**[The Volokh Conspiracy](http://www.volokh.com/)**

[HOME](http://www.volokh.com/)[ABOUT](http://www.volokh.com/about/)[DMCA CONTACT](http://www.volokh.com/dmca-contact/)[E-MAIL POLICY](http://www.volokh.com/e-mail-policy/)[LINK POLICY](http://www.volokh.com/linking-policy/)[STUFF](http://www.volokh.com/stuff/)[WHO ARE WE?](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#contact)[SEARCH](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#search)[SUBSCRIBE](http://volokh.com/2009/11/25/getting-blog-posts-for-the-conspiracy-and-other-rss-producing-blogs-in-your-outlook-inbox/)[TWITTER](http://twitter.com/VolokhC)[RSS](http://feeds.feedburner.com/volokh/mainfeed)[ARCHIVES](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#archives)

« [George Will on the People’s Rights Amendment](http://www.volokh.com/2012/05/05/george-will-on-the-peoples-rights-amendment/)

[“Obama Has Become More Hostile to Medical Marijuana Patients than Any President in U.S. History”](http://www.volokh.com/2012/05/06/obama-has-become-more-hostile-to-medical-marijuana-patients-than-any-president-in-u-s-history/) »

[**Is Contract Lawyering Doomed by Algorithm?**](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/)

[Kenneth Anderson](http://www.volokh.com/author/kenneth/) • May 5, 2012 7:24 pm

(Added: There are some excellent comments in the thread, well worth reading.) Walter Russell Mead and Tyler Cowen each point to [this article in New York Magazine,](http://nymag.com/daily/intel/2012/03/new-technology-may-spell-doom-for-new-lawyers.html) describing a decision by US Magistrate Andrew J. Peck (SDNY) to allow the “use of predictive coding, a computer-assisted document review that turns much of the legal grunt work currently done by underemployed attorneys over to the machines.” The work performed by armies of temporary “contract lawyers”?

The task of combing through mountains of emails, spreadsheets, memos and other records in the discovery process currently falls on a legion of “contract attorneys” who jump from one project to another, employed by companies like Epiq Systems. Many are recent grads who are unable to find full-time employment, or lawyers laid off during the recent recession.

Scan. Point. Click. Repeat. That’s the job. Contract attorneys are paid by the hour to sit in front of a computer and review a mind-numbing sequence of uploaded documents. There are cramped, sunless rooms in law firms throughout the city, with rows of computers piled one on top of the other, and constant uncertainty as to how long each particular stretch of employment will last.

Predictive coding promises to make this job much more efficient over time – and drastically reduce the amount of work and number of contract attorneys employed.

Using the technology, a senior attorney familiar with the intricacies of a specific case reviews and codes a “seed set” of documents. An algorithm then identifies properties among the manually reviewed documents to code and sort everything else. Each document is assigned a score to indicate the likelihood it’s correctly coded. Proponents say predictive coding is not only more accurate than using human reviewers, but also more efficient .... There’s no escaping the fact that as predictive coding is used more widely, the technology will reduce the overall number of documents to be reviewed and the attorneys needed to review them. Judge Peck noted the technology will require human review of less than 2 percent of all documents in an average case. His stamp of approval means that the document reviewer ranks may be culled sooner rather than later.

[Mead adds a comment](http://blogs.the-american-interest.com/wrm/2012/05/05/attention-young-lawyers-the-grim-reaper-is-here/) about what this means for the future of lawyering as a profession and as a safety ticket to the upper middle class. I would add the additional comment that survival lies in getting on the right side of algorithm production. That doesn’t necessarily mean that one has to write algorithms. We live in a highly complex economy with many niches. But whatever one’s niche, it helps a lot to be aligned on the side of the angels of creative destruction, rather than sitting in their path.

Software is getting smarter, and computers continue to grow more powerful. What we see now is only the beginning of a process by which the routine elements of legal work — and frankly speaking, that is where the bulk of the jobs have always been — can and will be automated. Not all young lawyers will be doomed. There will be some smart, entrepeneurial kids who figure out how all this computing power can allow a small, lightly capitalized firm to deliver high quality services at a breathtakingly low cost to selected clients. Those kids will do well.

Others will benefit from greater demand. Legal services are likely to get cheaper: there is a lawyer glut that is likely to grow, and the increasing capabilities of computers in the legal field mean that the amount of available legal brainpower will explode. Cheaper legal services mean that more people and firms will use the legal system and legal expertise in various ways: the lawyers and firms who figure out how to ride this wave will also do well. Brilliant and creative lawyers will continue to do well. So will the marketers, the deal makers and the connectors. But law isn’t going to be the kind of safety play ticket to the upper middle class that it used to be.

[Share on facebook](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Share on email](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Share on printfriendly](http://www.addthis.com/bookmark.php?v=300&winname=addthis&pub=agahe&source=tbx-300&lng=en-US&s=printfriendly&url=http%3A%2F%2Fwww.volokh.com%2F2012%2F05%2F05%2Fis-contract-lawyering-doomed-by-algorithm%2F&title=Is%20Contract%20Lawyering%20Doomed%20by%20Algorithm%3F&ate=AT-agahe/-/-/5201664fed014fe1/2&frommenu=1&uid=5201664f01a5eb36&ct=1&pre=https%3A%2F%2Fwww.google.com%2F&tt=0&captcha_provider=nucaptcha)

Categories: [Uncategorized](http://www.volokh.com/category/uncategorized/)

× Comments for this thread are now closed.

[**DavidGolden**](http://disqus.com/DavidGolden/)Predictive coding has more to do with eDiscovery vendors looking for a competitive advantage than any real advance.  My hunch is that the same expert that tunes the machine learning system could accomplish just as much by honing the set of key words used for traditional search. The result would be the same, but with more transparency.   As far as I'm aware, none of the so-called studies allegedly proving the predictive coding is more accurate than manual review were based on peer-reviewed research. For that matter, I'm not familiar with any recent studies done by a credible scientific or technical organization. Any judge that forced a party to pursue a process that was validated by merely an industry study or two is skating on very thin ice. If someone can point to credible, scientific evaluation of predictive coding, please add a reference to it on this thread.  Further, the down side of predictive coding is that opposing council can (and should) challenge the process, demanding that the entire seed set of documents (the "seed collection", which consists of 'like' responsive documents and 'unlike' non-responsive documents) used to train the prediction system be shared so that the machine learning process can be validated. Given that a seed collection would probably contain 10x-20x or more unresponsive documents for each responsive document (to achieve some sort of statistical significance in training), this represents the turnover of a rather large document collection to opposing counsel. Even with privilege review for the 10-20x unresponsive documents in the turnover (and keep in mind that privilege review is the most expensive review), and a clawback agreement just in case, an unnecessary and enormous liability is still  created. How would you like to explain to your client that the predictive coding technology that you recommended has resulted in the turnover of documents that exposed trade secrets or previously buried legal vulnerabilities?  And there's the real possibility of turning over a document that alerts opposing counsel to a future line of attack, that neither you nor your client even realized might be on the horizon.  As a side note, the 'expert' tuning the system is likely to be a senior attorney on the case. As this attorney is now part of the 'process,' no different from the technicians that collected and processed the electronic stored information, I would expect that the work product by that attorney is discoverable. This is clearly not a comfortable position for Outside Counsel, and the use of less skilled resources places the entire process under a cloud of suspicion.

[**ohara5.0**](http://disqus.com/google-a7f2a35d9c9bfeccecc5a23c3f4497d3/)As an MS Physics, MSEE, I predict that in the future lawyers will be my minions.  Minions have to be almost intelligent, capable of following directions and not very self directed.

[**Fat\_Man**](http://disqus.com/Fat_Man/)The alternative to having computers do that work was not going to be hiring hundreds of associates at $165,000 per year and charging the clients $330/hr for their work  (2400 hrs/yr).  The alternative to computers is India, or Nigeria, or Kenya.

[**Visitor\_Again**](http://disqus.com/Visitor_Again/)Question from someone who never worked for a big firm:  How do you maintain the attorney-client privilege if you farm out work to India, Nigeria or Kenya?

[**Fat\_Man**](http://disqus.com/Fat_Man/) If there is enough money in it. They will figure it out

[**DeeR**](http://disqus.com/yahoo-RS76D5SIYNMOLFQX6W3QIQUBLY/)Ten years ago document review was interesting because unfiltered email contained personal messages and very private details about the sender's lives. Today, document review has been reduced to looking at only relevant messages in which all sordid details and non relevant data have been stripped out. Recently, firms have been outsourcing to inexpensive Indian document reviewers and using more predictive coding of documents. Attorneys are now mostly reviewing for privilege and not relevance. The huge market for lowly,  inexperienced attorneys as document coders is quickly drying up. Soon firms will employ tiny staffs of attorneys experienced in document review, privilege logs, privilege duping etc to oversee the document review product produced by computers and third world outsourcing. The legal employment bubble is about to burst and leave a tiny demand for lawyers of any skill level.

[**Herb Ritts**](http://disqus.com/yahoo-3IOO6RXJHIVJZFUTUEPFJLSANY/)Tech like this is going to drastically change the practice of law. I wish I understood exactly how better, as I would be better positioned to make a mint or, more humbly and probably, retain a job. Discussions of this general topic generally draw a distinction between "low value" lawyers with little expertise (e.g., contract attorneys on doc review) and "high value" lawyers with very rare, very human expertise (e.g., the lead litigation or transactional partner making strategic decisions). I'm a corporate lawyer, so I'll use a corporate law context for the rest of this post. My question is, without the scut work of doc review or due diligence, how will the next generation of "high value" lawyers learn the skills that make them "high value" in the first place? In my experience, most attorneys teach more junior attorneys when a specific issue arises. I've never had a lawyer senior to me say, in the absence of a live issue, "Lad, let's take an hour to discuss the enforceability of offer letters in State X in case you, as a potential 'high value' lawyer, need to know this later."  I have had the discussion after actually reviewing badly written offer letters during diligence. In turn, I've never grabbed a junior associate and said, "Look, it's 8 pm on Friday night; we're both off the clock. This seems as good a time as any to teach you when, in a merger, to worry about an ambiguous non-assingment clause and when not to, even though you've never actually read a non-assignment clause. Because the algorithms do that." To be clear, I'm not saying that this is necessarily a GOOD way to train lawyers. I am saying that this IS the way that lawyers are trained now.  And I'm also saying that I've seen lots of articles about how the computers will replace the "low value" lawyers, but never an article that explained how we will continue to have "high value" lawyers. Put another way, I've never seen an article that explained how "high value" lawyers will gain the expertise they need to give that value in the absence of seeing a lot of "stuff," most of it through antiquated doc review.Would love to hear folks' thoughts.    For context, I've practiced 10 years at an AmLaw 100 firm, with my practice focused on securities offerings, M&A and public company reporting.

[**Agent\_Polsky**](http://disqus.com/Agent_Polsky/)My reading in this area seems to draw the same conclusions, yet offers no real solutions.  A Professor I know well maintains that the law is devolving into a "two tier" profession, with a top layer dominated by those "high level" strategic thinkers that you identify, and the lower tier for the "blue collar" "super-paralegals" who will do all the grunt work and be paid at like $25 / hour or something like that.  This is in a sense dangerous for the profession -- at least much of the commentary I've read on the issue thinks that it is -- but then again this is old news.  The law has been going through a "crisis of professionalism" since the nineties it seems, at least from the reading I've done.  Lawyers are becoming (or have already become) in a sense mere cogs in a machine, and the law is becoming nothing but another commodity, bought and sold on the open market in discrete packages, like any other service.  The articles I've read compare this current to a sort of former 'golden age' when communities looked up to lawyers as trusted leaders of the community.  Of course, that may be somewhat of a gloss on history, but nonetheless I think there is some validity to the idea.  Lawyers weren't always hated or despised in the popular culture, as they seem to be today; not that you cannot find exceptions to this, but in general.  It is just much more of a scramble now, and there is no getting around it.  The law requires an apprenticeship period to really turn out good lawyers, and if the clients are no longer willing to pay for this, then other mechanisms will have to somehow pick up the slack.  As to what those "other mechanisms" are, your guess is as good as mine -- you have a lot of experience in practice, so how did you learn your profession?  The information I looked at about this "crisis of professionalism" in the law discussed this problem as far back as the nineties -- so this has been and probably will be an ongoing problem.  It goes to the culture in general as well.  But then again this has always been a tension in the capitalist West.  Look at the Charlie Brown Christmas special, the real good one from the sixties, where Charlie Brown is complaining about the commercializing of culture.   The law is now a "brittle" profession, with big firms being forced to shrink the size of their staff in order to remain competitive.  This has just rippled out.  For example, medium sized firms in smaller markets now have access to talent, such as former Supreme Court clerks, that they never would have had ten or perhaps even five years ago.  This of course pushes everyone out a little further towards the edge of the plate.  If you are already on the edge, then sayonara.  I think technology is the key, and as someone just entering the profession, I am taking steps to figure out where that technology is going.  Of course, it's no easy task.  It seems that the law is no longer a secure middle class profession, or in any case, will not be for much longer.  I was discussing with a professional how at the time he graduated law school, there was no thought of NOT getting a legal job, unless you somehow didn't want one.  Now, I can say that the perception is very real that a job is not guaranteed, especially at a lower tier school.  The really interesting thing is that higher LSAT scorers are figuring this out, and going to business school instead.  I think the latest study indicated that the highest drop in law school applications was in around the 160-170 LSAT range.  Not the highest end of the curve, but smart people nonetheless.  Smart enough to figure out that this is no longer a secure market, especially if you are not at the highest end of the curve, and the costs are too high relative to the return.  This of course will force some higher ranking schools to dip into the lower LSAT scores, and the ripple effect may further strain lower tier schools...  Personally, I'm hoping my daughter takes to math and science.  It seems a much more secure way to earn a living...

[**kidmugsy**](http://disqus.com/kidmugsy/)How!  Quite soon it may be that the number of apprentice lawyers required will be so small that the requirement can be easily filled by hiring only the children of lawyers.

[**Fat\_Man**](http://disqus.com/Fat_Man/)I was a lawyer, my father was a lawyer, and my grandfather was a lawyer. None of my kids would consider going to law school. Few of my lawyer friends' children are going to law school.

[**kennethanderson1956**](http://disqus.com/kennethanderson1956/)Well, only if they do an unpaid internship for three or four years to weed out the people who have to repay student loans without parental support.  (Raising a tough ethical question: Lawyer parents or lawyer children - which first to go come the revolution?)

