Dissecting A Research Paper

Homa Alemzadeh

Electrical and Computer Engineering

Link Lab

University of Virginia

Research Papers

- Research is all about communication and critical thinking
  - Reading
  - Writing
  - Presenting
  - Reviewing

- Papers are the primary mechanism for **doing research**
  - Develop, crystallize, and convey ideas
  - Have dialogue with other researchers
  - Get feedback, critique, and collaborate

Components of a Research Paper

First questions to ask when reading, writing, or evaluating a paper:

1. **What is the problem and why is it important?**
   - Motivation, overall objective, and broader impacts (Who cares?)
2. **What are the technical challenges?**
   - Core technical problems to solve
3. **What is the state-of-the-art?**
   - Limitations of the current practice and gaps in research literature
4. **What the paper is proposing to do?**
   - Main contributions (Claims)
5. **How are the proposed methods/hypotheses evaluated?**
   - Evidence to support claims

Structure of a Research Paper

- Title (1000 readers)
- Abstract (4 sentences, 100 readers)
- Introduction (1 page, 100 readers)
- The problem (1 page, 10 readers) → **Problem Statement, Research Questions**
- The main idea (2 pages, 10 readers) → **Methodology**
- The details (5 pages, 3 readers) → **Experiments, Analyses, Results, Findings**
- Related work (1-2 pages, 10 readers)
- Conclusions and further work (0.5 pages)

"How to write a great research paper", Simon P. Jones, Microsoft Research: https://www.microsoft.com/en-us/research/people/simonpj/

