Research Article

Sangchul Park and Haksoo Ko*

Machine Learning and Law and Economics: A Preliminary Overview

<https://doi.org/10.1515/ajle-2020-0034>

Published online September 15, 2020

Abstract: This paper provides an overview of machine learning models, as compared to traditional economic models. It also lays out emerging issues in law and economics that the machine learning methodology raises. In doing so, Asian contexts are considered. Law and economics scholarship has applied econometric models for statistical inferences, but law as social engineering often requires forward-looking predictions rather than retrospective inferences. Machine learning can be used as an alternative or supplementary tool to improve the accuracy of legal prediction by controlling out-of-sample variance along with in-sample bias and by fitting diverse models to data with non-linear or otherwise complex distribution. In the legal arena, the past experience of using economic models in antitrust and other high-stakes litigation provides a clue as to how to introduce artificial intelligence into the legal decision-making process. Law and economics is also expected to provide useful insights as to how to balance the development of the artificial intelligence technology with fundamental social values such as human rights and autonomy.

Keywords: machine learning, artificial intelligence, natural language processing, algorithmic transparency, algorithmic fairness

1 Introduction

This paper provides an overview of machine learning (ML) models, as compared to traditional economic models. It also lays out new issues in law and economics that the ML methodology raises. In doing so, the Asian context is considered.

---

*Corresponding author: Haksoo Ko, Seoul National University, Seoul, The Republic of Korea,
E-mail: hsk@snu.ac.kr

Sangchul Park, Seoul National University, Seoul, The Republic of Korea

Open Access. © 2020 Sangchul Park and Haksoo Ko, published by De Gruyter. This work is licensed under the Creative Commons Attribution 4.0 International License.

ML is inherently a type of artificial intelligence (AI) that learns ‘by itself’ or ‘without being explicitly programmed.’¹ Following a paradigm shift from a rule-based, deductive approach of AI (such as expert systems)² to a data-driven, inductive approach (such as ML) around the 1990s, ML has recently become the prevailing form of AI. The term AI is thus often used interchangeably with ML in this paper; although AI is in general a broader concept than ML, we do not necessarily make a clear distinction in this paper other than that *ML learns from data*.³ This implies that an ML model is basically *a statistical model*. For an economist, a regression model (in particular, a logit model if the target variable is discrete and binary) could serve as a starting point for supervised learning. Empirical law and economics scholarship, which has relied mostly on regression models to test legal hypotheses, is thus better poised to adopt ML methodologies.

The paper is organized as follows. Section 2 discusses why law and economics scholarship should embrace ML models, in particular for future predictions; and how various ML algorithms can be deployed for law and economics studies. Section 3 provides an overview of newly emerging issues in AI and law in an Asian context. They include systematizing judicial decision-making with AI; addressing new problems arising from the algorithmic society; and facilitating the development and use of AI. Section 4 concludes.

2 Key Ideas of ML as Compared to Economic Models

2.1 Limitations of Traditional Econometrics Models: An Example of Logit

Law is social engineering (Pound 1954) and thus often requires forward-looking prediction. The law and economics literature has tried to apply traditional

---

¹ This definition of machine learning was coined in 1959 by Arthur L. Samuel, one of the pioneers of AI.

² Under the old paradigm, there had been several notable attempts to apply the rule-based system to legal problem-solving. Buchanan and Headrick (1970) identified four major legal problem-solving processes including finding conceptual linkage in pursuing goals; recognizing facts; resolving rule conflicts; and finding analogies, and reviewed the development of relevant computer systems applicable to each process, focusing on a program called Heuristic DENTRAL. McCarty (1977) presented the outcome of an experiment in AI and legal reasoning by utilizing the TAXMAN program, which was designed to provide advice on taxation in the context of corporate reorganization.

³ As an exception, a reinforcement learning model learns from processes rather than from preexisting data. However, to the extent that data are generated through an agent’s choice of actions based on states and rewards, it would not be far-fetched to say the model eventually learns from the data so generated.

econometric methods for the ‘prediction’ of future legal affairs or events at times. Econometrics is, however, optimized for *the inference of relationships between variables* (β ; coefficients). Doing so is typically achieved by way of minimizing *in-sample biases* through the application of ordinary least square (OLS) and other methodologies (Kleinberg et al. 2015, 492). Its problem is that the mean squared error (MSE), which is indicative of the quality of a predictor, is mathematically decomposed not only into an irreducible error or in-sample bias, but also into out-of-sample variances. In fact, econometrics may be ill-suited for *the prediction of future outcomes* ($\hat{y}$), because it does not control *out-of-sample variances* (Ibid, 492–3). The most fundamental dilemma that a predictive model may have to face is *the bias-variance tradeoff*, which means that techniques deployed to reduce in-sample biases may result in an increase in out-of-sample variances, and vice versa (Ibid, 492–3). Econometric models, which are fitted to minimize in-sample biases, are thus prone to *an overfitting problem*. The problem arises when the ‘production of an analysis which corresponds too closely or exactly to a particular set of data’ and, as a result, a model may ‘fail to fit additional data or predict future observations reliably.’⁴

To illustrate, assume that, in around 2000, a bankruptcy court in an Asian country tried to build a model for predicting the outcome of corporate reorganization proceedings (liquidation or emergence) based on the data for a 10-year period until 1999. Specifically, the available data has only two features: debt-to-equity ratios and net profit margins. The scatter and decision boundary plots below show that a traditional logit model is fitted well if outliers that were generated during the 1997 Asian Financial Crisis (points (3.5, 3) and (3.4, 3.1)) are not put into the dataset, but is slightly overfitted to the outliers if they are included in the dataset (see Figure 1).

2.2 Three Essential Techniques of ML: Train-Test Cycle, Regularization, and Cross-Validation

Due to the overfitting problem, the fitted model above may not be able to produce accurate predictions regarding the outcome of reorganization proceedings that take place after the end of the Financial Crisis. But how can we really determine whether the overfitted model above would not be able to predict accurately, without obtaining future data? To ensure *predictive accuracy*, the whole dataset needs to be split into a train set and a test set. That way, the model can first be fitted to the train set and then tested on the test set for accuracy. This is called *a train-test*

---

⁴ Definition of ‘overfitting’ at lexico.com (Oxford 2020), <https://www.lexico.com/definition/overfitting>.