[**jrlowery**](http://disqus.com/jrlowery/)I think the larger issue is not the loss of contract attorneys but the continued loss of need for young associates.  As a new lawyer starting out in a big firm, you typically filled your time doing document review and other drudgery while getting some opportunities to help out with or sit in on depositions, brief-writing, client-advising, trial-running, and the rest of the more glamorous legal work that can't be replaced by machines, and that most would consider to be the valuable tools that clients pay for.  Combine software replacement of the drudgery and clients' now-typical request not to pay for "training" of new associates, and you really have to wonder how big firms are going to support 20-40 entry-level associates that know nothing to turn into the needed bunch of mid-level/senior associates.  I would think there could be changes to the pyramid structure of modern firms and focus on recruiting and then retaining a much smaller number of new associates.  The continued reduction of entry-level positions will be the hardest thing though.  As is its hard enough to break into lawyering, and at least a few years experience doing scutwork at a firm used to be one of the more stable ways to do so, even if you didn't plan on staying at a firm long-term.

[**Harry Huntington**](http://disqus.com/Harry_Huntington/)Predictive coding is very useful in certain those kinds of litigation where you know the documents for which you are looking at the time you devise the coding scheme.  Some times life is not so simple.  Some times you would rather have the computer read every document and identify for you the universe of words used in the documents and the frequency with which words and phrases are used.  For a multi-million page document dump, frequency analysis (or whatever term your vendor uses) can be surprisingly useful.  You let the computer tell you what words the other folks commonly use when discussing an issue and read those documents first.  Then you go back and devise a coding scheme that relies (in part) on your target's jargon.  In other cases you want to know who is speaking and to whom.  You want to map patterns and volume of email traffic (or telephone traffic) between folks.  It can be quite instructive when this analysis shows that two senior company managers trade email all the time, but there is not a single email between them on the subject of your litigation.  Smart managers no longer use email on the most sensitive topics; these days you usually do not expect to find key documents on sensitive matters no matter how you code.  However, you expect to find spread sheets created (often deleted and never emailed) on hard drives of staff people who report to those senior managers who avoid speaking to one another in email.  The most sensitive analysis in many cases are done by running a computer model on a personal computer that is never saved, but whose results might be written in part by hand (and never copied).  The best work never goes on the Sharepoint or the shared hard drive.  The future of electronic discovery in major litigation lies with technical specialists (likely ex-NSA types) who will understand how to do a complete electronic capture of all data that could lead to the discovery of evidence.  Evidence is not just documents understood traditionally, but can include text messages, phone records, cell phone records, images (sent from your camera phone), credit card receipts, ATM records (or similar documents that establish a person's location at a point in time), and other electronic files.  Evidence may also be what was loaded (and then deleted) from thumb drives.  Discovery battles will reach into people's personal electronics.  Who has not sometimes sent email from a personal Gmail account?  Or isn't the personal Gmail sent to a spouse "relevant" when it reads: "Honey, will be home late tonight, big things at work."  What senior executive does not have two computers in the office: the work computer (perhaps a Dell or Lenovo) and also a personal Ipad or Macbook. Shouldn't you be capturing their personal hard drive to see if they ran the analysis on their Macbook to avoid leaving a bad electronic trail on their work computer?  Shouldn't you be discovering the personal email accounts of all of the business people as well as the work accounts?  Shouldn't you be tracking their personal cell phones and their work cell phones?  The work for discovery counsel in major litigation going forward is to capture all of the information that can be captured to establish every fact about how people were communicating during a relevant time period.  Although there may be less space for contract lawyers in the new age, you will need many more actual lawyers who understand how to wade through all of the electronic information, and how to make a case in a world where savvy business people understand how to walk across treacherous ground and avoid leaving any electronic footprints.  Just to be clear here, if it is not already true, it will be the case shortly that corporate boards at the smartest companies will prohibit all corporate officers (and other senior managers) from using email at all.  Better to have no electronic record.  The usual deposition answer will be: "No I don't use email at all.  If people send me things, they go to my assistant.  My assistant prints items out and gives me the ones I need to read.  My assistant deletes all items as they are printed.  So it is fair to say that just because someone sent it, I may never have seen it.  I would have no way of knowing what my assistant did not print."  And in response to the next question: "After I read items, I hand them back to my assistant who shreds them. I save nothing. I have no files.  I keep some hand written notes and a calendar.  I discard the hand written notes regularly when my working folder gets too stuffed and I no longer need the notes. I shred the notes I discard."  Then the next set of answers: "We have a great office.  When I need to talk to people I knock on their doors.  I avoid the phone for in-office calls.  I rarely use my cell phone for any calls of substance."

[**Andy Kent**](http://disqus.com/disqus_iCp3HNFLmj/)All that is great in theory, but if you're working in a modern corporation and your work is related to discoverable documents, but can be conducted without you actually interacting with those documents regularly, you're probably a prime candidate for downsizing. The job isn't "avoid leaving evidence that can be used in a lawsuit," it's "do what you were hired to do," and usually those are not the same thing.  The whole premise of electronic discovery is that people are producing documents in the course of ordinary business. You can't expect it to turn up things that are carefully hidden by a cabal of careful geniuses any more than you'd be able to prove the existence of paper memos that were buried in the back yard without being seen by others. (Let's be honest, if your targets are smart enough to be this careful, they're going to be smart enough to use things like TrueCrypt hidden volumes and data destruction tools that even the closest forensic analysis can't possibly beat anyway.)  On top of that, actually recovering stuff from personal e-mail accounts, personal computing devices not on the corporate network, deleted data from drives, etc., requires experts whose time is expensive as hell, and even then has a very low chance of discovering anything useful. Too expensive to go fishing in (and good luck convincing a judge in ordering that the other side conduct the computer equivalent of proctology exams just because you have a hunch).

[**OldToby**](http://disqus.com/OldToby/)To add some concrete verbiage to Harry’s excellent note, there will be “standard AIs” for many issues. Class action law suits may use them to filer millions of potential participants.   There will also be smaller subsets, and custom needs. I can imagine a subset of law based on crafting of custom sets of rules and frameworks, what the CS set calls ontologies. Considerable effort is underway in academia and business to create formal ontologies, and these efforts rely on far more technologies than the relatively well-known OWL.   Many white-shoe firms or other variants of Big Law got big (as I understand it – I welcome correction) because they had a particular understanding of a particular field of law, and a special theory on how to apply law within that small sub-field. I could argue with little urging that each legal such legal theory and subfield can be described by a formal ontology, and that small differences in ontology make all the difference not only in efficacy of document discovery, but in application of said documents (sorting) to build an argument.    If all the value of Big Law is in a few proprietary ontologies, those will be guarded closely. There will be another discipline, of how to craft searches based on an ontology that do not reveal the full ontology; no one will want that value stream to walk out the door. I can even imagine little poison pills, additions to the search ontology to mislead competitors about the core underlying ontology.    It has been said that each business has its own value proposition, i.e., its own ontology. Internal auditors might use these same approaches to look for transactions that appear to violate the business ontology as markers that might indicate fraud.   Many drudge-positions will be lost, but there will be new ones, as Harry suggests, ones that are quite stimulating to envision.

[**kennethanderson1956**](http://disqus.com/kennethanderson1956/)Thanks for this - really interesting and perceptive.

[**kidmugsy**](http://disqus.com/kidmugsy/)How!  Will bankruptcy lawyers do well out of this Brave New World?

[**Andy Kent**](http://disqus.com/disqus_iCp3HNFLmj/)There are several problems with this technology.  First, yes, of course it has trouble with anything where the text isn't easily machine-readable. Even OCR's only about 90% reliable. You can't rely on simple searches being 100% effective at finding terms in OCRed documents, forget complicated stuff.  On top of that, predictive coding has massive defensibility problems. "Well, sir, we reviewed a thousand documents, and put them in a black box, and the black box told us that only these three thousand were responsive, out of our document set of five million." Were the thousand documents -in any way- representative of that five million? Were there concepts in the greater universe of documents that didn't show up in your training set, but are plenty responsive to the request for discovery? More to the point, can you explain what's going on to a skeptical judge who's itching to slap your firm with sanctions in the six-digit range?  Granted, this kind of thing IS useful for culling out some of the chaff - while saying "hey, these documents are responsive, anything that's like these docs is also responsive but nothing else is" is a bit of a stretch, saying "hey, these documents are crap, anything that looks anything like them is likewise crap" can cut down on the time spent reviewing the spam your client didn't filter out, the birthday party invites, the daily newsletters from unrelated third parties, that sort of thing. But even for those, you don't really need predictive coding, just a good process that allows your reviewers to report that stuff for mass searching and deletion.  One judge approving it is noteworthy, sure, but it's hardly a sea change moment. That will come when there are established standards of predictive coding such that your firm doesn't get sanctioned even when your use of predictive coding causes you to massively underproduce - and we're a long, long way from that.

[**jrlowery**](http://disqus.com/jrlowery/)It's not just one judge.  The Federal Circuit (and the Eastern District of Texas) have promoted model discovery orders focused on email discovery limiting it to 5-8 custodians and 5-10 search terms per custodian.  And parties are increasingly agreeing to software-based review for responsiveness when both sides acknowledge the ridiculousness of e-discovery costs, which helps with defensibility.  There will be lawyers and clients that try to game the system or screw over their opponents through misuse of this stuff, but I don't see how that's a change from how things currently are.

[**CounselorJason**](http://disqus.com/CounselorJason/)As important as this case is, the more important decision (IMO) came put of Virgnia state court about a week and a half ago. In that case -- unlike the case before Judge Peck -- the parties disagreed on the very premise of using predictive coding. In the Virginia case -- unlike in Peck's DaSilva Moore case -- the parties disagreed on the very premise of using predictive coding. Defendant wanted to use it; plaintiffs did not want to. The Virginia Court -- via minute order, so there's not much to analyze -- authorized the use of predictive coding. That's the case that is more likely to influence litigants to use predictive coding because thats the case that says "you don't need to convince the other side to trust in this technology; even if they won't accept it, the courts will.". Well, at least one court will, but whee one does, others will (hopefully follow).  As for the contract attorneys, they don't need to go looking for new jobs quite yet. As I note below, this technology is no substitute for humans. What it will do, however, is make review more efficient. For some time now, document review costs have comprised an obscene portion of the total costs of litigating a commercial matter. This technology has the potential to move the needle back in the right direction.

[**David Welker**](http://disqus.com/facebook-681863557/)I cannot say that I see anything that eliminates mindless grunt work as anything other than progress.  What should really be in question is the requirement of three years of law school and the massive amounts of debt necessary nowadays to become an attorney. It really makes you feel bad to see someone invest so much in a legal education only to be unable to find meaningful work as a lawyer. The barriers to entry probably ought to be lower and perhaps more specialized. Contract lawyers who do document review might have jobs, but certainly not especially meaningful ones. You might as well try to do a job that actually benefits society and also makes better use of your education.  Obviously, there are no easy answers for the legions of attorneys who ALREADY have six figures of debt and are actually probably grateful to have contract employment opportunities available, as undesirable as they may be. However, rather than even thinking of resisting innovation, I would suggest the solution would be allowing student loans to be discharged in bankruptcy, which would allow an actual fresh start for impacted individuals.  One thing that is interesting is that people often distinguish between mortgages and education degrees when it comes to loans by making the point that an education, unlike a house, cannot be seized by a lender. But what about when a house burns down? In that case, the lender cannot take the house (and if anyone wants to quibble by saying they can take the land, think of a more extreme natural disaster that makes the land valueless as well). Well, what about educations that are rendered obsolete by creative destruction? Aren't such degrees more analogous to burned down houses? After all, it is not like the debtor is in a great position to benefit either.  In any case, whether student loans are ever again made dischargeable in bankruptcy or not, I welcome such innovation. In terms of technology, we are all very lucky to be witnessing this era. I think the recent drop in law school applications is a strong indication that students are starting to understand that law school is probably NOT the best place to try to wait out current economic hard times. People should go to law school only if you truly are committed and have a strong idea of what you actually want to accomplish by doing so. After all, law school is not the ticket to a fairly stable upper middle class existence than it used to be. And realistically, as technological progress continues, things are probably going to go even more strongly in that direction.

[**David M. Nieporent**](http://disqus.com/dnieporent/)About once in a cicada's life cycle, I agree with Welker, but he's right here (\*).  This development, to the extent it becomes widespread, is not destroying good attorney jobs; it's destroying terrible attorney jobs.  At one point document review may have been entry-level drudgery at a law firm on the way to a more meaningful assignment, but now it's just dead-end contract work that's mindless and pays terribly, and provides no useful transferable experience.  All it really does is allow an unemployed young law school grad to save face by allowing him to tell people (and himself) that he's working as an attorney instead of as a barista.  (\*) Except for his confused point about housing burning down.  Lenders don't give mortgages on houses unless the houses are insured.  If the house burns down, the lender is covered.  As far as I know, one can't take out employment insurance against one's job becoming obsolete.  EDIT: I see Kazinski already made that last point.

[**Kazinski**](http://disqus.com/Kazinski/) You should think your analogies through a little better.  Banks require insurance on any building that they write a mortgage on.  So if the building burns down the bank still gets paid.  Student loans also have an Income Based Repayment Plan which adjusts repayment levels based on income and family size.

[**David Welker**](http://disqus.com/facebook-681863557/)Okay. Imagine an act of war, which insurance policies generally do not cover.  I knew SOMEONE would quibble with the analogy. It is not really the point.  In terms of IBR program, as you may or may not know, that federal program is NOT available for private student loans. It is only available for federal student loans.

[**blaher**](http://disqus.com/blaher/)The idea that lawyering is still a golden ticket to riches or at least a steady living seems to be very alive and well with the people I talk to. I knew people attending or with eager aspirations toward nonranked schools that are only known locally if at all.  The general public seems oblivious, and it is only on places like this that I get a different story.  Maybe  the Eugenes of the world should count their lucky stars. For every hotshot professor there are countless drudges slaving at a keyboard or over a frappuchino and countless more bright eyed graduates marching to replace them. I guess its better than a liberal studies degree at least.

[**Bruce Hayden**](http://disqus.com/google-33950f2bffcb8bb2794c1d3b48388332/)I think that EV is the wrong example here - he apparently got his CS/math degree at an age when most of us were worrying about our first prom date.

