Opening the Black Box: Confronting Software-Based Evidence

Jeanna Matthews
Clarkson University/Data and Society

Questioning Forensics: Lawyers, Damn Lawyers, and Statistics
November 2, 2018
Software in the Criminal Justice System

• Software/algorithmic decision making used increasingly throughout the criminal justice system
  – Investigation/policing decisions, pretrial decisions, evidence at trial, sentencing decisions, parole decisions

• Often black boxes for which trade secret protection is claimed to be more important than rights of individual defendants or citizens to understand the decisions
Caution needed during a stop

Types of policing needed

Areas to police

People to stop

Gang and affiliation databases

People likely to become victims
Evidence gathering

- ShotSpotter
- Cell-site simulators (Stingray)
- Facial recognition
- Automated License Plate Readers (ALPRs)
- CP investigation (P2P, IP tracing)
- Network Investigation Techniques (NITs)
- Mobile device cracking (Cellebrite, Gray Key)
Probabilistic genotyping
Facial recognition
Latent prints (AFIS)
Social media analytics
Ballistics and toolmarks
Breath alcohol (Alcotest)
Individualized assessment

Bail determinations (flight risk)

Parole determinations (reoffense risk)

Sentencing

Parole/probation monitoring
Software Bugs

• We have all used software with bugs.
  – Company did a lot of testing before release, generally works but still had problems for you
  – Perhaps a problem others had difficulty reproducing but that is clearly happening in your unique system

• Forensic software has no special immunity from the bugs and mistakes that plague software in other fields.

• Software and complex systems need an iterative process of debugging and improvement!
  – Anyone who has used technology knows that there are glitches and bugs and unintended consequences!
  – Anyone who builds technology knows how easy it is for there to be substantial bugs you did not find!
• How can we find bugs and fix problems if the answer is always “you can’t question” and “you are just complaining because you are guilty”? 
Validation studies?

• What they are and what they are not

• Developers of the system do their own testing and publish results

• They have vested interest to demonstrate the system is working NOT to find bugs.

• What would our daily computing would look like with only this?

• What is the forcing function for debugging? Especially when interests of customers are different than interests of those being decided about?
Debugging?

• Huge advantages to independent, third-party testing aimed at finding bugs!

• If only those with interests in the success of software see the details, we have a huge problem and a recipe for injustice!

• Shifting to individual defense teams?

• People report errors and the system gets better, but what if all “bug reports” dismissed as just guilty people complaining?

• Doomed to run our criminal justice system on buggy software?
• What types of review might attorneys and judges seek in understanding software-based/computer-based evidence?

• Why law and public policy require disclosure of these materials to the public and independent experts?
What does it mean to evaluate software?

What Information Do Defense Experts Need to Evaluate Forensic Software?
## Executables

A screenshot showing a window with tabs for Sample Files, Reference Files, Profile Summary, Analysis, Sensitivity Analysis, Non-contributor Test, Reports, and About. The window contains a table labeled "LRmixStudio - example" with columns for Locus, Rep1, Rep2, and Rep3.

