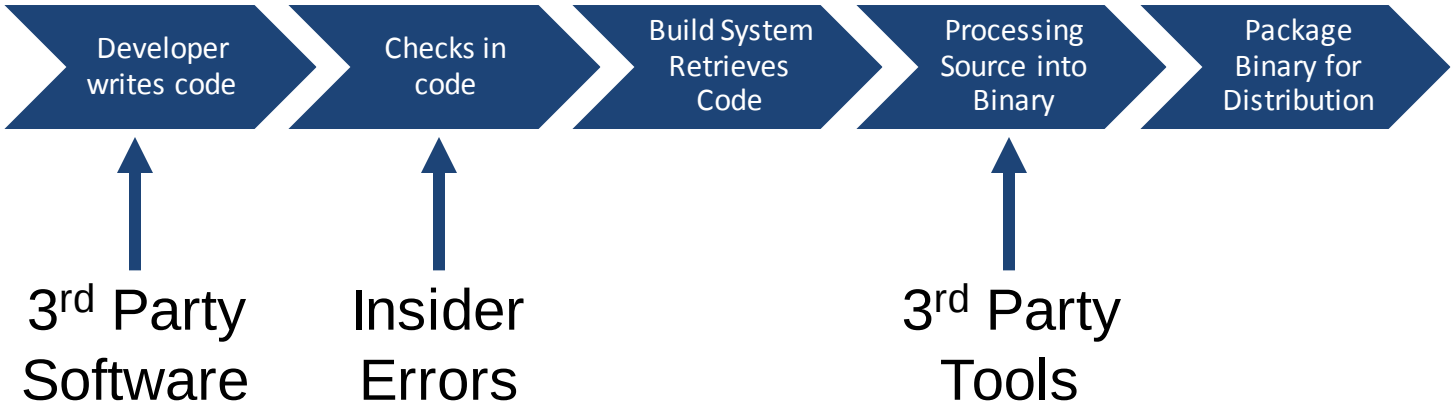
Development Process and Supply Chains



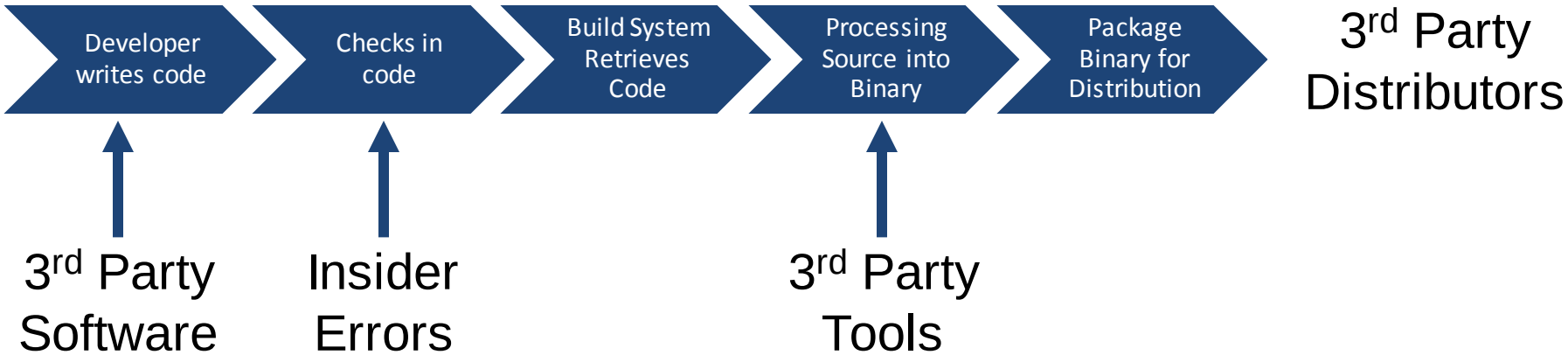
Development Process and Supply Chains



Development Process and Supply Chains



Development Process and Supply Chains



Corrupting the Development Process – Reliving SolarWinds

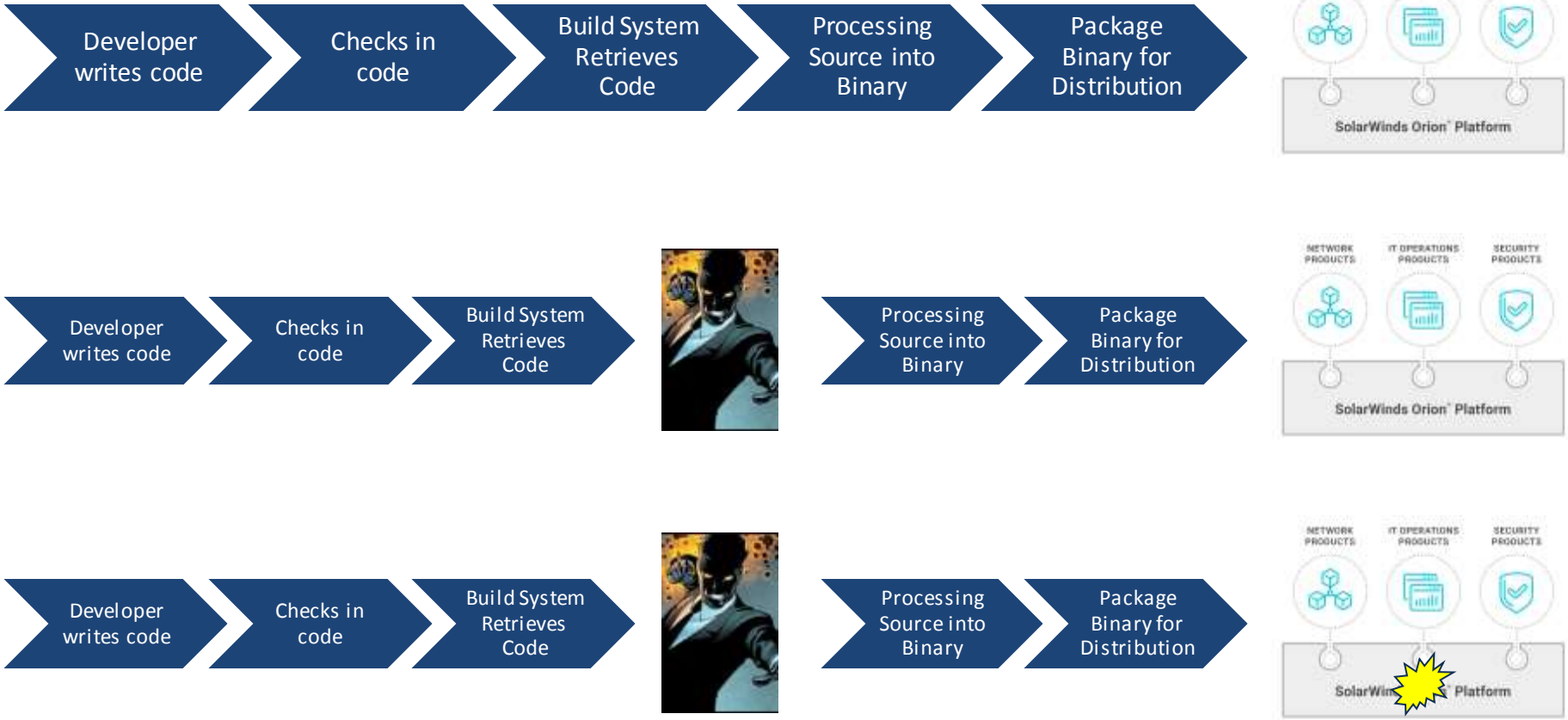
Corrupting the Development Process – Reliving SolarWinds

Developer writes code → Checks in code → Build System Retrieves Code → Processing Source into Binary → Package Binary for Distribution