[**blaher**](http://disqus.com/blaher/)I didn't know that but that still doesn't chance the point that law is a risky profession to take.

[**David Welker**](http://disqus.com/facebook-681863557/)Yes, but I think many people choose less economically remunerative majors thinking all along that they will eventually probably go to law school. This sort of plan made a lot of sense at one time, but much less so nowadays.  On the other hand, it sort of makes one sad to think of universities changing to focus too excessively on STEM majors. But what can you do? All sorts of knowledge is fascinating and important, but it is not all equally marketable.

[**Bruce Hayden**](http://disqus.com/google-33950f2bffcb8bb2794c1d3b48388332/)I disagree that all knowledge is equally important. Rather, I would argue that knowledge that ultimately results in a betterment of our physical well being is, at least now still, more important, because it allows us the opportunity for experiencing other types of knowledge. Think of it maybe as addressing [Maslow's hierarchy of needs](http://en.wikipedia.org/wiki/Maslow%27s_hierarchy_of_needs).  Ask this, whether STEM or, say, the humanities, have allowed us to have the Internet that we are using today, the cars that we drive everywhere, the planes for long distances, etc. Are there any recent novels that have impacted our country (and, indeed, the rest of the world), to the extent that, for example, the personal computer has?

[**David Welker**](http://disqus.com/facebook-681863557/)I never said anything about knowledge being equally important. I think that question is sort of nonsensical, because it must be qualified. Equally important for WHAT?  Maybe my goal in life is to really understand people better. Because my priority in life is to have a better relationship with my significant other. Knowledge of computer science is not going to be very valuable WHEN that is my goal. Knowledge of psychology, literature, and poetry might be more useful.  What knowledge is most important depends on exactly what you want to accomplish. And it varies from individual to individual and within the same individual at different times in their lives.

[**blaher**](http://disqus.com/blaher/) STEM has problems of its own. Bio is flooded with failed doctor wannabes compared to other STEM majors and the graduate/postdoc system is broken and abusive. Comp Sci has outsourcing and age discrimination to deal with. Not as bad as I hear it is for humanities majors but still there are difficulties.

[**David Welker**](http://disqus.com/facebook-681863557/)I think a lot of so-called age discrimination that arises in computer science is because people get comfortable in dead-end jobs and fail to learn new technologies. This can arise because someone may end up specializing in maintaining some legacy technology (especially when you have the problem of vendor lock stopping a particular organization for upgrading to more modern technologies) that is simply no longer very good. There are very few industries that are as meritocratic as computer science; people generally can really tell if you know your stuff or not in interviews.  In terms of outsourcing, salaries in computer science are still relatively high despite outsourcing. The reason is because the demand for improved technology is insatiable. After all, computer science is the one field that makes all other fields more efficient and more effective. Also, it is very difficult to outsource certain components, since it is difficult to communicate requirements and monitor whether an outsourced team is going in the right direction or not. That said, outsourcing visual components where one can tell whether the product is going in the right direction or not right away is probably a good cost saving strategy. (And one that is good for humanity; who says all of the good jobs should remain in the United States when there are talented people all over the world??? If the cure for some sort of cancer comes from India, that is good for everybody. Economic development in the BRIC countries benefits all of us.)  You are absolutely right about biology. I do not think that is a good major for an undergraduate interested in becoming a doctor, unless it is paired with computer science (a.k.a. bioinformatics) since people need to realize that there is a really high chance that they will not make it in the cut throat competition for slots in medical school. There are very few jobs available for biologists relative to their numbers. And the ones that do exist (in labs at least) tend to be underpaid and are probably too repetitive to be enjoyable for most people. On the other hand, biology paired with computer science (a.k.a. bioinformatics) can and will eventually totally change the world as we know it.  In the future, I think knowledge of computer science will probably be a prerequisite for an ever increasing number of people who aspire to some sort of middle class life.

[**Bruce Hayden**](http://disqus.com/google-33950f2bffcb8bb2794c1d3b48388332/)I would mostly agree with David here, except  that some of the age discrimination is real. The assumption maybe that people can't learn these newer technologies, even if they have.  As far as your suggestion that everyone should learn CS, again I come back to aptitude. Not everyone can learn all that much CS. For example, back when I was doing software engineering, I specialized in OS internals and data communication software design. Both required the ability to visualize the real time interactions of multiple processes and processors, to detect timing windows, and to eliminate them. I was surprised to find that even many of those working in this niche couldn't do this very well. And, yes, I was surprised that most of the people in our IT group in the firm I was in, couldn't do simple Visual Basic programming (or, at least simple to me - which caused innumerable problems, since they were supposed to be the experts, and I was one of the attorneys they were supposed to be supporting).  I do agree though that some CS is essential these days to thrive in our society. Just not what I would consider very much.  Finally, let me suggest that the ability to code these searches may not be all that much hard core CS, but rather, depend on the basic understanding of how computers work combined with an understanding of how people think and communicate. Which means that a social science major with a legal background combined with a knoweldge of how computers and computer searching operate might do a much better job than those of us with strong CS backgrounds.

[**Chris Tompkins**](http://disqus.com/chris_tompkins/)I think quite a few lawyers are people who didn't expect to get rich, they were just arts grads who couldn't do math and therefore believed that law was their only professional option.

[**David Welker**](http://disqus.com/facebook-681863557/)I think the problem of people not being able to do math is probably going to change. One common reason people do not do well in math is because they miss important concepts when they are sick or make some other sort of transition and math tends to build on top of itself. With the advent of online education, people will be increasingly able to fill in the gaps.  When more lucrative STEM majors start to face increasing competition, it may result in lower wages. However, this may be counteracted by the tendency of STEM jobs to build on themselves. (With the invention of each new technologies, there are a set of related technologies that adjust themselves in response. And in the realm of information technology, human desires to build better systems seem to be nearly infinite rather than subject to satiation.)

[**Bruce Hayden**](http://disqus.com/google-33950f2bffcb8bb2794c1d3b48388332/)Maybe that is somewhat the problem with STEM, but there is an awful lot of aptitude involved. Watched my kid go through high school into college, and the problems of their friends keeping up in math and physics. Of course, part of it is that my kid has some math aptitude from all four grandparents, and on my side, the expectation was that if you din't do well in math, you weren't trying very hard, and I would love to help out (hint, hint - esp. since I had taught them derivatives in middle school).  To be a science/math major, there is invariably a class that separates the wheat from the chaf. When I was an undergraduate, it was Organic Chemistry  for chem majors and Linear Algebra for math majors. At my kid's school, for the later, it was a class that they called "Bridge to Lower Self Esteem". I would have floundered in OChem, but enjoyed Linear Algebra. Calculus is concrete. The math separation classes determine whether or not you can think abstractly, and if you can, how easy it is for you. And, most people don't seem to be able to make that transition, no matter how hard they try.  I doubt that we are going to see much movement towards STEM in this country, because of the reality that they are much more objectively based (i.e. you can't BS a good grade) and with the exception of math, require significantly more work and dedication than most other majors (which is why my undergraduate degree was in in math and not physics). Our problem here is that there seem to be much higher percentages of the population who are willing to work that hard in countries like China and India, each of which have 3-4 times our population.  In any case, I think that you need a combination of of aptitude, opportunity, and desire to survive a STEM major,  and you are only talking about opportunity, when you talk about kids being sick, and that is the easiest aspect to address.

[**Unemployed\_Northeastern**](http://disqus.com/Unemployed_Northeastern/)I'll take all you said about the requirements and rigors of being a science/math major.  So what?  Does this somehow insulate or protect science/math majors from the harsh realities of outsourcing, automation, H1B hiring preferences, and the like?  Two examples:  1) I live a few miles from one of the country's largest math-software firms.  It's an open secret that essentially their entire workforce (minus the management) came in on a H1B work visa, not necessarily because of higher aptitude so much as much lower labor costs.  2) Dr. Keith Devlin - a mathematician at Stanford - penned an interesting story on HuffPo (bear with me) a few months ago entitled "All the Math Taught At University Can Be Outsourced.  Now What?"  Selected quotes from his article: "In particular, procedural mathematics (solving differential equations, optimizing systems of inequalities, etc.) can be outsourced....  With a few keystrokes, a designer or a CTO in New York or San Francisco can send a mathematical problem to India at 5:00 PM and by 9:00 AM the next morning the solution is back, ready to be used.   In fact, this is happening now, with companies such as Infosys, Tata Consultancy, Cognizant, HCL, Wipro, and iGate Patni. For example, iGate Patni is a Silicon Valley headquartered, Indian IT outsourcing company with over 26,000 well-educated employees who perform such tasks as writing smartphone apps, handling complex financial matters, and optimizing business logistics processes.   The outsourcing of mathematics and mathematics-dependent STEM activities is only going to increase. It's a question of sheer numbers. In China, with a population of 1.3 billion, and India, population 1.1 billion, there is enormous pressure on children (both parental and self-motivational) to secure a good education leading to a secure future, and that will inevitably produce more and more highly able mathematicians, scientists, and engineers. The US, with a total population of 300 million, less than a third of each of those two giants, cannot possibly compete -- even if we were to completely overhaul our STEM education."

[**SoronelHaetir**](http://disqus.com/SoronelHaetir/)I would argue that the aptitude itself can be taught if you get ahold of the kids early enough.  I've taught some very basic derivitive calculus to quite a few 4th and 5th graders over the years (mostly using classic problems like how fast a pile of sand that models a perfect cone is growing and that sort of thing).  I will agree that there is a point where the ability simply isn't present to do that sort of work (at least at that age), but I would also say that far more children are capable of it than most people realize.   Based on my experience I suspect that all but about the lower 1/4 to 1/3 could learn such material, at least in a one on one setting.  It of course becomes very much more difficult if you are trying to do it with a group like school teachers invariably must.  And of course, if you are talking about students near that cutoff then they are at least somewhat unlikely to have parents that can or will fill in the gaps.

[**Chris Tompkins**](http://disqus.com/chris_tompkins/)The author of the great book "Innumeracy" wrote that people can proclaim themselves to be mathematically challenged, as if that were a cute, endearing trait, but they'd never dream of making a similar comment about their impaired ability to read or write, because people would consider them to be an uneducated dolt. In other words, he's saying that there's enormous cultural pressure to be able to read and write fluently, but very little pressure to be skilled at mathematics. It makes one wonder how many capable STEM grads we'd have if the cultural pressure were there.

[**Mitch l.**](http://disqus.com/yahoo-NGSFBIJTCSMYFRFWV3BOXSLZPY/)Now we have to wonder:  as virtual AI programs and their less elegant robot counterparts come into being, will law school give them preference in student admissions and faculty hiring for clearly being minority? Or, will they be thought to possess a distinct, techno-hegemonic advantage over the typical flesh and blood legal mind unit, and so be either mainstreamed or marginalized for their superior ability, much like Asian applicants these days?

[**Bill\_Woods**](http://disqus.com/Bill_Woods/)They won't be a minority for long....

[**chokanson**](http://disqus.com/chokanson/) yes, I for one, welcome our new robotic overlords!

[**Bill\_Woods**](http://disqus.com/Bill_Woods/)From a (gulp) 30-year-old novel,  "Here we are, replacing doctors, lawyers, priests, and bureaucrats, freeing humanity from the insolence of humans who use place and property as a means to dominate their fellows. And you complain?"

[**BRobCleveland**](http://disqus.com/BRobCleveland/)Here is the problem -- the issue is (1) screening documents for relevant, produceable documents, and (2) screening privileged documents out of that group of  relevant, produceable  documents. Sometimes you do not know whether a document is privileged until it is viewed, mulled over, analyzed, and questions asked of people. Perfect example -- you can used screening words like "lawyer," "attorney," and "counsel" to flag privileged documents, but what if it is an e-mail with an attorney's surname? Or if the surname is misspelled in an e-mail, for instance. Will the screening program pick that up from context, when a program may not "get" context?  Yeah, I sat through a CLE on electronic discovery. Yes, they all say how it is more efficient and cheaper than using attorneys. (What else do you expect them to say?) But just let a couple choice privileged documents get through and have to explain that to the client . . . .

[**CounselorJason**](http://disqus.com/CounselorJason/)Studies have actually found that technology assisted review (aka predictive coding) is more accurate than human review.  Consider -- a large document review these days may include millions of documents, the electronic equivalent of thousands of bankers boxes. Lawyer 1 may review a document on Day 1 of the review. Lawyer 2 may see that same document on Day 30. How confident are you that the identical document will be classified e same way b nth lawyers?  This is not to suggest that technology assisted review is a pancea, or that humans can be removed from the equation. Neither is true. However, what technology assisted review does is enhance consistency and increase the opportunities for triage. Five years ago, any two documents selected at random from the review group were equally likely to be responsive, equally likely to be "hot," equally likely to be privileged. With technology assisted review, you can put in a little time up front in the review and figure out which documents are more likely to be the key documents. Then you can focus on those with your more capable reviewers. Or you can reach agreement with the other side that you don't have to review the ones that fall below a certain likelihood of being responsive. The technology doesn't replace humans. But it does enhancer their productivity and their accuracy.

[**Mark Lyon**](http://disqus.com/marklyon/)With a proper process in place, you can be comfortable that coding among multiple reviewers and across multiple days (excluding, of course, an interim change in protocol) can be accurate and consistent.  At my organization, we guarantee 95% accuracy across our review.  Very rarely are there significant concerns when the process is followed.

[**Gordon V. Cormack**](http://disqus.com/google-f67f91cba44c8d390623d4c439f6797c/) 95% accuracy is, in many situations, inadequate.  That's why information scientists measure recall, not accuracy.  Imagine a collection in which 5% of all documents are responsive.  If you code them all as non-responsive, you get 95% accuracy, but 0% recall.  "Accuracy" is not a useful measure.

[**Mark Lyon**](http://disqus.com/marklyon/)Yes, Professor, 0% recall would be unacceptable.    As a practical matter, though, you're not going to get 0% recall (and, if you allowed such a result unchecked, it would likely be a career-limiting event) if relevant documents are in the collection. Here, we're helped by the fact that most human review isn't limited to a 'relevant' and 'not relevant' decision.  We end up with multiple decisions per document, each one a separate opportunity to correctly or incorrectly categorize a document.  For added fun, some of those decisions are more important than others.  Privilege and Responsiveness defects are far more of a concern than most other tagging.  They get more weight.One of the things I learned quite some time ago is that explaining precision and recall to non-CS folks tends to result in a glazed, faraway look. It's like introducing pointers - the concept immediately clicks with some but leaves many completely bewildered. I find that it is better to go with terms that can be more easily understood, combining precision and recall into an objective "accuracy" measure; an F-Score, if you like.  This single measure is far more useful to people trying to understand the reliability of the coding than many other metrics, and better reflects their understanding of the process.

