

POWER / PLAY

Listening Post is an experiment in taking the apparatus of surveillance technology and repurposing its mechanisms for the intention of play rather than the reinforcement of power. It raises questions about the role of surveillance in contemporary society and it explores the ways in which the tools of the trade are inherently biased and flawed. Simultaneously, it investigates the theory that these tools can also be applied as open-ended techniques to produce a multitude of interesting outcomes, apart from the purposes for which the tools were originally designed; they are mounds of clay of a certain texture and consistency, but are ultimately formless, waiting to be shaped by a sculptor.

In the following pages I will discuss some of the fundamentals of machine learning, and will introduce the concept of inductive bias as an inextricable component in the framework of intelligent computer systems. I will examine how this bias represents an abstract danger with very real social and political consequences. Finally I will talk about how my project *Listening Post* attempts to invert this construct to transform a medium of control and coercion into a medium for creative praxis.

Machine Learning

There are many approaches to designing artificially intelligent systems, but every one of those systems is a model designed by an engineer to reflect how s/he believed a specific type of human knowledge could be most effectively represented to and learned by a computer. As models, each of these AI systems necessarily contains a simplified abstraction of reality, condensed according to the designer's perspective.

For the computer scientist, learning is generalization from experience in a way that improves a system's performance on successive iterations of training and evaluation (Luger, 352). These systems work by beginning with a large dataset; for example, when competitors for the Netflix Grand Prize¹ set out to find the best algorithm for learning users' movie preferences, they created models from a database of 100 million ratings from 480,000 users (Koren, 1). After an initial analysis of the data, the engineers attempt to decide which type of learning model will be most appropriate for the type of information under analysis and the desired outcome. Once the learning framework is determined, an algorithm is chosen and software is written. The software initially creates an a priori model that goes through a period of training, during which it is presented with example after example from the dataset. This type of repetition learning is similar to how a child will prepare for a spelling bee by reciting spellings for the same words over and over until they consistently get most of them correct. From the experience of repetition, the model becomes shaped to group types of data into categories based on similarity -- for example, pronunciations of the word "apple," or users who like the movies *Chinatown* and *Taxi Driver*. As the model sees more and more data, it makes stronger boundaries or correlations between data points, and eventually converges at a point where the rate at which it is learning has decreased significantly, indicating that little information remains which it is capable of extracting from the dataset.

After training comes a period of evaluation. Usually with any large dataset, the majority of the data is used for training and a small fold of the data is "held out" and used for evaluation. In the case of the Netflix contest this was an additional 4.2 million ratings (Koren, 1). The hope for

¹ The Netflix Grand Prize was a contest, launched in 2006, which challenged competitors to beat the existing Netflix "cinematch" algorithm at predicting user ratings of movies by a margin of 10%. The winner, "BellKor's Pragmatic Chaos," was announced in June of 2009 and received a prize of 1 million dollars.

evaluation is that the model has generalized well enough from the training data to successfully categorize new data of the same format.

If evaluation results are poor, the engineer will generally re-assess the algorithm chosen or some of the parameters used, and experiment until good numbers are achieved. The numbers say little about how a model will handle unexpected and messy data from real life, just as students who have learned how to take tests rather than engage with the learning process often have little understanding of how to implement their knowledge in the real world.

Inductive Bias

The type of machine learning discussed above is known as generalization and by definition involves *induction*, or the theorization of general facts from particular experiences. Classic examples include “All ice I have ever touched is cold, therefore all ice is cold.” Or “all swans I have ever seen are white, therefore all swans are white.”

At the foundation of machine learning algorithms is the idea that individuals have categories in their minds that allow them to make generalizations from their world of particular experiences; machine learning rests on the principles that we have a world of essential forms describing the world around us, impressed upon our minds through daily activities, and that this process of impression can be accomplished quickly and efficiently by a computer. The perspective of the computer program, therefore, is just as subjective as that of a human being. Just as we are shaped by our experience, for example of how English is spoken, or what suspicious people look like, the computer necessarily has the same set of biases and stereotypes built in. The philosophical quandary known as the Empiricist’s dilemma, which dates back to the days of Plato, asks not only how we can learn anything, but also how we can even know that we have learned at all, without some form of bias. I will not attempt to tackle this epistemological minefield here, but I want to point out that the same questions and prejudices that apply to human learning are applicable to an even greater extent in machine learning, where a human engineer, with all her biases and conceptions of the world, shapes a whole system in her likeness.