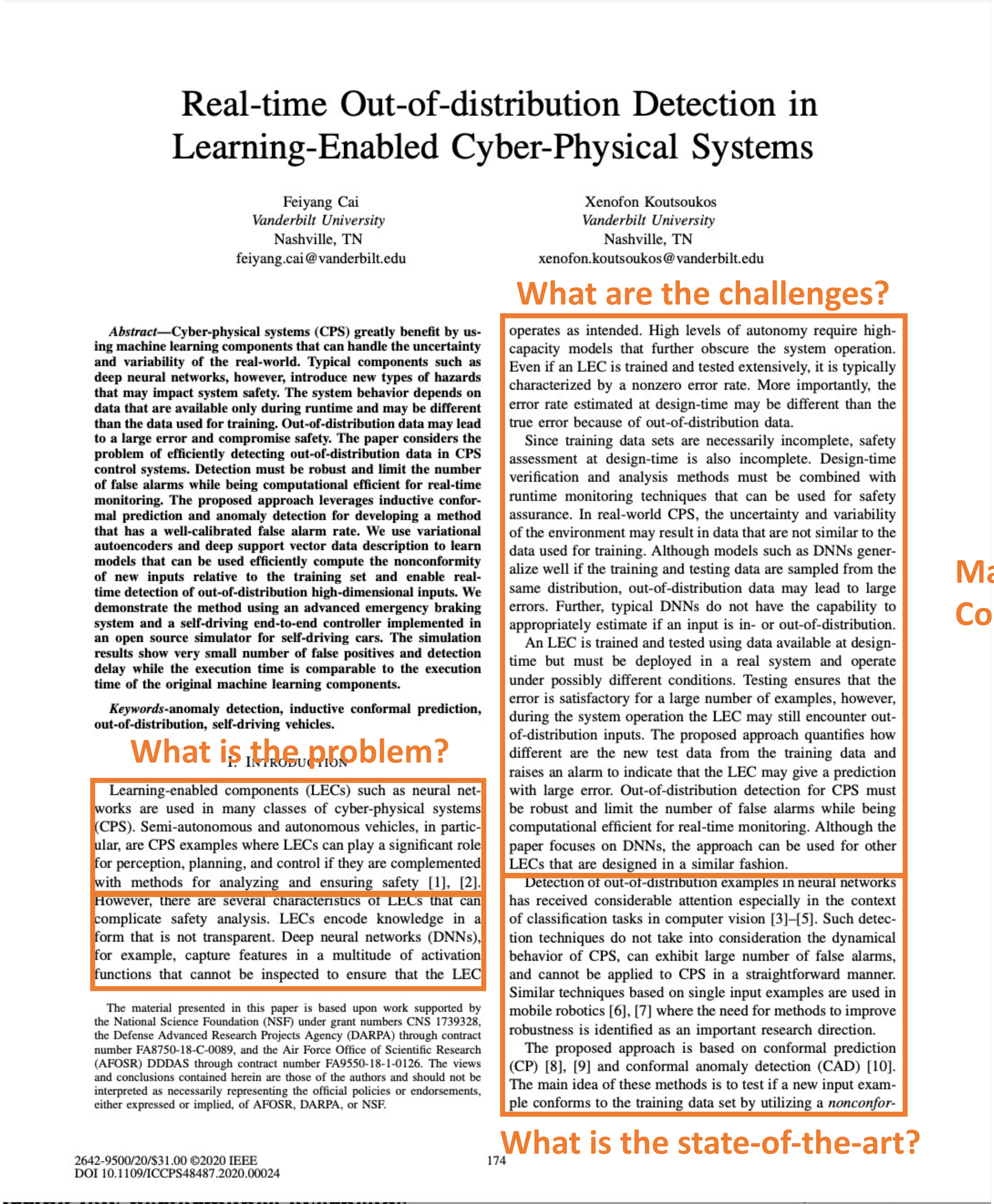
Claims and Evidence

- The **introduction** makes **claims**
- The **body** of the paper provides **evidence** to support each claim
- “Claims” can be about:
  - Hypothesis or research questions
  - New ideas, methods, theories, findings
  - New datasets, testbeds, metrics
- “Evidence” can be:
  - Analysis and comparison
  - Theorems
  - Measurements
  - Case studies
  - Discussion of related work

“How to write a great research paper”, Simon P. Jones, Microsoft Research: https://www.microsoft.com/en-us/research/people/simonpj/

Template Introduction

1. **What is the problem and why is it important? (~1 paragraph)**
    - State the overall objective of the paper
2. **What are the technical challenges? (~1 paragraph)**
    - Describe core technical problems with reference to literature
3. **What is the state-of-the-art? (~1 paragraph)**
    - Limitations and gaps in the literature by reference and evidence
4. **What the paper is proposing to do? (~1 paragraph)**
    - A list of contributions with forward-referenced evidence
5. **How are the proposed methods/hypotheses evaluated? (~1 paragraph)**
    - Summary of experiments, analysis, and important results and findings



Real-time Out-of-distribution Detection in Learning-Enabled Cyber-Physical Systems

Feiyang Cai
Vanderbilt University
Nashville, TN
feiyang.cai@vanderbilt.edu

Xenofon Koutsoukos
Vanderbilt University
Nashville, TN
xenofon.koutsoukos@vanderbilt.edu

***Abstract*—Cyber-physical systems (CPS) greatly benefit by using machine learning components that can handle the uncertainty and variability of the real-world. Typical components such as deep neural networks, however, introduce new types of hazards that may impact system safety. The system behavior depends on data that are available only during runtime and may be different than the data used for training. Out-of-distribution data may lead to a large error and compromise safety. The paper considers the problem of efficiently detecting out-of-distribution data in CPS control systems. Detection must be robust and limit the number of false alarms while being computational efficient for real-time monitoring. The proposed approach leverages inductive conformal prediction and anomaly detection for developing a method that has a well-calibrated false alarm rate. We use variational autoencoders and deep support vector data description to learn models that can be used efficiently compute the nonconformity of new inputs relative to the training set and enable real-time detection of out-of-distribution high-dimensional inputs. We demonstrate the method using an advanced emergency braking system and a self-driving end-to-end controller implemented in an open source simulator for self-driving cars. The simulation results show very small number of false positives and detection delay while the execution time is comparable to the execution time of the original machine learning components.**

***Keywords*-anomaly detection, inductive conformal prediction, out-of-distribution, self-driving vehicles.**

What is the problem?

I. INTRODUCTION

Learning-enabled components (LECs) such as neural networks are used in many classes of cyber-physical systems (CPS). Semi-autonomous and autonomous vehicles, in particular, are CPS examples where LECs can play a significant role for perception, planning, and control if they are complemented with methods for analyzing and ensuring safety [1], [2]. However, there are several characteristics of LECs that can complicate safety analysis. LECs encode knowledge in a form that is not transparent. Deep neural networks (DNNs), for example, capture features in a multitude of activation functions that cannot be inspected to ensure that the LEC

What are the challenges?

operates as intended. High levels of autonomy require high-capacity models that further obscure the system operation. Even if an LEC is trained and tested extensively, it is typically characterized by a nonzero error rate. More importantly, the error rate estimated at design-time may be different than the true error because of out-of-distribution data.

Since training data sets are necessarily incomplete, safety assessment at design-time is also incomplete. Design-time verification and analysis methods must be combined with runtime monitoring techniques that can be used for safety assurance. In real-world CPS, the uncertainty and variability of the environment may result in data that are not similar to the data used for training. Although models such as DNNs generalize well if the training and testing data are sampled from the same distribution, out-of-distribution data may lead to large errors. Further, typical DNNs do not have the capability to appropriately estimate if an input is in- or out-of-distribution.