Figure 1: Distribution of Data.

cycle. To illustrate, from the above example of bankruptcy proceeding, we randomly extract 28 train data (70%) from the 40 data (Figure 2), leaving the remaining 12 data (30%) for testing (Figure 3). The models fitted on the 28 train data are shown in Figure 2.

Figure 2 shows more intuitively that, to predict accurately, it is not enough to minimize in-sample biases only. Rather, a further measure needs to be taken in order to control out-of-sample variances by penalizing outliers. This is called *regularization*. Recall that a (two-dimensional) logit model $h_{\theta}(x) = \frac{1}{1+e^{-(\theta_0+\theta_1x_0+\theta_2x_1)}}$ where $\theta = (\theta_0, \theta_1, \theta_2)$ and $x = (x_0, x_1)$ is fitted by maximizing the log likelihood that the training set appears at a given θ . As such, $\hat{\theta}$ is obtained by solving:

$$\hat{\theta} \equiv \operatorname{argmin}_{\theta} \left(- \sum_{i=1}^N y^{(i)} \log h_{\theta}(x^{(i)}) + (1 - y^{(i)}) \log (1 - h_{\theta}(x^{(i)})) \right)$$

To control variance, either of the following two types of regularizers is commonly added:

$$\hat{\theta} \equiv \operatorname{argmin}_{\theta} \left(- \sum_{i=1}^N (y^{(i)} \log h_{\theta}(x^{(i)}) + (1 - y^{(i)}) \log (1 - h_{\theta}(x^{(i)}))) + C \|\theta\|_1 \right)$$

$$\hat{\theta} \equiv \operatorname{argmin}_{\theta} \left(- \sum_{i=1}^N (y^{(i)} \log h_{\theta}(x^{(i)}) + (1 - y^{(i)}) \log (1 - h_{\theta}(x^{(i)}))) + C \|\theta\|_2^2 \right)$$

The first one, which uses $C \|\theta\|_1 = C(|\theta_0| + |\theta_1| + |\theta_2|)$ as a penalty for outliers, is called the lasso regression (L_1 regularization). The second one, which uses

Figure 2: Models Fitted on the Train Set.

$C||\theta||_2^2 = C(\theta_0^2 + \theta_1^2 + \theta_2^2)$ as a penalty for outliers, is called the ridge regression (L_2 regularization). These regularizers make the model less fitted to outliers and thus better reflect the underlying logic of the data.

The parameter C is optimized so that it minimizes training errors. To that end, *cross-validation* is implemented. Here, k -fold cross-validation is applied with $k = 4$. The train dataset is randomly split into four disjoint subsets (having seven samples), and for each of the disjoint subsets (which is called a validation set), training

Figure 3: Data Reserved for the Test Set.

is done on all the train data except for the validation set and test them on the validation set to get the validation error. This process is repeated until an optimal C is reached, which minimizes average validation errors. The best fitted lasso and ridge models (at optimal C 's) are as shown in Figure 4.

The plot in the figure indicates that the ridge regularizer places a heavy penalty on the outliers (points (3.5, 3) and (3.4, 3.1)) to the extent that they are almost suppressed, whereas the lasso regularizer plays a limited role in controlling variance.

The final step is to compute *predictive accuracy*. By matching the fitted curves in the above example with 12 train sets, we get the following outcome (see Figure 5).

We get 58.33% test accuracy (seven correct predictions) for (unregularized) logit; 66.67% (eight correct predictions) for lasso; and 83.33% (10 correct predictions) for ridge. Note that regularization methods including lasso and ridge do not necessarily work in a way that improves predictive accuracy. In fact, ML developers go through heuristic processes to find out best fitted models and adjustments for a given dataset. Despite the limitations, this illustrates how the regularization methods, coupled with train-test split and cross-validation, can help enhance predictive accuracy by controlling variances.

2.3 Various ML Algorithms

As noted above, logit models have often been used for the empirical law and economics literature to answer discreet legal questions such as win or lose, guilty

Figure 4: Lasso and Ridge Models Best Fitted on the Train Set.

Figure 5: Measuring Predictive Accuracy on the Test Set.

or innocent, and liable or non-liable. Recently, however, with the development of ML methodologies, a variety of non-linear ML models have become available, providing enhanced prediction capabilities from datasets with complicated non-linear patterns.

2.3.1 Supervised Learning

A supervised learning model makes predictions based on a training sample $(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots (x^{(N)}, y^{(N)})$ of previously solved cases, where the joint values of all of the variables are known (Hastie, Tibshirani, and Friedman 2009, 485). A metaphor of ‘learning with a teacher’ can be used to explain the underlying mechanism (Ibid, 485). In this metaphor, the student ‘presents an answer $y^{(i)}$ for each $x^{(i)}$ in the training sample,’ and the teacher then ‘provides either the correct answer and/or an error associated with the student’s answer’ (Ibid, 485). Here, the error is characterized by a loss function $L(y, \hat{y})$, which is to be minimized to approximate the answer (Ibid, 485). An example of the loss function includes $L(y, \hat{y}) = (y - \hat{y})^2$, which is used under the method of least squares. To formalize, supervised learning tries to discover a function h that approximates the true function f , given a training set of N example input-output pairs $(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots (x^{(N)}, y^{(N)})$ where each $y^{(i)}$ was generated by an unknown function $y = f(x)$ (Russell and Norvig 2010, 695). In terms of the underlying statistical logic, supervised learning is not much different from conventional methodologies employed for empirical law and economics.

2.3.1.1 Logit and Softmax

Logit is still widely used as an ML classifier, if the target variable is discrete and binary ($y^{(i)} \in \{0, 1\}$). To control variance, logit with L_1 regularization (lasso) or L_2 regularization (ridge) may be used, as we have seen above. If the target variable is discrete but non-binary (for example, $y^{(i)} \in \{\text{“dog”}, \text{“cat”}, \text{“deer”}\}$ in an image recognition model), *softmax* can be used.

2.3.1.2 Support Vector Machine (SVM) and the Kernel Trick

SVM is considered to be among the best off-the-shelf supervised learning algorithms (Ng 2018). The intuition behind SVM is simple: it separates two groups of data points by drawing the best borderline between them (Ibid). More accurately, SVM classifies data by finding the ‘best hyperplane (or boundary)’ that separates data points of different classes, where the best hyperplane means a hyperplane with the largest margin between the two classes (Ibid).