These learning algorithms are like recipes. They contain some mathematical equations and some adjustable parameters, but by and large the most important component of an AI system is the input data. As with human beings, but to a much greater extent, the model’s initial knowledge of the world shapes how it will continue seeing the world forever after; if it is trained to recognize types of trees all the world will be trees, and a person or a bird will just appear to the model as a type of unrecognized tree, to be grouped into the default “other” category. A system trained to parse text and extract grammar will only be able to recognize the parts of speech and grammatical structures it has seen before.

Furthermore, if an AI system is trained in Europe to recognize swans it will learn that *all swans are white*. Now suppose the Australian government licenses this technology to use at home, assured that it has a 95% accuracy point. Will the software ever recognize a single Australian black swan? The answer really depends on how the system was designed, how many features it abstracted, and how important the experience of whiteness is to the model’s impression of the essential swan.

The algorithms chosen, the initial input data selected, and the way this information is transformed into bits and presented to the learning model form what is called the *inductive bias* of the system. This bias is the crucial inculcation that shapes every correlation, perception, recognition, or deduction that the model will make going forward. In the section following I will discuss some publicly available and commonly used AI systems, how they work, and how I imagine inductive bias shapes their performance.

Speech Recognition

Dragon NaturallySpeaking is popular consumer dictation software. In order for the application to work well, the user first needs to train it to recognize his or her individual voice by reciting a series of stock phrases. The program then creates a voice profile for this particular user, mapping features of user-specific pronunciations into numerical representations. If another person logs in using the same user profile and attempts to use the software, her success will be limited by the similarity of her voice to the initial training set, and chances are that the system will make plenty of mistakes through confusion. The features of the voice learned during training are stored in the software's language model, and this training forms one-third of the inductive bias of the program; it will always recognize voice through the lens of the voice it was trained on.

The second third of the inductive bias comes from the language model and dictionary chosen, a set of common words and pronunciations that the software uses to try to map what the user is saying to a closest match. For example, if the system hears a series of sounds that it decodes as "MO TO RN," depending on the language model this may be recognized as the word "motor" (if it has a dictionary of words relating to machines), or it may be recognized as "Motown" (if it has a dictionary related to music).

In AI systems, there is generally a direct correlation between accuracy and specificity. In this example, the smaller the dictionary, the greater the chance that the dictation software will recognize words correctly, while the broader the scope, the more likely it is to make mistakes. In the speech recognition field this is known as the *word error rate*. Sphinx, a popular speech recognition engine used under the hood in many telephony applications, reports a low word error rate of .168 for a small vocabulary model of 11 words. Increasing the vocabulary to just 1000 words, still a very limited set, increases the error rate to 2.88.

As with all the examples I will discuss, an implicit but invisible third of the inductive bias comes directly from Dragon's team of software engineers. The type of predictive model they chose, the parameters they allowed, the type and number of features that Dragon extracts from voice data, the extent of the system's ability to continuously learn or adapt to the speaker over time, and the choice to make a system speaker dependent or independent all add up to a big chunk of creative influence.

Social Web

Collaborative filtering is another increasingly common use of AI. From Netflix to Amazon to YouTube and Digg, collaborative filtering is the technique used to create data-profiles of users and thereby discover content to recommend. The speech recognition techniques described above create probabilistic models shaped by experience; in a similar fashion, collaborative filters employ user ratings to guess what content a given user is likely to enjoy. Just as the speech recognition system stored a set of numbers representing how a person pronounces a particular word, these systems store a set of numbers representing how a user rates movies, books or news. These numeric sets are then easily compared as matrices to find similarities between users. If John and Mary have similar numbers and John loves the new zombie movie, then recommend it to Mary - she will probably like it too.