An LEC is trained and tested using data available at design-time but must be deployed in a real system and operate under possibly different conditions. Testing ensures that the error is satisfactory for a large number of examples, however, during the system operation the LEC may still encounter out-of-distribution inputs. The proposed approach quantifies how different are the new test data from the training data and raises an alarm to indicate that the LEC may give a prediction with large error. Out-of-distribution detection for CPS must be robust and limit the number of false alarms while being computational efficient for real-time monitoring. Although the paper focuses on DNNs, the approach can be used for other LECs that are designed in a similar fashion.

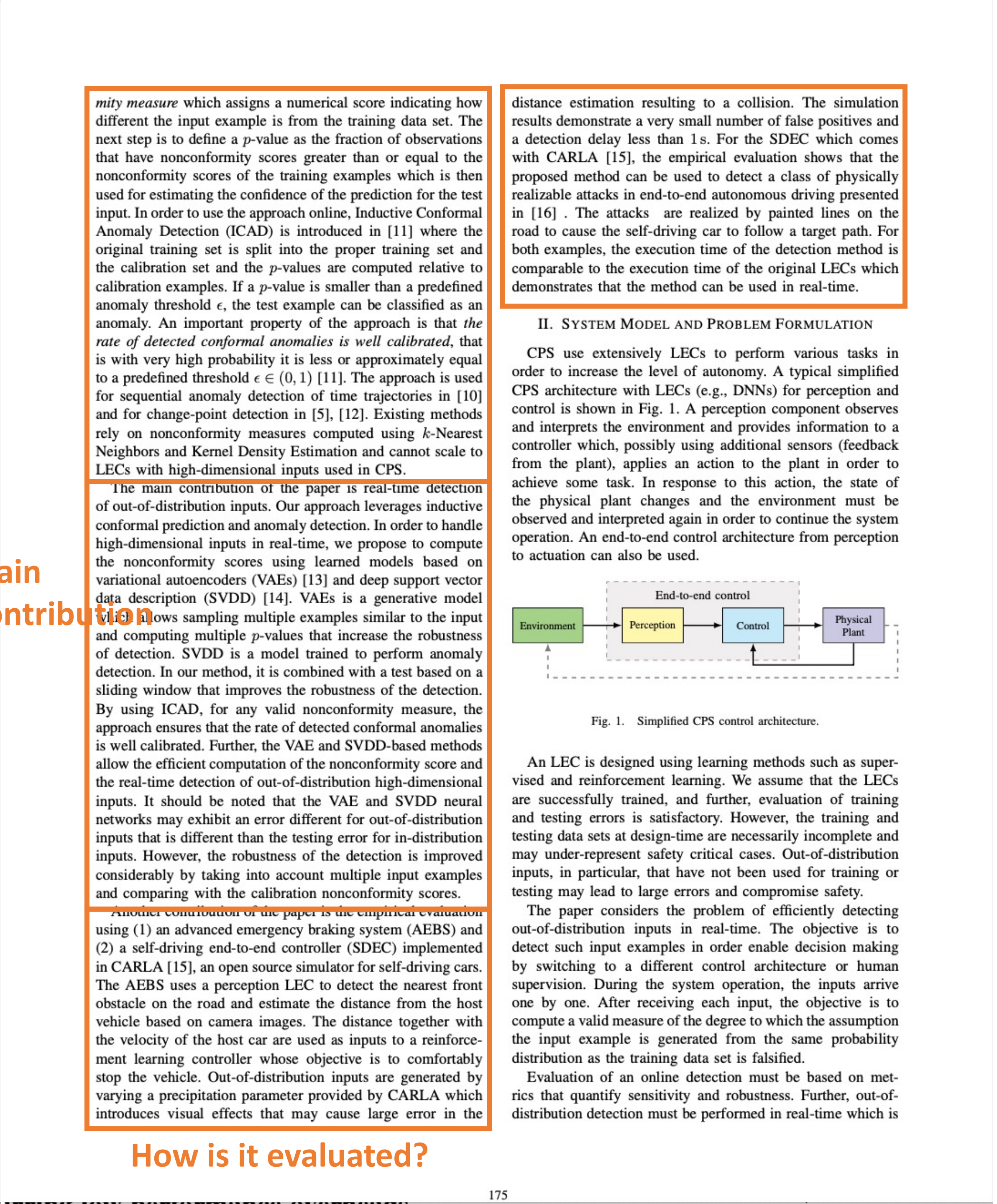
Detection of out-of-distribution examples in neural networks has received considerable attention in the context of classification tasks in computer vision [3]–[5]. Such detection techniques do not take into consideration the dynamical behavior of CPS, can exhibit large number of false alarms, and cannot be applied to CPS in a straightforward manner. Similar techniques based on single input examples are used in mobile robotics [6], [7] where the need for methods to improve robustness is identified as an important research direction.