As an illustration, suppose that, in 20 precedent cases in a training dataset, courts decided whether a certain copper pipe product conforms to the buyer’s requirements and that the courts’ decisions served as a basis for the buyer’s right to reject non-conforming goods. Suppose further that the deviation of the diameter and thickness of each product from the buyer’s requirements, in percentage terms, is distributed as shown in Figure 6.

A judge would wish to derive a consistent test regarding non-conforming products from these data compiled from precedents. An SVM model generally fits well with such legal line-drawing. SVM produces a separating hyperplane in the following steps. First, unlike a logit model, where 0 or 1 are assigned to each category, SVM starts with assigning -1 and 1 to each category (Ibid). Thus, in the above hypothetical example, the products that are judged to be conforming to the buyer’s requirements are assigned 1 , and those judged to be non-conforming are assigned -1 . We can then compile a training set $\{(x^{(i)}, y^{(i)}) ; i = 1, \dots, N\}$, where $y^{(i)} = 1$ when $x^{(i)}$ is a conforming product, and $y^{(i)} = -1$ when $x^{(i)}$ is a non-conforming product. Then, SVM’s job is, given such a training set, to get a decision boundary $w^T x + b = 0$ (where w is a parameter vector, and b is an intersection), which maximizes the margin (or distance) between the decision boundary and the nearest points (among $x^{(i)}$) (Ibid). The decision boundary so calculated, which is in fact a line, is plotted in Figure 7. However, we find that this linear decision boundary does not separate two groups well, as it is severely underfitted. To make the decision boundary better fit with the distribution of data, SVM uses a ‘kernel trick.’ Mapping data to new features through the Gaussian kernel $K(x, z) = e^{-\frac{\|x-z\|^2}{2\sigma^2}}$ before optimization, SVM produces the decision boundary as shown in Figure 8.

Figure 6: Distribution of a Training Dataset.

This decision boundary can serve as a basis upon which the judge draws a line between ‘conforming’ and ‘non-conforming’ cases.

SVM had been, since its development in 1994, widely recognized as the best performer for multiple purposes among various ML techniques, until deep learning made a spectacular revival around 2004 (Ibid). A drawback for SVM, however, would be the costs associated with the burdensome computation (Ibid). On the other hand, SVM is well suited to handle multi-dimensional calculations. Considering that many legal doctrines require multi-factor tests and sometimes utilize not clearly defined concepts such as the ‘totality of circumstances,’ SVM could prove to be exceedingly useful.

2.3.1.3 Decision Trees and Ensemble Methods (Bagging and Boosting)

To directly produce non-linear hypothesis function (without using, for instance, the kernel trick), a decision tree model is sometimes used. In a *decision tree*, each internal node, branch, and leaf node represents a test on an attribute, its outcome, and the final decision, respectively.

Ensemble methods combine several weak learners to get an effect of having a complex model. They are called a strong learner or ensemble model. The methods work particularly well with decision trees. *Bagging (bootstrap aggregation)* has weak learners learn in parallel and combines them under the majority voting or other averaging processes. Random forest is one of the most commonly used bagging algorithms. *Boosting* learns weak learners sequentially. That is, a

Figure 7: Linear SVM Without Kernelization.

subsequent weak learner learns from the output of the previous weak learner and combines them under a preset strategy.

2.3.2 Unsupervised Learning

If supervised learning is similar to ‘learning with a teacher,’ *unsupervised learning* is analogous to ‘learning without a teacher’ (Hastie, Tibshirani, and Friedman 2009, 486). We sometimes need to categorize (or, ‘cluster’) items into one or more groups based on the difference (or, more formally, ‘distance’) among these items without being told a standard for determining such difference (Ibid, 486). Whereas supervised learning works based on the premise that there is a clear measure of success or failure (or, more precisely, expected loss over the joint distribution $P(X, Y)$), there is no such a measure in unsupervised learning (Ibid, 486). As such, heuristic judgments are made in order to assess the quality of the result (Ibid, 486–7). To formalize, unsupervised learning aims to directly infer the properties of the probability density $P(X)$ of observations $(x^{(1)}, x^{(2)}, \dots, x^{(N)})$ in X (Ibid, 486).

Since, in many cases, a legal judgment eventually leads to a yes or no determination, the usefulness of clustering techniques of unsupervised learning for law and economics might be limited. These techniques could, however, be usefully

Figure 8: SVM with Gaussian Kernel.

deployed for certain specialized purposes. For instance, the Principal Component Analysis is widely used for purposes of preprocessing datasets for dimensionality reduction before running a regression model.

2.3.3 Reinforcement Learning

The problem of supervised learning is that it could be difficult and sometimes unwieldy to provide explicit supervision for sequential decision-making and control problems (Ng 2018). *Reinforcement learning* is useful in overcoming such a problem. In order to do so, reinforcement learning uses observed rewards, instead of preexisting data, to learn an optimal or nearly optimal policy for the environment (Russell and Norvig 2010, 830). Ng (2018) illustrates this using a four-legged robot as an example: a programmer would like it to walk but it is all but impossible to use supervised learning to supervise its behavior and to make it walk (Ibid). In such circumstances, a reward function can be used. That is, the programmer can provide the four-legged robot with a walking algorithm in the form of a reward function. This algorithm would tell the learning agent which behavior is desirable or undesirable, and then the agent will choose its action over time for enhanced rewards through a trial and error process (Ibid).

Lots of contemporary reinforcement learning algorithms are modeled as a Markov Decision Process, a discrete-time state-transition system which finds an optimal policy that maximizes the expected value of the total discounted rewards (Ibid). Under this process, an agent is thrown into an environment, continues to perceive states from the environment, and takes actions based on the states, in turn affecting the environment (Ibid). The agent takes actions without any built-in or explicit strategy. The agent first explores the environment by making random decisions based on, for instance, a brute force algorithm. Yet it repeats trials and errors, and the reward function maps the agent's actions and environment to payoffs. The agent continues to choose its actions over time for large rewards through a repeated game process. Reinforcement learning is particularly well suited with a game that is played within a closed environment. As such, it was only natural that reinforcement learning was effectively applied to the game of Go, beating world-class human Go masters.

From this explanation, law and economics scholars could realize that reinforcement learning is more akin to *agent-based simulation* than to conventional empirical analysis. In order to gain more useful law and economics insights, perhaps more attention should be paid to *multi-agent reinforcement learning (MARL)*, which has the potential for significantly improving the prediction of multiple agents' strategic behaviors, in particular in the game theory context.