The inductive bias in these types of systems arises from some very fundamental decisions made in the design process. For example, are votes binary (yes/no) or scaled (1-5 stars)? Is voting positive only or negative as well? How do you parameterize a user? What makes users similar? How many features categorize the content? What kind of information is relevant and how do you quantify subjective aesthetic qualities like degrees of liking something? Some people tend to like everything, some people tend to be very critical; are these user-specific patterns part of the system as well?

Depending on the answers to these questions, certain kinds of users influence the concept of what a user is, creating a definitive bias at a very low, almost invisible and untraceable level.

These biases pose the most danger, because the people using the system, and sometimes even the people designing the system, are unaware of their presence.

Search

Finally, perhaps the most ubiquitous AI these days is Google. If any company and any algorithm shape our perception of the web, and of each other as mediated through the web, it's Google's well-guarded search algorithm. The best-known feature of the Google system is what they call "PageRank Technology," a system of ranking web pages based on their "importance" as determined by how many other pages link to them and how much traffic they receive. Google also uses something they call "hypertext-matching analysis," which means describing as much "relevant" information about a page as possible and comparing this set of features to the user's query. If these descriptions have a familiar ring, they should; judgments of importance and relevance definitely constitute inductive bias.

But Google wouldn't be Google if that was all there was to it. As Google puts it in their "Organic Search" patent application:

"This GenericScore may not appropriately reflect the site's importance to a particular user if the user's interests or preferences are dramatically different from that of the random surfer. The relevance of a site to a user can be accurately characterized by a set of profile ranks, based on the correlation between a site's content and the user's term-based profile..."

In other words, search algorithms also depend on your data profile, the kind of things you search for and tend to click on. If you have an email account, use Google docs, Google groups, etc. then the kind of things you write about and your demographic information provide further data about *what kind of user you are*. This model raises the same questions of bias discussed with collaborative filtering, but more importantly, it brings us into very questionable ethical territory; when a company has this much information about its clients, how do we know what they will do with it? We already know this information goes into determining what links will be available to us and what ads will appear on the page, and that when asked Google will hand this information over to the government. We tend to believe that they do a great job and that we really are getting the most important and relevant information, but then again, how would we know if we weren't? And how much is a good ad suggestion really worth?

Surveillance

In the next few paragraphs I will describe some of the current surveillance techniques used by the US government and corporations. I will then describe how I believe the inductive bias described above proves beyond any reasonable doubt that these systems are inherently and inescapably discriminatory and error-prone, and how this bias affects, or could affect, everyone subject to analysis by surveillance systems.

Images

Anyone who lives in a major city in the US has probably noticed at least one video camera in a public place recently. According to the ACLU there are at least 8000 cameras in New York City and the NYPD is currently adding another 3000 in lower Manhattan and is planning to add thousands more, blanketing midtown between 30th and 60th streets. As the city is currently cutting back dramatically on the number of new officers being trained (Huguenin, Winston), it raises the question: if no new officers are being hired, who will be watching all those cameras?

The NYPD privacy policy describes the new "Domain Awareness System" as "technology deployed in public spaces as part of the counterterrorism program of the NYPD's Counterterrorism Bureau, including: NYPD- owned and Stakeholder-owned closed circuit television cameras (CCTVs) providing feeds into the Lower Manhattan Security Coordination Center; License Plate Readers (LPRs); *and other domain awareness devices, as appropriate*" (Public Security Privacy

Guidelines, 2, emphasis my own).

The mention of “other devices” in the NYPD policy clearly seems intended to leave a backdoor open for future developments in video recognition technology. Mayor Bloomberg, in a recent attempt to garner support for his re-election campaign, confirmed this suspicion when he proposed that the Real Time Crime Center in New York use facial recognition software to analyze surveillance camera footage as part of a multi-fold plan to “keep New York safe (by) using technology to fight crime.” As he put it, “in this way, the image of a killer caught on a bodega video camera, for example, can be instantly compared to tens of thousands of mug shots already housed in police files” (Fink, Bloomberg).