[**danheskett**](http://disqus.com/danheskett/)I think a computer scientist would find your use of qualifying phrases a little silly.  For one, there is the inbuilt assumption that 'the process is followed' in the manual review.  That's a scary thought.  From a CS perspective, the starting point is "process is followed", and there is no alternative case.   Second, the concept of an interim change in policy is interesting.  In a well-run document automation system, the valuations of the documents are stored, and so a criteria or policy change can be rapidly re-applied to a huge document set.   In one document system I am familiar with, a single processing environment can process over a million pages of data an hour.  By scaling out to more (cheaply available) nodes, there is no practical limit to the number of documents that can be processed.    Finally, out of curosity, what happens when you need to go from 95% to 99%?  Is that an option that can be offered?

[**Mark Lyon**](http://disqus.com/marklyon/)Processing is a bit different from predictive coding.  It's also possible for automated systems to not execute the desired process correctly.   With many predictive coding options, someone - who must be essentially perfect during every coding decision - codes sample documents from clusters or groupings developed by the system until a certain level of confidence is achieved.  I've tried many of these systems and they can help cut through some of the fluff with broad decisions, but don't do so well in a review with large numbers of coding decisions.  For cutting through a massive population or prioritizing work to meet tight deadlines, though, they should certainly be considered.When comparing human review to machine review, it's important to understand what you're comparing.  There is a very process-driven way to do human review which results in high accuracy.  There is also a "process" - far too often employed - which is essentially a large group of people doing work at random with minimal or nonexistent quality controls.  After all, you're going to put some really smart people on the review - they should be able to know how to handle documents correctly.  Those reviews often result in some very scary work product.  Yes, increasing to 99% is possible.  It does cost a bit more and isn't always necessary or practical.

[★ 0 ](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

**70 comments**

[Newest](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Best](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Newest](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Oldest](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Community](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[⚙](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Login](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Send Feedback](https://www.surveymonkey.com/s/5RBPTTZ?c=http%3A%2F%2Fwww.volokh.com%2F2012%2F05%2F05%2Fis-contract-lawyering-doomed-by-algorithm%2F)

[Share ⤤](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

Share this discussion on

[Twitter](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Facebook](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/) [DavidGolden](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-521405647) • 1 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

Predictive coding has more to do with eDiscovery vendors looking for a competitive advantage than any real advance.

My hunch is that the same expert that tunes the machine learning system could accomplish just as much by honing the set of key words used for traditional search. The result would be the same, but with more transparency.

As far as I'm aware, none of the so-called studies allegedly proving the predictive coding is more accurate than manual review were based on peer-reviewed research. For that matter, I'm not familiar with any recent studies done by a credible scientific or technical organization. Any judge that forced a party to pursue a process that was validated by merely an industry study or two is skating on very thin ice. If someone can point to credible, scientific evaluation of predictive coding, please add a reference to it on this thread.

Further, the down side of predictive coding is that opposing council can (and should) challenge the process, demanding that the entire seed set of documents (the "seed collection", which consists of 'like' responsive documents and 'unlike' non-responsive documents) used to train the prediction system be shared so that the machine learning process can be validated. Given that a seed collection would probably contain 10x-20x or more unresponsive documents for each responsive document (to achieve some sort of statistical significance in training), this represents the turnover of a rather large document collection to opposing counsel. Even with privilege review for the 10-20x unresponsive documents in the turnover (and keep in mind that privilege review is the most expensive review), and a clawback agreement just in case, an unnecessary and enormous liability is still  created. How would you like to explain to your client that the predictive coding technology that you recommended has resulted in the turnover of documents that exposed trade secrets or previously buried legal vulnerabilities?  And there's the real possibility of turning over a document that alerts opposing counsel to a future line of attack, that neither you nor your client even realized might be on the horizon.

As a side note, the 'expert' tuning the system is likely to be a senior attorney on the case. As this attorney is now part of the 'process,' no different from the technicians that collected and processed the electronic stored information, I would expect that the work product by that attorney is discoverable. This is clearly not a comfortable position for Outside Counsel, and the use of less skilled resources places the entire process under a cloud of suspicion.

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[ohara5.0](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-521346988) • 1 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

As an MS Physics, MSEE, I predict that in the future lawyers will be my minions.  Minions have to be almost intelligent, capable of following directions and not very self directed.

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Fat\_Man](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-521311978) • 0 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

The alternative to having computers do that work was not going to be hiring hundreds of associates at $165,000 per year and charging the clients $330/hr for their work  (2400 hrs/yr).  The alternative to computers is India, or Nigeria, or Kenya.

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Visitor\_Again](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#) [> Fat\_Man](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-521311978)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-521371224) • 1 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

Question from someone who never worked for a big firm:  How do you maintain the attorney-client privilege if you farm out work to India, Nigeria or Kenya?

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Fat\_Man](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#) [> Visitor\_Again](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-521371224)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-525193728) • 0 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

 If there is enough money in it. They will figure it out

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[DeeR](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-521309983) • 0 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

Ten years ago document review was interesting because unfiltered email contained personal messages and very private details about the sender's lives. Today, document review has been reduced to looking at only relevant messages in which all sordid details and non relevant data have been stripped out. Recently, firms have been outsourcing to inexpensive Indian document reviewers and using more predictive coding of documents. Attorneys are now mostly reviewing for privilege and not relevance. The huge market for lowly,  inexperienced attorneys as document coders is quickly drying up. Soon firms will employ tiny staffs of attorneys experienced in document review, privilege logs, privilege duping etc to oversee the document review product produced by computers and third world outsourcing. The legal employment bubble is about to burst and leave a tiny demand for lawyers of any skill level.

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Herb Ritts](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-521286980) • 2 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

Tech like this is going to drastically change the practice of law. I wish I understood exactly how better, as I would be better positioned to make a mint or, more humbly and probably, retain a job. Discussions of this general topic generally draw a distinction between "low value" lawyers with little expertise (e.g., contract attorneys on doc review) and "high value" lawyers with very rare, very human expertise (e.g., the lead litigation or transactional partner making strategic decisions). I'm a corporate lawyer, so I'll use a corporate law context for the rest of this post. My question is, without the scut work of doc review or due diligence, how will the next generation of "high value" lawyers learn the skills that make them "high value" in the first place? In my experience, most attorneys teach more junior attorneys when a specific issue arises. I've never had a lawyer senior to me say, in the absence of a live issue, "Lad, let's take an hour to discuss the enforceability of offer letters in State X in case you, as a potential 'high value' lawyer, need to know this later."  I have had the discussion after actually reviewing badly written offer letters during diligence. In turn, I've never grabbed a junior associate and said, "Look, it's 8 pm on Friday night; we're both off the clock. This seems as good a time as any to teach you when, in a merger, to worry about an ambiguous non-assingment clause and when not to, even though you've never actually read a non-assignment clause. Because the algorithms do that." To be clear, I'm not saying that this is necessarily a GOOD way to train lawyers. I am saying that this IS the way that lawyers are trained now.  And I'm also saying that I've seen lots of articles about how the computers will replace the "low value" lawyers, but never an article that explained how we will continue to have "high value" lawyers. Put another way, I've never seen an article that explained how "high value" lawyers will gain the expertise they need to give that value in the absence of seeing a lot of "stuff," most of it through antiquated doc review.Would love to hear folks' thoughts.

For context, I've practiced 10 years at an AmLaw 100 firm, with my practice focused on securities offerings, M&A and public company reporting.

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Agent\_Polsky](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#) [> Herb Ritts](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-521286980)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-521597355) • 0 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

My reading in this area seems to draw the same conclusions, yet offers no real solutions.  A Professor I know well maintains that the law is devolving into a "two tier" profession, with a top layer dominated by those "high level" strategic thinkers that you identify, and the lower tier for the "blue collar" "super-paralegals" who will do all the grunt work and be paid at like $25 / hour or something like that.  This is in a sense dangerous for the profession -- at least much of the commentary I've read on the issue thinks that it is -- but then again this is old news.  The law has been going through a "crisis of professionalism" since the nineties it seems, at least from the reading I've done.  Lawyers are becoming (or have already become) in a sense mere cogs in a machine, and the law is becoming nothing but another commodity, bought and sold on the open market in discrete packages, like any other service.  The articles I've read compare this current to a sort of former 'golden age' when communities looked up to lawyers as trusted leaders of the community.  Of course, that may be somewhat of a gloss on history, but nonetheless I think there is some validity to the idea.  Lawyers weren't always hated or despised in the popular culture, as they seem to be today; not that you cannot find exceptions to this, but in general.  It is just much more of a scramble now, and there is no getting around it.  The law requires an apprenticeship period to really turn out good lawyers, and if the clients are no longer willing to pay for this, then other mechanisms will have to somehow pick up the slack.  As to what those "other mechanisms" are, your guess is as good as mine -- you have a lot of experience in practice, so how did you learn your profession?  The information I looked at about this "crisis of professionalism" in the law discussed this problem as far back as the nineties -- so this has been and probably will be an ongoing problem.  It goes to the culture in general as well.  But then again this has always been a tension in the capitalist West.  Look at the Charlie Brown Christmas special, the real good one from the sixties, where Charlie Brown is complaining about the commercializing of culture.

The law is now a "brittle" profession, with big firms being forced to shrink the size of their staff in order to remain competitive.  This has just rippled out.  For example, medium sized firms in smaller markets now have access to talent, such as former Supreme Court clerks, that they never would have had ten or perhaps even five years ago.  This of course pushes everyone out a little further towards the edge of the plate.  If you are already on the edge, then sayonara.  I think technology is the key, and as someone just entering the profession, I am taking steps to figure out where that technology is going.  Of course, it's no easy task.  It seems that the law is no longer a secure middle class profession, or in any case, will not be for much longer.  I was discussing with a professional how at the time he graduated law school, there was no thought of NOT getting a legal job, unless you somehow didn't want one.  Now, I can say that the perception is very real that a job is not guaranteed, especially at a lower tier school.

The really interesting thing is that higher LSAT scorers are figuring this out, and going to business school instead.  I think the latest study indicated that the highest drop in law school applications was in around the 160-170 LSAT range.  Not the highest end of the curve, but smart people nonetheless.  Smart enough to figure out that this is no longer a secure market, especially if you are not at the highest end of the curve, and the costs are too high relative to the return.  This of course will force some higher ranking schools to dip into the lower LSAT scores, and the ripple effect may further strain lower tier schools...

Personally, I'm hoping my daughter takes to math and science.  It seems a much more secure way to earn a living...

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[kidmugsy](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-521154497) • 2 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

How!  Quite soon it may be that the number of apprentice lawyers required will be so small that the requirement can be easily filled by hiring only the children of lawyers.

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Fat\_Man](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#) [> kidmugsy](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-521154497)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-521313543) • 0 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

I was a lawyer, my father was a lawyer, and my grandfather was a lawyer. None of my kids would consider going to law school. Few of my lawyer friends' children are going to law school.

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[kennethanderson1956](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#) Conspirator [> kidmugsy](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-521154497)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-521191720) • 2 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

Well, only if they do an unpaid internship for three or four years to weed out the people who have to repay student loans without parental support.  (Raising a tough ethical question: Lawyer parents or lawyer children - which first to go come the revolution?)

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[jrlowery](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520963117) • 2 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

I think the larger issue is not the loss of contract attorneys but the continued loss of need for young associates.  As a new lawyer starting out in a big firm, you typically filled your time doing document review and other drudgery while getting some opportunities to help out with or sit in on depositions, brief-writing, client-advising, trial-running, and the rest of the more glamorous legal work that can't be replaced by machines, and that most would consider to be the valuable tools that clients pay for.

Combine software replacement of the drudgery and clients' now-typical request not to pay for "training" of new associates, and you really have to wonder how big firms are going to support 20-40 entry-level associates that know nothing to turn into the needed bunch of mid-level/senior associates.  I would think there could be changes to the pyramid structure of modern firms and focus on recruiting and then retaining a much smaller number of new associates.  The continued reduction of entry-level positions will be the hardest thing though.  As is its hard enough to break into lawyering, and at least a few years experience doing scutwork at a firm used to be one of the more stable ways to do so, even if you didn't plan on staying at a firm long-term.

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Harry Huntington](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520877609) • 3 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

Predictive coding is very useful in certain those kinds of litigation where you know the documents for which you are looking at the time you devise the coding scheme.  Some times life is not so simple.  Some times you would rather have the computer read every document and identify for you the universe of words used in the documents and the frequency with which words and phrases are used.  For a multi-million page document dump, frequency analysis (or whatever term your vendor uses) can be surprisingly useful.  You let the computer tell you what words the other folks commonly use when discussing an issue and read those documents first.  Then you go back and devise a coding scheme that relies (in part) on your target's jargon.  In other cases you want to know who is speaking and to whom.  You want to map patterns and volume of email traffic (or telephone traffic) between folks.  It can be quite instructive when this analysis shows that two senior company managers trade email all the time, but there is not a single email between them on the subject of your litigation.  Smart managers no longer use email on the most sensitive topics; these days you usually do not expect to find key documents on sensitive matters no matter how you code.  However, you expect to find spread sheets created (often deleted and never emailed) on hard drives of staff people who report to those senior managers who avoid speaking to one another in email.  The most sensitive analysis in many cases are done by running a computer model on a personal computer that is never saved, but whose results might be written in part by hand (and never copied).  The best work never goes on the Sharepoint or the shared hard drive.

The future of electronic discovery in major litigation lies with technical specialists (likely ex-NSA types) who will understand how to do a complete electronic capture of all data that could lead to the discovery of evidence.  Evidence is not just documents understood traditionally, but can include text messages, phone records, cell phone records, images (sent from your camera phone), credit card receipts, ATM records (or similar documents that establish a person's location at a point in time), and other electronic files.  Evidence may also be what was loaded (and then deleted) from thumb drives.  Discovery battles will reach into people's personal electronics.  Who has not sometimes sent email from a personal Gmail account?  Or isn't the personal Gmail sent to a spouse "relevant" when it reads: "Honey, will be home late tonight, big things at work."  What senior executive does not have two computers in the office: the work computer (perhaps a Dell or Lenovo) and also a personal Ipad or Macbook. Shouldn't you be capturing their personal hard drive to see if they ran the analysis on their Macbook to avoid leaving a bad electronic trail on their work computer?  Shouldn't you be discovering the personal email accounts of all of the business people as well as the work accounts?  Shouldn't you be tracking their personal cell phones and their work cell phones?  The work for discovery counsel in major litigation going forward is to capture all of the information that can be captured to establish every fact about how people were communicating during a relevant time period.