2.3.4 Deep Learning

Deep learning refers to an ML methodology that learns from a hierarchical representation of data. It is a technique that stacks an interconnected group of nodes. To illustrate how it works, suppose there is an ML model which determines whether a particular use of a copyrighted material constitutes fair use under copyright law.

In Figure 9, each node represents a neuron and each arrow represents a connection from the output of a neuron to the input of another. In this hypothetical example, fair use ultimately depends on four derived features: substantiality, effect, purpose of use, and nature of work (see Ng 2018). This supposes that we already have the insight that these features may determine fair use under copyright law (Ibid). Yet a surprising aspect of deep learning is that we only need to know the input features x and the output y . Neural networks will, through a process called 'end-to-end learning,' figure out what would be in the middle by itself (Ibid). In Figure 9, five input features are connected to four hidden internal neurons. These five features are: the degree of similarity, change in the frequency of use, commercial? art?, and fictional?. These hidden neurons are connected to the output

Figure 9: An Imaginary Deep Learning Model Predicting Fair Use.

layer which outputs whether the use of copyrighted work constitutes fair use (1) or not (0). The goal of the deep learning model is to automatically determine the hidden features such that they can make a prediction about fair use and, in order to do so, we only need to have a sufficient number of training examples $(x^{(i)}, y^{(i)})$ (Ibid). Every junction between the layers has a parameter (or weight) which constitutes an element of the vectors of weights W , and the activation function $g(z)$ (in most cases, $g(z) = \max(z, 0)$ (ReLU function) is used for hidden layers) converts the weighted sum $z = W^T x$ to the values to be sent to the next layers ($a = g(z)$) (Ibid). We place training examples into the neural network one by one and compute the losses of the neural network based on the difference between the predicted output $\hat{y}^{(i)}$ and actual $y^{(i)}$ (1 if courts recognized fair use in past cases, and 0 otherwise). After the final loss of the neural network is computed, the chain rule is recursively applied to compute gradients all the way back to the inputs and to update the weights in a manner that the loss is minimized and the neural network thus fits the data best (Ibid). This process is called *backpropagation*. Once the model is trained, it is tested on the test set to measure predictive accuracy.

Due to the difficulties in understanding the underlying features that deep learning models have created, they are often called a black box (Ibid). So, while deep learning has produced numerous promising and exhilarating results, the opaqueness and inexplicability of deep learning algorithms have raised concerns that the algorithms, if applied to affect legitimate human interests, may undermine human autonomy.

2.4 Natural Language Processing (NLP)

NLP is the application of AI to interactions with natural language (which means human language as opposed machine-readable language) in order to analyze a large amount of natural language data. NLP covers syntactic works such as lemmatization, parsing, sentence breaking, and word segmentation, as well as semantic and/or pragmatic works such as information retrieval, information extraction, question answering, and machine translation. As applied to legal areas, the NLP technology is already capable of reliably handling simpler classifications such as classifying contract provisions per their headings, searching for keywords (related to a smoking gun) during e-discovery, and supporting intelligent case search. However, it has not yet reached a level of replacing lawyers' cognitive power and legal reasoning capabilities. That is, the following functions can be conducted with NLP techniques, but only with limited capability: reading and understanding arguments in briefs; evaluating evidence; applying relevant statutes and cases to a factual situation; and drafting a decision. Given the rapid pace of developments of the NLP technology, however, it may soon become mature enough so that NLP can more reliably be used in the legal context.

3 Debates on AI and Law

We have so far discussed how to improve legal prediction by introducing ML approaches to empirical legal studies. As the understanding of this positive aspect of AI and law requires some familiarity with empirical research methods, most lawyers have paid more attention to normative issues involving the application of AI for legal practice or new social problems arising in the algorithmic society. That said, we will see that these issues carry no less profound implications for the economic analysis of law. There are three broad strands of debates on point. The first is regarding how to improve the judicial decision-making by applying AI models so that it becomes more efficient, consistent, and foreseeable. The second is regarding how to cope with new social issues or ramifications that arise with the advent of the algorithmic society. The third is regarding how to facilitate the development and use of AI within the legal system.

3.1 Systematizing Judicial Decision-Making with AI

The general public often embraces the idea of introducing and adopting an impartial and efficient 'AI judge' (see, for instance, Ulenaers 2020). In this vein, there have

been a few well-publicized experiments for replacing some of judges' tasks with AI, such as the introduction of Robot Judge in Estonia and the automation of e-Court judgment by default in debt collection proceedings in Netherlands (Ibid).

There are, however, two major hurdles in trying to automate the judicial decision-making process. The first is a *legal theoretical limitation* which would manifest itself in the process of automated legal reasoning. While a group of legal positivists have proposed to transform the legal system into a 'legal automaton,' a closed logical system which makes a decision based on preestablished rules (Hart 1958, 601–2), the proponents of the natural law theory or legal realism have tended to espouse a human judge's role to find moral norms or prevailing social interests, respectively. Using terminology that is more familiar to law and economics scholars, the substitution of the legal automaton for a human judge would often require *the substitution of rules for standards* (Fagan and Levmore 2019, 31–3). That can be suboptimal when empirical limitations such as overfitting, Simpson's Paradox, and omitted variables make it hard to measure data (Ibid, 14–28). For this reason, there is a growing support for a view that the legal automaton's role would be not to replace a human judge but to support her judgment in the form of an expert opinion. The law and economics scholarship has a long history of presenting an expert opinion based on regression analysis in antitrust and other high-stakes litigation. In the U.S., doing so was first recognized by a circuit court as a reliable scientific method (Petruzzi's IGA Supermarkets v. Darling-Delaware, 998 F.2d 1224 (3d Cir 1993)). The admissibility of an expert opinion based on AI models in judicial proceedings could be discussed in a similar context.

The second hurdle is that law is composed of *natural language* that machine is hard to read. To get over this problem, a few legal scholars, including a mathematician and lawyer Gottfried Leibniz, proposed to transform law into a machine-readable logic system (Wolfram 2018, 103–4). The development of such 'machine-readable' or 'computational' law, however, has not yet reached a sufficient level of maturity. As noted, we also need further development of NLP techniques to mimic human cognition, discretion, and intuition, as applied to legal reasoning.