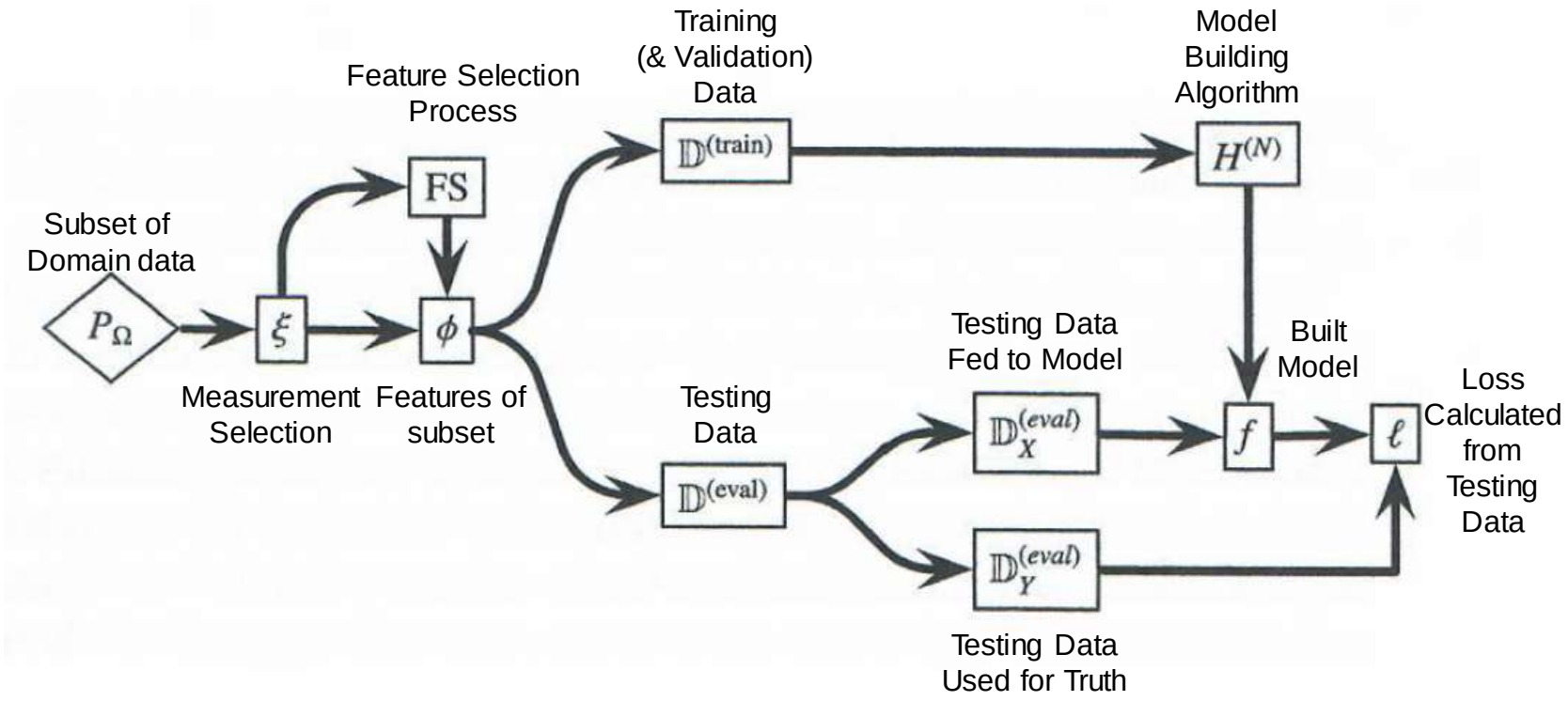
Developer writes code → Checks in code → Build System Retrieves Code

Processing Source into Binary → Package Binary for Distribution

Corrupting the Development Process – Reliving SolarWinds

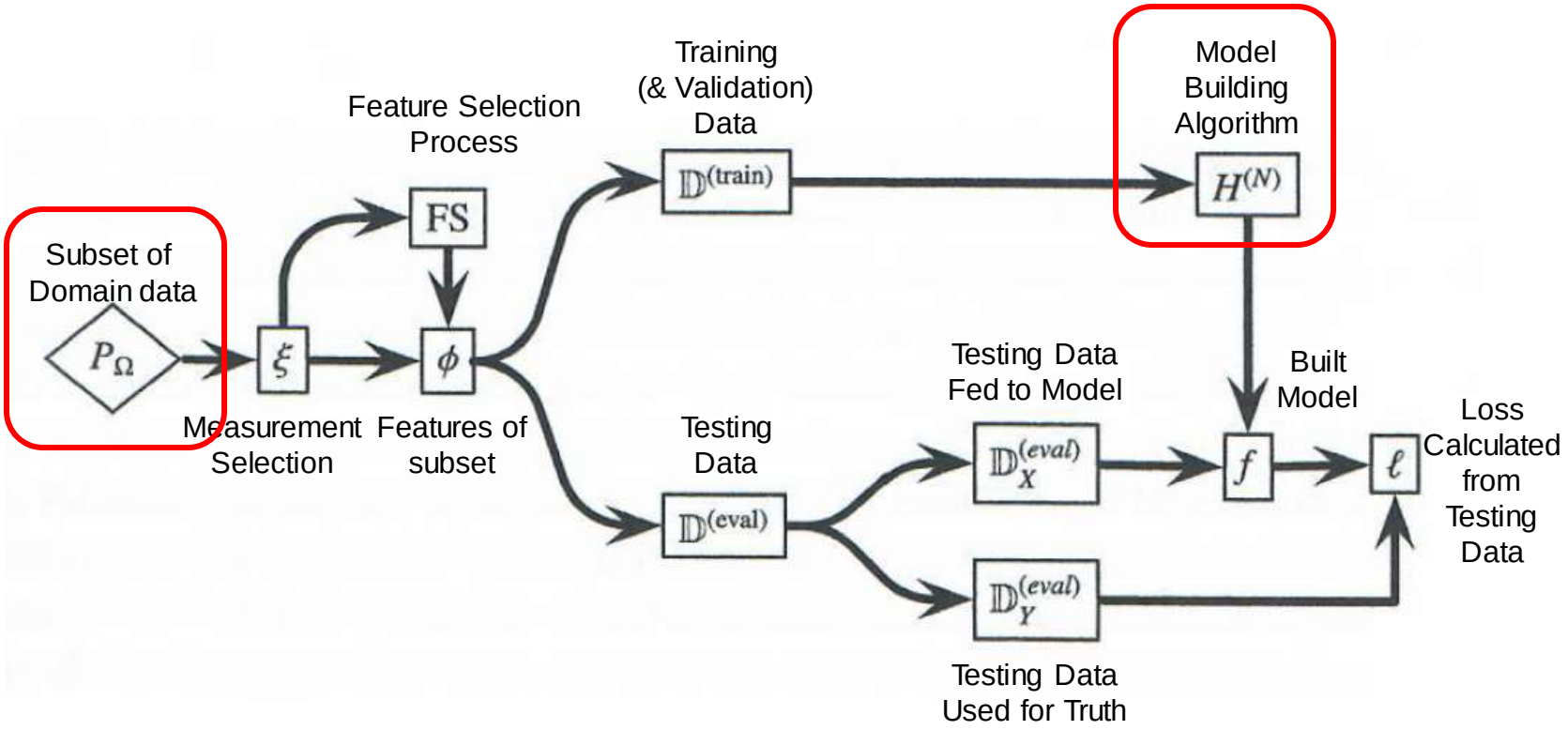


Developing a Machine Learning Application



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

FaceForensics Benchmark

Benchmarks ▾ Data and Documentation About Submit

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
MesoNet		0.873	0.562	0.612	0.407	0.726	0.660
XceptionNet Full Image		0.745	0.759	0.709	0.733	0.510	0.624
Bayar and Stamm		0.845	0.737	0.825	0.707	0.462	0.616
Rahmouni		0.855	0.642	0.563	0.607	0.500	0.581
Recasting		0.855	0.679	0.738	0.780	0.344	0.552
Steganalysis Features		0.736	0.737	0.689	0.633	0.340	0.518

