Social and Ethical Attitudes Towards Trait Prediction via Genomic Samples

A Research Paper submitted to the Department of Engineering and Society Presented to the Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the
Degree Bachelor of Science, School of Engineering

Paul Vann
Spring, 2023
Department of Computer Science

Signature _____ Date _____
Paul Vann

Signature _____ Date _____
Joshua Earle, Department of Engineering and Society

Introduction

Over the last few decades, genetic research has expanded significantly. Companies like 23andMe and Ancestry.com use DNA samples to trace back family trees, find lost family members, and identify people's origins through DNA tracing. In the field of medical research, DNA samples are being used to identify specific medications for individuals that fit that person's specific health needs. Furthermore, law enforcement agencies continue to use genomic samples from crime scenes to attempt to identify criminals, keeping databases of DNA to hopefully identify criminals in the future. Now a new use for genetic data has emerged that likely has the most invasive impact on genetic privacy: trait-predicting algorithms. Trait-predicting algorithms take genetic information as input and can output specific phenotypes or physical features associated with that genetic data. More specifically this means that with access to a person's genetic information, someone could predict what a person looked like, or what certain physical traits they may have (Pośpiech 2022).

The upside to these trait-predicting algorithms is the benefit that they can have from a law enforcement and forensic investigation standpoint. With an algorithm like this, law enforcement agencies would no longer need to have a criminal's DNA in their database to classify them. Rather, law enforcement could be able to use the algorithm to identify key features of a suspect such as their hair color, or eye color. While this isn't fool-proof, when used correctly, it could be a step in the right direction towards crime fighting as well as preventing innocent people from going behind bars. Consider an algorithm that can predict eye color, skin color, and hair color for example. While this algorithm cannot paint a perfect picture of a culprit, it is still able to significantly narrow down the pool of criminals who may have committed the crime.

On the other hand of all of this though are the ethical implications of a trait-predicting algorithm. Genetic data is extremely sensitive and private and because of this, it raises the question of how ethical a trait-predicting algorithm is. For one, a trait-predicting algorithm requires significant genomic samples with related traits to train the algorithm. This means a lot of people would have to release genomic data, running the risk of making it public. Another issue is that in the wrong hands, the algorithm could be used to identify who specific genetic samples belong to which is a very big violation of privacy (Erlich, 2014). Another major ethical issue is the possibility that such an algorithm becomes biased towards specific racial groups. The way that technology behaves is majorly determined by its creator, and when created by individuals with poor intentions, technology can in turn reflect those poor decisions (Benjamin, 2019). For example, imagine a trait predicting algorithm with the purpose of identifying criminals, thats training is weighted on a specific racial group. That algorithm would in turn make biased decisions that negatively impact that group (Benjamin, 2019).

Methods

In order to evaluate the current state of genetic privacy laws and regulations within companies and governments, as well as get individual's opinions on the matter of trait predicting algorithms, there are two methods: A legislation evaluation and a Poll of University of Virginia Students.

Legislation and Regulation Evaluation

This method is the evaluation of legislation and regulation surrounding genetic privacy. Both governmental and company-specific regulations will be evaluated for their effectiveness of protecting genetic information and protecting the distribution of genetic information. More specifically, the United States legislation as well as specific State legislation will be evaluated.

Because all 50 states' legislation cannot be evaluated, California's legislation will be specifically looked at as they are one of the most progressive states in genetic privacy legislation. Furthermore, 5 companies who handle genetic information will be evaluated for their genetic privacy regulations. These companies include:

- 23andMe
- Futura Genetics
- Veritas Genetics
- Myriad Genetics
- MyDNA

These companies all handle genetic information in some manner, whether that be individuals sending in samples for lineage analysis, or for medical purposes and research. The evaluation of these companies' regulations, as well as governmental legislation will take the form of numeric rankings (from 0-10) in the following categories: Anonymity, Sharing, Storage, and Privacy Promises in regards to genetic information. Each of these categories will be ranked based on the way they are conveyed in each unique regulation. These numeric evaluations will be followed by a conclusive analysis, highlighting the overall state of genetic legislation and regulation.

Poll of Students at UVA

The main goal of this poll is to evaluate as a whole the overall opinion of students at the University of Virginia on trait predicting algorithms, as well as the sharing and privacy of genetic information and data. The poll is framed to be fairly short so that participants can quickly complete it, however the questions are written in order to get the most valuable and accurate information. There are 7 questions on the poll, all framed as a "On a scale of 1-10" question. At

the beginning of the poll, students will be given context on what a trait predicting algorithm is and any other information that is reasonable to provide but does not impact the results of the study. The questions on the poll are the following:

- On a scale of 1-10 how likely would you be to share your DNA sample with a company like Ancestry.com or 23andMe?
- How likely on a scale of 1-10 would you be to share your DNA sample today if it were to be used to create a breakthrough in genetic technology?
- On a scale of 1-10 how uncomfortable would you be if your anonymized DNA sample was shared with a private company if a trait predicting algorithm did NOT exist?
- On a scale of 1-10 how uncomfortable would you be if your anonymized DNA sample was shared with a private company if a trait predicting algorithm did exist?
- On a scale of 1-10 how strongly would you like there to be legislation AGAINST a trait predicting algorithm?
- On a scale of 1-10 how strongly would you like there to be innovation in the field of trait predicting algorithms?
- On a scale of 1-10 how well do you think genetic privacy laws currently protect United States citizens to the best of your knowledge?