That said, there are a few legal areas where features (X) are already machine-readable without a need to deploy NLP techniques, and the accuracy of prediction ($\hat{y}$) is verifiable on observable data within a short period of time. A striking example is a criminal justice system, where categorical or numerical attributes (such as age, sex, and financial status) of suspects, defendants, or convicts can be collected through investigation, and where the accuracy of human prediction and of machine prediction can be compared based on observable outcomes such as repeated crime or a failure to appear at mandatory judicial proceedings.

For *recidivism prediction*, several U.S. states have adopted risk assessment instruments (RAIs) based on regression models. One of the well-publicized

examples is correctional offender management profiling for alternative sanctions (COMPAS). More often than before, a COMPAS report is attached to a Presentencing Investigation Report (PSI), allegedly having impact on the court's sentencing. The use of COMPAS reports have been controversial, however, and there have been constant challenges against their use. Also, an experimental research reported that COMPAS, which takes account of the 137 features that it collects, produces no better results than laymen's rough guesses or a result from a simple linear classifier with only two features (Dressel and Farid 2018). COMPAS is also suspected of overrating the recidivism risk of African-American defendants. Several defendants challenged the use of COMPAS in criminal proceedings based on their due process rights. In 2016, the Wisconsin Supreme Court held that the trial court's use of COMPAS in sentencing did not violate due process principles, but required giving warning before the use of algorithmic risk assessment tools in sentencing, and in 2017, the U.S. Supreme Court denied the writ of certiorari (*Loomis v. Wisconsin*, 881 N.W. 2d 746 (Wis. 2016), *cert. denied*, 137 S.Ct. 2290 (2017)). Several Asian countries introduced RAIs, and may possibly experience similar controversies. For example, Korea developed and has used the Korean Sex Offender Risk Assessment Scale (KSORAS) to decide the electronic monitoring of adult sex offenders, and the Korean Risk Assessment System (KORAS-G) to assess recidivism risk of general offenders.

Another strand is the use of algorithm for *bail decision*. In New Jersey, 38.5% percent of those incarcerated were found to lack the capability be to post bail (12% percent due to inability to pay \$2500 or less) (VanNostrand 2013, 13). And, starting from January 2017, a bail reform was launched to replace bail (for nonviolent defendants) with the Public Safety Assessment (PSA) tool. The PSA tool would make predictions regarding (i) failure to appear for court events (FTA), (ii) new criminal activity (NCA), or (iii) new violent criminal activity (NVCA), based on statistical analysis of nine risk factors. In a year after the bail overhaul, 81.3% of defendants were released pretrial, dropping the pretrial jail population by 20%.⁵ The Third Circuit, in its recent decision in *Holland v. Rosen*, 895 F.3d 272 (3d Cir. 2018), *cert denied*, 139 S Ct 440 (2018), rejected a constitutional challenge against the PSA, ruling that criminal defendants do not have a constitutional right that guarantees them the option to pay cash bail. Kleinberg et al. (2018) study pretrial release decisions made in New York City and find that, by replacing the human judge decision with an ML model, crime can be reduced by up to 24.8% with no change in jailing rates, or jail populations can be reduced by 42.0% with no increase in criminal rates.

---

5 New Jersey Judiciary, 2017. "2017 Report to the Governor and the Legislature." pp. 15, 19. <https://www.njcourts.gov/courts/assets/criminal/2017cjrannual.pdf> (Accessed July 14, 2020).

3.2 Addressing New Problems Arising from the Algorithmic Society

The advent of the algorithmic society is expected to bring forth novel social issues and, in order to address them, fresh legal and ethical frameworks would be needed. The increased awareness of the relevant issues fueled a global boom in articulating and promulgating AI ethics principles. In a related vein, in the U.S., to discuss algorithmic transparency, fairness, and accountability, along with other ethical concerns, several executive orders and reports were issued such as the National AI Research and Development Strategic Plan (2016 and 2019), the Executive Order on Maintaining American Leadership in AI (2019), and Using Artificial Intelligence and Algorithms (2020), while the EU appears to have set forth even more guidelines and reports: Communication: AI for Europe (2018), Ethics Guidelines for Trustworthy AI (2019), Liability for AI and Other Emerging Digital Technologies (2019), Commission Report on Safety and Liability Implications of AI, IoT and Robotics (2020), and White Paper on AI (2020).

To keep pace, East Asian countries issued guidelines that discuss, among others, algorithmic transparency, fairness, and accountability. Some of these include: China's Next Generation AI Development Plan (2017), Three-Year Action Plan for Facilitating Next Generation AI Industry Development (2018–2020), and White Paper on Standardization of AI (2018); Japan's (Draft) AI Development Guideline (2017), AI Utilization Guideline (2018), Principle of Human-Centric AI Society (2019); and Korea's Mid- and Long-Term Comprehensive Countermeasure for Intelligence Information Society (2016), Ethics Guideline and Charter for Intelligence Information Society (2018), and Principle on User-Centric Intelligence Information Society (2019). In 2020, Korea went on to set forth the publicness, accountability, controllability, and transparency of 'intelligence information technology' in a statute (Article 62 of the Framework Act on Intelligence Informatization).

3.2.1 Algorithmic Transparency

A primary issue that these AI ethics guidelines try to address is that the *opaqueness and inexplicability* of AI algorithms (in particular, deep learning as a 'black box' algorithm) could, unless properly managed, undermine *human autonomy and control*. Some of these guidelines propose technological measures to make algorithm more explicable and grant a right to the users to request explanation as to how an algorithm works. Some also contain a proposal to audit

the process of algorithmic decision-making, or to establish a mechanism to contest the outcome.

Several jurisdictions have gone further and legislated regulations over *algorithmic transparency and explicability*. Under Article 22 of EU's General Data Protection Regulation (GDPR), the data subject is not subject to a decision based solely on automated processing, including profiling, without her consent, unless the decision is necessary for contracting or authorized by EU or member state laws. The data subject is also granted the right to obtain human intervention in the automated processing, to express her viewpoint, and to contest the decision. Under Korea's Act on the Use and Protection of Credit Information (Credit Information Act) (amended in February and effective in August 2020), a data subject, who is subject to an automated credit scoring by personal credit bureaus or financial institutions, has the right to request the explanation of the outcome, standard, and underlying data of the automated scoring, and to contest the scoring by submitting advantageous information or requesting the correction, removal, or reevaluation of underlying data (Credit Information Act, Article 36-2).