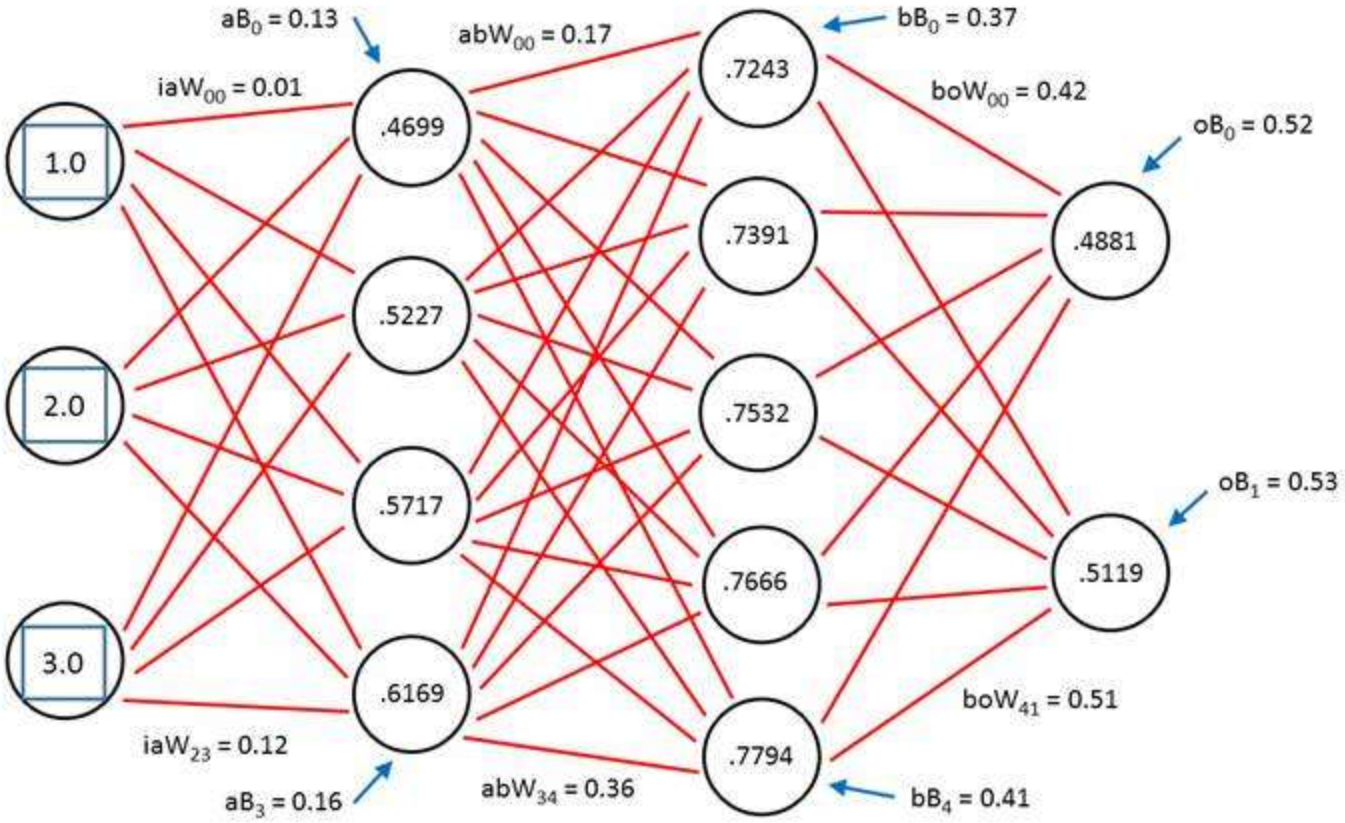
Cannot reliably verify training data

Fake data easy to introduce into a model

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

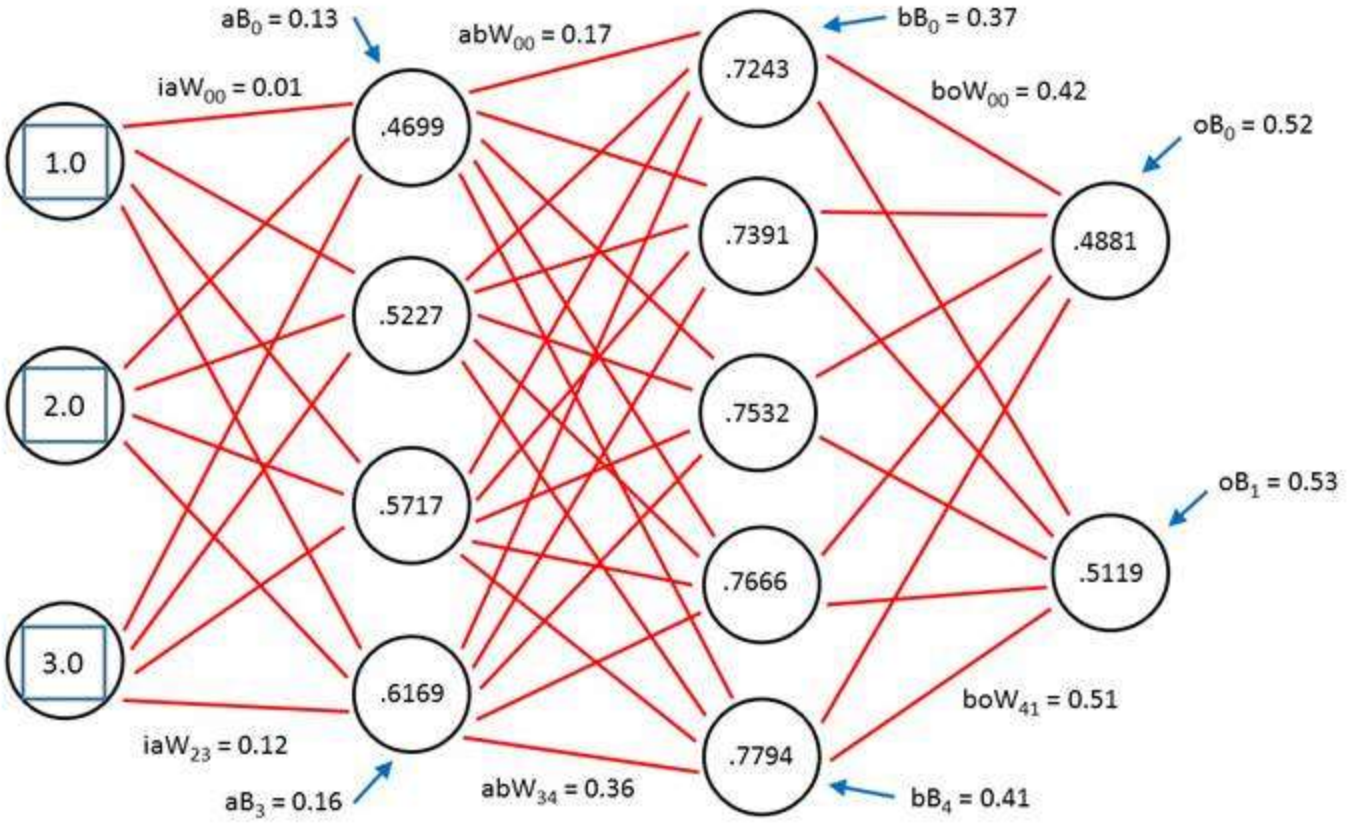
Source: http://kaldir.vc.in.tum.de/faceforensics_benchmark/index.php (as of 9/25/19)

Trained Deep Neural Network



Sources: Sergey Golubev, Deep Neural Networks: A Getting Started Tutorial, Part #1, 30 June 2014, https://www.mql5.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

