THE BALK STOPS HERE: STANDARDS FOR THE JUSTICIABILITY OF GERRYMANDERING IN THE COMING AGE OF ARTIFICIAL INTELLIGENCE

DOUGLAS RUDEEN*

ABSTRACT

At the time of this writing, the Supreme Court seems to have abandoned establishing an objective test for drawing nonpartisan districts altogether—leaving this task largely to state-level courts and legislatures in the aftermath of Rucho v. Common Cause.[1] *Many have seen this deferment to the states as the latest in a series of unsatisfactory 'balks,' and have openly wondered what redistricting laws will look like in the next several years as a result of the Court's general refusal to intervene in this area outside of Voting Rights Act[2] litigation.*

This paper will argue that there is at least one foreseeable outcome—that (in the absence of mathematically rigorous legal standards for gerrymandering) state-level redistricting processes will be left vulnerable to abuse through the use of sophisticated artificial intelligence platforms.

To make this argument, the paper will first provide a brief and beginner-friendly primer on the basics of AI and machine learning. Then it will detail how AI is uniquely suited to perpetrate gerrymanders in ways that computer systems would not have been able to during the 2010 redistricting cycle. Finally, it will conclude with a policy recommendation—that the only reliable way to forestall gerrymandering in the age of AI is to employ a form of fully-automatic redistricting, or a novel semi-automatic redistricting process I will call the "Rawlsian Default."

TABLE OF CONTENTS

---

* UCLA Class of 2018. Currently with Anderson, Julian, & Hull, Boise.

1. 139 S. Ct. 2484 (2019).
2. 52 U.S.C. §§ 10101–10702 (2018).

I. INTRODUCTION

The law of gerrymandering has always been untidy, and nothing in recent years has changed this fact very much. In cases of partisan, racial, and other kinds of gerrymandering, determining what the applicable law is—and even whether a legal violation has occurred—is daunting.

In the case of partisan gerrymandering the Supreme Court's recent decisions in *Gill v. Whitford*[3] and *Rucho v. Common Cause*[4] have not been universally received as a satisfying end to years of costly fights. Some commentators had hoped that Justice Kennedy's concurrence in 2004's *Vieth v. Jubelirer*[5] had opened the door to the possibility of a quantitative outer limit on partisan gerrymandering.[6] However, with Kennedy's retirement and the *Rucho* decision, it seems unlikely this dream will be realized anytime soon.

With respect to racial gerrymandering, the Court's efforts in cases like *Shelby County v. Holder*[7] and *Cooper v. Harris*[8] seem to have also made application of the Voting Rights Act in this space more inexact. Moreover, racial and partisan gerrymanders have become increasingly difficult to tease apart as Americans "self-sort," adding new confusion to an already tenuous distinction.[9]

While criticisms of these outcomes are many and various, this paper will discuss one criticism that I believe has gone underrecognized. Namely, the ease with which new, inexpensive AI platforms can be used to affect gerrymandering—even as more stringent legal requirements are placed on the redistricting process.

To illustrate this, the paper will engage in a rudimentary (and as mathematically uninvolved as possible) discussion of how AI and machine learning platforms operate. The objective is to give readers with little or no experience working with or studying AI a crash course in what it actually does. Next, the paper will explore areas of vulnerability in the redistricting process that AI platforms are uniquely capable of exploiting. Finally, the paper will suggest a form of semi-automated redistricting that could help abate AI's undemocratic influence; a novel redistricting schema called the "Rawlsian Default."

---

3. 138 S. Ct. 1916 (2018).
4. 139 S. Ct. 2484 (2019).
5. 541 U.S. 267, 306–17 (2004) (Kennedy, J., concurring).
6. Nicholas Stephanopoulos & Eric McGhee, *Partisan Gerrymandering and the Efficiency Gap*, 82 U. CHI. L. REV. 831 (2015).
7. 570 U.S. 529 (2013).
8. 137 S. Ct. 1455 (2017).
9. *See* Jowei Chen & Johnathan Rodden, *Unintentional Gerrymandering: Political Geography and Electoral Bias in Legislatures*, 8 Q.J. POL. SCI. 239 (2013).

II. A LOW-MATH PRIMER ON AI

In order to predict how the prevalence of AI will affect gerrymandering, it is first necessary to broadly describe how AI works. Contrary to some popular belief – AI is nothing magical, it is simply a series of refined statistical calculations. I contend that by understanding these calculations (even at a very rudimentary level), it becomes possible to make better predictions about what sophisticated AI can do.

AI, generally, refers to any computer system that "imitate[s] intelligent behavior."[10] More specifically, AIs are computer systems that superficially resemble the functional properties of neurons in humans and animals, and that allow computers to engage in certain tasks that would be inefficient or impossible using other kinds of computer algorithms.[11] Ultimately, this process permits computers to engage in two problem-solving tasks that are especially germane to a discussion of gerrymandering. Namely, these two tasks are sophisticated pattern recognition and the optimization of heuristic algorithms.[12]

A. Artificial Neural Nets

The conceptual heart of modern AI is the Artificial Neural Net (ANN, hereinafter) and its utility for "supervised machine learning."[13] wherein the neural net is trained to recognize new data based on training data. This process (in some ways) simulates the neurons and synapses that characterize the nervous systems of most animals (see figure 1).[14]

---

10. *See Artificial Intelligence*, MERRIAM-WEBSTER, https://www.merriam-webster.com/dictionary/artificial%20intelligence (last visited May 22, 2020). This definition is, in the view of this author, overbroad—since many machines "simulate intelligent behavior" without actually trying to simulate the underlying biological processes that are thought to give animals and humans the faculty of "intelligence." Compare, for example, a sophisticated automaton that "simulates intelligent behavior" via a complicated system of gears and cams but has nothing functionally in common with a living thing.

11. *See, e.g.*, Hassabis et al., *Neuroscience-Inspired Artificial Intelligence*, 95 NEURON 245 (2017).

12. *See* discussion *infra* Part C.2. for a discussion of what is meant by "heuristic algorithm."

13. *See* GAVIN EDWARDS, *Machine Learning – An Introduction*, TOWARDS DATA SCI. (Nov. 18, 2018), https://towardsdatascience.com/machine-learning-an-introduction-23b84d51e6d0#ca90. Comprehensive coverage of the various applications of machine learning would be a herculean task. However, Edwards' article is a good additional primer on supervised machine learning (including several topics not covered in this writing).