The proposed approach is based on conformal prediction (CP) [8], [9] and conformal anomaly detection (CAD) [10]. The main idea of these methods is to test if a new input example conforms to the training data set by utilizing a *nonconformity measure* which assigns a numerical score indicating how different the input example is from the training data set. The next step is to define a p-value as the fraction of observations that have nonconformity scores greater than or equal to the nonconformity scores of the training examples which is then used for estimating the confidence of the prediction for the test input. In order to use the approach online, Inductive Conformal Anomaly Detection (ICAD) is introduced in [11] where the original training set is split into the proper training set and the calibration set and the p-values are computed relative to calibration examples. If a p-value is smaller than a predefined anomaly threshold ϵ, the test example can be classified as an anomaly. An important property of the approach is that *the rate of detected conformal anomalies is well calibrated*, that is with very high probability it is less or approximately equal to a predefined threshold $\epsilon \in (0,1)$ [11]. The approach is used for sequential anomaly detection of time trajectories in [10] and for change-point detection in [5], [12]. Existing methods rely on nonconformity measures computed using k-Nearest Neighbors and Kernel Density Estimation and cannot scale to LECs with high-dimensional inputs used in CPS.

What is the state-of-the-art?

The material presented in this paper is based upon work supported by the National Science Foundation (NSF) under grant numbers CNS 1739328, the Defense Advanced Research Projects Agency (DARPA) through contract number FA8750-18-C-0089, and the Air Force Office of Scientific Research (AFOSR) DDDAS through contract number FA9550-18-1-0126. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of AFOSR, DARPA, or NSF.

2642-9500/20/$31.00 ©2020 IEEE
DOI 10.1109/ICCPS48487.2020.00024



Main Contribution

The main contribution of the paper is real-time detection of out-of-distribution inputs. Our approach leverages inductive conformal prediction and anomaly detection. In order to handle high-dimensional inputs in real-time, we propose to compute the nonconformity scores using learned models based on variational autoencoders (VAEs) [13] and deep support vector data description (SVDD) [14]. VAEs is a generative model which allows sampling multiple examples similar to the input and computing multiple p-values that increase the robustness of detection. SVDD is a model trained to perform anomaly detection. In our method, it is combined with a test based on a sliding window that improves the robustness of the detection. By using ICAD, for any valid nonconformity measure, the approach ensures that the rate of detected conformal anomalies is well calibrated. Further, the VAE and SVDD-based methods allow the efficient computation of the nonconformity score and the real-time detection of out-of-distribution high-dimensional inputs. It should be noted that the VAE and SVDD neural networks may exhibit an error different for out-of-distribution inputs that is different than the testing error for in-distribution inputs. However, the robustness of the detection is improved considerably by taking into account multiple input examples and comparing with the calibration nonconformity scores.

Another contribution of the paper is the empirical evaluation using (1) an advanced emergency braking system (AEBS) and (2) a self-driving end-to-end controller (SDEC) implemented in CARLA [15], an open source simulator for self-driving cars. The AEBS uses a perception LEC to detect the nearest front obstacle on the road and estimate the distance from the host vehicle based on camera images. The distance together with the velocity of the host car are used as inputs to a reinforcement learning controller whose objective is to comfortably stop the vehicle. Out-of-distribution inputs are generated by varying a precipitation parameter provided by CARLA which introduces visual effects that may cause large error in the distance estimation resulting to a collision. The simulation results demonstrate a very small number of false positives and a detection delay less than 1 s. For the SDEC which comes with CARLA [15], the empirical evaluation shows that the proposed method can be used to detect a class of physically realizable attacks in end-to-end autonomous driving presented in [16] . The attacks are realized by painted lines on the road to cause the self-driving car to follow a target path. For both examples, the execution time of the detection method is comparable to the execution time of the original LECs which demonstrates that the method can be used in real-time.

How is it evaluated?