A paradox in this type of approaches is that, in general, the more transparent an ML model is made, the less functional and less accurate the model may become. For example, if a complete formula for credit scoring is made public, loan applicants may try to submit the features which are found to have higher correlation with the outcome of credit scoring and which are conducive to enhance the outcome of credit scoring. Such adaptive and exploitative behaviors are likely to impair the functionality of the ML model as a classifier. This problem would be particularly serious when the ML model was introduced to expand the opportunity of the financially distressed (e.g., an ML model that analyzes social network service can be deployed to expand the opportunities of those having a thin credit file like the young generation). Moreover, in practical terms, conducting automated processing, at the full exclusion of human intervention, appears to be uncommon in practice and, as such, the actual scope of applicability of these regulations can be much more limited than initially expected.⁶ Therefore, we need more thorough law and economics studies to find an optimal point where the social benefit that can be derived from a well-functioning ML model is balanced against human autonomy and other fundamental social values.

---

⁶ Article 29 Data Protection Working Party's Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 (2018) makes clear that the 'decision based solely on automated processing' under Article 22(1) of the GDPR means that 'there is no human involvement in the decision process,' although this cannot be evaded by 'fabricating human involvement'

3.2.2 Algorithmic Fairness

The data, on which ML models heavily rely on, are often biased and may not represent the whole population properly. An ML model trained on the biased data can cause direct discrimination (or *disparate treatment*) or indirect discrimination (or *disparate impact*) when applied to different groups of people. An AI agent trained on historical data can, for instance, overlook the recent growth of gender equality and reveal gender biases when deployed for automated recruiting or credit scoring. This resulted in the debates on how to ensure *algorithmic fairness* vis-à-vis the protected group, and many of the ethics guidelines mentioned above deal with algorithmic fairness.

The discussions on algorithmic fairness are truly transdisciplinary, and there is already extensive literature in computer science, law, economics, and public policy. From an economics viewpoint, the consideration of algorithmic fairness can be perceived as *constrained utility maximization* (Corbett-Davies et al. 2017). Numerous ways of defining this constraint have been proposed. Verma and Rubin (2018) categorize them into (i) definitions based on predicted outcome, (ii) definitions based on predicted and actual outcomes, and (iii) definitions based on predicted probabilities and actual outcomes. Among them, Corbett-Davies et al. (2017) identify the three most popular definitions: (i) *statistical parity* (an equal proportion in each group receives the same classification), (ii) *conditional statistical parity* (an equal proportion in each group receives same classification if a set of legitimate risk factors are controlled), and (iii) *predictive equality* (false positive rates are made even across different groups). The first and second definitions are based on predicted outcomes, while the third is based on predicted and actual outcomes. Paying attention to the definitions based on predicted probabilities and actual outcomes instead, Kleinberg, Mullainathan, and Raghavan (2017) identify three key elements of algorithmic fairness: (i) *calibration within groups* (people with the same predicted probability have the same probability to be classified in the positive class regardless of the group they belong to; for example, same acceptance rate across different sexes given the same merit); (ii) *balance for the negative class* (the members of the negative class from different groups have same average predicted probability; for example, male and female applicants rejected have the same merits); and (iii) *balance for the positive class* (the members the positive class from different groups have same average predicted probability; for example, male and female applicants accepted have the same merits), but at once prove that except in highly constrained special cases, no algorithm can simultaneously satisfy the three conditions. In fact, as there is a tradeoff between the ability to classify accurately and the fairness of the resulting data (Feldman et al. 2015), we need to pay attention to the marginal decrease in utility for an ML classifier in return for fairness. A more normative

strand of the literature has paid attention to the due process aspect in the presence of the conscious 'masking' of the discriminatory intent under the veil of opaque algorithm (Barocas and Selbst 2016, 692–3, 712–3).

As illustrated in the above example of the COMPAS system, this issue of algorithmic fairness is likely to develop in parallel with the increased use of algorithms in the judiciary or in the public administrative processes. That said, unlike the U.S. (see the Civil Rights Act of 1964) and the EU (see, for example, the Race and Framework Directives and Title III of the Charter for Fundamental Rights), most Asian countries do not appear to have enacted omnibus anti-discrimination legislation that inhibits discrimination in the private sector and, instead, some Asian countries have targeted regulations aimed at narrower areas such as equal employment. As such, this issue would be closely associated with the development of anti-discrimination laws that govern the private sector in general and the expansion of the constitutional principle of equality to civil relationships.

3.2.3 Algorithmic Accountability

There are ongoing debates on how to reform tort, product liability, and safety regulation regimes to effectively address the harms that could be caused by robots such as self-driving cars, medical robots, and drones or other AI agents by holding right persons *accountable* to the harm. Initial solutions have been sought from extending traditional liability regimes (such as respondeat superior liability theory, vicarious liability, or strict liability) to hold stakeholders liable or conversely shielding stakeholders from liability by granting an AI agent the status of electronic personhood. From the perspective of law and economics, however, the key task would be to identify which of various stakeholders (including developers, controllers, manufacturers, sellers, service providers, platforms, and users) can avoid relevant harms at the least costs and to allocate liabilities to the parties so identified. At the same time, to lower the costs of enforcing legal remedies by ensuring the traceability of accountable parties, appropriate technical governance mechanisms, industry standards, and audit systems should also be devised. Separate from this, in order to prevent undue chilling effects arising from potential liability burdens, discussions on algorithmic accountability should be coupled with the discussions on the structure of risk pooling (by way of insurance, for instance) and the scope of immunity.

3.2.4 Addressing Potential Economic Harm from Algorithmic Pricing Agents

Antitrust scholarship has debated on the potential anticompetitive effect from *price discrimination* by way of behavioral targeting and personalized pricing. In

addition, economic harms from the ‘*tipping*’ or *convergence* between actions by multiple algorithmic agents, such as stock trading bots or dynamic pricing agents, has drawn attention. In particular, a concern that price-setting algorithms might facilitate collusion in oligopolistic markets (*‘algorithmic tacit collusion conjecture’* (Ittoo and Petit 2017)) hard-hit antitrust scholarship, shortly after first proposed by Mehra (2014) and elaborated by Ezrachi and Stucke (2016). At the basis of their conjecture stand implicit suppositions: (i) a causation or correlation between a heightened price transparency or frequency/speed of interaction and a heightened risk of tacit collusion and (ii) a direct impact of the use of the same or similar algorithms or self-learning algorithms, leading to tacit collusion (without the mediating effect of market concentration) (Ibid). Their conjecture, however, has some theoretical weaknesses such as: (i) the transparency on the customer side (unlike that on the supplier side) can rather make it harder for the suppliers to collude; (ii) there is no theoretical or empirical ground for asserting that the use of the same or similar algorithms would facilitate tacit collusion; and (iii) in a heterogeneous product market, the agents can evolve in a way to effectuate price discrimination, even with no interests or intention to collude.