14. *See* MICHAEL A. NIELSEN, NEURAL NETWORKS AND DEEP LEARNING (2019), http://neuralnetworksanddeeplearning.com/index.html; 3Blue1Brown, *But What Is a Neural Network? | Deep Learning, Chapter 1*, YOUTUBE (Oct. 5, 2017), https://www.youtube.com/watch?v=aircAruvnKk&list=PLZHQObOWTQDNU6R1_67000Dx_ZCJB-3pi&index=2&t=0s; 3Blue1Brown, *Gradient Descent, How Neural Networks Learn | Deep Learning, Chapter 2*, YOUTUBE (Oct. 16, 2017), https://www.youtube.com/watch?v=IHZwWFHWa-

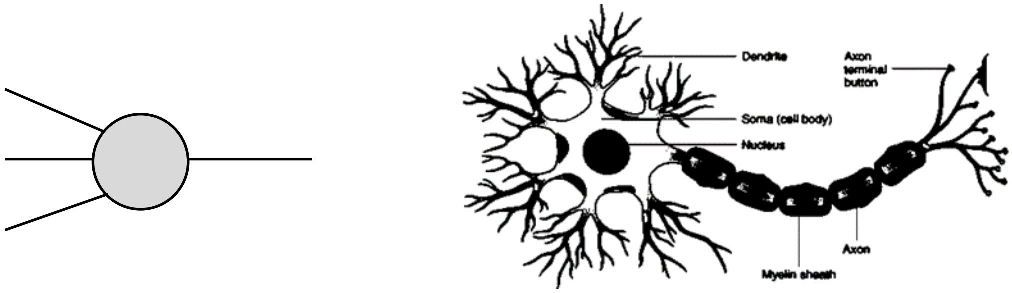


Figure 1 : On the left is a "node" (the circle) and "edges" (the attached lines). Nodes and edges like the one pictured above will be used for the remainder of this section to illustrate how ANNs operate. At right, see a drawing of a nerve cell as would be seen in the nervous system of an animal or human being. Science has shown that the axons/dendrites and somas of a nerve cell send electrochemical impulses to one another in a way that may be analogous to the way digital signals pass between different edges and nodes in an ANN, respectively.[15]

The clearest illustration of how ANNs are employed in pattern recognition tasks (and how they operate generally) is probably through a basic description of a hypothetical ANN called the "multilayer perceptron."[16] This machine was proposed as early as the 1950s and 60s, but only found widespread practical application within the last fifteen years.[17]

---

w&list=PLZHQObOWTQDNU6R1_67000Dx_ZCJB-3pi&index=3&t=0s; 3Blue1Brown, *What Is Backpropagation Really Doing? | Deep Learning, Chapter 3*, YOUTUBE (Nov. 3, 2017), https://www.youtube.com/watch?v=Ilg3gGewQ5U&list=PLZHQObOWTQDNU6R1_67000Dx_ZCJB-3pi&index=4&t=0s; 3Blue1Brown, *Backpropagation Calculus | Deep Learning, Chapter 4*, YOUTUBE (Nov. 3, 2017), https://www.youtube.com/watch?v=tIeHLnjs5U8&list=PLZHQObOWTQDNU6R1_67000Dx_ZCJB-3pi&index=5&t=0s.

15. The diagram of the nerve cell is based on an online answer to a problem from Chapter Thirty-seven of a Biology textbook published by Chegg. SYLVIA S. MADER, BIOLOGY ch. 37, prob. 2RC (10th ed. 2009); *see also We Have Solutions For Your Book!* CHEGG, https://www.chegg.com/homework-help/describe-structure-neuron-give-function-part-mentioned-name-chapter-37-problem-2rc-solution-9780077274337-exc (last visited May 22, 2020).

16. MICHAEL A. NIELSEN, NEURAL NETWORKS AND DEEP LEARNING (2019) at ch. 1. The machine described in this section is often referred to as a multilayer perceptron (or "MLP"), but this is a bit of a misnomer. Technically a "perceptron" is composed of neurons that only give binary output (zeros or ones) while the types of neurons used in a pattern recognition task like the one I'm describing can give any output between one and zero, and depend on an activation function—more on that later. *See infra* Figure 5 and Figure 6 and accompanying text.

17. Nielsen, and others, point to innovations in "deep learning" technology as the primary reason why perceptron-like computer programs have found new life. For detailed descriptions of what "deep learning" is and why it has improved the overall efficiency of ANNs, see NIELSEN, *supra* note 14, at ch. 1.

Imagine that we would like our machine to recognize single handwritten numbers like the eights pictured in figure 2. Because the numbers are non-uniform, we need some way to systematically recognize their general characteristics. Then, from those characteristics, we need to render some probability that the number is any integer between 0 and 9.

Figure 2: Handwritten number eights at different resolutions. While it is easy for us to recognize that these all represent an eight, we might have some trouble with the one in the middle.[18]

The ANN is uniquely suited to this task. The first stage of the ANN's operation involves breaking up the handwritten number into a grid of pixels, and then detecting the relative greyscale of each pixel.[19] Each pixel value is then assigned its own node so that each node has a value between zero and one (see figure 2).[20] In figure 3 the input nodes are aligned vertically to form a "layer."

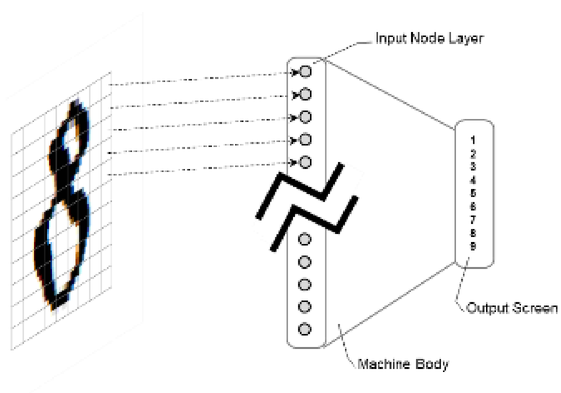


Figure 3: One of the ambiguous number eights from the earlier figure is placed under a ten-by-ten grid. Each pixel in the grid is assigned to one of one-hundred nodes in the input layer – which we see vertically aligned and with a break mark in the center to save space. The goal is to get the machine to recognize which number is being displayed and cause a corresponding

18. *See* NIELSEN, *supra* note 14, at ch. 1.

19. *See id.* In this example I use "pixel" to mean any one-dimensional square division of an image. In the example "100%" is totally black and "0%" is totally white.