The new “enhanced” driver’s license, required for crossing land borders into the US without a passport, has a Radio Frequency Identification (RFID) chip and stores a unique ID which is an index to your personal biometric data (AKA facial features), which are stored in the national Homeland Security database when you get your card. If this kind of driver’s license becomes the standard driver’s license, the government will have a huge, ready-made database of the faces of its citizens. Interpol and Canadian and United States law enforcement agencies recently proclaimed interest in using this database, in combination with a network of CCTV cameras and facial recognition software, to catch criminals attempting to cross the border (Parsons).

While all the implications of these interlocking stories are intimidating, the aspect that I find most alarming is that facial recognition systems are simply not very accurate. Teaching a computer to see is one of the holy grails of AI, but anyone in the field will tell you that we are simply not there yet.

Several cities and companies that have attempted to use these systems in the past have already discovered this flaw. In 2001, Tampa, Florida installed a citywide facial recognition and surveillance camera system as an experiment in using the technology to spot wanted criminals. After two years during which not a single criminal was found, the system was taken down (CNET News). A similar system installed in Virginia Beach also failed to produce any recognition.

The UK is the world leader in government-funded CCTV camera installations. They have 4 million cameras countrywide, with 1 million in London alone. A recent report from their Metropolitan Police department admitted that the ratio of cameras to arrests was 1000 to 1. For every 1000 cameras the police installs they make a single arrest. For a system whose installation cost 500 million British pounds over 10 years, this news is pretty disappointing. Even more interesting was the revelation that even when people are caught committing crimes on camera, the video evidence does not necessarily translate into arrests; only half of the footage turns out to be of suitable quality to convict in court. And these results are from a system where the footage is analyzed by actual people. (BBC)

A 2005 US government report exploring the effectiveness of facial recognition compared to fingerprints found that recognition systems had error rates as high as 50% in poor lighting conditions. New 3-D camera recognition systems seem to promise lower error rates, but implementation of these systems requires expensive specialized technology and willing subjects, making it feasible only for controlled environments such as entry to a secure office, rather than general street surveillance.

Video recognition works poorly under the best of circumstances, but with the kind of degraded data that real surveillance cameras produce - low resolution, side views of faces, patches of shadow and over-exposure – any accurate recognition would be nothing short of amazing. Furthermore, the people operating the recognition systems never talk to the people who make them. Even if the engineer of a given software system understands some of the inductive bias it contains, what is the likelihood that the police officer manning the computer that runs that software will? While the subjectivity of these systems may be evident at the level of design, the end user’s impression of software is usually a system that is objective and reliable, more so than a human would be. Technology companies and research institutes love to proclaim that their

algorithm is better at identifying a face than a human being, and readers, both technologists and lay-people alike, love to believe it. (Williams)

Ultimately, the invisible effect of visibly installing thousands of new cameras will be that systems not nearly robust enough for the job will be rolled out and used for recognition and tracking by people who have no deep understanding of how the technology works.

Voice

In light of the illegal-made-legal domestic wiretapping conducted under the Bush administration, and now approved by the Obama administration as well (Jones), I think it is safe to assume that any American communication through any medium may be surveyed. It is always interesting to peruse the Department of Advanced Research Projects Agency (DARPA)² website's list of current programs, just to get an idea of exactly what research they are funding. The information available online is just an overview – a glimpse – but still provides insight. If nothing else, it often serves as proof that your suspicions are correct... they really are working on a giant all-terrain robotic dog... Hmm.

In my research for the *Listening Post* project I stumbled upon a DARPA project aptly titled GALE, Global Autonomous Language Exploitation, which I found progressively more disturbing as I learned more and more about speech recognition technology. The stated goal of the program is to:

apply computer software technologies to absorb, translate, analyze, and interpret huge volumes of speech and text in multiple languages, eliminating the need for linguists and analysts, and automatically providing relevant, concise, actionable information to military command and personnel in a timely fashion. (GALE website)

And another page describes how:

GALE technology will make it possible for English speakers to do much of what foreign language operators and analysts do now: convert large quantities of "undecipherable" foreign language data into timely, actionable intelligence in English. (ibid)

GALE is a multi-tiered AI system that comprises an automatic transcription phase, followed by automatic translation into English, and then content extraction of relevant information. The idea that a natural language processing system would be capable of doing all of these things reliably and accurately enough to create "actionable information" for the military is both absurd and extremely dangerous. Each individual component of the system is unachievable on its own; in combination the synergy of errors would be tremendous.