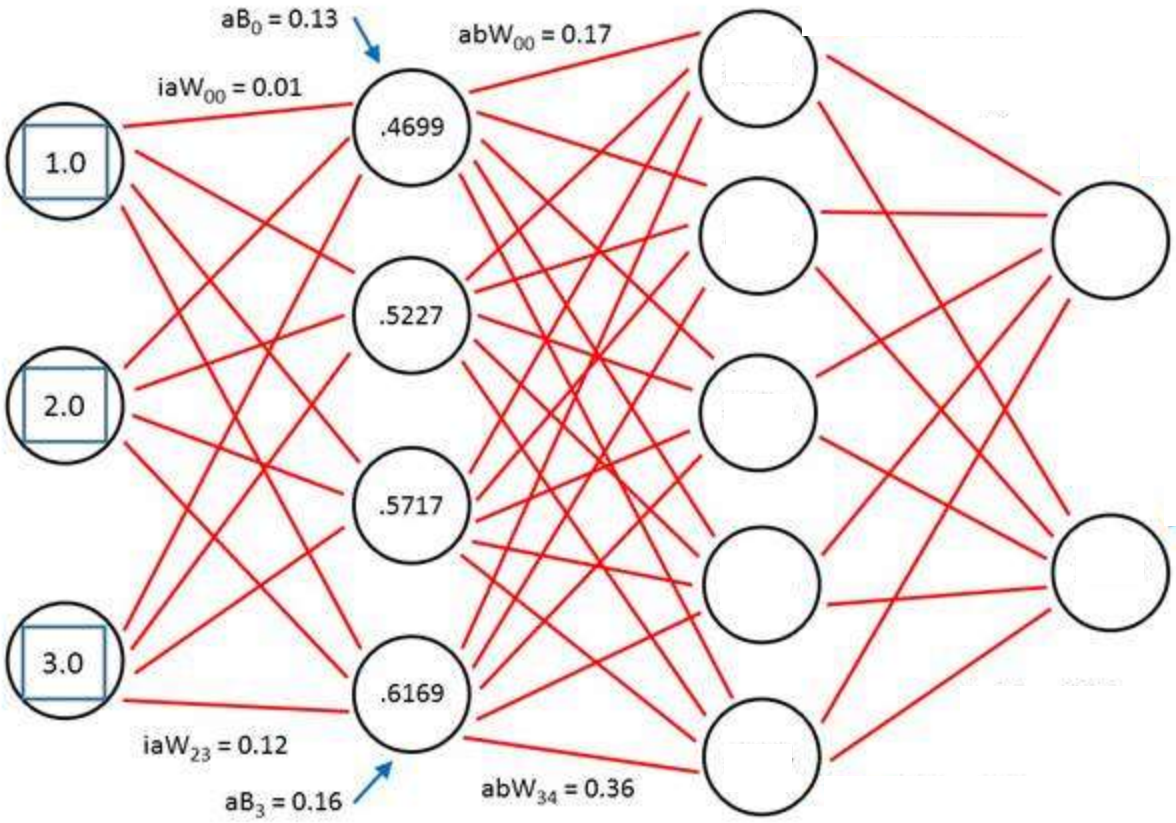
Trained Deep Neural Network



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

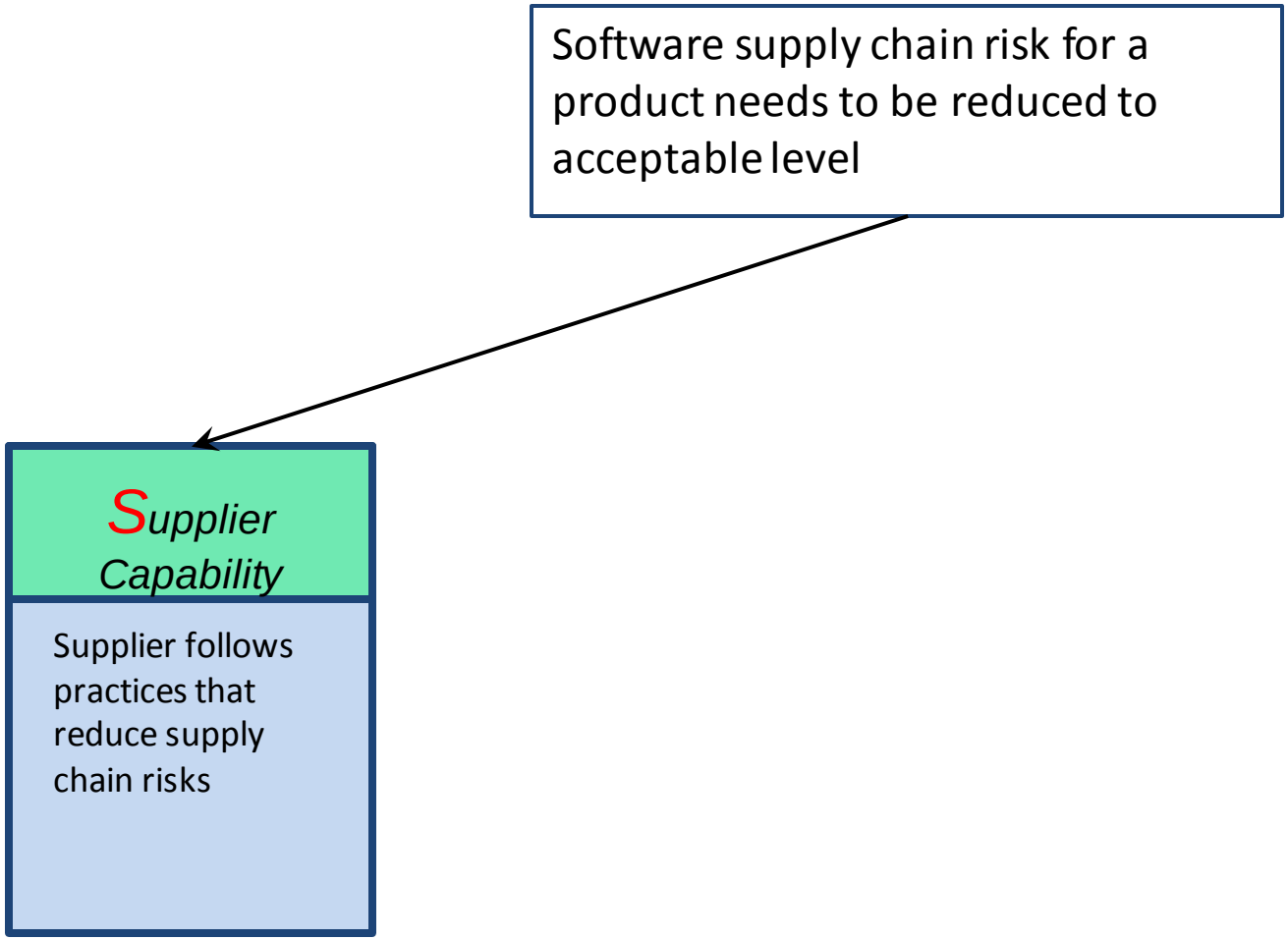
Sources: Sergey Golubev, Deep Neural Networks: A Getting Started Tutorial, Part #1, 30 June 2014, https://www.mql5.com/en/blogs/post/203 ; Brown, et al, "Language Models are Few-Shot Learners," July 22, 2020, https://arxiv.org/pdf/2005.14165.pdf ; Lucian Constantin, "How data poisoning attacks corrupt 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