20. *Id.*

value on the output screen to light up. Increasing the number of input nodes (and therefore the relative resolution of the image) can improve the accuracy of the machine.[21]

Within the machine body lie a series of so-called "hidden layers" that the input layer is connected with (see figure 3).[22] Each node in the hidden layers has a series of branching edges that connect to every other node in the subsequent layer.[23] After a signal is received from the input layer, it is relayed through the hidden layers along the edges until there are only a few nodes remaining.[24]

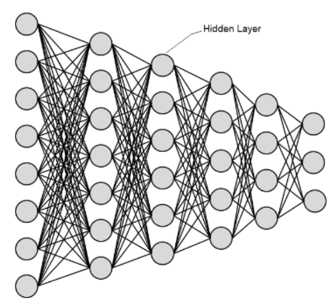
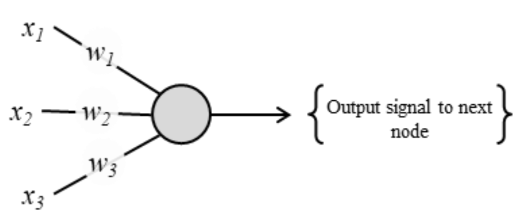


Figure 4: The nodes and edges that lie within the machine body in figure 4, above. As digital signals pass from node to node along the edges, they are multiplied (and thereby weighted) by certain values.[25]

Each node within the hidden layers has several values assigned to it: values that represent weighted averages of the values relayed to it from all the preceding nodes, as well as a "bias," which is a number subtracted from the weighted average of the signals arriving from the preceding nodes (see figure 5).[26]



21. *See* NIELSEN, supra note 14, at ch. 1.
22. *See id.*
23. *See id.*
24. *See id*.
25. *See* NIELSEN, supra note 14, at ch. 1.
26. *See id.*

Figure 5: The values from the previous nodes ($x_{1...3}$) are multiplied by a series of weights ($w_{1..3}$) before entering the node, where they are processed using an "activation function"—a mathematical function that converts all the preceding values into a single number between 0 and 1, and then outputs that function to every node in the adjacent hidden layer.[27]

Inside the node, the weights and the bias value are combined using a multiplicative "activation function."[28] The possible output of the activation function may vary somewhat depending on our task, but the output always follows some kind of curve relative to the value of the combined weights it accepts as input and subtracts from a bias versus what it outputs (see figure 6).[29]

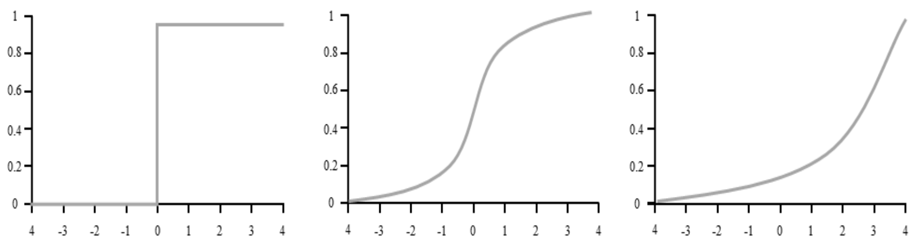


Figure 6: Graphs of several possible output functions of nodes.[30] *They include a step function (left), a sigmoid or logistic function (center) and a logarithmic function (right).*[31] *Depending on the combined value of the weights and bias (the x-axis), the ultimate output of the nodes (the y-axis) will fall at some point along whichever function curve we think is best suited for our task – outputting a number between one and zero.*[32]

The original grayscale values from the image are multiplied by the weights and biases of each node until there are only a few values remaining.[33] In our case, we want these remaining values to correspond with percentage likelihoods that the handwritten image is a particular number; then have the appropriate number display on the output screen (see figure 3, *supra*). If the weights, biases, and activation functions are properly calibrated, the ANN will be capable of "recognizing" handwritten numbers within a certain margin of error.

27. *See id.*
28. *See id.*
29. *See id.*
30. *See* NIELSEN, supra note 14, at ch. 1.
31. *Id.*
32. *Id.*
33. *Id.*

B. Gradient Descent and Optimization Algorithms

Clearly, for the ANN to recognize numbers correctly, it is necessary for the weights, biases, and output functions associated with each node to be highly refined. However, trying to "dial in" the weights and biases of each node on an ad hoc basis until accuracy results for all ten numbers (zero through nine) would be incredibly inefficient. Fortunately, several optimization algorithms exist to automatically make these adjustments.[34] "Gradient Descent"[35] is likely the most straightforward of these algorithms.

Gradient Descent involves several steps. First, it quantifies the extent to which the ANN is failing to make accurate identifications from the training data by way of a linear model (illustrated in figure 7, left by a correlation graph) that compares the ANN's guesses (x) relative to an accurate result (y).[36] Next, the algorithm iterates through possible slopes and y-intercepts to identify a line of best fit in that correlation graph.[37] Some y-intercept and some slope provide the least error-prone (or least "costly") line of best fit – the point at which the ANN is least likely to make an error.[38]

An efficient way to find this optimal y-intercept and slope involves graphing a curved line on a plane that has possible y-intercepts as its x-axis, and the relative error rate as its y-axis (see figure 7, right).[39] As the line of best fit becomes more accurate, the error rate decreases along the curve (in figure 7, right). By iterating through a series of derivatives taken along this line, the algorithm approaches a minimum point along the curve, representing the y-intercept of the line of best fit.[40]

Finding this line of best fit is significant – it means we have also located the "least costly" model for identifying the data in our training dataset.[41] We can be assured that the weights and biases of our ANN – when they are attenuated to the

34. See generally AURÉLIEN GÉRON, HANDS-ON MACHINE LEARNING WITH SCIKIT-LEARN AND TENSORFLOW: CONCEPTS, TOOLS, AND TECHNIQUES TO BUILD INTELLIGENT SYSTEMS 297 (2017) for several examples.

35. For the remainder of this section I will be using "Gradient Descent" as shorthand for "Batch Gradient Descent" – probably the most accessible and illustratively valuable machine learning algorithm.

36. GÉRON, *supra* note 34, at 110–111, 113. *See also* 3Blue1Brown, *Gradient Descent, How Neural Networks Learn | Deep Learning, Chapter 2*, YOUTUBE (Oct. 16, 2017), https://www.youtube.com/watch?v=IHZwWFHWa-w&list=PLZHQObOWTQDNU6R1_67000Dx_ZCJB-3pi&index=3&t=0s.