The results of this poll will be synthesized and conclusions will be discussed in the results section.

STS Frameworks

In regards to trait predicting algorithms and genetic privacy laws, this paper takes a look at technological determinism and social construction of technology. Both of these STS frameworks frame the issue of genetic privacy and technological bias very well in their own

ways, helping to better understand the potential best approach to a middle ground between privacy and innovation. As a whole the end goal in regards to this issue is to be able to work with and understand trait predicting algorithms, while also being able to guarantee and protect individual's privacy.

Technological Determinism is a theory that suggests that technology and the way it is designed and constructed has an impact on societal and cultural development (Smith, 1994). More specifically to the issue at hand, the way that an algorithm is built directly has an impact on societal and cultural development. Genetic research really started to blossom in the 20th century with the discovery of the double helix by Watson, Crick, and Franklin. Past that point, genetic information and engineering has become an instrumental part of society, more than most people realize. Into the late 20th century and 21st century specific research on individual genomes and genetic data began taking place, investigating how to predict traits from an individual's genome (Pośpiech, 2022). In the early 21st century GMOs (Genetically Modified Organisms) began being sold in grocery stores as they yielded more food and were overall cheaper to produce (National Geographic, 2022). As discussed earlier, companies like 23andMe emerged analyzing individuals' DNA and tracing their lineage back to certain locations and nationalities. However it is important to note that there have been controversies surrounding the accuracy of such genetic tests, sparking a lawsuit in 2013 against 23andMe (Munro, 2013). As a whole, the emergence of genetic engineering in the 20th century changed how we as humans eat, learn more about ourselves, and add to genetic research as a whole. Therefore in reference to genetic algorithms, it is important to consider how a tool like a trait predicting algorithm is built, as the impact that could follow its creation could be much more significant than just impacting forensic analysis or biology. Even more important is to consider what this impact will be, and whether it will be

positive or negative. While GMOs have yielded much higher food growth and cheaper food, they are also not good for the ecosystem or the human body compared to natural foods (National Geographic, 2022). So while some see this societal impact of genetic research as a good thing, it can also be seen as bad. For example, the Center for Food Safety looks to limit the sale and production of genetically engineered foods, and encourages consumers to limit their purchasing of them (Center for Food Safety, 2022).

Social construction of technology is a theory that suggests the use and development of technology is driven by social and cultural factors, causing some technologies to be very socially-positive and others to be very socially-negative (Yousefikhah, 2017). In reference to genetic algorithms such as a trait predicting algorithm this is extremely important. Consider a trait predicting algorithm that can almost perfectly predict an individual's facial features. When used with good intentions, this could be used by law enforcement agencies to heavily reduce violent crime and make the United States, and other countries around the world a safer place. This indicates a socially-positive use of the technology, and a reason why heavy legislation may not be the best choice for such an algorithm. On the other hand though, when used with bad intentions, a malicious person who's gained access to anonymized genomic samples could use such an algorithm to deanonymize every genomic sample. This would compromise the privacy of every single individual whose genomic sample that individual had obtained, and could potentially indicate a lot of information about that person's health in the future. This is an example of a socially-negative use of the technology, and a reason why many individuals have discussed heavy legislation of these algorithms. As a whole, both the positive and negative sides are possible scenarios and neither can be avoided. The question at hand is how best to handle this dilemma and also, what are people's thoughts about the matter.

Background

Genetic Algorithms

Genetic trait-predicting algorithms are very new in the field of genetic research and computer science. Trait prediction has been done manually for many years by genetic researchers analyzing specific components of a person's genome to determine phenotypes such as eye color and hair color. However, these manual processes are very time-consuming and are only able to capture a very few of the simplest phenotypes. With recent advances in machine learning these algorithms have become possible (Kristin, 2009). Machine learning algorithms can be trained or sometimes even train themselves on data on a specific task, getting better and better over time. Machine learning models can look very different from one another, with some being neural networks and others being reinforcement learning algorithms. These models require input data, predicted output data, and a model structure to be effectively trained. Prior to recent years, machine learning algorithms have been able to produce better results than the manual processing of genetic data. However, it was not until recently with the rise of neural networks that these algorithms started to become possible.

Neural networks are a newer field in machine learning that involve giving a machine a set of input data and passing it through "neurons", similar to those in the brain, until they are able to solve a given problem (Holbrook, 2022). This type of learning is perfect for trait-predicting algorithms because over time the machine will slowly learn where to look in the genomic data to identify certain phenotypes. Genetic data factors into these algorithms in two ways. When training a trait-predicting algorithm, the model needs two sets of data. A set of genomic data, and a set of corresponding traits (phenotypes). For example, one data point would be a person's genetic information and an image of their facial features, such as eye color, hair color, etc. This

genetic information is then passed into the algorithm as well as the predicted output. If the algorithm can correctly predict the facial features, then positive feedback is given, and if not then negative feedback is given (Yoo, 2022). This cycle continues until almost no negative feedback is given, and the algorithm has a very high accuracy rate.

One major concern with these algorithms is that the data that is used to train them could be leaked, or be made public. An example of this has been pointed out by cybersecurity company Kaspersky, which highlights the risk of using personal data to train machine learning models (Root, 2023). However, there are privacy algorithms and controls that can be put in place for both reinforcement learning algorithms and other machine learning algorithms. One of these privacy models is called Differential Privacy and is a mathematical framework for guaranteeing the privacy of training data within a model (Dwork, 2006). Methods like this and others can be used to significantly lower the risk of leaking genetic training data.