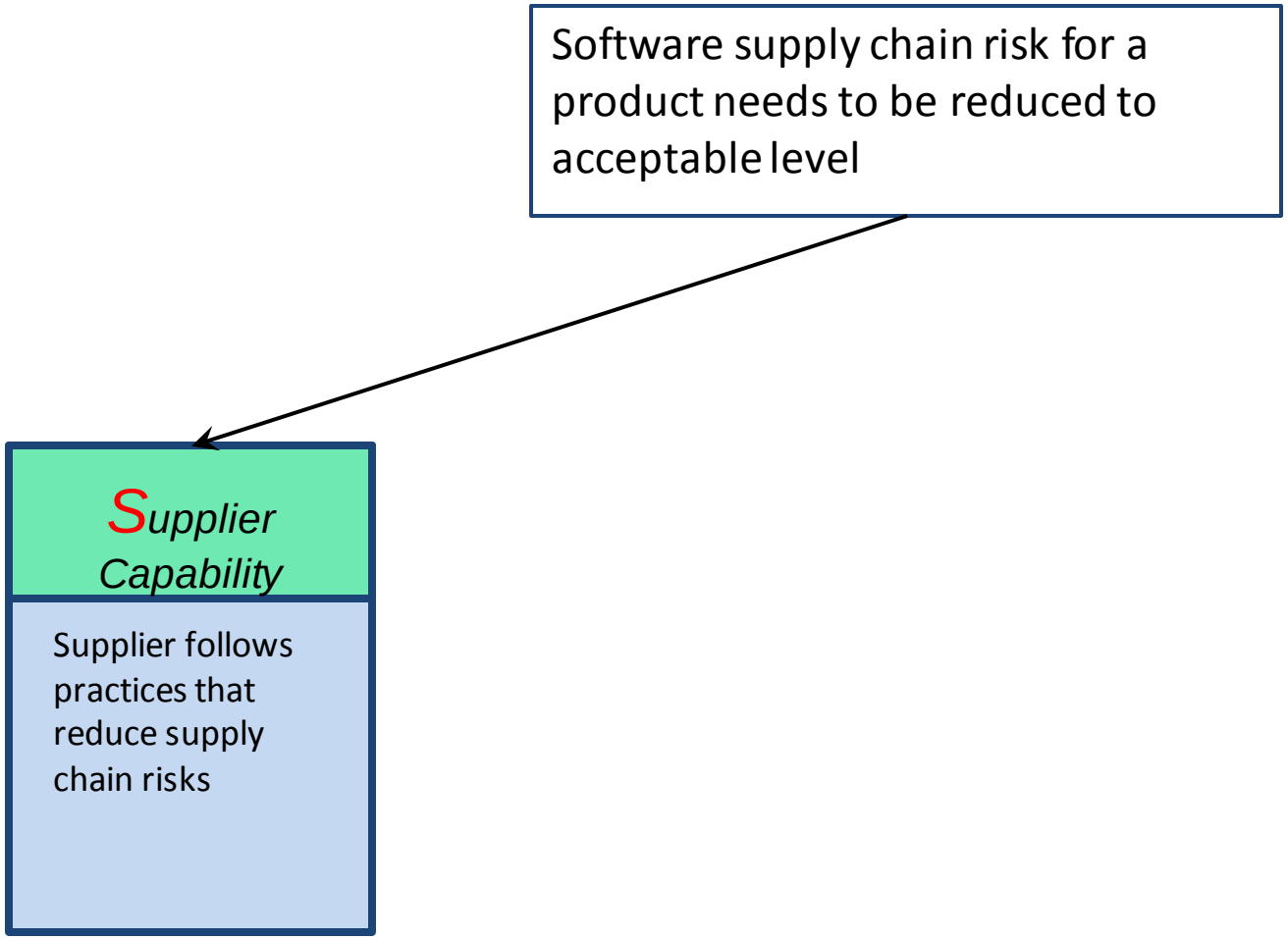


Reusing model layers saves time and resources in training but is a vector for hidden processing