Although there may be less space for contract lawyers in the new age, you will need many more actual lawyers who understand how to wade through all of the electronic information, and how to make a case in a world where savvy business people understand how to walk across treacherous ground and avoid leaving any electronic footprints.

Just to be clear here, if it is not already true, it will be the case shortly that corporate boards at the smartest companies will prohibit all corporate officers (and other senior managers) from using email at all.  Better to have no electronic record.  The usual deposition answer will be: "No I don't use email at all.  If people send me things, they go to my assistant.  My assistant prints items out and gives me the ones I need to read.  My assistant deletes all items as they are printed.  So it is fair to say that just because someone sent it, I may never have seen it.  I would have no way of knowing what my assistant did not print."  And in response to the next question: "After I read items, I hand them back to my assistant who shreds them. I save nothing. I have no files.  I keep some hand written notes and a calendar.  I discard the hand written notes regularly when my working folder gets too stuffed and I no longer need the notes. I shred the notes I discard."  Then the next set of answers: "We have a great office.  When I need to talk to people I knock on their doors.  I avoid the phone for in-office calls.  I rarely use my cell phone for any calls of substance."

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Andy Kent](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#) [> Harry Huntington](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520877609)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-521193527) • 3 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

All that is great in theory, but if you're working in a modern corporation and your work is related to discoverable documents, but can be conducted without you actually interacting with those documents regularly, you're probably a prime candidate for downsizing. The job isn't "avoid leaving evidence that can be used in a lawsuit," it's "do what you were hired to do," and usually those are not the same thing.

The whole premise of electronic discovery is that people are producing documents in the course of ordinary business. You can't expect it to turn up things that are carefully hidden by a cabal of careful geniuses any more than you'd be able to prove the existence of paper memos that were buried in the back yard without being seen by others. (Let's be honest, if your targets are smart enough to be this careful, they're going to be smart enough to use things like TrueCrypt hidden volumes and data destruction tools that even the closest forensic analysis can't possibly beat anyway.)

On top of that, actually recovering stuff from personal e-mail accounts, personal computing devices not on the corporate network, deleted data from drives, etc., requires experts whose time is expensive as hell, and even then has a very low chance of discovering anything useful. Too expensive to go fishing in (and good luck convincing a judge in ordering that the other side conduct the computer equivalent of proctology exams just because you have a hunch).

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[OldToby](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#) [> Harry Huntington](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520877609)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520995596) • 0 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

To add some concrete verbiage to Harry’s excellent note, there will be “standard AIs” for many issues. Class action law suits may use them to filer millions of potential participants.   There will also be smaller subsets, and custom needs. I can imagine a subset of law based on crafting of custom sets of rules and frameworks, what the CS set calls ontologies. Considerable effort is underway in academia and business to create formal ontologies, and these efforts rely on far more technologies than the relatively well-known OWL.   Many white-shoe firms or other variants of Big Law got big (as I understand it – I welcome correction) because they had a particular understanding of a particular field of law, and a special theory on how to apply law within that small sub-field. I could argue with little urging that each legal such legal theory and subfield can be described by a formal ontology, and that small differences in ontology make all the difference not only in efficacy of document discovery, but in application of said documents (sorting) to build an argument.    If all the value of Big Law is in a few proprietary ontologies, those will be guarded closely. There will be another discipline, of how to craft searches based on an ontology that do not reveal the full ontology; no one will want that value stream to walk out the door. I can even imagine little poison pills, additions to the search ontology to mislead competitors about the core underlying ontology.    It has been said that each business has its own value proposition, i.e., its own ontology. Internal auditors might use these same approaches to look for transactions that appear to violate the business ontology as markers that might indicate fraud.   Many drudge-positions will be lost, but there will be new ones, as Harry suggests, ones that are quite stimulating to envision.

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[kennethanderson1956](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#) Conspirator [> Harry Huntington](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520877609)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520890039) • 0 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

Thanks for this - really interesting and perceptive.

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[kidmugsy](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520855291) • 0 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

How!  Will bankruptcy lawyers do well out of this Brave New World?

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Andy Kent](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520822192) • 0 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

There are several problems with this technology.

First, yes, of course it has trouble with anything where the text isn't easily machine-readable. Even OCR's only about 90% reliable. You can't rely on simple searches being 100% effective at finding terms in OCRed documents, forget complicated stuff.

On top of that, predictive coding has massive defensibility problems. "Well, sir, we reviewed a thousand documents, and put them in a black box, and the black box told us that only these three thousand were responsive, out of our document set of five million." Were the thousand documents -in any way- representative of that five million? Were there concepts in the greater universe of documents that didn't show up in your training set, but are plenty responsive to the request for discovery? More to the point, can you explain what's going on to a skeptical judge who's itching to slap your firm with sanctions in the six-digit range?

Granted, this kind of thing IS useful for culling out some of the chaff - while saying "hey, these documents are responsive, anything that's like these docs is also responsive but nothing else is" is a bit of a stretch, saying "hey, these documents are crap, anything that looks anything like them is likewise crap" can cut down on the time spent reviewing the spam your client didn't filter out, the birthday party invites, the daily newsletters from unrelated third parties, that sort of thing. But even for those, you don't really need predictive coding, just a good process that allows your reviewers to report that stuff for mass searching and deletion.

One judge approving it is noteworthy, sure, but it's hardly a sea change moment. That will come when there are established standards of predictive coding such that your firm doesn't get sanctioned even when your use of predictive coding causes you to massively underproduce - and we're a long, long way from that.

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[jrlowery](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#) [> Andy Kent](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520822192)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520956445) • 1 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

It's not just one judge.  The Federal Circuit (and the Eastern District of Texas) have promoted model discovery orders focused on email discovery limiting it to 5-8 custodians and 5-10 search terms per custodian.  And parties are increasingly agreeing to software-based review for responsiveness when both sides acknowledge the ridiculousness of e-discovery costs, which helps with defensibility.

There will be lawyers and clients that try to game the system or screw over their opponents through misuse of this stuff, but I don't see how that's a change from how things currently are.

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[CounselorJason](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520784760) • 1 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

As important as this case is, the more important decision (IMO) came put of Virgnia state court about a week and a half ago. In that case -- unlike the case before Judge Peck -- the parties disagreed on the very premise of using predictive coding. In the Virginia case -- unlike in Peck's DaSilva Moore case -- the parties disagreed on the very premise of using predictive coding. Defendant wanted to use it; plaintiffs did not want to. The Virginia Court -- via minute order, so there's not much to analyze -- authorized the use of predictive coding. That's the case that is more likely to influence litigants to use predictive coding because thats the case that says "you don't need to convince the other side to trust in this technology; even if they won't accept it, the courts will.". Well, at least one court will, but whee one does, others will (hopefully follow).

As for the contract attorneys, they don't need to go looking for new jobs quite yet. As I note below, this technology is no substitute for humans. What it will do, however, is make review more efficient. For some time now, document review costs have comprised an obscene portion of the total costs of litigating a commercial matter. This technology has the potential to move the needle back in the right direction.

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[David Welker](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520756192) • 1 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

I cannot say that I see anything that eliminates mindless grunt work as anything other than progress.

What should really be in question is the requirement of three years of law school and the massive amounts of debt necessary nowadays to become an attorney. It really makes you feel bad to see someone invest so much in a legal education only to be unable to find meaningful work as a lawyer. The barriers to entry probably ought to be lower and perhaps more specialized. Contract lawyers who do document review might have jobs, but certainly not especially meaningful ones. You might as well try to do a job that actually benefits society and also makes better use of your education.

Obviously, there are no easy answers for the legions of attorneys who ALREADY have six figures of debt and are actually probably grateful to have contract employment opportunities available, as undesirable as they may be. However, rather than even thinking of resisting innovation, I would suggest the solution would be allowing student loans to be discharged in bankruptcy, which would allow an actual fresh start for impacted individuals.

One thing that is interesting is that people often distinguish between mortgages and education degrees when it comes to loans by making the point that an education, unlike a house, cannot be seized by a lender. But what about when a house burns down? In that case, the lender cannot take the house (and if anyone wants to quibble by saying they can take the land, think of a more extreme natural disaster that makes the land valueless as well). Well, what about educations that are rendered obsolete by creative destruction? Aren't such degrees more analogous to burned down houses? After all, it is not like the debtor is in a great position to benefit either.

In any case, whether student loans are ever again made dischargeable in bankruptcy or not, I welcome such innovation. In terms of technology, we are all very lucky to be witnessing this era. I think the recent drop in law school applications is a strong indication that students are starting to understand that law school is probably NOT the best place to try to wait out current economic hard times. People should go to law school only if you truly are committed and have a strong idea of what you actually want to accomplish by doing so. After all, law school is not the ticket to a fairly stable upper middle class existence than it used to be. And realistically, as technological progress continues, things are probably going to go even more strongly in that direction.

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[David M. Nieporent](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#) [> David Welker](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520756192)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-521140313) • 0 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

About once in a cicada's life cycle, I agree with Welker, but he's right here (\*).  This development, to the extent it becomes widespread, is not destroying good attorney jobs; it's destroying terrible attorney jobs.  At one point document review may have been entry-level drudgery at a law firm on the way to a more meaningful assignment, but now it's just dead-end contract work that's mindless and pays terribly, and provides no useful transferable experience.

All it really does is allow an unemployed young law school grad to save face by allowing him to tell people (and himself) that he's working as an attorney instead of as a barista.

(\*) Except for his confused point about housing burning down.  Lenders don't give mortgages on houses unless the houses are insured.  If the house burns down, the lender is covered.  As far as I know, one can't take out employment insurance against one's job becoming obsolete.

EDIT: I see Kazinski already made that last point.

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Kazinski](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#) [> David Welker](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520756192)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520764097) • 1 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

 You should think your analogies through a little better.  Banks require insurance on any building that they write a mortgage on.  So if the building burns down the bank still gets paid.

Student loans also have an Income Based Repayment Plan which adjusts repayment levels based on income and family size.

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[David Welker](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#) [> Kazinski](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520764097)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520765352) • 0 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

Okay. Imagine an act of war, which insurance policies generally do not cover.

I knew SOMEONE would quibble with the analogy. It is not really the point.

In terms of IBR program, as you may or may not know, that federal program is NOT available for private student loans. It is only available for federal student loans.

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[blaher](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520725209) • 1 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

The idea that lawyering is still a golden ticket to riches or at least a steady living seems to be very alive and well with the people I talk to. I knew people attending or with eager aspirations toward nonranked schools that are only known locally if at all.  The general public seems oblivious, and it is only on places like this that I get a different story.

Maybe  the Eugenes of the world should count their lucky stars. For every hotshot professor there are countless drudges slaving at a keyboard or over a frappuchino and countless more bright eyed graduates marching to replace them. I guess its better than a liberal studies degree at least.

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Bruce Hayden](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#) [> blaher](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520725209)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520889450) • 0 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

I think that EV is the wrong example here - he apparently got his CS/math degree at an age when most of us were worrying about our first prom date.

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[blaher](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#) [> Bruce Hayden](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520889450)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-521041107) • 0 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

I didn't know that but that still doesn't chance the point that law is a risky profession to take.

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[David Welker](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#) [> blaher](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520725209)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520757508) • 0 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

Yes, but I think many people choose less economically remunerative majors thinking all along that they will eventually probably go to law school. This sort of plan made a lot of sense at one time, but much less so nowadays.

On the other hand, it sort of makes one sad to think of universities changing to focus too excessively on STEM majors. But what can you do? All sorts of knowledge is fascinating and important, but it is not all equally marketable.

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Bruce Hayden](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#) [> David Welker](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520757508)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520912050) • 1 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

I disagree that all knowledge is equally important. Rather, I would argue that knowledge that ultimately results in a betterment of our physical well being is, at least now still, more important, because it allows us the opportunity for experiencing other types of knowledge. Think of it maybe as addressing [Maslow's hierarchy of needs](http://en.wikipedia.org/wiki/Maslow%27s_hierarchy_of_needs).

Ask this, whether STEM or, say, the humanities, have allowed us to have the Internet that we are using today, the cars that we drive everywhere, the planes for long distances, etc. Are there any recent novels that have impacted our country (and, indeed, the rest of the world), to the extent that, for example, the personal computer has?

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[David Welker](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#) [> Bruce Hayden](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520912050)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-521171902) • 2 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

I never said anything about knowledge being equally important. I think that question is sort of nonsensical, because it must be qualified. Equally important for WHAT?

Maybe my goal in life is to really understand people better. Because my priority in life is to have a better relationship with my significant other. Knowledge of computer science is not going to be very valuable WHEN that is my goal. Knowledge of psychology, literature, and poetry might be more useful.

What knowledge is most important depends on exactly what you want to accomplish. And it varies from individual to individual and within the same individual at different times in their lives.

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[blaher](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#) [> David Welker](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520757508)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520761737) • 1 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

 STEM has problems of its own. Bio is flooded with failed doctor wannabes compared to other STEM majors and the graduate/postdoc system is broken and abusive. Comp Sci has outsourcing and age discrimination to deal with. Not as bad as I hear it is for humanities majors but still there are difficulties.

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[David Welker](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#) [> blaher](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520761737)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520770538) • 0 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

I think a lot of so-called age discrimination that arises in computer science is because people get comfortable in dead-end jobs and fail to learn new technologies. This can arise because someone may end up specializing in maintaining some legacy technology (especially when you have the problem of vendor lock stopping a particular organization for upgrading to more modern technologies) that is simply no longer very good. There are very few industries that are as meritocratic as computer science; people generally can really tell if you know your stuff or not in interviews.

In terms of outsourcing, salaries in computer science are still relatively high despite outsourcing. The reason is because the demand for improved technology is insatiable. After all, computer science is the one field that makes all other fields more efficient and more effective. Also, it is very difficult to outsource certain components, since it is difficult to communicate requirements and monitor whether an outsourced team is going in the right direction or not. That said, outsourcing visual components where one can tell whether the product is going in the right direction or not right away is probably a good cost saving strategy. (And one that is good for humanity; who says all of the good jobs should remain in the United States when there are talented people all over the world??? If the cure for some sort of cancer comes from India, that is good for everybody. Economic development in the BRIC countries benefits all of us.)