Privacy Legislation

As genetic algorithms have been growing in popularity, legislation has increased significantly over the last few decades. Both federal and state governments have made a point to regulate the use of genetic data and in some cases genetic algorithms. Many States have begun implementing legislation that requires companies that use genetic data to maintain certain privacy standards. This includes being transparent with how they keep genetic data secure, maintaining certain security measures, and receiving written consent from donors of genetic samples (Annas, 1999). Take the GIPA or Genetic Information Protection Act which was signed into law in California in 2021 for example. This act gives individuals the ability to access and delete their genetic information, as well as know how their data is being used within that company (Compton, 2021). Many other states have implemented similar regulations and laws

that prevent companies like 23andme and Ancestry.com from having free reign over citizens' data. Currently, 24 states in the United States legally require informed consent to release or use genetic information that was not requested by the individual (Johnson, 2023). This has been a growing trend in the last decade, as state governments have realized that genetic information is going to continue to be used for more and more applications.

Genetic legislation does not just exist at the state level but also extends to federal protections. While not specific to genetic law, HIPAA protects all medical data that is not anonymized or cannot be anonymized. Therefore regarding genetic law, because everyone has a different genomic sequence it is very difficult to anonymize, meaning that HIPAA provides some pretty strong protections in terms of keeping genetic data private (NIH, 2021). Other than HIPAA, there is not any actual federal law that protects the privacy of an individual's genetic information when given to a company such as 23andMe. However, companies like 23andMe are not able to make any misleading claims regarding what they will do with the data, otherwise, they can get into trouble with the FTC (NIH, 2021).

As discussed above, there are varying levels of legislation on genetic privacy but as a whole there is fairly strong protection on the use of genetic data. However, this does not include much legislation on the trait predicting algorithms and their use. There are many stances and arguments for and against laws against these algorithms. One of the main arguments against strong legislation that negatively affects these algorithms is the benefit that it could do for police organizations and crime fighting. In the United States there are over 60,000 murders unsolved each year, many of which have DNA from the perpetrator(s) on the scene (Hargrove, 2021). With algorithms like these, the number of murders that go unsolved in the United States each year could drop significantly. Furthermore, there are a lot of really interesting research use cases for

these algorithms that don't even involve humans. For example, there are new fossils and DNA samples found every year by archaeologists and biologists that an algorithm like this could be used to study and predict more what certain species may have looked like.

On the other hand however, there are many arguments for why there should be strong legislation against these algorithms. One of these arguments is because there are so many means of sharing genetic information in today's day and age. With all of these genetic information companies, and medical companies collecting genetic data, there is simply too much genetic information floating around our very interconnected world to have an algorithm that connects this genetic information back to a person. Another argument for strong legislation against these algorithms is the fact that it involves a lot of genetic data for training. This is in combination with the fact that a lot of individuals are skeptical of the privacy guarantees that differential privacy and other algorithms have to offer. As a whole, both sides are valid arguments and until legislation is created on these algorithms we will not be able to determine which side is correct. These next sections will explore viewpoints on trait-predicting algorithms as well as an evaluation of current legislation in terms of trait-predicting algorithms.

Bias in Genetic Algorithms

Another major issue with trait predicting algorithms is the ability for bias to be introduced to them. As discussed earlier, the way that technology behaves is dependent on the way it was built, and more specifically, who built it. This bias is prevalent in technologies all around us on a day to day basis. In Simone Browne's, *Dark Matters: On the Surveillance of Blackness*, she highlights the many ways in which surveillance both in the past and through technology in the present has a bias towards blackness (Browne, 2015). Browne notes that not only has blackness been a key factor in surveillance outside of technology, but also in

technologies like airport surveillance where TSA agents and security monitor for suspicious activity with a bias on certain racial groups (Browne, 2015). This ties into trait predicting algorithms in that if a trait predicting algorithm is trained with a weight on specific racial groups, it could be used to incorrectly classify a suspect in a crime based on their race, something that we have seen being done even without the use of technology in recent times. Furthermore, if any data from such an algorithm is manipulated or tampered with, the algorithm's output can change drastically, potentially affecting a specific group of people negatively (Benjamin, 2019).

This idea of bias in algorithms ties into the legislation discussed above as well. As O'Neil notes in her piece, *Weapons of Math Destruction*, algorithms in today's day and age are unregulated and opaque in the way that they operate (O'Neil, 2016). Without proper regulation of technologies like trait predicting algorithms, or the ability to see how they are actually trained or operate, we lose the ability to see whether there is inherent bias in these algorithms and further risk the ability for this bias to be present. Furthermore, the introduction of genetic data into algorithms as a whole in the past has caused much furor surrounding race. As genetic algorithms begin to provide more information about people's backgrounds, their race, and impact their future, it is changing the way that many people view both their past and their future (Nelson, 2018).

As a whole, racial bias is something that is very important to consider when considering the ethicality of trait predicting algorithms. While there is much room for these algorithms to have a positive impact, if built incorrectly or deliberately designed with bias, they can do much more harm than good.