<table>
<thead>
<tr>
<th>Locus</th>
<th>Rep1</th>
<th>Rep2</th>
<th>Rep3</th>
</tr>
</thead>
<tbody>
<tr>
<td>D10S1248</td>
<td>13 14 16</td>
<td>13 14 15 16</td>
<td>13 14 15 16</td>
</tr>
<tr>
<td>VWA</td>
<td>15 17 18 19</td>
<td>15 16 17 18 19</td>
<td>15 16 17 18 19</td>
</tr>
<tr>
<td>D165539</td>
<td>11 13</td>
<td>10 11 12 13</td>
<td>10 11 12 13</td>
</tr>
<tr>
<td>D2S1338</td>
<td>17 18 19</td>
<td>17 18</td>
<td>17 18 19</td>
</tr>
<tr>
<td>D8S1179</td>
<td>10 12 14 15</td>
<td>10 12 14 15</td>
<td>10 11 12 14 15</td>
</tr>
<tr>
<td>D21S11</td>
<td>28 29 30 33.2</td>
<td>28 29 30 32.2 33.2</td>
<td>28 30 33.2</td>
</tr>
<tr>
<td>D18S51</td>
<td>14 15 17</td>
<td>10 14 15 17</td>
<td>14 15 18</td>
</tr>
<tr>
<td>D22S1045</td>
<td>11 12 15 16</td>
<td>12 15 16</td>
<td>11 12 15 16</td>
</tr>
<tr>
<td>D19S433</td>
<td>13 14</td>
<td>13 14</td>
<td>13 14</td>
</tr>
<tr>
<td>TH01</td>
<td>6 7 8 9.3</td>
<td>7 8 9.3</td>
<td>7 8 9.3</td>
</tr>
<tr>
<td>FGA</td>
<td>18 21 24</td>
<td>18 19 22 23 24</td>
<td>18 21 22 23 24</td>
</tr>
<tr>
<td>D25441</td>
<td>10 11</td>
<td>10</td>
<td>10 11 11.3</td>
</tr>
<tr>
<td>D3S1358</td>
<td>16 17</td>
<td>15 16 17</td>
<td>15 16</td>
</tr>
<tr>
<td>D1S1656</td>
<td>12 15 16 18</td>
<td>11 12 15 16</td>
<td>12 16</td>
</tr>
<tr>
<td>D12S391</td>
<td>15 17 19 20 23</td>
<td>15 17 19 20</td>
<td>15 17 19 20</td>
</tr>
</tbody>
</table>
/// This function checks for the total frequencies according to races and removes the alleles from calculation
/// if the sum of frequencies are greater than 0.97.
/// </summary>
public void CheckFrequencyForRemoval(DataTable dtFrequencies)
{
    // if our db connection isn't initialized, do it. then, get all the ethnicities (races)
    myDb = myDb ?? new Database();
    DataTable raceTable = myDb.getAllEthnics();
    int intsr = 0;
    string[] srem = new string[comparisonLoci.Count];

    // we go through all the comparison loci and check whether the sum of the frequencies for that locus is greater than 0.97.
    // if it is, we remove the locus. frequencies are only used for the alleles in the evidence replicates.
    for (int i = 0; i < comparisonLoci.Count; i++)
    {
        bool blRemove = false;
        // get a CSV list of alleles for all the replicates at a locus
        IEnumerable<string> unknownPair = EvidenceAllelesAtLocus(evidenceAlleles[comparisonLoci[i]]);
        // check if the frequency is greater than 0.97 for any of the races. Frequencies are values for an allele at a locus for a certain race
        foreach (DataRow eachRow in raceTable.Rows)
        {
            string raceName = eachRow.Field<string>("EthnicName");
            float freqSum = GetFrequencySum(unknownPair, comparisonLoci[i], raceName, dtFrequencies);

            if (freqSum >= 0.97)
            {
                blRemove = true;
                break;
            }
        }
        if (blRemove)
        {
            srem[intsr] = comparisonLoci[i];
        }
    }
Information from the Development Process

• Design documents

• Testing plans and results
Experience with Deployed Software

• Bug reports

• Change logs

• Revision Control
Machine Learning Systems

• Training sets

• Learn from, but don’t reproduce, the past
What are the arguments that law and public policy require disclosure of these materials to the public and independent experts?
Due Process

• Due process entitles the defense to review the prosecution’s evidence
• Due process prohibits burden shifting to the defense.
• Requiring the defense to make a showing of materiality before granting access to the forensic software materials runs contrary to basic constitutional guarantees
• Commitment to open and public trials
Disclosure of Trade Secret Information

• Disclosure of trade secrets under protective orders is common in civil cases
  – Even when the parties are direct competitors with an interest in profiting from the proprietary information of the other