37. GÉRON, *supra* note 34, at 113–116. In order to keep this section relatively free of pure mathematical explanation, I have omitted the formal equation for the gradient descent algorithm. However, for readers who would like to study the algorithm in greater detail, I recommend Michael A. Nielsen's *Neural Network and Deep Learning* and 3Blue1Brown's video, "What is Backpropagation Really Doing? | Deep Learning, Chapter 3." NIELSEN, *supra* note 14; http://neuralnetworksanddeeplearning.com/index.html; 3Blue1Brown, *What is Backpropagation Really Doing? – Deep Learning, Chapter 3*, YOUTUBE (Nov. 3, 2017), https://www.youtube.com/watch?v=Ilg3gGewQ5U&list=PLZHQObOWTQDNU6R1_67000Dx_ZCJB-3pi&index=4&t=0s.

38. GÉRON, *supra* note 34, at 113–116

39. *Id.*

40. The slope of the line of best fit can also be determined this way by adding additional dimensions to the graph of the curve. *See id.*; *see also* StatQuest with Josh Starmer, *Gradient Descent, Step-by-Step*, YOUTUBE (Feb 5, 2019), https://www.youtube.com/watch?v=sDv4f4s2SB8.

41. GÉRON, *supra* note 34, at 113–116

point returning the line of best fit - are likely optimized for recognizing new data relative to the training data we have already provided it.[42]

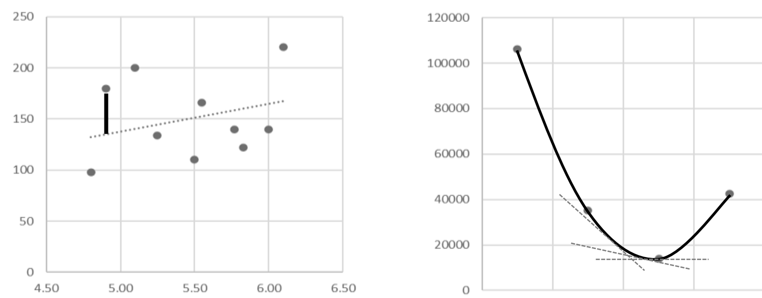


Figure 7: Graphs showing examples of two functions essential to gradient descent.[43] On the right is graph guesses an ANN might make about the value of y relative to x, with a line of best fit (dashed) showing where accurate guesses would fall with respect to a set of training data. A single error bar (thick black line, left) indicates how far off a particular guess was from the training data. The graph on the right shows a possible curve resulting from plotting the sum of the squared error bars from the chart on the left at different y-intercepts for the line of best fit. The minimum point on this curve represents the most accurate guesses that the machine can make about the y-intercept. Derivatives are shown at right using dashed lines.[44]

Gradient descent is – as noted above – just one example of an optimization algorithm, and several others exist that are more efficient, powerful, or well-suited for other applications of machine learning besides supervised learning.[45]

C. Other Topics in Machine Learning

Virtualized versions of ANNs conceptually similar to the multilayer perceptron described above can be scaled up and modified in a variety of ways, allowing them to engage in very complex pattern recognition tasks.[46] Just as the machine described above can detect ambiguous numerical shapes with a high degree of accuracy, more complicated AIs (consisting of more nodes, edges, and hidden layers, and employing more-sophisticated optimization functions) are capable of

42. *See id.*; see also StatQuest, *supra* note 40.

43. Images above are my own, primarily based on examples in StatQuest, *supra* note 40, and 3Blue1Brown, *Gradient Descent, How Neural Networks Learn | Deep Learning, Chapter 2*, YOUTUBE (Oct. 16, 2017), https://www.youtube.com/watch?v=IHZwWFHWa-w&list=PLZHQObOWTQDNU6R1_67000Dx_ZCJB-3pi&index=3&t=0s.

44. *See id*.

45. *See* GÉRON, *supra* note 34 .

46. *See* GÉRON, *supra* note 34, at xiii, 523–531.

recognizing patterns in our movie preferences, photos, purchasing habits, and even (topically enough) our voting behavior.[47]

Many of these more-complex functions require significantly altered applications of the sort of processes described previously.[48] This section will briefly describe two of these concepts, particularly unsupervised learning and the optimization of heuristic algorithms. In later sections, the importance of these concepts with respect to gerrymandering will be explored further.

i. Unsupervised Learning

There are certain situations where the use of large sets of training data (like the pairs of handwritten and digital numbers in the example in Sections II.A) are either impractical or undesirable.[49] In these situations, other forms of machine learning are employed. One common form is "unsupervised learning," which does not require training data at all.[50]

In unsupervised learning, a computer is only given a raw dataset and is charged with recognizing patterns within those data based just on the data itself (i.e., no training data is involved).[51] The typical example of this is a process called "clustering," wherein a computer organizes data around sets of common traits that it determines reduce the data's complexity or dimensionality.[52]

For example, it may be helpful to visualize unsupervised learning as a process that asks a machine to cluster together handwritten numbers relative to their *similarity to one another*, instead of associating them together based on what the number *should* look like depending on training data.[53] So, in addition to grouping together handwritten digits that look like eights—the computer might also place groups of sixes and nines near the eights.[54] This, because the computer would observe that all three numbers resemble two stacked closed loops (i.e. six looks like an eight with a broken upper loop and nine resembles an eight with a broken lower loop).[55]

This form of learning is particularly useful in situations where we want to try and detect new and novel patterns and where preconception about what the computer should be looking out for can actually hinder the ability to recognize new patterns.[56] Medical diagnostics are a ready example.[57] An unsupervised clustering process of patients with a particular disease based on many hundreds of different criteria could potentially reveal underlying risk factors.[58]

---

47. *See id.*; *see also* MATT JONES, *Should Forecasters turn to Artificial Intelligence to Predict the US Election?,* NS TECH, (Nov. 7, 2016), https://tech.newstatesman.com/big-data/forecasters-turn-artificial-intelligence-predict-us-election.
48. *See supra* Sections II(A)–II(B).
49. *See, e.g.*, Edwards, *supra* note 13.
50. *See, e.g.*, Edwards, *supra* note 13.
51. *See, e.g.*, *id.*
52. *See id.*
53. *See id.*
54. *See id.*
55. *See id.*
56. *See* Edwards, *supra* note 13.
57. *See id.*
58. *See id.*

ii. Heuristic Algorithms

Another use of AI technology is the optimization of heuristic algorithms.[59] A "heuristic" can be generally described as a decision-making shortcut.[60] If, for example, someone is having a difficult time choosing between watching TV and exercising, they may weigh the relative merits of both and decide which one is a better use of their time. Since this person cannot consider every possible contingency that might make one activity better than the other, they might rely on some kind of general principle that produces reliable results (e.g., looking at the weather forecast). While using this general principle may not be optimal or perfect, it is good enough to be relied upon—this is a heuristic.