Results

Legislation Evaluation

The following table highlights the results of the legislation and regulation analysis discussed above. Each of the scores were determined by looking through specific documentation, regulation, and legislation pertaining to that entity's genetic privacy policies. The following describes the meaning of each numerical ranking from 1-10, in regards to how the score was derived:

1-3: A score of 1-3 is given for one of these categories if there is specific documentation within the organization's policies or legislation that go against this category. For example, an organization may receive a 1 in anonymity if there is a policy that states that genetic data will not be kept anonymous in any way. To determine where in this range a score is given, we look at how severely the policy goes against the category.

4-6: A score of 4-6 is given for one of these categories if there is not specific legislation or policies going against the category, but there is also not very much legislation in support of the category. For example, an organization may receive a 5 in sharing if they don't indicate anywhere that they will share your data, however also don't indicate anywhere that they don't intend to share data or have very little information there. To determine where in this range a score is given, we look at whether there is more documentation for or against the category in the policy.

7-10: A score of 7-10 is given for one of these categories when there is specific legislation or policy that supports a category. For example, an organization may receive a 10 in storage if they specifically state that all genetic data will be encrypted with the most up to date encryption algorithm, and access even from employees will be prevented. To determine where in this range a score is given, we look at how well the policy supports the category.

A score of "-" simply means that we did not have, or could not find enough information to derive a reasonable result.

	Anonymity	Sharing	Storage	Promises
U.S. Federal Government	6	2	2	2
California State	7	8	7	7
23andMe	7	10	10	8
Futura Genetics	5	7	7	5
Veritas Genetics	-	9	7	7
Myriad Genetics	-	5	7	5
myDNA	10	10	8	9

Fig. 1 Genetic Privacy Legislation and Regulation Evaluation

As a whole, aside from United States federal regulations relating to genetic privacy, every company evaluated and California state legislation ranked fairly well in terms of giving individuals basic privacy protections. However there were a few important trends that I noticed in the data that were quite interesting. Firstly, while some companies like 23andMe were very open about letting individuals control how their data is used, other companies outline in their regulations that they can use this genetic information for internal and in some instances, external research. While this may not be a concern for some people, especially since the data is anonymized, a trait predicting algorithm could make it so that this data that is being shared externally becomes de-anonymized, threatening an individual's privacy. This leads to my second point, that some of these companies seem to tread a fine line right along with what federal and state legislation requires of them. For example, many of them only provide slightly more protection than federally regulated HIPAA. On the other hand however, companies like myDNA

provide extensive protections for individuals, guaranteeing encryption, anonymity, and even limiting employee access to data. Moving forward, this is the direction that federal and state legislation should go in in order to protect genetic privacy in the future.

Poll of Students at UVA

The following diagrams highlight the results of the UVA student poll. 35 students responded to this survey over the course of two days and were only provided a definition of trait predicting algorithms along with the questions.