The literature has tried to run reinforcement learning models (in particular, multi-agent Q-learning models) to verify the algorithmic tacit collusion conjecture. The first actual implementation of the algorithm is found in Calvano et al. (2018), which concludes that their two-agent independent Q-learning model, built on the environment of logit demand and constant marginal costs, ‘systematically learn to collude’ after an average of 165,000 iterations. Klein (2019)’s experiments with a two-agent independent Q-learning model appears to show that Q-learning can learn to price above static levels in a sequential competition situation. These experiments appear to support the algorithmic collusion conjecture at a first glance, but their findings are based on strong simplifying assumptions and thus are ‘largely suggestive’ (Deng 2018, 91). One of the overly strong assumptions of these experiments might be that there exist only two players, given that one of Ezrachi and Stucke’s key intuitions is that algorithmic agents can collude even in the absence of market concentration (Ezrachi and Stucke 2017). Overall, this conjecture remains to be a theoretical conjecture not based on solid empirical grounds. This conjecture and other related discussions, nonetheless, have made significant contributions in that reinforcement learning models have been designed and applied to analyze actual and potential social harms from these discussions. More broadly, the deployment of AI models in businesses and its impact on market dynamics have emerged as an important area of research.

3.2.5 Heightened Privacy Concerns

As ML-based image classifiers such as convolutional neural network achieve outstanding predictive accuracy, AI-based facial recognition through closed-circuit television, satellites, and drones has got the potential to be used for *predictive policing* – the ‘use of historic crime data to identify individuals or geographic areas with elevated risks for future crimes, in order to target them for increased policing’ (Asaro 2019) – or more direct *surveillance* over a specific group or person. This is just one example where the use of an AI model could have serious privacy implications. Since today’s AI is predicated upon the extensive use of data – often personal data – developing and deploying an AI model often have ramifications on privacy and, as such, how to find a balance is an important issue.

3.3 Facilitating the Development and Use of AI

The last area is how to reform the legal system so that the development and use of AI can be facilitated. Following the paradigm shift to the data-driven AI, the quality of an AI model has become heavily dependent on the availability of good quality data. Particular attention has thus been paid to how to facilitate an AI developer’s access to the trove of data held by the private and public sector entities.

In Asia, one of the biggest hurdles has been laws and regulations in data protection which are largely transplanted from the EU regime and are based on the consent principle. Thus, a data subject’s consent is crucial for collection, use, and sharing of personal data. Following the advent of the data-driven AI, there is a growing demand for data, which would help realize the ever increasing economic value of personal data. In this context, pseudonymization is emerging as an important candidate to achieve a balance between data protection and proper utilization (see Articles 5(1)(b) and 89 of the GDPR, which exempt, from purpose limitation, the processing for (i) archiving in the public interest, (ii) scientific or historical research, or (iii) statistical purposes).

In February 2020, Korea made amendments to major laws in the area of data protection, including the Personal Information Protection Act and the Credit Information Act (effective as of August 5, 2020), in order to, among others, promote the utilization of pseudonymized personal data by allowing processing of the data for archiving, scientific research, or statistical purposes without consent from data subjects. In June 2020, Japan amended the Act on the Protection of Personal Information (expected to be effective in 2022), which, among others, allows the use of pseudonymized data without consent for the internal use of the business operator. India’s Personal Data Protection Bill 2019, which is currently pending at

the Indian Parliament, also stipulates that its data protection agency can exempt research, archiving or statistical processing from any provisions of the law if certain conditions including de-identification are met. These waves of regulatory reform call for a dramatic resurrection of information economics-based approaches to privacy (see Stigler 1980) that had long waned in the presence of zero risk-minded normative approaches. At once, a more thorough economic analysis of how to balance between protection and utilization, based on the statistic value of privacy, would be needed.

How to make data held by the public sector available to the private sector is another pivotal issue. In 2013, Korea enacted the Act on the Facilitation of Sharing and Use of Public Data, which requires each government agency to share public data unless the data falls under non-disclosable data under Korea's freedom of information law or is proprietary.

Separately, intellectual property could be an issue. That is, how to apply intellectual property law regimes to an invention or creation by an AI agent has also garnered attention. In China, the People's Court of Nanshan District of Shenzhen, in its March 2020 decision, held that Shanghai Yingmo Technology's copying of an investment research report written by Tencent's AI agent titled 'Dreamwriter' infringed Tencent's copyright.⁷ Unlike physical property rights, the intellectual property right is one of various legal devices consciously designed to help internalize positive externalities from invention or creation (including direct R&D subsidy or facilitation of the venture capital market). A dogmatic approach based on statutory interpretation or analogy to traditional invention or creation works may, from a policy perspective, lead to an erroneous decision. The ongoing economic debates in the context of intellectual property as to how to strike a balance between giving incentives to creators and giving access to the users to encourage utilization (see Posner 2005) need to be revitalized to provide a solid solution based on the degree of traceability, if any, of each stakeholder's contribution to an AI agent's works and the resulting allocation of incentives.

4 Conclusions

As more technological advances take place and more data become available, the usefulness of ML for legal prediction will naturally be enhanced. In that process,

---

7 People's Court of Nanshan District of Shenzhen, 2020. "Nanshan Court Judged China's First Case where the AI-Generated News Article Constitutes an Original Work of Authorship." <http://nsqfy.chinacourt.gov.cn/article/detail/2020/03/id/4860346.shtml> (Accessed July 14, 2020).

ML can also benefit from concepts that have been used and evolved in econometrics such as confounding variables, natural experiments, explicit experiments, regression discontinuity, and instrumental variables (Varian 2014). Past experiences of applying econometrics in the legal context (in the area of antitrust and other high-stakes litigation) can also help avoid repeating the same type of errors.

On a broader level, AI needs to be further demystified so that rational approaches can replace both unquestioning faith in AI and unreasonable anxiety about AI. In order to do that, analytic toolboxes that law and economics has honed so far can usefully be deployed to help the legal system reach an optimal point, where AI technologies can be developed without undue restrictions while addressing various social, legal, and policy issues appropriately.