In mathematics and computer science,[61] an analogous situation arises in situations where quickly finding an optimal or "best possible" solution is exceedingly complex, or impossible.[62] In these situations, certain algorithms exist to find solutions that—while potentially sub-optimal—are good enough to be very useful.

A classic example is the travelling salesman problem. This problem asks us to imagine a door-to-door salesman who must pick the shortest direct path between a scattered array of locations.[63] While this problem is easy to solve for limited numbers of points, the more points on a map the salesman must visit, the more complex the problem becomes.[64] It becomes so complex, in fact, that it appears to become incalculably hard, even for the fastest computers.[65]

AIs can be useful in this context because they can learn to dynamically ignore certain proposed solutions to the problem that would be a waste of time to test.[66] This, as opposed to some traditional computer algorithms that would have to iterate through nearly every possible solution before arriving at an acceptable one.[67]

59. *See* Micah Altman, *The Computational Complexity of Automated Redistricting: Is Automation the Answer?*, 23 RUTGERS COMPUTER & TECH. L.J. 81, 90 (1997). Heuristic algorithms, as a general matter, are particularly germane to a discussion of the mathematics of gerrymandering. *Id.* at 90–91, 123–126.

60. James Chen and Roger Wohlner, *Heuristics*, INVESTOPEDIA (Jul 31, 2020), https://www.investopedia.com/terms/h/heuristics.asp.

61. Particularly many of the mathematical problems that tend to arise in the context of gerrymandering. See Altman, *supra* note 59, at 90–91.

62. Particularly the class of problems described as NP-Complete or NP-Hard. *See* Micah Altman & Michael McDonald, *The Promise and Perils of Computers in Redistricting*, 5 DUKE J. CONST. L. & PUB. POL. 69, 81 (2010).

63. *See* Suzanne Ma, *Understanding the Travelling Salesman Problem (TSP)*, ROUTIFIC (Jan. 2, 2020), https://blog.routific.com/travelling-salesman-problem.

64. *See id.*

65. *See* Altman, *supra* note 59 (citing Ronald V. Book, *Relativizations of the P=? NP and Other Problems: Developments in Structural Complexity Theory*, SIAM Review 36(2), 157, 159–60 (1994)).

66. *See* Altman, *supra* note 59, at 90–95.

67. *See id.*

III. NOWHERE TO HIDE—THE VULNERABILITY OF THE CURRENT LEGAL FRAMEWORK

Understanding AI, in a broad conceptual sense, may have some general utility for any policymaker or jurist, especially as it becomes more common. But what does it amount to in the context of gerrymandering? Why should the various functions of AI described in the previous section *especially* matter to policymakers charged with drawing electoral maps? In this section, I will explore several reasons why AI should be of extreme interest to mapmakers, courts, and other stakeholders in the redistricting process, given some of its unique capabilities.

First, this section will explore two primary ways that malevolent AIs could affect the map drawing process directly. These include a) optimizing the creation of maps that—while technically compliant with existing legal standards—are still heavily gerrymandered; and b) through training superficially unbiased AIs with biased datasets. Second, this section will explore some of the secondary or indirect ways in which AI might be put to use to attack the redistricting process.

Before proceeding, it is worth mentioning that I take for granted three premises in this section:

1) That without quantitative definition, any given standard for limiting gerrymandering (e.g., requirements of compactness, contiguity, and even the "equal population" standard) can be applied in more than one way.[68]
2) If none of the standards are truly mutually exclusive, then combinations of these different factors into multi-factor legal tests can also be applied in multiple ways.[69]
3) Some subset of these various applications can serve different political goals. I.e. defining ephemeral criteria like "competitiveness" one way may serve one political goal, while defining it another way may serve another.

If these background facts are taken as true, I will argue that it is reasonable to expect certain AI platforms, either during the coming cycle of redistricting in 2020, or at most the upcoming cycle in 2030, will capably exploit the existing set of anti-gerrymandering laws to serve differing political goals – thereby appreciably distorting voter preference.[70] In this section, I will briefly overview the current legal

---

68. I will refer to standards capable of various application as "ambiguous" for the remainder of the section.

69. As a corollary to this statement, it should be noted that having one prong of an anti-gerrymandering test be quantitative is still an insufficient guarantor of consistent application when the other parts of the test are not. For example, Arizona's redistricting laws between 2000 and 2010 seemed to include a general proviso that "compactness" was to be defined according to the "Polsby-Popper" test for compactness – a rigorous quantitative definition. However, this seemed to have been treated as a guideline rather than a stringent requirement, and was not the ultimate determiner of compactness. *See* Justin Levitt, *All About Redistricting*, LOY. L. SCH., https://redistricting.lls.edu/states-AZ.php#criteria (last visited May 22, 2020); for the original Polsby-Popper test, *see* Daniel D. Polsby & Robert D. Popper, *The Third Criterion: Compactness as a Procedural Safeguard Against Partisan Gerrymandering*, 9 YALE L. & POL'Y REV. 301 (2015).

70. This should be broadly interpreted to include provisions of Federal statutory law directed at preventing racial gerrymandering, *e.g.* Voting Rights Act, 52 U.S.C. §§ 10301, 10304 (1965).

standards and their shared vulnerabilities, and will discuss "direct distortion" and "secondary distortion," two broad means by which AI could exploit these legal vulnerabilities.

A. The Current Legal Standards

The existing set of redistricting laws in the United States have a long history of untidiness.[71] Very early in the nation's history there is evidence that the political branches were struggling with how to create geographic partitions of the several states to achieve fair representation, and whether or not this was an improvement over at-large elections.[72] Constitutional guidance on partisan gerrymandering is scant,[73] and federal interventions in the area—including the "one-person one vote" standard and post-incorporation interpretation of the 14th Amendment equal protection clause—have done little to set categorical prohibitions on the practice.[74] Racial gerrymandering has been more clearly prohibited. The Voting Rights Act (VRA) and associated cases have made it clear that the government does have a compelling interest in preventing this practice.[75] However, guidance on how exactly the VRA is to be applied in many situations is also murky.[76]

Presently, the practice of partisan gerrymandering is largely controlled by state law as a result of *Rucho v. Common Cause*.[77] State-level gerrymandering rules, in turn, typically employ some combination of the following five factors:

71. *See generally* ELMER CUMMINGS GRIFFITH, THE RISE AND DEVELOPMENT OF THE GERRYMANDER (Kessinger Legacy Reprints, 2010) (1907); ERIK J. ENGSTROM, PARTISAN GERRYMANDERING AND THE CONSTRUCTION OF AMERICAN DEMOCRACY, 16, 23–27 (2013).