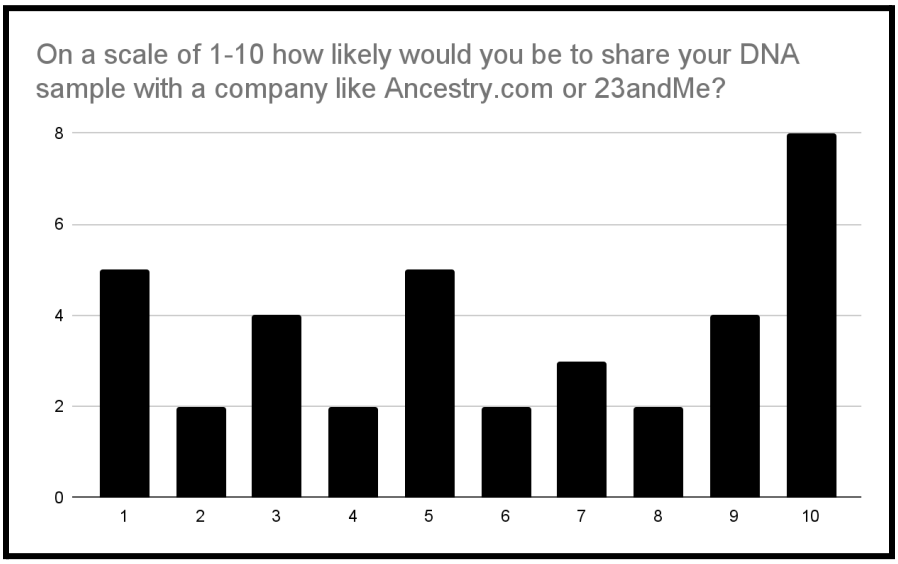

Fig 2. Poll Question #1 Results Histogram

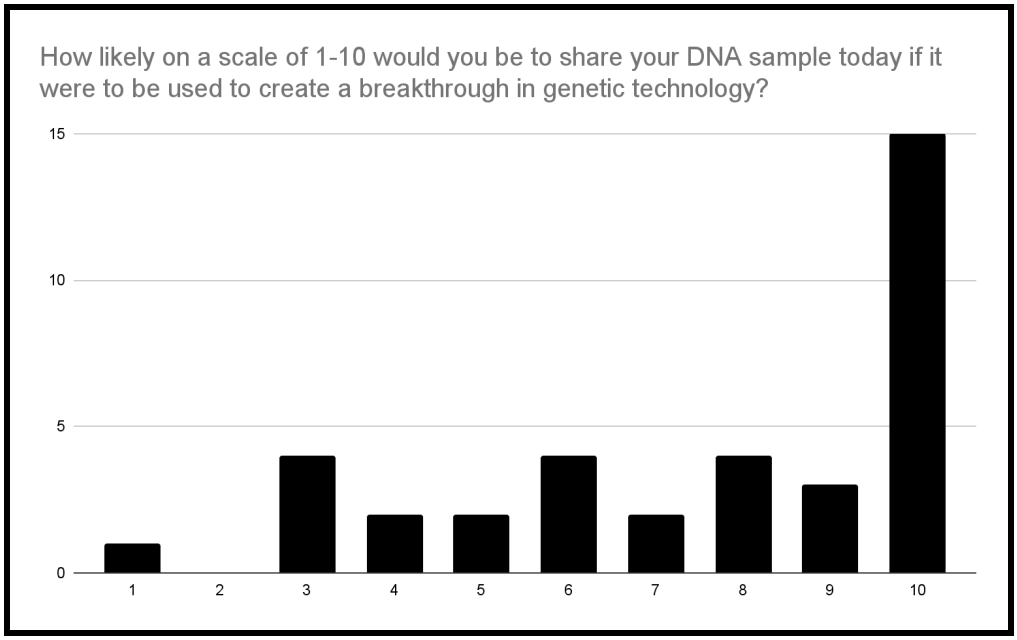

Fig 3. Poll Question #2 Results Histogram

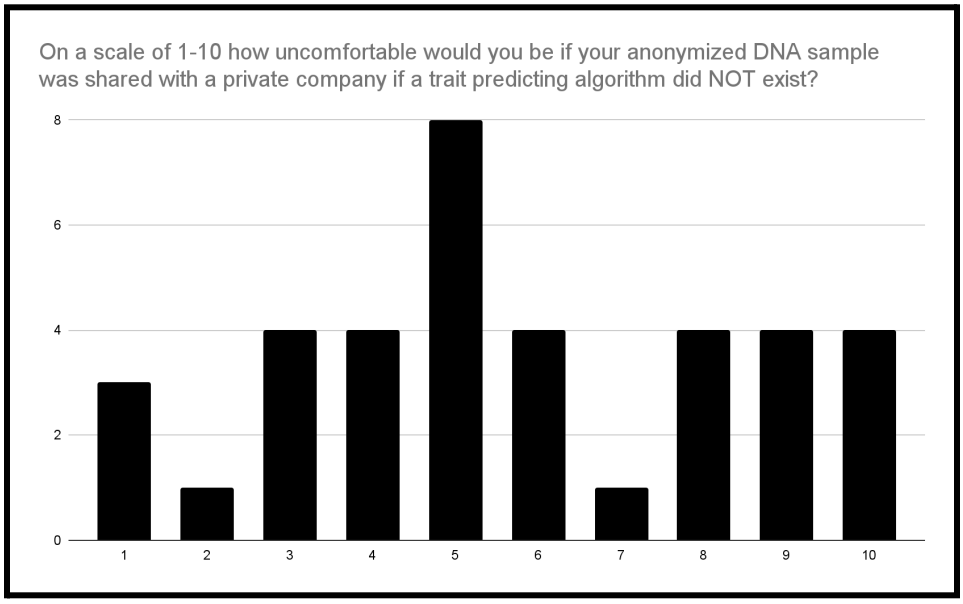

Fig 4. Poll Question #3 Results Histogram

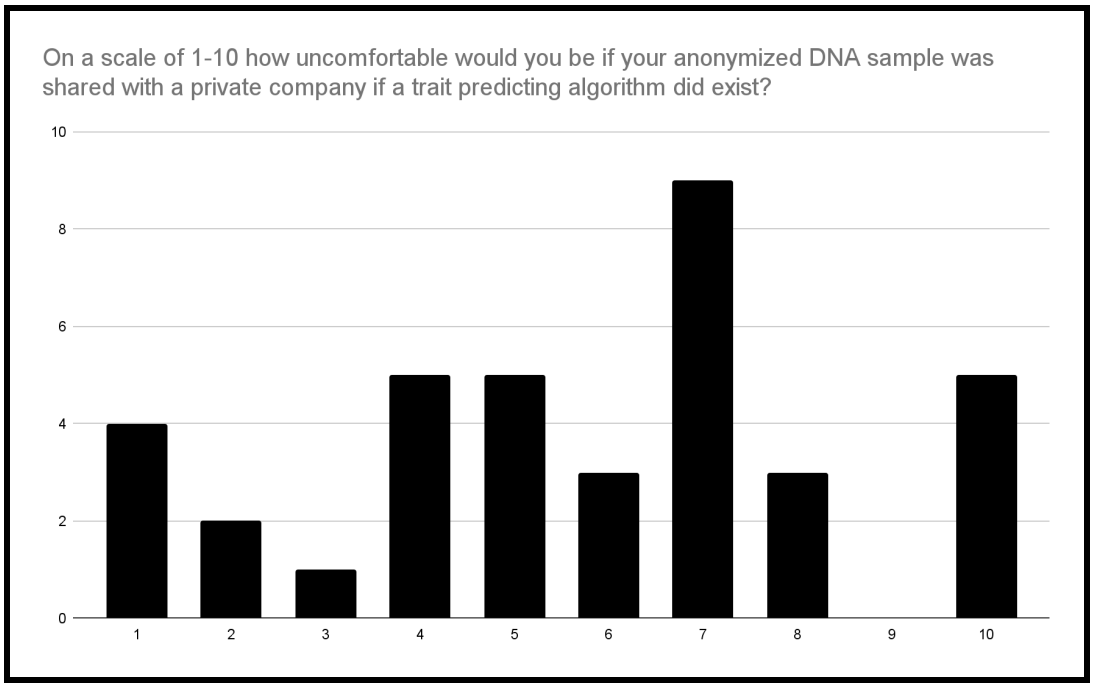

Fig 5. Poll Question #4 Results Histogram

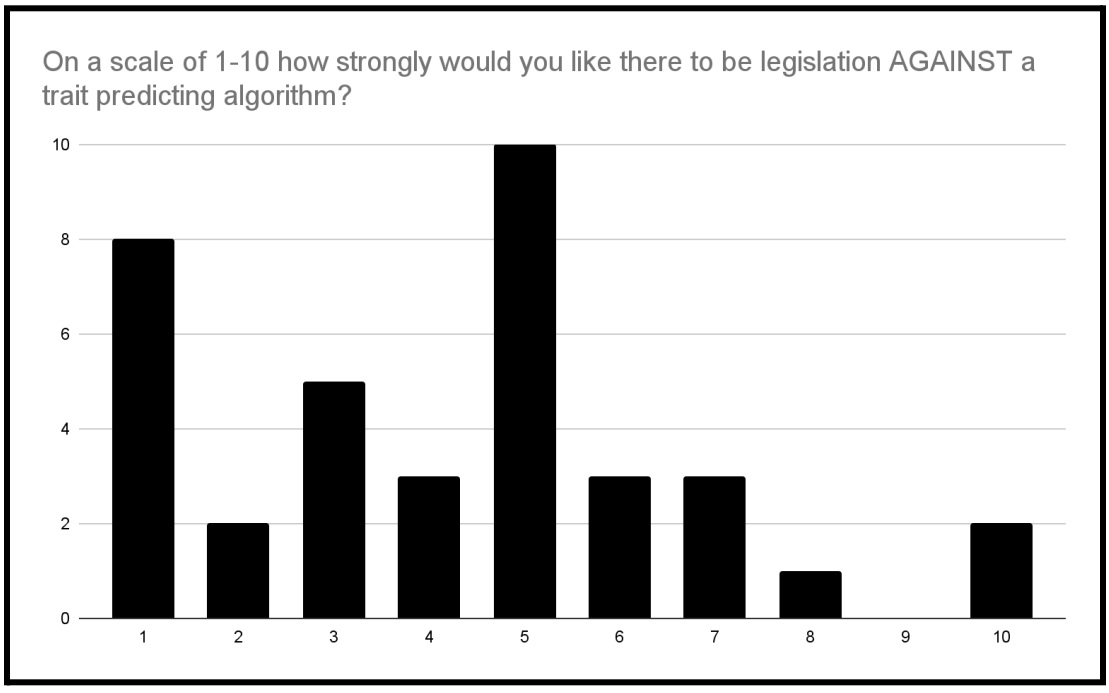

Fig 6. Poll Question #5 Results Histogram

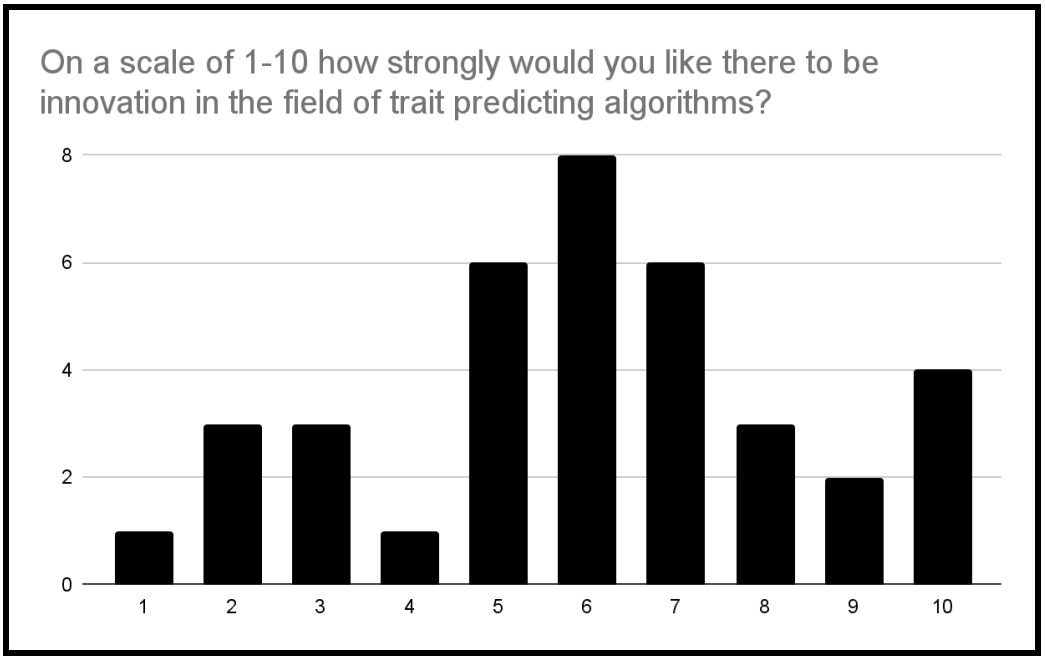

Fig 7. Poll Question #6 Results Histogram

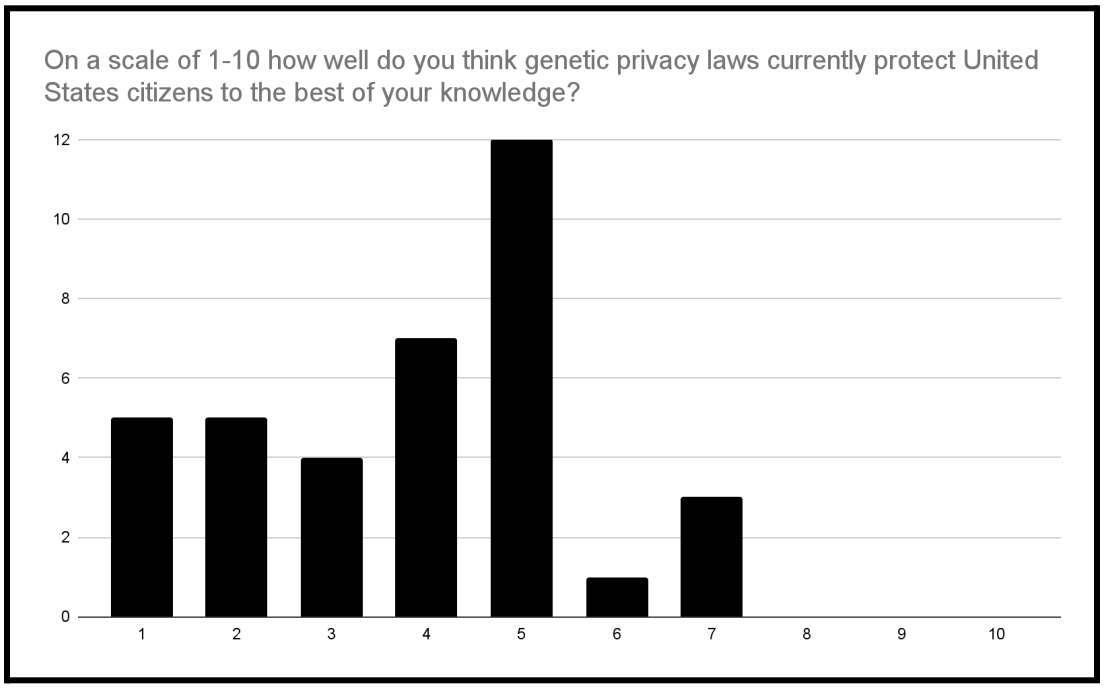

Fig 8. Poll Question #7 Results Histogram

As portrayed above, this poll revealed numerous interesting trends about UVA student's stances on genetic privacy, trait predicting algorithms, and privacy legislation. There were three

main trends that I identified as important from this poll. The first is that students seemed to be more uncomfortable with their anonymized data sample being present if a trait predicting algorithm existed. Looking at Fig. 4, you see a normal distribution with the average response being approximately 5 when asked: “On a scale of 1-10 how uncomfortable would you be if your anonymized DNA sample was shared with a private company if a trait predicting algorithm did NOT exist?”. However in Fig 5, when asked about if a trait predicting algorithm did exist, this distribution shifted slightly to the right indicating that a trait predicting algorithm would make them slightly more uncomfortable.