I have several reasons for believing this to be true. The first is my own experience with both speech recognition and natural language processing. In my experience, systems of this kind only have acceptable accuracy rates over very limited domains: for example, recognizing a set of 10 commands spoken in a quiet room, or identifying what language a user is typing in when discussing specific information. These very limited applications are successful precisely because they are limited. And even these simple applications make mistakes, and plenty of them.

I draw further evidence from my consumer experience. I am aware of products on the market that claim to accomplish speaker-independent voice recognition, such as Dragon's \$20,000 audio-mining API. But even this corporate product confesses publicly (on its own promotional website): "Accuracy levels depend upon the quality of the recording. Studio-based content will provide higher accuracy levels, but the system also provides a reasonable level of

²<http://www.darpa.mil/ipto/programs/programs.asp>

accuracy for telephone, public presentation and broadcast content.” What is a “reasonable level of accuracy” for providing actionable military information?³

If systems of this kind were available, or even possible, the corporate sector would be offering them up. Consumers would be the first to know. As the subjects of these technologies, we are aware of the state of the art and we know its limitations. Even high-cost automatic transcription systems commonly have 5-10% error rates in controlled conditions on voices they are trained to recognize. Additionally, due to the contextual way in which these systems work, once they transcribe one word wrong they tend to interpret the words that surround it relative to that wrong word, with the result that the whole sentence quickly goes wrong. This is acceptable for dictating letters, but not military intelligence.

Add this system to the current government wiretaps on all foreign communications, and we have a perfect recipe for uninformed decision-making, which could lead to unreasonable detentions, unnecessary interrogations, and even more alarmingly, misinformed military strategies.

Inductive bias revisited

In his book *Artificial Intelligence*, George Luger describes the contradiction in the field of machine learning between the demands of efficiency and expressiveness, and suggests that inductive bias is a necessary side effect of heuristic search for any complex problem (Luger, 383-383).

These are precisely the issues that will lead to mistakes in noisy data situations, and in the real world outside the lab, all data is noisy. You will never get a perfectly lit, perfectly framed shot of a person's face, or a voice sample without background noise; degradation of quality is inherent to communication media.

Furthermore, the end users of these systems are not the designers and will never meet the designers. So most system users are not aware of the biases that constitute the system design; sometimes even the designers remain unaware, because the biases occur at such a deep and fundamental level that they are invisible. If a person gives us an opinion, we might be able to weigh our knowledge of that person and his prejudices and opinions, and then take what he says with a grain of salt; but when a computer pronounces, we tend to think its output is both objective and infallible. Unlike our friends, whose conclusions we can understand through talking with them about their experiences in life, and whose processes of learning we imagine to be much like our own, computers appear as black boxes of magic. AI software in particular is so shrouded in mystery that it might as well be handed down directly from above. Even engineers refer to these techniques as a “black art.”

Real world AI systems are imperfect; they operate under the worst scenarios imaginable, often pushed to perform feats they were never intended to perform, with the worst input data available, by people who don't understand how they work. Can we really trust our criminal justice system, our military, and our corporations with a weapon of such ambiguous validity?

Listening Post: inductive bias as artistic statement

From the very beginning of my research in this area I noticed that machine learning - with its focus on essential forms, abstraction and generalization - had a lot in common with artistic practice. This parallel initially attracted me to this field and has sustained my interest, in varying forms, for the past seven years. Inductive bias, which is such a burden and philosophical

³ According to the GALE website: The ultimate performance targets are to translate Arabic and Chinese speech and text with 95% accuracy and an extremely high degree of consistency (90-95%), and to extract and deliver key information with proficiency matching or exceeding that of humans. A high target, but what will they settle for if they can't achieve it?

confusion to the computer scientists, becomes an asset to the artist, who is free to use the algorithms as a medium to form an artistic statement.