II. System Model and Problem Formulation

CPS use extensively LECs to perform various tasks in order to increase the level of autonomy. A typical simplified CPS architecture with LECs (e.g., DNNs) for perception and control is shown in Fig. 1. A perception component observes and interprets the environment and provides information to a controller which, possibly using additional sensors (feedback from the plant), applies an action to the plant in order to achieve some task. In response to this action, the state of the physical plant changes and the environment must be observed and interpreted again in order to continue the system operation. An end-to-end control architecture from perception to actuation can also be used.

Fig. 1. Simplified CPS control architecture.

An LEC is designed using learning methods such as supervised and reinforcement learning. We assume that the LECs are successfully trained, and further, evaluation of training and testing errors is satisfactory. However, the training and testing data sets at design-time are necessarily incomplete and may under-represent safety critical cases. Out-of-distribution inputs, in particular, that have not been used for training or testing may lead to large errors and compromise safety.

The paper considers the problem of efficiently detecting out-of-distribution inputs in real-time. The objective is to detect such input examples in order enable decision making by switching to a different control architecture or human supervision. During the system operation, the inputs arrive one by one. After receiving each input, the objective is to compute a valid measure of the degree to which the assumption the input example is generated from the same probability distribution as the training data set is falsified.

Evaluation of an online detection must be based on metrics that quantify sensitivity and robustness. Further, out-of-distribution detection must be performed in real-time which is

The Three Pass Method

- **The first pass** (~ 5-10 mins)
  - Bird’s eye view to decide on doing more passes or not => Hook the reader/reviewer
  - Get a general idea about the paper’s category, context, correctness, contributions, clarity
  - **Try to find the answer to the 5 key questions.**
- **The second pass** (~ 1 hour)
  - Grasp the paper’s content, but not its details
  - Summarize the main thrust with supporting evidence
  - **Evaluate the 5 components and the clarity of writing**
- **The third pass** (~ 4-5 hours)
  - Full understanding of the paper, specially if you are the reviewer
  - Pinpoint implicit assumptions, missing citations, issues with analytical and experimental methods
  - **Identify the strengths and weaknesses**

“How to Read a Paper,” S. Keshav, University of Waterloo https://web.stanford.edu/class/ee384m/Handouts/HowtoReadPaper.pdf

The Three Pass Method

- **The first pass** (~ 5-10 mins)
  - Carefully read title, abstract, and introduction
  - Read the section and subsection headings and look at figures, tables, illustrations
  - **Try to find the answer to the 5 key questions.**
- **The second pass** (~ 1 hour)
  - Read the paper carefully but ignore details
  - Pay careful attention to identify common mistakes in analysis, results, illustrations
  - **Evaluate the 5 components and the clarity of writing**
- **The third pass** (~ 4-5 hours)
  - Virtually re-implement the paper using the same assumptions (compare virtual vs. actual)
  - Identify and challenge every assumption in every statement
  - **Identify the strengths and weaknesses**

“How to Read a Paper,” S. Keshav, University of Waterloo https://web.stanford.edu/class/ee384m/Handouts/HowtoReadPaper.pdf

Critical Evaluation of Research

- **Intellectual merit**
  - Significance of technical problem and novelty of ideas
  - Potential to advance knowledge and impact the field or other fields
- **Evidence to support claims**
  - Evaluate the evidence provided for each claim
  - Quality and validity of claims and evidence
- **Design of experiments**
  - Validity with respect to the claims and research questions
  - Metrics and success criteria
- **Threats to validity**
  - Assumptions (Both stated and unstated)
  - Limitations (Unanswered and new questions)

Group Activity: 2 Phase Reading

- You are given with five example CPS papers
  - Best paper awardees from ICCPS, IOTDI, IPSN, DSN Conferences 2020-2021
- Do a quick first pass on all
  - Pick one paper based on the category, context, clarity that is interesting to you
  - Answer the key 5 questions
- Do a short second pass on your selected paper
  - Evaluate the 5 components and the clarity of writing
- Present a short summary

References

- “Heilmeier Catechism”, https://medium.com/art-of-the-start/the-heilmeier-catechism-a-recipe-for-managing-innovation-41511be748a
- “You and Your Research”, Richard W. Hamming, https://www.youtube.com/watch?v=a1zDuOPkMSw
- Research Skills, Simon Peyton Jones, https://www.microsoft.com/en-us/research/people/simonpj/
- “How to Read a Paper,” S. Keshav, https://web.stanford.edu/class/ee384m/Handouts/HowtoReadPaper.pdf
- “Dissecting Research Articles,” https://campustools.capella.edu/BBCourse_Production/PhD_Colloquia/Track_1/phd_t1_u02s6_h01_dissect.html
- “Reading Research Papers,” Andrew Ng, https://www.youtube.com/watch?v=733m6qBH-jI&t