72. *See* ENGSTROM, *supra* note 71, at 21–22.; *see also* Vieth v. Jubelirer, 541 U.S. 267, 304 (2004). Interestingly, several states attempted to avoid this issue outright by holding at-large elections prior to the Decennial Apportionment Act of 1842. *See generally* ENGSTROM, *supra* note 71.

73. *See* U.S. CONST. art. I, § 5; *see generally* ENGSTROM, *supra* note 71, at 22.

74. *See generally* U.S. CONST. art. I, § 5; Bryan H. Wildenthal, *The Lost Compromise*, 61 OHIO ST. L. J. 1051 (2000); *see also* Gitlow v. New York, 268 U.S. 652 (1925); Chicago B. & Q. R.R. Co. v. City of Chicago, 166 U.S. 226 (1897), seminal incorporation doctrine cases.

75. *See* Voting Rights Act, 52 U.S.C. §§ 10301, 10304 (1965); Cooper v. Harris, 137 S. Ct. 1455 (2017); Thornburg v. Gingles, 478 U.S. 30 (1986).

76. For a case study in the troubled application of Sections 2 and 5 of the VRA, *see* Galen Druke, *Is Gerrymandering the Best Way to Make Sure Black Voters Are Represented?*, FIVETHIRTYEIGHT (Dec. 14, 2017, 3:24 PM), https://fivethirtyeight.com/features/is-gerrymandering-the-best-way-to-make-sure-black-voters-are-represented/; *see also* a discussion of the application of the *Gingles* factors in *Cooper*, 137 S. Ct. 1455; *see generally* Shelby Cty. v. Holder, No. 10-0651(JDB), 2013 WL 10509740 (D.D.C. Oct. 11, 2013) (creating a general lack of clarity for states on the application of Section 5 of the VRA).

77. Rucho v. Common Cause, 139 S. Ct. 2484 (2019). The requirement that districts have relatively equal populations is still Constitutionally controlled by Article I, section 5, but the stringency of this requirement varies between congressional districts and state and state and local legislative districts. *See* Justin Levitt, *Where are the Lines Drawn?*, LOY. L. SCH., http://redistricting.lls.edu/where-state.php#compactness.

1) Compactness;
2) Contiguity;
3) Competitiveness;
4) Preservation of Communities of Interest;
5) Preservation of Existing Political Boundaries;[78]

Beyond these factors, certain jurisdictions are still charged with the creation of majority-minority districts pursuant to Section 2 of the VRA.[79] Certain states have also taken it upon themselves to create additional rules directed towards ensuring electoral fairness.[80]

Detailed discussion of what each of these factors can potentially entail has been treated in significant depth by other authors.[81] For our purposes, it is enough to note that each standard is relatively ambiguous. For instance, whether "compactness" is purely a function of geography, or a function of residents' mean distance from a given centroid is debatable, and both have been proposed to measure this concept.[82] Contiguity is another example. It should ostensibly be easy to measure—it probably ought to mean that the district is not divided into disjointed islands. But are districts that are strung together with parcels of land where no one resides "contiguous"? If true, is there a way of quantifying how densely populated each portion of a district must be for it to be considered part of a contiguous district? For each of the standards itemized above, questions like these abound.[83]

B. Direct Distortion

"Direct distortion" is a term I will use to describe a process wherein an ANN is directly programmed to create maps that comply (within an acceptable margin of error) with sets of ambiguous legal standards, but still affect gerrymanders. Ironically, as policymakers layer factor upon factor in the interest of trying to stop gerrymandering, they are likely doing just the opposite—creating a safe haven for malevolent AIs to work unnoticed.

Imagine, for example, that we wanted to draw a map that was "compact" but lacked an unambiguous standard for compactness. An AI could be designed to use different segments of an ANN to test multiple different accounts of what "compactness" means. The AI could efficiently test a series of maps containing population data about our subject jurisdiction, and it could theoretically recognize

78. *See e.g.* CAL. ELEC. CODE § 21500 (West 2020).

79. Voting Rights Act, 52 U.S.C. § 10301 (1965).

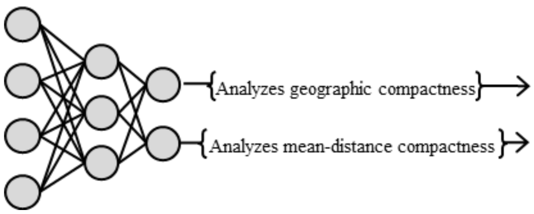
80. *E.g.* preventing drawing districts that "unduly" favor a candidate or political party. *See, e.g.*, HAW. CTY. CHARTER art. III, § 3–17(f)(1) (2015).

81. *See* Levitt, *supra* note 77.

82. *Compare* Brian Olson, *Redistricter*, BITBUCKET.ORG, https://bitbucket.org/bodhisnarkva/redistricter/src/default/ (last visited May 22, 2020) (Brian Olson's proposed use of mean distance), *with* Lukas Svec, Sam Burden, & Aaron Dilley, *Applying Voronoi Diagrams to the Redistricting Problem*, 28 UMAP J. 313 (2007) (proposing use of weighted Voronoi diagrams, a geographically based measure of compactness) *and* Chung-I Chou & S.P. Li, *Taming the Gerrymander—Statistical Physics Approach to Political Districting Problem*, 369 PHYSICA A. 799 (2006) (advocating a similar approach).

83. *See* Levitt, *supra* note 77.

maps that were any combination of "geographically compact" and/or "mean-distance compact." If our legal distinction is ambiguous, the AI could suggest a greater array of possible maps (see figure 8), thereby increasing the likelihood that a legally-compliant (in this case "compact") but gerrymandered map would be within the solution set.