Reducing software supply chain risk factors

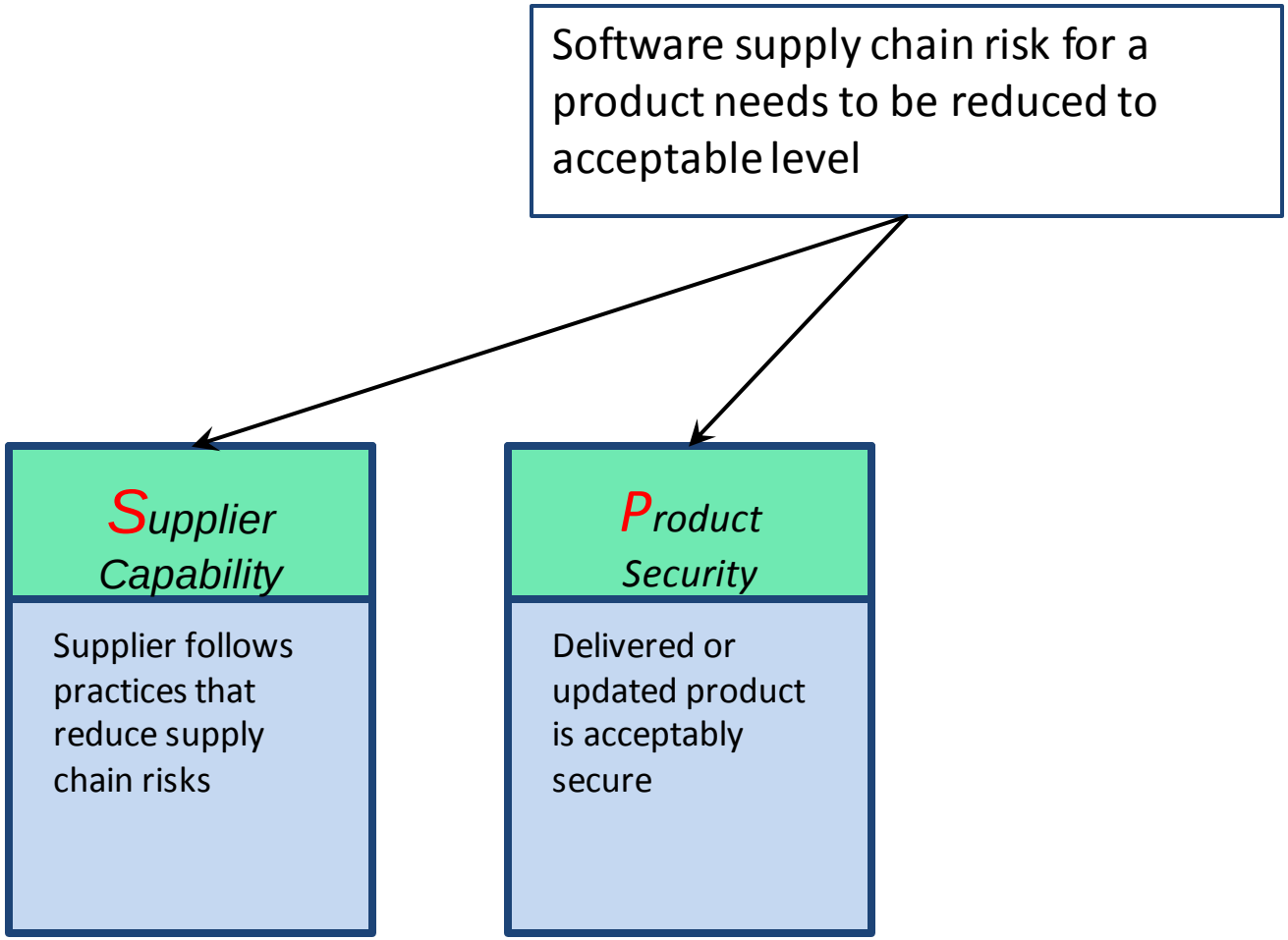


Reducing software supply chain risk factors



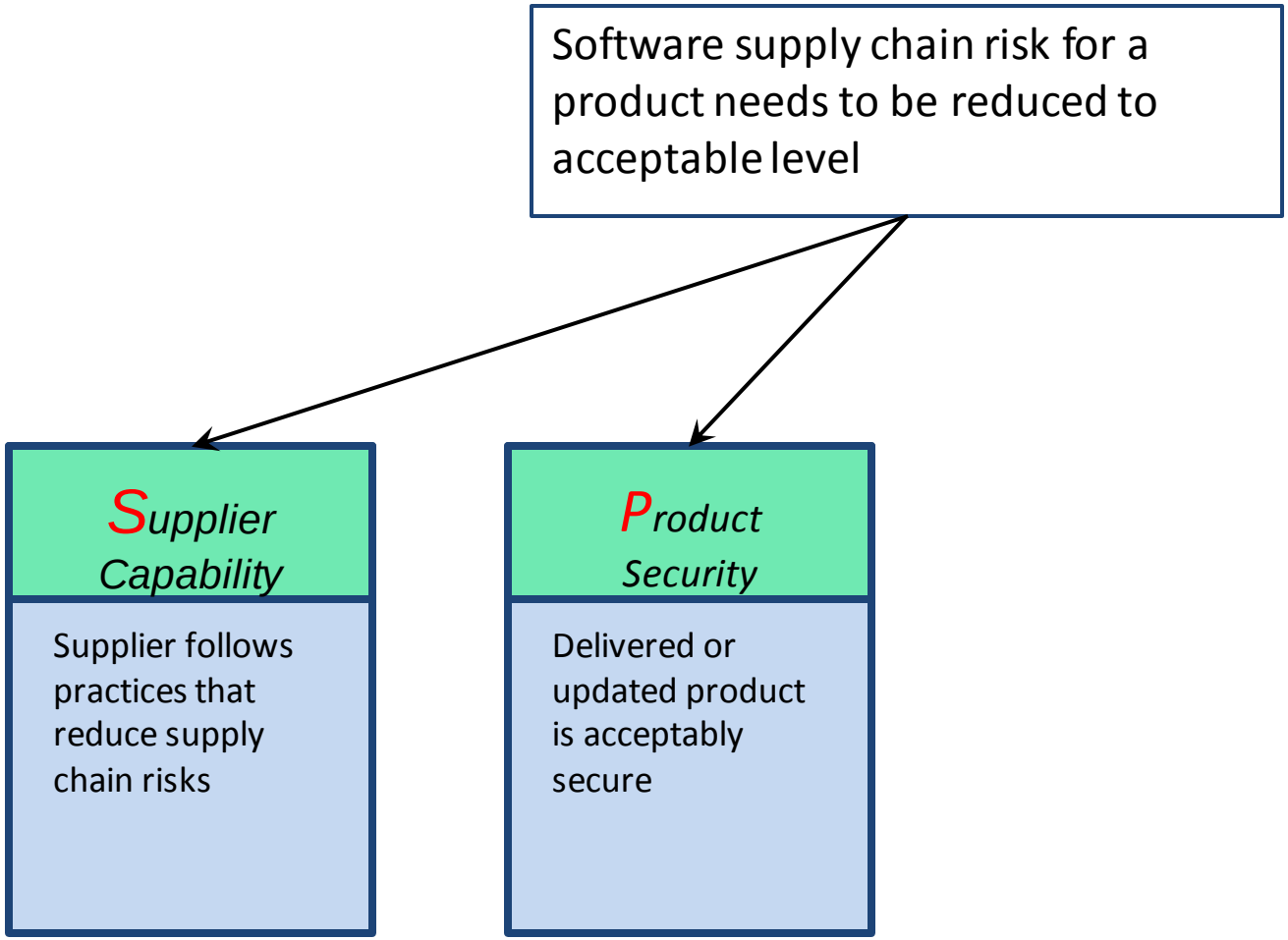
Repeatable Builds

Reducing software supply chain risk factors