You are absolutely right about biology. I do not think that is a good major for an undergraduate interested in becoming a doctor, unless it is paired with computer science (a.k.a. bioinformatics) since people need to realize that there is a really high chance that they will not make it in the cut throat competition for slots in medical school. There are very few jobs available for biologists relative to their numbers. And the ones that do exist (in labs at least) tend to be underpaid and are probably too repetitive to be enjoyable for most people. On the other hand, biology paired with computer science (a.k.a. bioinformatics) can and will eventually totally change the world as we know it.

In the future, I think knowledge of computer science will probably be a prerequisite for an ever increasing number of people who aspire to some sort of middle class life.

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Bruce Hayden](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#) [> David Welker](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520770538)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520923588) • 1 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

I would mostly agree with David here, except  that some of the age discrimination is real. The assumption maybe that people can't learn these newer technologies, even if they have.

As far as your suggestion that everyone should learn CS, again I come back to aptitude. Not everyone can learn all that much CS. For example, back when I was doing software engineering, I specialized in OS internals and data communication software design. Both required the ability to visualize the real time interactions of multiple processes and processors, to detect timing windows, and to eliminate them. I was surprised to find that even many of those working in this niche couldn't do this very well. And, yes, I was surprised that most of the people in our IT group in the firm I was in, couldn't do simple Visual Basic programming (or, at least simple to me - which caused innumerable problems, since they were supposed to be the experts, and I was one of the attorneys they were supposed to be supporting).

I do agree though that some CS is essential these days to thrive in our society. Just not what I would consider very much.

Finally, let me suggest that the ability to code these searches may not be all that much hard core CS, but rather, depend on the basic understanding of how computers work combined with an understanding of how people think and communicate. Which means that a social science major with a legal background combined with a knoweldge of how computers and computer searching operate might do a much better job than those of us with strong CS backgrounds.

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Chris Tompkins](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#) [> blaher](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520725209)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520744742) • 1 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

I think quite a few lawyers are people who didn't expect to get rich, they were just arts grads who couldn't do math and therefore believed that law was their only professional option.

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[David Welker](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#) [> Chris Tompkins](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520744742)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520758684) • 0 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

I think the problem of people not being able to do math is probably going to change. One common reason people do not do well in math is because they miss important concepts when they are sick or make some other sort of transition and math tends to build on top of itself. With the advent of online education, people will be increasingly able to fill in the gaps.

When more lucrative STEM majors start to face increasing competition, it may result in lower wages. However, this may be counteracted by the tendency of STEM jobs to build on themselves. (With the invention of each new technologies, there are a set of related technologies that adjust themselves in response. And in the realm of information technology, human desires to build better systems seem to be nearly infinite rather than subject to satiation.)

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Bruce Hayden](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#) [> David Welker](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520758684)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520907928) • 1 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

Maybe that is somewhat the problem with STEM, but there is an awful lot of aptitude involved. Watched my kid go through high school into college, and the problems of their friends keeping up in math and physics. Of course, part of it is that my kid has some math aptitude from all four grandparents, and on my side, the expectation was that if you din't do well in math, you weren't trying very hard, and I would love to help out (hint, hint - esp. since I had taught them derivatives in middle school).

To be a science/math major, there is invariably a class that separates the wheat from the chaf. When I was an undergraduate, it was Organic Chemistry  for chem majors and Linear Algebra for math majors. At my kid's school, for the later, it was a class that they called "Bridge to Lower Self Esteem". I would have floundered in OChem, but enjoyed Linear Algebra. Calculus is concrete. The math separation classes determine whether or not you can think abstractly, and if you can, how easy it is for you. And, most people don't seem to be able to make that transition, no matter how hard they try.

I doubt that we are going to see much movement towards STEM in this country, because of the reality that they are much more objectively based (i.e. you can't BS a good grade) and with the exception of math, require significantly more work and dedication than most other majors (which is why my undergraduate degree was in in math and not physics). Our problem here is that there seem to be much higher percentages of the population who are willing to work that hard in countries like China and India, each of which have 3-4 times our population.

In any case, I think that you need a combination of of aptitude, opportunity, and desire to survive a STEM major,  and you are only talking about opportunity, when you talk about kids being sick, and that is the easiest aspect to address.

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Unemployed\_Northeastern](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#) [> Bruce Hayden](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520907928)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-521822270) • 0 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

I'll take all you said about the requirements and rigors of being a science/math major.  So what?  Does this somehow insulate or protect science/math majors from the harsh realities of outsourcing, automation, H1B hiring preferences, and the like?  Two examples:

1) I live a few miles from one of the country's largest math-software firms.  It's an open secret that essentially their entire workforce (minus the management) came in on a H1B work visa, not necessarily because of higher aptitude so much as much lower labor costs.

2) Dr. Keith Devlin - a mathematician at Stanford - penned an interesting story on HuffPo (bear with me) a few months ago entitled "All the Math Taught At University Can Be Outsourced.  Now What?"  Selected quotes from his article: "In particular, procedural mathematics (solving differential equations, optimizing systems of inequalities, etc.) can be outsourced....

With a few keystrokes, a designer or a CTO in New York or San Francisco can send a mathematical problem to India at 5:00 PM and by 9:00 AM the next morning the solution is back, ready to be used.

In fact, this is happening now, with companies such as Infosys, Tata Consultancy, Cognizant, HCL, Wipro, and iGate Patni. For example, iGate Patni is a Silicon Valley headquartered, Indian IT outsourcing company with over 26,000 well-educated employees who perform such tasks as writing smartphone apps, handling complex financial matters, and optimizing business logistics processes.

The outsourcing of mathematics and mathematics-dependent STEM activities is only going to increase. It's a question of sheer numbers. In China, with a population of 1.3 billion, and India, population 1.1 billion, there is enormous pressure on children (both parental and self-motivational) to secure a good education leading to a secure future, and that will inevitably produce more and more highly able mathematicians, scientists, and engineers. The US, with a total population of 300 million, less than a third of each of those two giants, cannot possibly compete -- even if we were to completely overhaul our STEM education."

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[SoronelHaetir](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#) [> Bruce Hayden](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520907928)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520925158) • 0 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

I would argue that the aptitude itself can be taught if you get ahold of the kids early enough.  I've taught some very basic derivitive calculus to quite a few 4th and 5th graders over the years (mostly using classic problems like how fast a pile of sand that models a perfect cone is growing and that sort of thing).  I will agree that there is a point where the ability simply isn't present to do that sort of work (at least at that age), but I would also say that far more children are capable of it than most people realize.   Based on my experience I suspect that all but about the lower 1/4 to 1/3 could learn such material, at least in a one on one setting.  It of course becomes very much more difficult if you are trying to do it with a group like school teachers invariably must.  And of course, if you are talking about students near that cutoff then they are at least somewhat unlikely to have parents that can or will fill in the gaps.

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Chris Tompkins](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#) [> SoronelHaetir](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520925158)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520981126) • 3 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

The author of the great book "Innumeracy" wrote that people can proclaim themselves to be mathematically challenged, as if that were a cute, endearing trait, but they'd never dream of making a similar comment about their impaired ability to read or write, because people would consider them to be an uneducated dolt. In other words, he's saying that there's enormous cultural pressure to be able to read and write fluently, but very little pressure to be skilled at mathematics. It makes one wonder how many capable STEM grads we'd have if the cultural pressure were there.

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Mitch l.](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520720987) • 2 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

Now we have to wonder:  as virtual AI programs and their less elegant robot counterparts come into being, will law school give them preference in student admissions and faculty hiring for clearly being minority? Or, will they be thought to possess a distinct, techno-hegemonic advantage over the typical flesh and blood legal mind unit, and so be either mainstreamed or marginalized for their superior ability, much like Asian applicants these days?

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Bill\_Woods](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#) [> Mitch l.](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520720987)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-521130015) • 0 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

They won't be a minority for long....

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[chokanson](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#) [> Bill\_Woods](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-521130015)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-521195593) • 1 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

 yes, I for one, welcome our new robotic overlords!

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Bill\_Woods](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#) [> chokanson](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-521195593)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-521250344) • 2 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

From a (*gulp*) 30-year-old novel,

"Here we are, replacing doctors, lawyers, priests, and bureaucrats, freeing humanity from the insolence of humans who use place and property as a means to dominate their fellows. And you *complain*?"

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[BRobCleveland](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520704404) • 2 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

Here is the problem -- the issue is (1) screening documents for relevant, produceable documents, and (2) screening privileged documents out of that group of  relevant, produceable  documents. Sometimes you do not know whether a document is privileged until it is viewed, mulled over, analyzed, and questions asked of people. Perfect example -- you can used screening words like "lawyer," "attorney," and "counsel" to flag privileged documents, but what if it is an e-mail with an attorney's surname? Or if the surname is misspelled in an e-mail, for instance. Will the screening program pick that up from context, when a program may not "get" context?

Yeah, I sat through a CLE on electronic discovery. Yes, they all say how it is more efficient and cheaper than using attorneys. (What else do you expect them to say?) But just let a couple choice privileged documents get through and have to explain that to the client . . . .

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[CounselorJason](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#) [> BRobCleveland](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520704404)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520782924) • 1 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

Studies have actually found that technology assisted review (aka predictive coding) is more accurate than human review.

Consider -- a large document review these days may include millions of documents, the electronic equivalent of thousands of bankers boxes. Lawyer 1 may review a document on Day 1 of the review. Lawyer 2 may see that same document on Day 30. How confident are you that the identical document will be classified e same way b nth lawyers?

This is not to suggest that technology assisted review is a pancea, or that humans can be removed from the equation. Neither is true. However, what technology assisted review does is enhance consistency and increase the opportunities for triage. Five years ago, any two documents selected at random from the review group were equally likely to be responsive, equally likely to be "hot," equally likely to be privileged. With technology assisted review, you can put in a little time up front in the review and figure out which documents are more likely to be the key documents. Then you can focus on those with your more capable reviewers. Or you can reach agreement with the other side that you don't have to review the ones that fall below a certain likelihood of being responsive. The technology doesn't replace humans. But it does enhancer their productivity and their accuracy.

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Mark Lyon](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#) [> CounselorJason](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-520782924)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-521060574) • 0 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

With a proper process in place, you can be comfortable that coding among multiple reviewers and across multiple days (excluding, of course, an interim change in protocol) can be accurate and consistent.  At my organization, we guarantee 95% accuracy across our review.  Very rarely are there significant concerns when the process is followed.

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Gordon V. Cormack](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#) [> Mark Lyon](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-521060574)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-524302442) • 0 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

 95% accuracy is, in many situations, inadequate.  That's why information scientists measure recall, not accuracy.  Imagine a collection in which 5% of all documents are responsive.  If you code them all as non-responsive, you get 95% accuracy, but 0% recall.  "Accuracy" is not a useful measure.

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Mark Lyon](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#) [> Gordon V. Cormack](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-524302442)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-524465974) • 0 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

Yes, Professor, 0% recall would be unacceptable.

As a practical matter, though, you're not going to get 0% recall (and, if you allowed such a result unchecked, it would likely be a career-limiting event) if relevant documents are in the collection. Here, we're helped by the fact that most human review isn't limited to a 'relevant' and 'not relevant' decision.  We end up with multiple decisions per document, each one a separate opportunity to correctly or incorrectly categorize a document.  For added fun, some of those decisions are more important than others.  Privilege and Responsiveness defects are far more of a concern than most other tagging.  They get more weight.One of the things I learned quite some time ago is that explaining precision and recall to non-CS folks tends to result in a glazed, faraway look. It's like introducing pointers - the concept immediately clicks with some but leaves many completely bewildered. I find that it is better to go with terms that can be more easily understood, combining precision and recall into an objective "accuracy" measure; an F-Score, if you like.  This single measure is far more useful to people trying to understand the reliability of the coding than many other metrics, and better reflects their understanding of the process.

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[danheskett](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#) [> Mark Lyon](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-521060574)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-521289433) • 0 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

I think a computer scientist would find your use of qualifying phrases a little silly.  For one, there is the inbuilt assumption that 'the process is followed' in the manual review.  That's a scary thought.  From a CS perspective, the starting point is "process is followed", and there is no alternative case.

Second, the concept of an interim change in policy is interesting.  In a well-run document automation system, the valuations of the documents are stored, and so a criteria or policy change can be rapidly re-applied to a huge document set.   In one document system I am familiar with, a single processing environment can process over a million pages of data an hour.  By scaling out to more (cheaply available) nodes, there is no practical limit to the number of documents that can be processed.

Finally, out of curosity, what happens when you need to go from 95% to 99%?  Is that an option that can be offered?

[see more](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Mark Lyon](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#) [> danheskett](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-521289433)• [a year ago](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#comment-521293660) • 0 upvotes

[−](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

[Flag as inappropriate](http://www.volokh.com/2012/05/05/is-contract-lawyering-doomed-by-algorithm/#)

Processing is a bit different from predictive coding.  It's also possible for automated systems to not execute the desired process correctly.

With many predictive coding options, someone - who must be essentially perfect during every coding decision - codes sample documents from clusters or groupings developed by the system until a certain level of confidence is achieved.  I've tried many of these systems and they can help cut through some of the fluff with broad decisions, but don't do so well in a review with large numbers of coding decisions.

For cutting through a massive population or prioritizing work to meet tight deadlines, though, they should certainly be considered.When comparing human review to machine review, it's important to understand what you're comparing.  There is a very process-driven way to do human review which results in high accuracy.  There is also a "process" - far too often employed - which is essentially a large group of people doing work at random with minimal or nonexistent quality controls.  After all, you're going to put some really smart people on the review - they should be able to know how to handle documents correctly.  Those reviews often result in some very scary work product.

Yes, increasing to 99% is possible.  It does cost a bit more and isn't always necessary or practical.