When I work with algorithms, I start out with a vague goal in mind. In this case, I wanted to work with speech recognition, and the more I learned about the technology and its social context, the more interested I became in subverting it. This interest steered me away from a project that might use the technology more blatantly, for example by interpreting the words of gallery-goers to control a robot. I didn't want to simply *showcase the technology*. As I worked on the project I envisioned a community language forming from bits and pieces of speakers passing by. I imagined that I would be able to parse out exactly what people were saying, the part of speech of each word they spoke, and from there I could parse out the grammatical structure of phrases and begin to save these structures as templates. I wanted the algorithm to waltz through the data in an almost Frankensteinian way, reassembling grammatical structures with words and sounds torn from their original context and stitching together the voices of countless different speakers.

As I worked on the project I learned how fallible the speech recognition technology actually was. Every time I took a step closer to what my actual input data would be (8k audio from a noisy city street with background music and a fountain nearby) I surrendered accuracy exponentially. And as I continued on, tweaking the system, dedicating my hours to changing parameters and attempting to squeeze out every last drop of accuracy, the humor of the whole endeavor struck me: why should I try to save this technology from its own pitfalls, when I could just allow it to naturally expose itself as the hilariously erroneous technique that it is?

In this way, one of the central ideas behind *Listening Post* emerged: an attempt to play with the biases and mistakes native to machine learning systems. By highlighting the errors of the system, the project would turn a technology normally used to (attempt to) overhear conversations, extract transcriptions, and mine them for intelligible information into a piece of music, a community portrait of the sounds of the street. *Listening Post* pokes fun at or disarms the technology by turning it on its head; rather than extracting interesting phrases or patterns in behavior to market a product or find a criminal, it tears the data apart and reassembles it in such a way that no individual voice is distinguishable, no one person is ever talking, and a voice is just a single point in a cluster of sounds.

The online surveillance-evading tool TrackMeNot attempts to distract search engines from storing meaningful information by hiding your real queries in mountains of randomized and nonsensical surrounding searches. The search engine software is still tracking you but its data becomes useless. Similarly, *Listening Post* obscures the original data, evolving over time to consist of infinite sequences of speech-like sounds that never quite make sense. The sounds are, or really were, real people's voices, but as they pass through the system they become one voice, one flow of sounds patterned after the way sounds flow.

Listening Post is a conceptual exploration of a technology and a medium. It is political but also fetishistic. It describes both my fascination and disgust with a field, towards which I am continually fighting to balance my attraction and repulsion. It captures both my enthusiasm to create a really good speech recognition system and engage with the theoretical underpinnings of such an endeavor, as well as my desire to break down the system of AI-monitored surveillance and invert its power.

Perhaps it is precisely up to those of us who find ourselves so attracted to these technologies to provide the critical framework for their repulsion.

Heather Dewey-Hagborg, September 2009

Works Cited

BBC News. "1,000 cameras 'solve one crime.'" BBC News Online, 24 August 2009.

Bloomberg, Mike. *Keeping NYC Safe: Using Technology to Fight Crime*. Campaign pamphlet, 2009.

Bowman, Lisa. "Tampa drops face-recognition system." CNET News, 21 August 2003.

Fink, Jason. "NYPD going high-tech, as some fear 'Big Brother.'" *AMNY*, 21 September 2009.

GALE website: <http://www.darpa.mil/ipto/programs/gale/gale.asp>

Huguenin, Patrick. "With surveillance cameras around New York City, you're being watched," *New York Daily News*, 15 July 2009.

Jones, Tim. "In Warrantless Wiretapping Case, Obama DOJ's New Arguments Are Worse Than Bush's." Electronic Frontier Foundation, 7 April 2009.

Koren, Yelda. *The BellKor Solution to the Netflix Grand Prize*. Published online, August 2009.

Luger, George F. *Artificial Intelligence Structures and Strategies for Complex Problem Solving*, 4th Edition. Harlow: Addison-Wesley, 2002.

Parsons, Christopher. "Driving Your Liberties Away: Biometrics and 'Enhanced' Drivers Licenses." *Atlantic Free Press*, 10 Nov. 2008.

Williams, Mark. "Better Face-Recognition Software." *MIT Technology Review*, 30 May 2007.

Winston, Ali. "Curious about the 'Ring of Steel' in lower Manhattan? The NYPD has made its privacy guidelines public." *City Limits Weekly* #678, 16 March 2009.