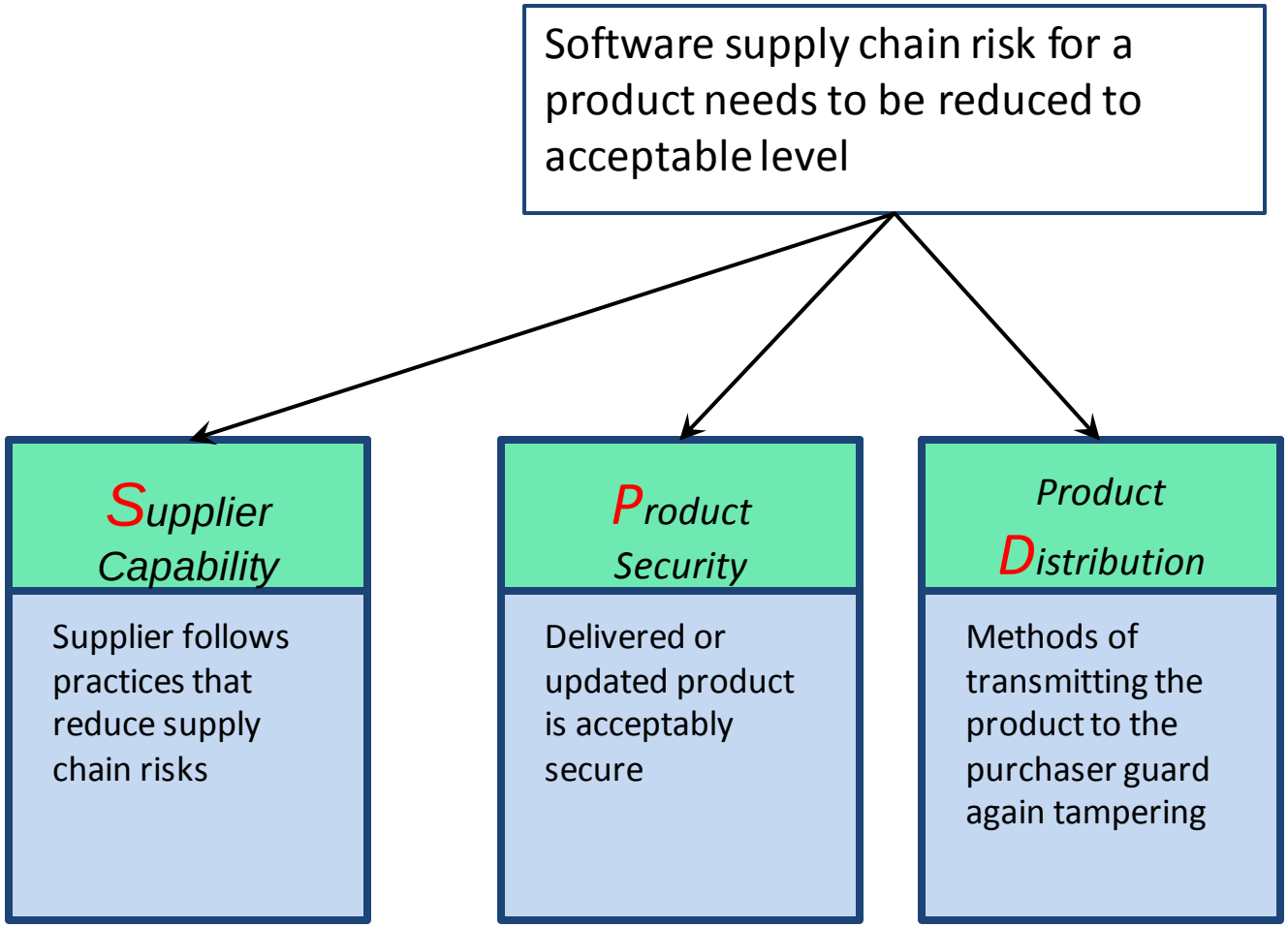


Repeatable Builds

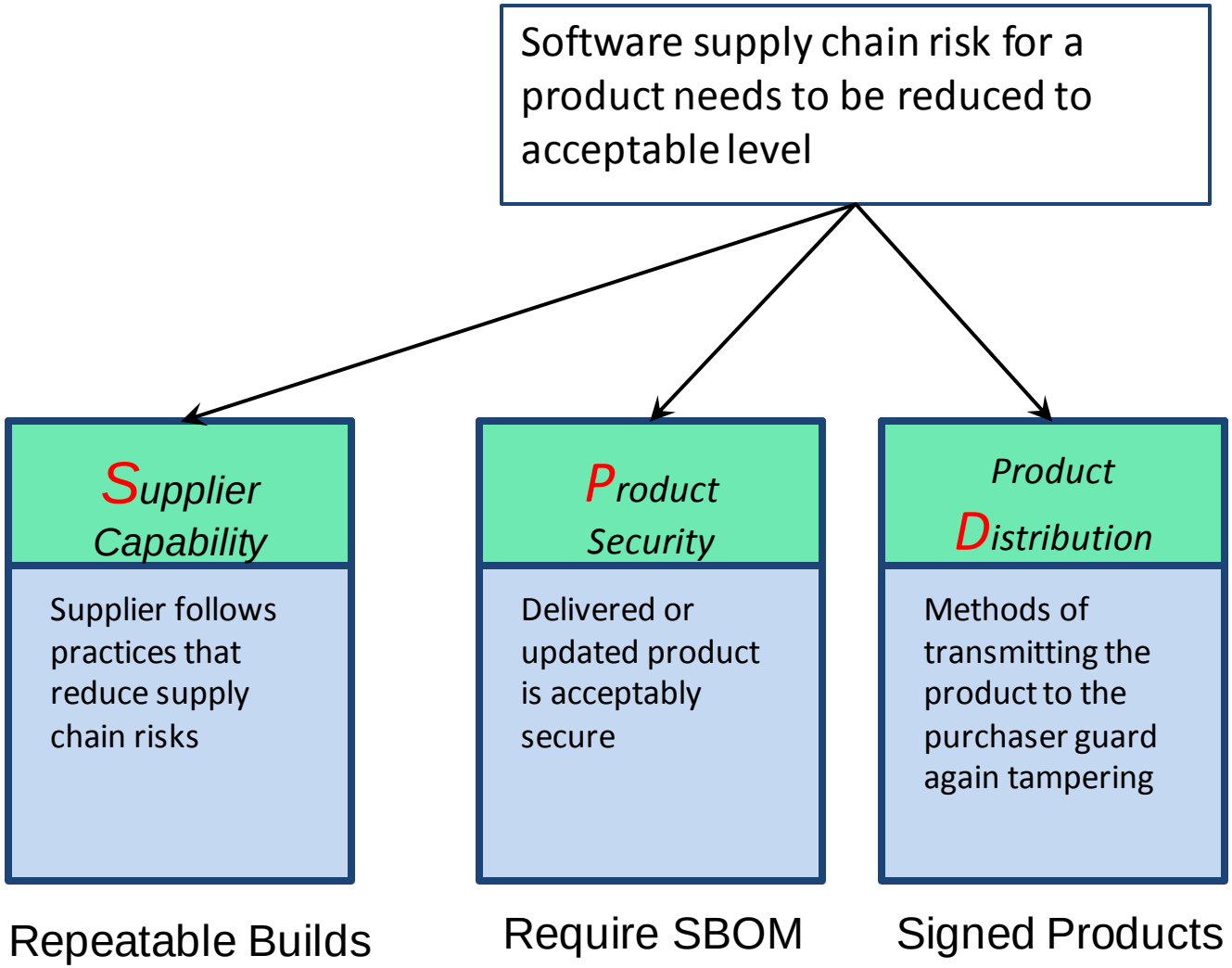
Reducing software supply chain risk factors



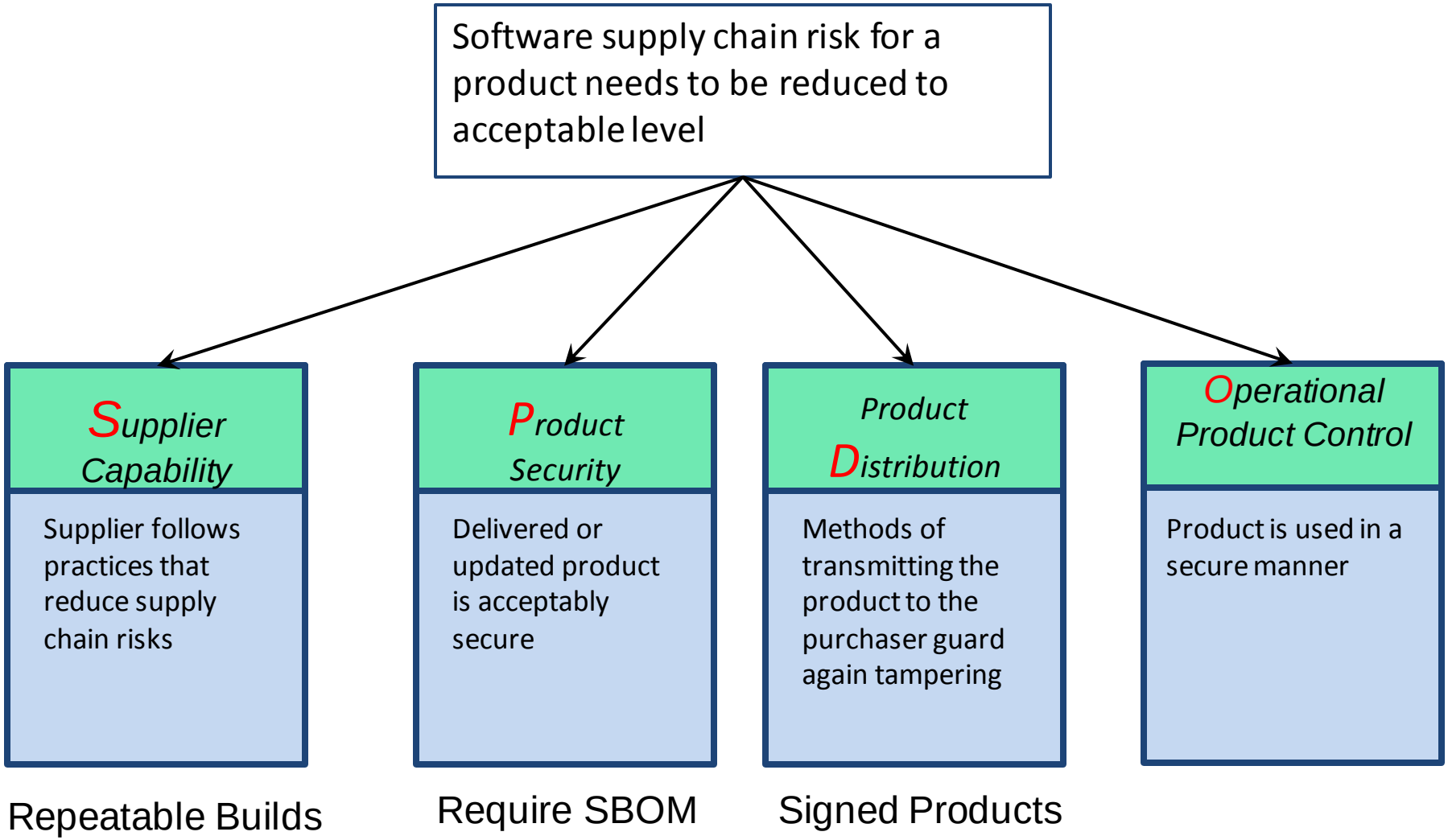
Reducing software supply chain risk factors