This leads to the next important trend from the poll which is that students seemed to be against legislation against trait predicting algorithms, and in favor of innovation in the field in fact (Fig. 6, 7). This trend along with the first one can be supported by the social construction of technology STS framework. It seems like students who participated in the poll see the negative ways that the technology could be used to de-anonymize anonymized genetic data, but also see the positives of such a technology and the positive impact it could have on society.

Finally, the last important trend that I noticed in the polling data was the lack of confidence in genetic privacy legislation (Fig. 8). As seen in the legislation evaluation, it is already clear that there is more that can be done in terms of genetic privacy regulation and legislation. However, to further that point, the results from this poll reveal that there is much more to be done from society’s perspective as well, and not just a research perspective.

Conclusion

In conclusion, genetic information and data will continue to be used extensively in many fields, including in the field of trait predicting algorithms. There are both positives and negatives to such an algorithm that target issues of both innovation and ethicality. Evaluating the current

legislation and regulation surrounding genetic privacy provides significant insight into how this legislation can be improved and where it currently stands. We see that on a federal level legislation is lacking more than anywhere else, and that state governments are starting to take action, but definitely can do more to protect individuals' genetic privacy. We also see that some companies provide very significant privacy protections while others are treading right along non-strict federal regulations. Furthermore, by polling University of Virginia students on their opinions on genetic privacy, genetic legislation, and trait predicting algorithms, we can make informed decisions to promote socially positive outcomes of a trait predicting algorithm. It can be concluded that students are non-confident in current genetic privacy regulations and aren't sure about sharing their anonymized genetic information, especially in the presence of a trait predicting algorithm. However it can also be concluded that students want to see innovation in this field. The most important thing to consider is that a technology like trait predicting algorithms has many avenues to be misused, in terms of both privacy and bias. While it could be very advantageous to see such an innovation come to fruition, it is important that both governments and organizations further regulate these algorithms and become more opaque about how they are designed, and weighted. Otherwise, it is impossible to determine if the way they are functioning is ethical.

Bibliography

Annas, G. J. (1999). Genetic Privacy: There Ought to Be Law. Texas Review of Law & Politics, 4(1), 9-16.

Azencott C.-A. 2018 Machine learning and genomics: precision medicine versus patient Privacy Phil. Trans. R. Soc. A.3762017035020170350 http://doi.org/10.1098/rsta.2017.0350

Benjamin, R. (2019). Race after technology: Abolitionist tools for the new Jim code. Polity.