• Should be easier for a defendant trying to defend their life and liberty to access and assess forensic software, as compared to a party with a mere economic interest

• Prosecution Typically Cannot Establish That Disclosure Subject to a Protective Order Would Cause Harm
ARTICLE

Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System

Rebecca Wexler*
Brown Institute Magic Grant: Decoding Differences in Forensic DNA Software

ProPublica Seeks Source Code for New York City’s Disputed DNA Software

We're asking a federal court for the code behind a technique that critics say may have put innocent people in prison.

by Lauren Kirchner, Sept. 25, 2017, 7:54 p.m. EDT
Methods

• Independent, third-party, adversarial testing and review
  – Automated testing harnesses
  – Common file formats and settings
  – Source code analysis

• Recommendations
  – Clear advice for judges, defense attorneys, journalists
  – Sample requirements for software systems, targeting the procurement phase
Procurement Phase Wishlist

- When public money used for criminal justice software, require! or at least give credit for:
  - Software artifacts: bug reports, internal testing plans and results, software requirements and specifications, risk assessments, design documents, etc.
    - Lack of software standards in traditionally non-computing fields (e.g. DNA)
  - Source code
  - No clauses preventing third party review or publishing of defects found
  - Access to executables for third party testing
  - Scriptable interfaces to facilitate automated testing
  - Bug bounties

- Reward/Fund/Incentivize non-profit third party entities to do independent testing and find problems!
Case of Forensic Statistical Tool (FST)

- Probabilistic genotyping software
- Developed by Office of the Chief Medical Examiner (OCME)
- NY State Commission on Forensic Science approves FST for use in casework
Problems/Changes

• April 2011 online, within weeks offline
• June 2011 back online after repairs/modifications

• For some samples reanalyzed post-modification, likelihood ratio “values were slightly modified as expected.”
  – Quality Control Test of Forensic Statistical Tool (FST) Version 2.0, June 30, 2011

• “Because this modification did not affect the methodology of the program, it did not require submission to the Commission on Forensic Science or the DNA Subcommittee.”
  – Affidavit of Eugene Lien, OCME Assistant Director, July 17, 2017
What were the changes?

• In retrospect we know
• ~70 line function, checkFrequencyForRemoval
/// This function checks for the total frequencies according to races and removes the alleles from calculation
/// if the sum of frequencies are greater than 0.97.
/// </summary>

public void CheckFrequencyForRemoval(DataTable dtFrequencies)
{
    // if our db connection isn't initialized, do it. then, get all the ethnicities (races)
    myDb = myDb ?? new Database();
    DataTable raceTable = myDb.getAllEthnics();
    int intsr = 0;
    string[] srem = new string[comparisonLoci.Count];

    // we go through all the comparison loci and check whether the sum of the frequencies for that locus is greater than 0.97.
    // if it is, we remove the locus. frequencies are only used for the alleles in the evidence replicates.
    for (int i = 0; i < comparisonLoci.Count; i++)
    {
        bool blRemove = false;
        // get a CSV list of alleles for all the replicates at a locus
        IEnumerable<string> unknownPair = EvidenceAllelesAtLocus(evidenceAlleles[comparisonLoci[i]]);
        // check if the frequency is greater than 0.97 for any of the races. frequencies are values for an allele at a locus for a certain race
        foreach (DataRow eachRow in raceTable.Rows)
        {
            string raceName = eachRow.Field<string>("EthnicName");
            float freqSum = GetFrequencySum(unknownPair, comparisonLoci[i], raceName, dtFrequencies);

            if (freqSum >= 0.97)
            {
                blRemove = true;
                break;
            }
        }

        if (blRemove)
        {
            srem[intsr] = comparisonLoci[i];
        }
    }
}
Dropping some data that supports the defense hypothesis without notice!