**Strict-ish liability? An experiment in the law as algorithm**

Published April 12, 2013 | By [Hannah Maslen](http://blog.practicalethics.ox.ac.uk/author/hannah-maslen/)

Some researchers in the US recently conducted an [‘experiment in the law as algorithm’](http://blogs.law.stanford.edu/werobot/files/2013/04/Shay-et-al_Lisa.pdf). (One of the researchers involved with the project was interviewed by *Ars Technia,* [here](http://arstechnica.com/tech-policy/2013/04/hate-red-light-cameras-youll-really-hate-speeding-ticket-robots/).) At first glance, this seems like quite a simple undertaking for someone with knowledge of a particular law and mathematical proficiency: laws are clearly defined rules, which can be broken in clearly defined ways. This is most true for strict liability offences, which require no proof of a mental element of the offence (the *mens rea*). An individual can commit a strict liability offence even if she had no knowledge that her act was criminal and had no intention to commit the crime. All that is required under strict liability statutes is that the act itself (the *actus reus*) is voluntary. Essentially: if you did it, you’re liable – it doesn’t matter why or how. So, for strict liability offences such as speeding it would seem straightforward enough to create an algorithm that could compare actual driving speed with the legal speed limit, and adjudicate liability accordingly.

This possibility of law as algorithm is what the US researchers aimed to test out with their experiment. They imagined the future possibility of automated law enforcement, especially for simple laws like those governing driving. To conduct their experiment, the researchers assigned a group of 52 programmers the task of automating the enforcement of driving speed limits. A late-model vehicle was equipped with a sensor that collected actual vehicle speed over an hour-long commute. The programmers (without collaboration) each wrote a program that computed the number of speed limit violations and issued mock traffic tickets.

Despite the seemingly clear-cut nature of what it means to break the speed limit, the experiment demonstrated that even relatively narrow and straightforward ‘rules’ can be problematically indeterminate in practice. Even though the programmers worked with quantitative data for both vehicle speed and the speed limit, the number of tickets issued varied from none to one per sensor sample above the speed limit. The results demonstrated significant deviation in number and type of tickets issued during the course of the commute, based on legal interpretations and assumptions made by programmers untrained in the law.

It is perhaps surprising that assumptions would bias an algorithm designed to indicate the frequency and magnitude of speeding offences. What assumptions could be involved when deciding whether the actual driving speed X is greater than the limit of Y? However, the researchers point out that laws were not created with automated enforcement in mind, and that even seemingly simple laws have subtle features that require programmers to make assumptions about how to encode them. For example:

*An automated system […] could maintain a continuous flow of samples based on driving behavior and thus issue tickets accordingly. This level of resolution is not possible in manual law enforcement. In our experiment, the programmers were faced with the choice of how to treat many continuous samples all showing speeding behavior. Should each instance of speeding (e.g. a single sample) be treated as a separate offense, or should all consecutive speeding samples be treated as a single offense? Should the duration of time exceeding the speed limit be considered in the severity of the offense? [p.11]*

When we manually enforce laws relating to speeding– or even when we use speed cameras – we know that these mechanisms capture only a fraction of the total number of instances of speeding. There is also usually a ‘buffer zone’ of a few miles per hour within which a driver might technically be speeding but would not get picked up. Particularly when police officers use speed guns to measure drivers’ speeds, there is room for discretion which cannot be built in to an algorithm. As the researchers say, bias can be encoded into the system but, once encoded, the code is unbiased in its execution. The researchers conclude that discretion after the fact may actually be important even for the simplest of offences, like speeding. Offences requiring the mental element in addition to commission of the prohibited act are likely to be even harder to effectively encode ex ante:

*The question arises, then: What is the societal cost of automated law enforcement, particularly when involving artificially-intelligent robotic systems unmediated by human judgment? Our tradition of jurisprudence rests, in large part, on the indispensable notion of human observation and consideration of those attendant circumstances that might call—or even mandate—mitigation, extenuation, or aggravation. When robots mediate in our stead either on the side of law enforcement or the defendant, whether for reasons of frugality, impartiality, or convenience—an essential component of our judicial system is, in essence, stymied. Synecdochically embodied by the judge, the jury, the court functionary, etc., the human component provides that necessary element of sensibility and empathy for a system that always, unfortunately, carries with it the potential of rote application, a lady justice whose blindfold ensures not noble objectivity but compassionless indifference. [p. 28]*

This, perhaps, is an unsurprising view when considering complex offences that require that the offender acted with intention or knowledge or recklessness. But it also raises interesting questions for strict liability. Might it be the case that strict liability statutes are not only enacted under the *assumption* but perhaps even the *hope* that not all volitions will be picked up? Is the lower resolution of manual law enforcement actually preferable for less serious offences? The answer to this will depend in part on the seriousness of the offence in question and the justifications for the attendant sanctions: Deterrence? Retribution? Generation of revenue?

There is, of course, an important difference between seeing the algorithm as inadequate because it gets something factually wrong and seeing it as inadequate because some discretion might be preferable. For example, the discretion involved in deciding how offences should be delineated as a driver meanders above and below the speed limit is something we might wish to preserve. Further, the experiment demonstrated that hilly terrain caused the vehicle to exceed the speed limit despite the cruise control being set at the speed limit. This inability for a driver to have precision control over her speed provides justification for a buffer zone. Thus, despite the conceptual simplicity of what it means to break the speed limit, the experiment in law as algorithm at least raises the possibility that, in some cases, strict*-ish* liability is actually what we optimally want.

|  |  |
| --- | --- |
| Like | 9 people like this. Be the first of your friends. |

[](http://www.facebook.com/share.php?u=http://blog.practicalethics.ox.ac.uk/2013/04/strict-ish-liability-an-experiment-in-the-law-as-algorithm/&t=Strict-ish%20liability?%20An%20experiment%20in%20the%20law%20as%20algorithm)[](http://twitter.com/home?status=Strict-ish%20liability?%20An%20experiment%20in%20the%20law%20as%20algorithm%20-%20http://bit.ly/YhaaZT)[](http://reddit.com/submit?url=http://blog.practicalethics.ox.ac.uk/2013/04/strict-ish-liability-an-experiment-in-the-law-as-algorithm/&title=Strict-ish%20liability?%20An%20experiment%20in%20the%20law%20as%20algorithm)

Posted in [Crime](http://blog.practicalethics.ox.ac.uk/crime-2/), [Current Affairs](http://blog.practicalethics.ox.ac.uk/current-affairs/), [Reflections](http://blog.practicalethics.ox.ac.uk/reflections/), [Regulation](http://blog.practicalethics.ox.ac.uk/regulation/), [Web/Tech](http://blog.practicalethics.ox.ac.uk/webtech/) | Tagged [strict liability; law](http://blog.practicalethics.ox.ac.uk/tag/strict-liability-law/)

**6 Responses to Strict-ish liability? An experiment in the law as algorithm**

*Professor Sandel* says: [April 12, 2013 at 10:09 am](http://blog.practicalethics.ox.ac.uk/2013/04/strict-ish-liability-an-experiment-in-the-law-as-algorithm/comment-page-1/#comment-57888) This is all very well, but your post neglects two things:  1. Law is an embedded constituent of our own communities and therefore the very notion of algorithm is not subject to the kind of platono-kantian abstractology to which you subject it.  (B) You haven’t solved free will and until you do that your whole argument is moot!

*Ian* says: [April 12, 2013 at 3:28 pm](http://blog.practicalethics.ox.ac.uk/2013/04/strict-ish-liability-an-experiment-in-the-law-as-algorithm/comment-page-1/#comment-57957) Is language an algorith?

*Professor Sandel* says: [April 12, 2013 at 6:00 pm](http://blog.practicalethics.ox.ac.uk/2013/04/strict-ish-liability-an-experiment-in-the-law-as-algorithm/comment-page-1/#comment-57973) In denying that language is an algorithm we deny our own community constituted non-abstract humanity. This constitutes an anti-proto-kantian approach to philosophy which feeds upon stripto-enlightenment canards.

*Anthony Drinkwater* says: [April 12, 2013 at 6:39 pm](http://blog.practicalethics.ox.ac.uk/2013/04/strict-ish-liability-an-experiment-in-the-law-as-algorithm/comment-page-1/#comment-57979) I guess that you studied with Jean-Baptiste Botul, professor Sandel. It must have been inspiring to be in the presence of the author of such classics as “La Vie Sexuelle d’Emanuel Kant”. Cf my unpublished monograph “Some thoughts on Botulism and contemporary algorithms, with particular reference to “soft” approaches to hegemony”.

*Ian* says: [April 13, 2013 at 11:07 am](http://blog.practicalethics.ox.ac.uk/2013/04/strict-ish-liability-an-experiment-in-the-law-as-algorithm/comment-page-1/#comment-58196) Language; what definition is being used and does a language alone always enable a complete reflection of all ongoing community constituted humanity. You appear to conflate the answer to that within free will so it seems to require clarification and this is not a Wittgensteinian response to Kant.  Although it may ready people for those issues in life, illustrating the tone of the DVLA advertising campaign when they amalgamated data to automate the enforcement process in various responses regarding free will still does not fully illuminate the potential breadth of complexity contained within free will in any given community.

*Martha Salomo* says: [May 12, 2013 at 7:54 am](http://blog.practicalethics.ox.ac.uk/2013/04/strict-ish-liability-an-experiment-in-the-law-as-algorithm/comment-page-1/#comment-62445) This was an especially interesting, given that last year at We Robot in Miami, some authors presented a paper on the problems of removing humans from the loop and turning law enforcement over to a computer (Confronting Automated Law Enforcement). This year, the authors took the question a step further. In an experiment, they look at what happens when you convert laws into algorithms.

[NEW ALGORITHM, OLD PROBLEMS](http://uculr.com/blog/2013/4/11/new-algorithm-old-problems)

*April 11, 2013*

**By Wakanene Kamau**

The issue of music piracy is certainly not new, but, after Google [launched an improved search algorithm](http://insidesearch.blogspot.com/2012/08/an-update-to-our-search-algorithms.html) in August of last year, it was believed to be on it’s way out. In the new algorithm, search rankings began to take into consideration the amount of copyright removal notices a given site had received in its tabulation. In doing so, Google intended users to find legal sources of content more easily. Ideally, by making legal avenues quicker and easier to find[, users would able to get content with less fuss and piracy would decrease](http://venturebeat.com/2012/08/10/watch-out-pirates-googles-new-search-changes-target-copyright-offenders/) naturally as a side effect. Futhermore, copyright holders would then have less to complain about. It was a win-win situation. However, we all know that all that glitters is not gold.

Fast forward six months and, in [a statement released by the Recording Industry Association of America](http://riaa.com/blog.php?content_selector=riaa-news-blog&content_selector=riaa-news-blog&blog_selector=Googles-Move-&news_month_filter=2&news_year_filter=2013) (RIAA), when compared with the old algorithm *“we have found no evidence that Google’s policy has had a demonstrable impact on demoting sites with large amounts of piracy. These sites consistently appear at the top of Google’s search results for popular songs or artists.”*

As a result of the Digital Millennium Copyright Act (DMCA), search engines are required to remove links upon request of the right holders who tell the engine the links lead to infringing content. If the engine does not remove the link, it could be liable for infringement.

The data on the matter is certainly convincing, as the statement from the RIAA [went on to say,](http://76.74.24.142/3CF95E01-3836-E6CD-A470-1C2B89DE9723.pdf)*“[w]hatever Google has done, it doesn’t appear to be working.”* The following graph produced by the RIAA shows the average percent of time a site for which Google had gotten over 100,000 DMCA removal requests was found among the top 10 search results for 50 pop songs.

Source: RIAA

Bearing in mind that Google itself is a giant media company and its own content hub, Google Play, which sells everything from music to TV shows to magazines, competes alongside authorized download sites such as iTunes and Amazon against pirate sites, why would Google rank highly the sites detrimental to its own business model?

The data and the conclusions that are deduced must be taken with a skeptical eye. The above graph was constructed with the methodology: “*For this analysis we performed searches for [artist] [track] mp3 and [artist] [track] download over a period of several weeks starting December 3, 2012”* Google’s business algorithm is built on a understanding what people are looking for when they search and then bringing that to them. When someone searches [artist] [track] mp3 and [artist] [track] download, they are clearly looking to download for free, not to buy; thusly the results show pirate sites. The problem with the RIAA’s line of thinking is that they suppose that the reason people pirate music is because they lack alternative methods of acquiring music. Even presented with options to buy the music, those who intend to pirate will pirate. If the methodology looked at, for example, [artist] [track] buy then perhaps the intended demotion of sites that receive DMCA takedown results could be more easily observed.

Perhaps what is most clear from all of this is that the issue of piracy on the web is still far from being solved. The interests of the search engine, the media industry and the will of the people must be weighted in a manner that is reasonable to all sides.  Google’s algorithm change cannot be expected to be a one-size-fits-all fix to the piracy issue. The RIAA, while good intentioned, may have their expectations set too high. At the very least, this recent fight has brought the piracy issue to attention of the public once again. Hopefully from this dialogue, a more lasting peace can be made.