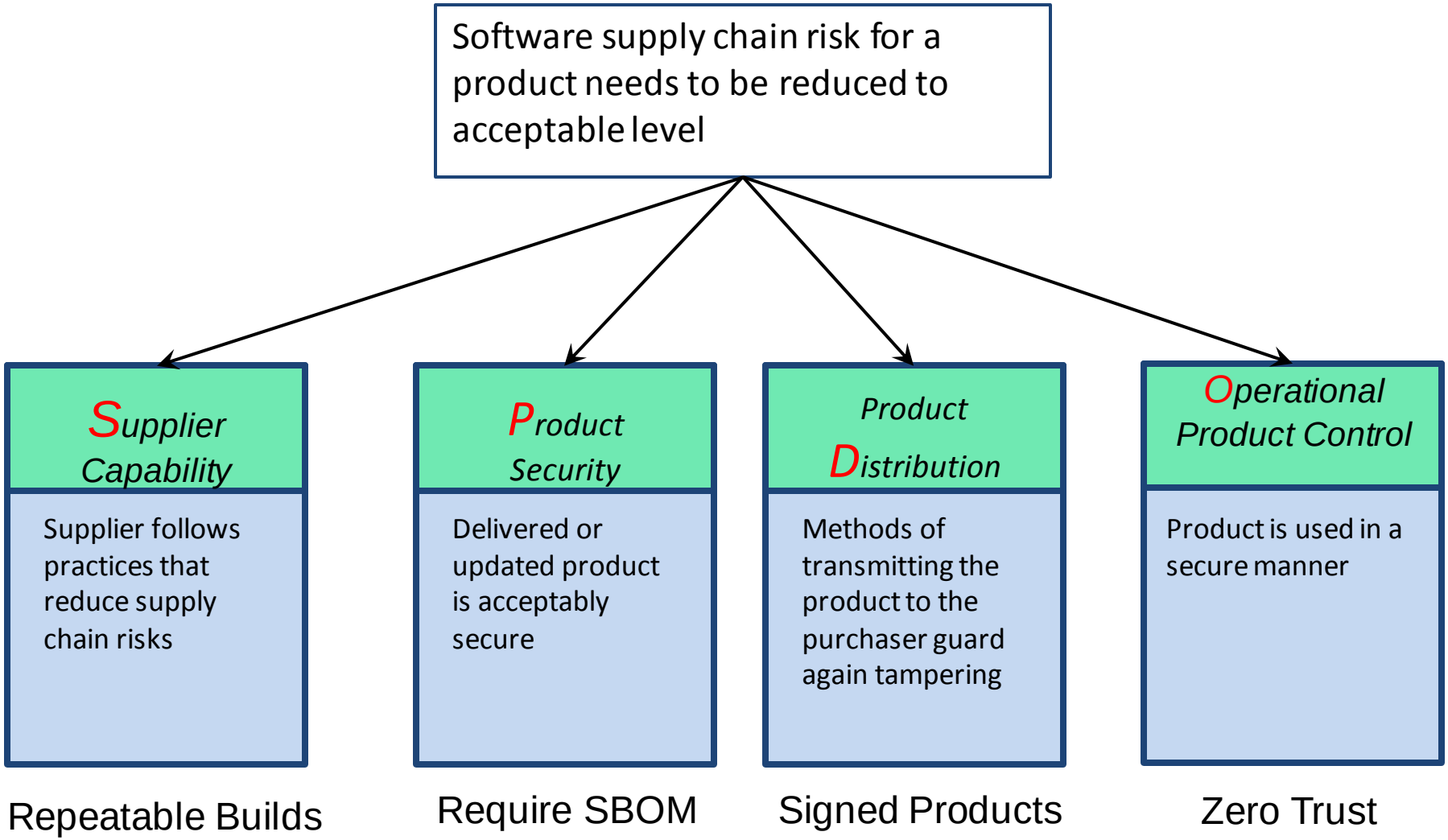
Reducing software supply chain risk factors



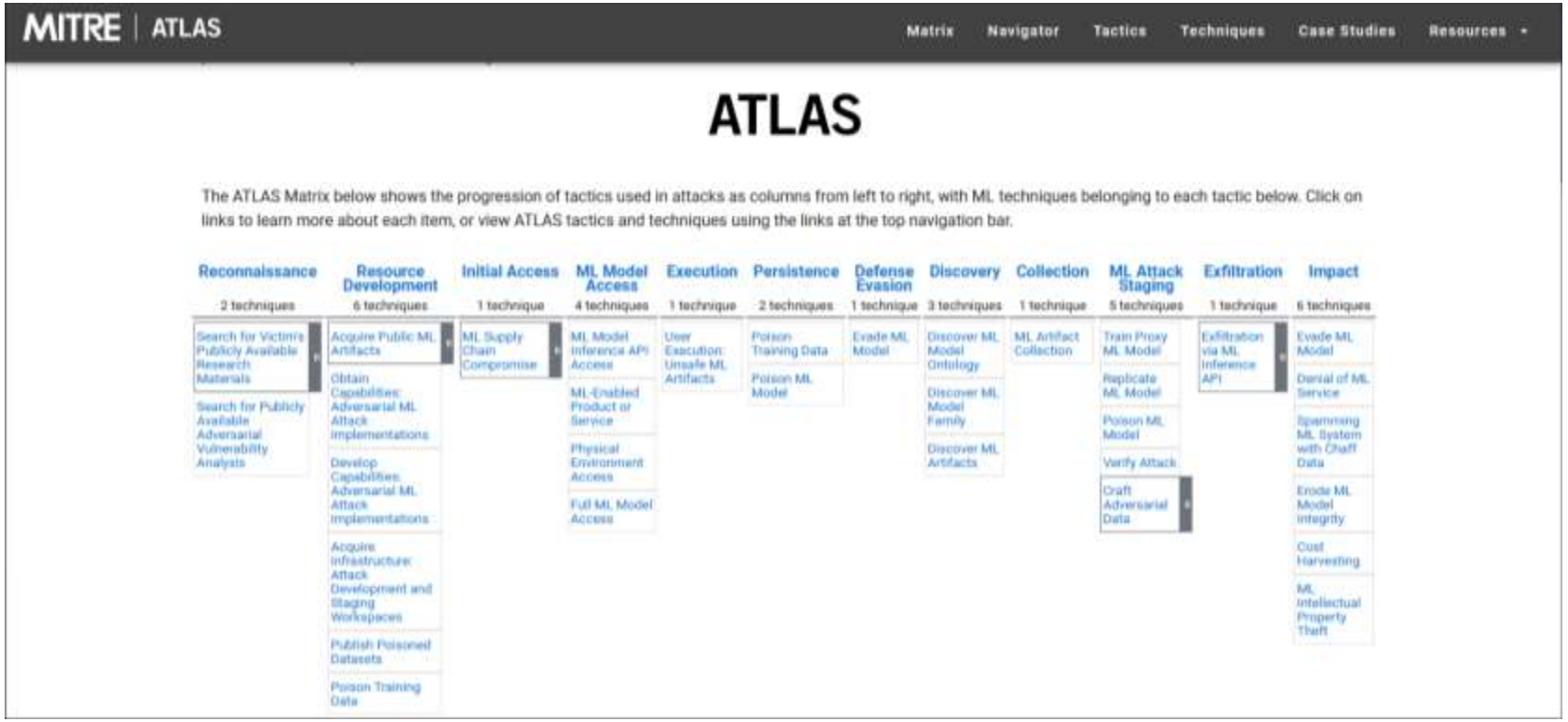
Reducing software supply chain risk factors



Reducing software supply chain risk factors



Additional Attacks on Machine Learning Applications



Summary

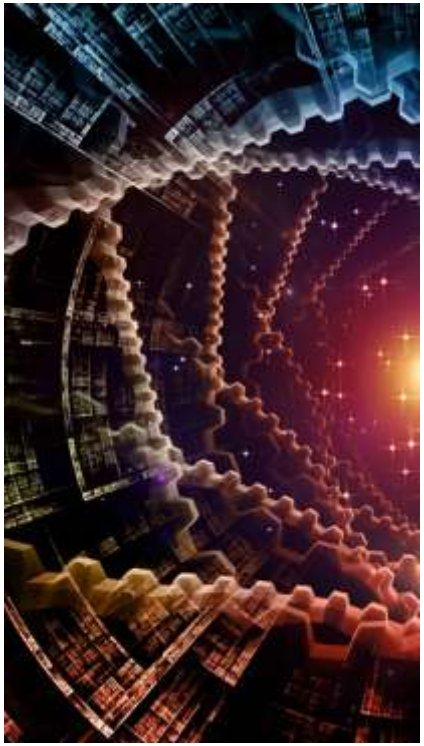
Development Processes for Machine Learning Applications are Complex

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