<table>
<thead>
<tr>
<th>Profile</th>
<th>D8S1179</th>
<th>D21S11</th>
<th>D7S820</th>
<th>CSF1PO</th>
<th>D3S1358</th>
<th>TH01</th>
<th>D13S317</th>
<th>D16S539</th>
<th>D2S1338</th>
<th>D19S433</th>
<th>vWA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evidence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>8, 12, 13, 14, 15</td>
<td>28, 29, 30, 30.2</td>
<td>8, 9, 10, 11</td>
<td>8, 9, 10, 12</td>
<td>8, 9, 10, 11</td>
<td>8, 9, 10, 11</td>
<td>8, 9, 10, 11</td>
<td>8, 9, 10, 11</td>
<td>8, 9, 10, 11</td>
<td>8, 9, 10, 11</td>
<td>8, 9, 10, 11</td>
</tr>
<tr>
<td>2</td>
<td>8, 10, 12, 13, 14, 15</td>
<td>28, 29, 30, 30.2</td>
<td>8, 9, 11</td>
<td>9, 10, 12</td>
<td>8, 9, 11</td>
<td>8, 9, 11</td>
<td>8, 9, 11</td>
<td>8, 9, 11</td>
<td>8, 9, 11</td>
<td>8, 9, 11</td>
<td>8, 9, 11</td>
</tr>
<tr>
<td>3</td>
<td>8, 13, 14</td>
<td>29, 30, 30.2, 31</td>
<td>8, 9</td>
<td>9, 10, 12</td>
<td>8, 9</td>
<td>8, 9, 10, 11</td>
<td>8, 9</td>
<td>8, 9</td>
<td>8, 9, 10, 11</td>
<td>8, 9, 10, 11</td>
<td>8, 9, 10, 11</td>
</tr>
</tbody>
</table>

**Comparison Result**

<table>
<thead>
<tr>
<th></th>
<th>Asian</th>
<th>Black</th>
<th>Caucasian</th>
<th>Hispanic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likelihood Ratio</td>
<td>3.03e+04</td>
<td>339.47</td>
<td>70.6</td>
<td>415.32</td>
</tr>
</tbody>
</table>

- LR > 1 supports inclusion as a contributor
- LR < 1 supports exclusion as a contributor

Removing data at 3 loci that is...

- Exclusionary: 0.53
- 3.1 1.3 Inclusionary
Where is debugging?

- What hope would an individual have of seeing whether a change like this impacted them?
- How likely to find at all without source code access?
- Work of debugging falling to individual defense teams
- You’re just complaining because you’re guilty?
Algorithmic Transparency and Accountability Efforts

• US-ACM/EUACM Statement on Algorithmic Transparency and Accountability
  – 7 principles
  – Awareness, access and redress, accountability, explanation, data provenance, auditability, validation and testing

• CACM article “Toward Algorithmic Transparency and Accountability”

https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf
AI Ethics and Anti-Patterns

• Upcoming article in AI Magazine and talk at 2019 AI For Good

• 10 Common Anti-Patterns
  – 1. Learn from the Past Without Remembering the Context
  – 2. Learning from Humans Without Remembering Human Bias and the Possibility of Malicious Training
  – 3. Using Data You Have Rather than the Data You Need
  – 4. Transparency without Accountability
  – …..
Human vs. computer-based decision making?

• Human decision making flawed as well *but* at least we recognize it

• Myth of logical, unbiased, nearly infallible computer decision making

• Possibility of more reproducible, accurate, fair decisions *but* to achieve that we will need investments in debugging, skepticism and explanation
  – Bail vs. automated risk assessments

• Important to preserve our ability to question in general and in a specific case
Decision-Making Landscape

• Big decisions about the lives of individuals are being made in a partnership between human decision-makers and computer systems.
• Fundamentally changing the landscape of our societal decision-making processes
• In the process of automation, moving to new platforms, are we undermining principles on which societies have been built?
• What are we doing to secure and debug these complex socio-technical systems? To build in incentives for debugging and improvement?