Acknowledgments: This research was supported by the 2020 Research Fund of the Seoul National University Asia-Pacific Law Institute, donated by the Seoul National University Law Foundation.

Research funding: Seoul National University Law Foundation.

References

- Asaro, P. M. 2019. "AI Ethics in Predictive Policing: From Models of Threat to an Ethics of Care." *IEEE Technology and Society Magazine* 38 (2): 40–53.
- Barocas, S., and A. D. Selbst. 2016. "Big Data's Disparate Impact." *California Law Review* 104 (3): 671–732.
- Buchanan, B. G., and T. E. Headrick. 1970. "Some Speculation about Artificial Intelligence and Legal Reasoning." *Stanford Law Review* 23: 40–62.
- Calvano, E., G. Calzolari, V. Denicolò, and S. Pastorello. 2018. "Artificial Intelligence, Algorithmic Pricing and Collusion." *SSRN*, <https://doi.org/10.2139/ssrn.3304991>.
- Corbett-Davies, S., E. Pierson, A. Feller, and S. Goel. 2017. "Algorithmic Decision Making and the Cost of Fairness." In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 797–806, <https://doi.org/10.1145/3097983.309809>.
- Deng, A. 2018. "What Do We Know about Algorithmic Tacit Collusion." *Antitrust* 33 (1): 88–95.
- Dressel, J., and H. Farid. 2018. "The Accuracy, Fairness, and Limits of Predicting Recidivism." *Science Advances* 4 (1): eaao5580.
- Ezrachi, S., and M. E. Stucke. 2016. *Virtual Competition*. Cambridge, MA: Harvard University Press.
- Ezrachi, A., and M. E. Stucke. 2017. "Algorithmic Collusion: Problems and Counter-measures." *OECD Roundtable on Algorithms and Collusion DAF/COMP/WD (2017)* 25: 2–35, <https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=DAF/COMP/WD%282017%2925&docLanguage=En> (Accessed July 14, 2020).
- Fagan, F., and S. Levmore. 2019. "The Impact of Artificial Intelligence on Rules, Standards, and Judicial Discretion." *Southern California Law Review* 93 (1): 1–36.
- Feldman, M., S. A. Friedler, J. Moeller, C. Scheidegger, and S. Venkatasubramanian. 2015. "Certifying and Removing Disparate Impact." *Proceedings of the 21st ACM SIGKDD*

- International Conference on Knowledge Discovery and Data Mining* 259–68, <https://doi.org/10.1145/2783258.2783311>.
- Hart, H. L. A. 1958. “Positivism and the Separation of Law and Morals.” *Harvard Law Review* 71 (4): 593–629.
- Hastie, T., R. Tibshirani, and J. Friedman. 2009. *The Elements of Statistical Learning*, 2nd ed. New York, NY: Springer New York.
- Ittoo, A., and N. Petit. 2017. “Algorithmic Pricing Agents and Tacit Collusion: A Technological Perspective.” *L’intelligence artificielle et le droit*, pp 241–56. Bruxelles: Larcier.
- Klein, T. 2019. Autonomous Algorithmic Collusion: Q-Learning under Sequential Pricing. Amsterdam Law School Research Paper No. 2018–15, <http://dx.doi.org/10.2139/ssrn.3195812>.
- Kleinberg, J. M., H. Lakkaraju, J. Leskovec, J. Ludwig, and S. Mullainathan. 2018. “Human Decisions and Machine Predictions.” *Quarterly Journal of Economics* 133 (1): 237–93.
- Kleinberg, J. M., Ludwig, J., Mullainathan, S., and Obermeyer, Z. 2015. “Prediction Policy Problems.” *American Economic Review* 105 (5): 491–5.
- Kleinberg, J. M., Mullainathan, S., and Raghavan, M. 2017. “Inherent Trade-Offs in the Fair Determination of Risk Scores.” *Proceedings of the 8th Conference on Innovations in Theoretical Computer Science* 43: 1–43.
- McCarty, L. T. 1977. “Reflections on TAXMAN: An Experiment in Artificial Intelligence and Legal Reasoning.” *Harvard Law Review* 90 (5): 837–93.
- Mehra, S. K. 2014. “De-Humanizing Antitrust: The Rise of the Machines and the Regulation of Competition.” Temple University Legal Studies Research Paper No. 2014–43, <https://doi.org/10.2139/ssrn.2490651>.
- Ng, A. 2018. CS229 Lecture Notes. Also available at <http://cs229.stanford.edu/notes/> (accessed July 14 2020).
- Posner, R. A. 2005. “Intellectual Property: The Law and Economics Approach.” *Journal of Economic Perspectives* 19 (2): 57–73.
- Pound, R. 1954. “The Lawyer as a Social Engineer.” *Journal of Public Law* 3: 292.
- Russell, S. J., and Norvig, P. 2010. *Artificial Intelligence: A Modern Approach*, 3rd ed. Upper Saddle River, NJ: Prentice Hall.
- Stigler, G. J. 1980. “An Introduction to Privacy in Economics and Politics.” *Journal of Legal Studies* 9 (4): 623–44.
- Ulenaers, J. 2020. “The Impact of Artificial Intelligence on the Right to a Fair Trial: Towards a Robot Judge?” *Asian Journal of Law and Economics* 11 (2), <https://doi.org/10.1515/ajle-2020-0008>.
- VanNostrand, M. 2013. *New Jersey Jail Population Analysis: Identifying Opportunities to Safely and Responsibly Reduce the Jail Population*. Luminosity. Also available at <http://www.ncjrs.gov/App/publications/abstract.aspx?ID=264950> (accessed July 14, 2020).
- Varian, H. R. 2014. “Big Data: New Tricks for Econometrics.” *Journal of Economic Perspectives* 28 (2): 3–28.
- Verma, S., and J. Rubin. 2018. “Fairness Definitions Explained. FairWare 2018.” In *Proceedings of the ACM/IEEE International Workshop on Software Fairness*. 1–7, <https://doi.org/10.1145/3194770.3194776>.
- Wolfram, S. 2018. “Computational Law, Symbolic Discourse, and the AI Constitution.” *Computational Law, Symbolic Discourse, and the AI Constitution. Data-Driven Law*. 103–26. Boca Raton, FL: CRC Press.