	Geographic Compactness	Mean Distance Compactness	Outcome
Map A	False	True	True
Map B	True	False	True
Map C	False	False	False
Map D	True	True	True

Figure 8: An abbreviated illustration of a hypothetical ANN with two sections running parallel to one another, one that has been trained to recognize geographic compactness in a series of legal maps, and one that recognizes mean-distance compactness in a series of legal maps. We can see that this creates the equivalent of a logical AND/OR relationship in the table at right. Accordingly, the possible number of acceptable maps has increased because of the ambiguity of our legal standard.[84]

Moreover, it is relatively easy to imagine a situation where the AI described above could be programmed to gerrymander by giving additional weight to maps that showed a high tendency to swing one way or another along party or racial lines. In these situations, the AI could be trained to pick whichever arbitrary definition of "compactness" best accomplishes this end.[85]

It is also easy to envision a situation where the kind of unsupervised learning described above,[86] could be employed to search out latent or discrete statistical characteristics among groups of likely voters that would correlate with them voting for a particular party, or were suggestive of their belonging to given racial groups. Using these characteristics as a proxy for either party affiliation or race would allow the AI to engage in a nearly identical process of direct distortion to that described

84. This image is my own, based on Altman, *supra* note 59.

85. One interesting example of this results when Brian Olsen's algorithm is used on a map of North Carolina. Aaron Bycoffe, Ella Koeze, David Wasserman, & Julie Wolfe, *The Atlas Of Redistricting*, GERRYMANDERING PROJECT (Jan. 25, 2018, 6:00 AM), https://projects.fivethirtyeight.com/redistricting-maps/north-carolina/. While the algorithm ostensibly lacks intent to engage in racial gerrymandering, using it on this map results in remarkably few districts that might be reasonably considered 'majority-minority.'

86. *See supra* Section II.C.1.

in the previous paragraph, even if the AI lacked specific data about voters' party registration or race.

A final form of "direct distortion" would obtain if an AI platform (perhaps one that a redistricting commission agrees to use to aide in their mapmaking beforehand) were somehow compromised to write biased maps. This is easy to conceptualize if we consider how training data can be employed to bias the kinds of solutions that an AI recognizes.[87] If an otherwise "objective" AI platform or library were surreptitiously retrained on a dataset of biased maps, it follows that it would begin to render maps exhibiting similar biases itself.

Ultimately, any of the forms of direct distortion described above, combined with AI's ability to employ heuristic algorithms to quickly refine the number of possible maps it can choose from, make it a formidable obstacle to the evenhanded application of existing legal standards.

C. Secondary Distortion

It is likely that the ability of AI to mimic the speech and writing of natural persons—even if imperfect—could make it an important tool for those trying to promote gerrymandered maps in bad faith, regardless of whether the maps themselves were originally drawn using an AI platform. I will refer to processes like this—where an AI is used as a gerrymandering tool in a way that is *secondary* to the actual drawing of maps—as "secondary distortion." This could be accomplished in several ways, but one particularly troubling way is the manufacture of fake testimony.

Several states rely heavily on public testimony as a feature of their redistricting processes, particularly if they rely on independent redistricting commissions.[88] In the early stages of redistricting, these commissions often solicit preliminary testimony from citizens regarding how they would like maps to be drawn within the states' legal frameworks.[89] In California's 2010 redistricting cycle, for example, citizens were given the chance to email the redistricting commission about ways in which they believed they were parts of "communities of interest" such that they should be permitted to vote as a bloc in certain elections.[90] During the 2010 cycle in Arizona, citizens were allowed to provide some electronic feedback to that state's commission about proposed maps.[91]

While soliciting information like this likely allowed the redistricting commissions to make well-informed decisions about how the subject maps were

87. *See supra* Section II.B.

88. *See generally* Galen Druke, *Even A Gerrymandering Ban Can't Keep Politicians From Trying to Shape Their Districts*, GERRYMANDERING PROJECT (Jan. 4, 2018, 3:28 PM) [hereinafter Druke, *Even A Gerrymandering Ban*], https://fivethirtyeight.com/features/even-a-gerrymandering-ban-cant-keep-politicians-from-trying-to-shape-their-districts/ (providing a profile of the redistricting process in California); Galen Druke, *Want Competitive Elections? So Did Arizona. Then the Screaming Started*, GERRYMANDERING PROJECT (Dec. 21, 2017, 10:03 AM) [hereinafter Druke, *Want Competitive Elections?*], https://fivethirtyeight.com/features/want-competitive-elections-so-did-arizona-then-the-screaming-started/ (providing a profile of the redistricting process in Arizona).

89. *See, e.g.*, Druke, *Even A Gerrymandering Ban*, *supra* note 88; Druke, *Want Competitive Elections?*, *supra* note 88.

90. *See* Druke, *Even A Gerrymandering Ban*, *supra* note 88.

91. *See* Druke, *Want Competitive Elections?*, *supra* note 88.

ultimately drawn,[92] they potentially opened the door for fake testimony favoring or disfavoring commission activity or particular maps. Given that fake testimony has been present in one form or another in earlier redistricting efforts,[93] the use of AI to improve the convincingness of those fakes should be heeded with extreme caution by redistricting commissions, or by any political body seeking to engage in analogous forms of open redistricting.[94]

IV. AUTOMATED REDISTRICTING—HAS ITS TIME FINALLY COME?

One appealing cure for some of the problems described in the previous sections is quantifying existing legal standards in such a way that redistricting can be fully automated (e.g. by implementing a redistricting statute that was merely a restatement of code executed on a computer to generate new maps).[95] The idea of automated redistricting is hardly new.[96] Even before personal computers were widely available, commentators still saw value in automation as a means of preempting gerrymandering outright.[97]

However, despite its superficial intuitiveness, automated redistricting has never caught on. Justifications for why it has never caught on abound, but they tend to gravitate around three general objections: First, there is a sense in which the automation of redistricting would still be an inherently political activity, and that any scheme of automated redistricting could mirror the biases of programmers.[98] Second, there is the idea that automating redistricting would remove some measure of democratic process from the act of redistricting—and that this is normatively undesirable.[99] Third, there are criticisms of automated redistricting

92. This at least seemed to be true in California, where the efficiency gap of the resulting map was very low. *See* Nicholas O. Stephanopoulos, *Communities and the California Commission*, 23 STAN. L. & POL'Y REV. 281, 282 (2012).

93. The 2010 Arizona commissioners, at least, observed that many of the emails they received contained largely duplicated text, and paid little attention to those emails. Druke, *Want Competitive Elections?*, *supra* note 91. If AI can generate sufficiently different, natural-sounding testimonial between fake emails, accordingly searching for duplicate text would lose reliability. David Daley also recounts an interesting example of fake testimonials given in Florida during 2011. DAVID DALEY, RATF**KED: WHY YOUR VOTE DOESN'T COUNT 125–127 (2016). It is also worth noting that map-drawing contests, like the one Ohio undertook in 2009, would also be susceptible to this kind of attack. *See* Altman & McDonald, *supra* note 62, at 100 (describing Ohio's map-drawing contest).