Browne S. (2015). Dark matters : on the surveillance of blackness. Duke University Press.

Center for Food Safety. (2022). About Ge Foods: | about Genetically Engineered Foods. Center for Food Safety. Retrieved April 29, 2023, from https://www.centerforfoodsafety.org/issues/311/ge-foods/about-ge-foods

Compton, L., & John, S. (2021). California's Senate Bill 41: The Genetic Information Privacy Act. Mintz. Retrieved March 11, 2023, from https://www.mintz.com/insights-center/viewpoints/2826/2021-10-19-californias-senate-bill-41-genetic-information-privacy

Dwork, C. (2006). Differential Privacy. In: Bugliesi, M., Preneel, B., Sassone, V., Wegener, I. (eds) Automata, Languages and Programming. ICALP 2006. Lecture Notes in Computer Science, vol 4052. Springer, Berlin, Heidelberg. https://doi.org/10.1007/11787006_1

Ellen Wright Clayton, Barbara J Evans, James W Hazel, Mark A Rothstein, The law of genetic privacy: applications, implications, and limitations, Journal of Law and the Biosciences, Volume 6, Issue 1, October 2019, Pages 1–36, https://doi.org/10.1093/jlb/lsz007

Erlich Y, Williams JB, Glazer D, Yocum K, Farahany N, Olson M, et al. (2014) Redefining Genomic Privacy: Trust and Empowerment. PLoS Biol 12(11): e1001983. https://doi.org/10.1371/journal.pbio.1001983

Geographic, N. (2022). Genetically Modified Organisms. National Geographic Education. Retrieved April 29, 2023, from https://education.nationalgeographic.org/resource/genetically-modified-organisms/

Gostin, L. O., & Hodge, J. G. (1999). GENETIC PRIVACY AND THE LAW: AN END TO GENETICS EXCEPTIONALISM. Jurimetrics, 40(1), 21–58. http://www.jstor.org/stable/29762629

Hargrove, T. (2021). Cold case homicide stats - project: Cold case. Project Cold Case.

Retrieved December 16, 2022, from https://projectcoldcase.org/cold-case-homicide-stats/

Holbrook, R. (2021). Learn Introduction to Deep Learning. Kaggle Learn. Retrieved April, 3, 2023, from https://www.kaggle.com/learn/intro-to-deep-learning

Johnson, A. (2023). State Genetic Privacy Laws. State Genetic Summary Table on Privacy Laws. Retrieved March 11, 2023, from http://pierce.wesleyancollege.edu/faculty/hboettger-tong/docs/hbt%20public%20folder/FYS/State%20Genetic%20Summary%20Table%20on%20Privacy%20Laws.htm

Kristin K. Nicodemus, James D. Malley, Predictor correlation impacts machine learning algorithms: implications for genomic studies, Bioinformatics, Volume 25, Issue 15, 1 August 2009, Pages 1884–1890, https://doi.org/10.1093/bioinformatics/btp331

Muhammad Naveed, Erman Ayday, Ellen W. Clayton, Jacques Fellay, Carl A. Gunter, Jean-Pierre Hubaux, Bradley A. Malin, and Xiaofeng Wang. (2015) Privacy in the Genomic Era. ACM Comput. Surv. 48, 1, Article 6 https://doi.org/10.1145/2767007

Munro, D. (2013). Class action law suit filed against 23andMe. Forbes. Retrieved April 29, 2023, from https://www.forbes.com/sites/danmunro/2013/12/02/class-action-law-suit-filed-against-23andme/

Nelson A. (2018). The social life of DNA: racial reconciliation and institutional morality after the genome. The British journal of sociology, 69(3), 522–537. https://doi.org/10.1111/1468-4446.12607

NIH. (2021). Privacy in Genomics. Genome.gov. Retrieved March 11, 2023, from https://www.genome.gov/about-genomics/policy-issues/Privacy

O'Neil, C. (2017). Weapons of math destruction. Penguin Books.

Pośpiech E, Teisseyre P, Mielniczuk J, Branicki W. Predicting Physical Appearance from DNA Data—Towards Genomic Solutions. Genes. 2022; 13(1):121. https://doi.org/10.3390/genes13010121

Root, E., Aver, H., Kaminsky, S., & Team, K. (2023). Neural networks reveal the images used to train them. Kaspersky Daily. Retrieved April 29, 2023, from https://www.kaspersky.com/blog/neural-networks-data-leaks/47992/

S. Jha, L. Kruger and V. Shmatikov, "Towards Practical Privacy for Genomic Computation,"

2008 IEEE Symposium on Security and Privacy (sp 2008), 2008, pp. 216-230, doi: 10.1109/SP.2008.34.

Smith, M. R., & Marx, L. (Eds.). (1994). Does technology drive history?: The dilemma of technological determinism. Mit Press.

Yoo, H. Y., Lee, K. C., Woo, J. E., Park, S. H., Lee, S., Joo, J., Bae, J. S., Kwon, H. J., & Park, B. J. (2022). A Genome-Wide Association Study and Machine-Learning Algorithm Analysis on the Prediction of Facial Phenotypes by Genotypes in Korean Women. Clinical, cosmetic and investigational dermatology, 15, 433–445. https://doi.org/10.2147/CCID.S339547

Yousefikhah, S. (2017). Sociology of innovation: Social construction of technology perspective. AD-minister, (30), 31-43.