94. For a detailed discussion of "open redistricting" and some novel forms it has taken, I recommend David Daley's book "Ratf**cked: Why Your Vote Doesn't Count." *See generally* DAVID DALEY, RATF**KED: WHY YOUR VOTE DOESN'T COUNT (2016).

95. See Altman & McDonald, s*upra* note 62, at 73, for a discussion of the meaning of "fully automated."

96. *See* Altman, *supra* note 59, at 83.

97. *See id.* at 84; William Vickrey, *On the Prevention of Gerrymandering*, 76 POL. SCI. Q. 105, 106 (1961).

98. *See* Altman, *supra* note 59, at 84–86.

99. *Id.*

that center on the sheer computational complexity of automatically selecting maps, even with a highly-refined set of criteria.[100]

These criticisms are well taken. However, it seems reasonable to this author that automating the redistricting process should be generally revisited by policymakers and ultimately implemented. This, if only for the fact that the threats posed by AI (at least including the direct distortion and secondary distortion described above) may now prove to outweigh whatever inherent democratic virtues went hand-in-hand with the traditional redistricting process. While an agreed-upon redistricting algorithm may incidentally show some level of bias – there seems to be merit in the assurance that this bias would be predictable, measurable, transparent, and identifiable as a product of the algorithm. This, in turn, may allow policymakers to offset the bias in other ways, instead of wondering how it arose and basing their corrective actions on incomplete information. Finally, the computational complexity of automated redistricting does seem to be manageable if parameters with which to limit it are agreed upon.[101]

A. The Rawlsian Default

My proposed solution to the problem of AI is a process I will call the "Rawlsian Default." This name stems from some existing literature on redistricting,[102] which posits that one potential advantage of automated redistricting is that it helps create a version of the so-called "veil of ignorance"—a thought experiment proposed by the philosopher John Rawls.[103] Rawls asks us to imagine a situation where individuals must choose the rules of a hypothetical society prior to knowing what their position within this society will be.[104] The supposition is that the actors will thereby produce rules that are mutually agreeable to all players.[105] Automated redistricting, in theory, mimics this by forcing political actors to adopt policy positions in ignorance of how their districts could swing relative to those positions.[106] It thereby tends to maximize responsiveness to voters - the underlying virtue of voters picking their politicians rather than the reverse.

In the "Rawlsian Default," the existing rules for redistricting in given jurisdictions are left intact. States with independent commissions would be allowed to engage in the normal set of hearings and solicitations of public testimony that they would engage in during any redistricting year. Once these processes are complete and the maps finalized, the first election would proceed as usual. If,

100. *See generally* Altman, *supra* note 59. The computational complexity of fully automated redistricting is largely a function of the NP-hardness or NP-Completeness of finding "optimal" maps in any given jurisdiction. *See id.* at 91–94, 104–108. I contend that viewing the process as a kind of algorithmic "hunt" for an optimal map from a possibly infinite number of possible maps in polynomial time is a counterproductive way of thinking about this problem.

101. *See* Altman, *supra* note 59.

102. *See id.* at 87–89.

103. *See* JOHN RAWLS, A THEORY OF JUSTICE 12, 19 (1971).

104. *Id*. at 16–22.

105. *Id*. at 117–20.

106. *See* Altman, *supra* note 59, at 87.

however, the results of that election returned an efficiency gap in excess of 8%,[107] the Rawlsian Default would kick in, and the jurisdiction would be forced to hold the remaining elections in the cycle relative to a map generated through an automated process.

For this automation, I would recommend utilizing a redistricting algorithm based on the one proposed by mathematicians Lukas Svec, Sam Burden, and Aaron Dilley in 2007.[108] First, this algorithm would find areas of population prominence in a map of the jurisdiction.[109] Then, it would place *n* centroids within the *n*-most prominent census tracts in the jurisdiction, and draw Voronoi distances relative to those centroids.[110] Next, the Voronoi distances could be weighted (i.e. incrementally expanded or contracted) until equality of population of each district was achieved within a certain threshold. This would at least ensure that forms of compactness, contiguity, and equal population were attended to. The Voronoi distances may also be Manhattanized, snapped to their respective Chebyshev or "chessboard" distances, and/or limited according to existing political boundaries in the course of the weighting process for ease of administration.

Finally, for an added layer of 'ignorance,' it could be left up to random chance whether the Rawlsian Default algorithm would be seeded relative to areas of population prominence, or relative to anti-prominent population valleys. In the former, areas of large population like cities are likely to remain intact, while any "cracking" would likely occur in suburban or rural areas. In the latter condition, cities would likely be cracked and rural areas left in intact voting blocks. This would force partisans or other bad-faith actors in the redistricting process to hedge their efforts to gerrymander relative to the likelihood that one of these two possible outcomes would thwart their purposes completely. Additionally, this process would act as a kind of fail-safe against a situation where the efforts of a redistricting commission were hijacked by malevolent actors employing sophisticated AIs.

107. See Gill v. Whitford, 138 S. Ct. 1916, 1933 (2018), for discussion of a previously proposed efficiency gap that would trigger scrutiny. For districts subject to the relevant provisions of the Voting Rights Act, racial efficiency gaps—where voters of particular minority groups are found to waste their votes at high rates—may be another potential trigger for the Rawlsian default.

108. Lukas Svec, Sam Burden, & Aaron Dilley, *Applying Voronoi Diagrams to the Redistricting Problem*, 28 UMAP JOURNAL 313 (2007).

109. Prominence itself can be measured in several ways, but this supposes a mutually agreed upon definition of prominence.

110. *See* Svec, Burden & Dilley, supra note 108. A Voronoi Distance, in simple terms, involves marking at least two points on a plane, and then diving the plane where the points are equidistant. Doing this for multiple points creates divisions of the plane that are optimally compact relative to the points. For an online tool useful in visualizing Voronoi Diagrams, see the Interactive Voronoi Diagram from Sarah Y. Greer's blog. Sarah Y. Greer, *Interactive Voronoi Diagram*, SARAH Y. GREER (Oct. 6, 2017), http://www.sygreer.com/projects/voronoi/.

V. CONCLUSION

The ultimate effect of AIs on the 2020 redistricting cycle and the cycles after that remains anyone's guess. However, given what we already know about AI, it seems reasonable to say that any failure by courts to promulgate clear, quantifiable legal standards for gerrymandering is now more ill-advised than ever. AI has created (and will continue to create) an unprecedented need for quantitative specificity in this area of law. The more that courts, legislatures, and other policymakers balk at writing quantitative gerrymandering rules, the more costly and destructive the effects of AI have the potential to become.