*Wakanene Kamau is a First Year in the College majoring in Biological Chemistry.*

|  |  |  |
| --- | --- | --- |
|  |  |  |

YLJO Essay

|  |  |  |  |
| --- | --- | --- | --- |
| [**Prometheus Rebound: Diagnostics, Nature, and Mathematical Algorithms**](http://www.yalelawjournal.org/the-yale-law-journal-pocket-part/intellectual-property/prometheus-rebound:-diagnostics,-nature,-and-mathematical-algorithms/) |  |  |  |

|  |
| --- |
| *Rebecca S. Eisenberg,* Monday, 01 April 2013 [[**View as PDF**]](http://www.yalelawjournal.org/images/pdfs/1145.pdf) |
| \*\*This Essay is part of a *Yale Law Journal Online* series called "**Summary Judgment**," featuring short commentaries on recent Supreme Court cases.\*\*  *The Supreme Court’s decision last Term in* Mayo v. Prometheus *left considerable uncertainty as to the boundaries of patentable subject matter for molecular diagnostic inventions.* *First, the Court took an expansive approach to what counts as an unpatentable natural law by applying that term to the relationship set forth in the challenged patent between a patient’s levels of a drug metabolite and the indication of a need to adjust the patient’s drug dosage. And second, in evaluating whether the patent claims add enough to this unpatentable natural law to be patent eligible, the Court did not consult precedents concerning the patentability of claims involving natural laws and natural products. Instead, it turned to two seemingly inconsistent decisions that reached opposing conclusions concerning the patent eligibility of industrial methods that used mathematical algorithms. The Court’s analysis invites challenges to many issued patents, while offering little guidance for resolving them. This Term, in the*Association for Molecular Pathology *case,* *the Court has another opportunity to clarify the meaning of its exclusion of natural phenomena from patent eligibility.*  The promise of personalized medicine cannot be delivered without new precision diagnostic tools for tailoring treatment interventions to the needs of individual patients. The recent decision by the Supreme Court in *Mayo Collaborative Services v. Prometheus Laboratories, Inc.*,[**1**](http://www.yalelawjournal.org/the-yale-law-journal-pocket-part/intellectual-property/prometheus-rebound:-diagnostics,-nature,-and-mathematical-algorithms/#) raises doubts about the eligibility of these diagnostic tools for patent protection and calls into question the validity of many previously issued patents. In the *Association for Molecular Pathology* case,[**2**](http://www.yalelawjournal.org/the-yale-law-journal-pocket-part/intellectual-property/prometheus-rebound:-diagnostics,-nature,-and-mathematical-algorithms/#) the Court has an opportunity to clarify the applicable rules by identifying the kinds of diagnostic inventions that fall within patentable subject matter as well as those that are excluded from patent protection. Otherwise, the federal courts are likely to face a stream of appeals on patentable subject matter in the years ahead.  The patent in *Mayo v. Prometheus* claimed a method of optimizing the dosage of thiopurine drugs for treatment of immune-mediated gastrointestinal disorders (such as irritable bowel syndrome). The method involved comparing a patient’s levels of two drug metabolites with reference values specified in the patent.[**3**](http://www.yalelawjournal.org/the-yale-law-journal-pocket-part/intellectual-property/prometheus-rebound:-diagnostics,-nature,-and-mathematical-algorithms/#) Observed levels below certain values would indicate a need to raise the dosage, while observed levels above different values would indicate a need to lower it.[**4**](http://www.yalelawjournal.org/the-yale-law-journal-pocket-part/intellectual-property/prometheus-rebound:-diagnostics,-nature,-and-mathematical-algorithms/#) The Supreme Court thought that the patent impermissibly claimed laws of nature, “namely, relationships between concentrations of certain metabolites in the blood and the likelihood that a dosage of a thiopurine drug will prove ineffective or cause harm.”[**5**](http://www.yalelawjournal.org/the-yale-law-journal-pocket-part/intellectual-property/prometheus-rebound:-diagnostics,-nature,-and-mathematical-algorithms/#) Other steps in the process recited in the claim (such as administering a thiopurine drug to a patient or determining the patient’s levels of drug metabolites) consisted of “well-understood, routine, conventional activity previously engaged in by scientists in the field.”[**6**](http://www.yalelawjournal.org/the-yale-law-journal-pocket-part/intellectual-property/prometheus-rebound:-diagnostics,-nature,-and-mathematical-algorithms/#) In the Court’s view, those process steps did not add enough to the natural laws to classify the claim as a patent-eligible *application* of the natural laws rather than an impermissible “patent upon the natural law itself.”[**7**](http://www.yalelawjournal.org/the-yale-law-journal-pocket-part/intellectual-property/prometheus-rebound:-diagnostics,-nature,-and-mathematical-algorithms/#)  Two key moves in the Court’s analysis cast a shadow of uncertainty over the validity of patents on diagnostic inventions*.* First, the Court took an expansive approach to what counts as a “law of nature” by attaching that label to the relationship set forth in the patent between a patient’s drug metabolite levels and the indication of a need to adjust the patient’s drug dosage. Because “laws of nature” are not patentable, the Court asked “whether the claims do significantly more than simply describe these natural relations.”[**8**](http://www.yalelawjournal.org/the-yale-law-journal-pocket-part/intellectual-property/prometheus-rebound:-diagnostics,-nature,-and-mathematical-algorithms/#) Second, to answer this question, the Court did not consider prior cases on patents that involved natural laws and natural products, but turned instead to two prior cases involving computer-implemented industrial processes—*Parker v. Flook*[**9**](http://www.yalelawjournal.org/the-yale-law-journal-pocket-part/intellectual-property/prometheus-rebound:-diagnostics,-nature,-and-mathematical-algorithms/#) and *Diamond v. Diehr*[**10**](http://www.yalelawjournal.org/the-yale-law-journal-pocket-part/intellectual-property/prometheus-rebound:-diagnostics,-nature,-and-mathematical-algorithms/#)—as the “the cases most directly on point.”[**11**](http://www.yalelawjournal.org/the-yale-law-journal-pocket-part/intellectual-property/prometheus-rebound:-diagnostics,-nature,-and-mathematical-algorithms/#) Those two cases reach opposing conclusions on similar facts and are difficult to reconcile, as Justice Stevens observed in dissent in *Diehr.*[**12**](http://www.yalelawjournal.org/the-yale-law-journal-pocket-part/intellectual-property/prometheus-rebound:-diagnostics,-nature,-and-mathematical-algorithms/#)Taken together, the Court’s twin moves invite patent challenges while offering only vague guidance for resolving them.  The Court’s characterization of the relationship between the observed metabolite levels and the need to adjust drug dosage as a “natural law” is puzzling. The Court acknowledged that it takes human action to administer a thiopurine drug to a patient and thereby trigger a “manifestation of this relation in a particular person.”[**13**](http://www.yalelawjournal.org/the-yale-law-journal-pocket-part/intellectual-property/prometheus-rebound:-diagnostics,-nature,-and-mathematical-algorithms/#) But the Court nonetheless asserted that “the relation itself exists in principle apart from any human action” because thiopurine compounds are metabolized by the body according to “entirely natural processes.”[**14**](http://www.yalelawjournal.org/the-yale-law-journal-pocket-part/intellectual-property/prometheus-rebound:-diagnostics,-nature,-and-mathematical-algorithms/#)  Perhaps what the Court meant was that a patient’s reaction to a drug is controlled by biological processes that follow certain natural laws. The same is true of any method of using a drug in medical treatment, and yet many decisions have upheld the patent eligibility of such methods. Indeed, as the Court noted, “a typical patent on . . . a new way of using an existing drug” is patentable on the ground that it is limited to “particular applications of [natural] laws” rather than an impermissible patent on the natural law itself.[**15**](http://www.yalelawjournal.org/the-yale-law-journal-pocket-part/intellectual-property/prometheus-rebound:-diagnostics,-nature,-and-mathematical-algorithms/#) The Court did not explain why a method of treatment that makes use of a patient’s biological response to a drug is a patent-eligible application of a natural law, while a diagnostic method that makes use of this same biological response is not.  The Court may see the diagnostic method as involving too little value added by humans beyond the observation of a natural biological phenomenon to qualify as a human invention.However, even if natural laws determine a patient’s response to drug therapy, nature does not determine when those consequences indicate a need to raise or lower the drug dosage. Nature does not specify when the miseries of irritable bowel syndrome outweigh the risks of myelosuppression and liver toxicity from the use of thiopurine drugs. At most, nature supplies the raw data, while human judgment is necessary to interpret the data and to guide medical intervention. The technological contribution of this particular invention is to quantify and systematize that judgment to improve treatment. Other methods are possible, and they might do a better or worse job of optimizing treatment. Indeed, after using the Prometheus Laboratories invention under license for a time, the defendant, Mayo Collaborative Services, decided to change the metabolite values that it thought called for adjusting the dosage (although the change was not enough to avoid infringement liability if the claims had been upheld). Those different views about what drug metabolite levels are problematic show that the levels recited in the claims represent a human technological choice that goes beyond mere recital of a natural law.  Many prior cases have struggled with the distinction between patent-eligible human inventions and patent-ineligible natural products and phenomena.[**16**](http://www.yalelawjournal.org/the-yale-law-journal-pocket-part/intellectual-property/prometheus-rebound:-diagnostics,-nature,-and-mathematical-algorithms/#) Rather than turning to those cases for guidance, the Court made its second puzzling move: it turned to two decisions from 1978 and 1981 concerning the patentability of methods that recite “mathematical algorithms”to resolve the patent eligibility of claims that recite “laws of nature.”  *Parker v. Flook* held that a method using a mathematical algorithm to update alarm limits for process variables in a catalytic conversion process was *not* patentable subject matter.[**17**](http://www.yalelawjournal.org/the-yale-law-journal-pocket-part/intellectual-property/prometheus-rebound:-diagnostics,-nature,-and-mathematical-algorithms/#) On the other hand, *Diamond v. Diehr* held that a method of operating a rubber molding press using a mathematical algorithm to repeatedly recalculate the cure time *was* patentable subject matter.[**18**](http://www.yalelawjournal.org/the-yale-law-journal-pocket-part/intellectual-property/prometheus-rebound:-diagnostics,-nature,-and-mathematical-algorithms/#)To the *Diehr* majority, the relevant distinction was that Flook had sought to patent a method of computing a number, whereas Diehr sought to patent a method of curing synthetic rubber.[**19**](http://www.yalelawjournal.org/the-yale-law-journal-pocket-part/intellectual-property/prometheus-rebound:-diagnostics,-nature,-and-mathematical-algorithms/#) But upon closer examination the primary difference seemed to be a matter of claim-drafting, as Justice Stevens pointedly noted in his dissent in *Diehr*.[**20**](http://www.yalelawjournal.org/the-yale-law-journal-pocket-part/intellectual-property/prometheus-rebound:-diagnostics,-nature,-and-mathematical-algorithms/#)  The Court in *Mayo v. Prometheus* identified a different distinction in its own paraphrase of these inconsistent holdings, explaining that the patent-ineligible claim in *Flook* merely recited a mathematical formula—“the equivalent of a natural law”[**21**](http://www.yalelawjournal.org/the-yale-law-journal-pocket-part/intellectual-property/prometheus-rebound:-diagnostics,-nature,-and-mathematical-algorithms/#)—followed by a bare instruction to “apply it.”[**22**](http://www.yalelawjournal.org/the-yale-law-journal-pocket-part/intellectual-property/prometheus-rebound:-diagnostics,-nature,-and-mathematical-algorithms/#) By contrast, the patent-eligible claim in *Diehr* recited additional steps that “apparently added . . . something that in terms of patent law’s objectives had significance” and “transformed the process into an inventive application of the formula.”[**23**](http://www.yalelawjournal.org/the-yale-law-journal-pocket-part/intellectual-property/prometheus-rebound:-diagnostics,-nature,-and-mathematical-algorithms/#) The claims in *Mayo v. Prometheus* struck the Court as tantamount to a recital of a natural law followed by a bare instruction to “apply it” because, once the Court sets aside as “natural laws” the metabolite levels specified in the claims, the other steps “add nothing specific to the laws of nature other than what is well-understood, routine, conventional activity, previously engaged in by those in the field.”[**24**](http://www.yalelawjournal.org/the-yale-law-journal-pocket-part/intellectual-property/prometheus-rebound:-diagnostics,-nature,-and-mathematical-algorithms/#) But that observation does not explain the distinction between *Diehr* and *Flook*:both of those cases involved methods that applied new mathematical algorithms to familiar process steps.[**25**](http://www.yalelawjournal.org/the-yale-law-journal-pocket-part/intellectual-property/prometheus-rebound:-diagnostics,-nature,-and-mathematical-algorithms/#) Moreover, the Court in *Diehr* explicitly rejected the approach of excluding conventional process steps from consideration in determining whether a claim recites patentable subject matter.[**26**](http://www.yalelawjournal.org/the-yale-law-journal-pocket-part/intellectual-property/prometheus-rebound:-diagnostics,-nature,-and-mathematical-algorithms/#) The Court in *Mayo v. Prometheus* did not make a serious effort to compare the additional steps in the claims before it to those in *Flook* and *Diehr*. Instead, it left future courts to puzzle over how far *Mayo v. Prometheus* has reanimated these ghosts from the past and how to resolve their inconsistencies.  In its day, *Diehr* marked a turning point between an earlier era of parsimonious patent protection for computer-implemented inventions and a new era of expanded patent eligibility.[**27**](http://www.yalelawjournal.org/the-yale-law-journal-pocket-part/intellectual-property/prometheus-rebound:-diagnostics,-nature,-and-mathematical-algorithms/#) Subsequent case law offers little guidance in resolving the inconsistency between the two cases because, at the time, *Diehr* functioned more as a claim-drafting guide to avoid the outcome in *Flook* than as a counterexample to explain the limitations of *Flook*.  By returning to the elusive distinction between *Flook* and *Diehr* after more than three decades and assigning to those cases the new task of discerning the boundaries of patent eligibility for all claims that recite “laws of nature,” the Court in *Mayo v. Prometheus* seemed to be on a new mission. Rather than restricting the reach of judicial limitations on patentable subject matter by affirming the claims before it, the Court appeared to be narrowing the boundaries of patentable subject matter in a field that has long taken for granted the availability of patent protection for its innovations. By broadly defining “laws of nature” to include human interpretation of biological responses to medical interventions, the Court seemed to call into question the validity of many previously allowed claims, inviting more litigation contesting patentable subject matter and drawing courts into the murky waters of *Diehr* and *Flook.*  The Court of Appeals for the Federal Circuit declined to step into those murky waters when it considered the patent eligibility of advances in medical diagnostics in *Association for Molecular Pathology v. U.S. Patent & Trademark Office*.[**28**](http://www.yalelawjournal.org/the-yale-law-journal-pocket-part/intellectual-property/prometheus-rebound:-diagnostics,-nature,-and-mathematical-algorithms/#) That case challenged the patent eligibility of claims to DNA sequences for the BRCA1 and BRCA2 genes associated with breast cancer susceptibility as well as claims to diagnostic methods and drug-screening methods that make use of these sequences. The Federal Circuit has ruled on the case twice, once[**29**](http://www.yalelawjournal.org/the-yale-law-journal-pocket-part/intellectual-property/prometheus-rebound:-diagnostics,-nature,-and-mathematical-algorithms/#) before the Supreme Court’s decision in *Mayo v. Prometheus* and again[**30**](http://www.yalelawjournal.org/the-yale-law-journal-pocket-part/intellectual-property/prometheus-rebound:-diagnostics,-nature,-and-mathematical-algorithms/#) on remand for reconsideration in light of that decision. Each member of the panel wrote separately each time, presenting a range of competing views to guide the Court in its consideration of the issues presented.  The Federal Circuit panel was divided on the patent eligibility of the claims to the BRCA1 and BRCA2 DNA molecules, the most prominent issue in the case and the only issue on which the Court has granted certiorari.[**31**](http://www.yalelawjournal.org/the-yale-law-journal-pocket-part/intellectual-property/prometheus-rebound:-diagnostics,-nature,-and-mathematical-algorithms/#) Yet each panel member joined Judge Lourie’s analysis of the method claims, unanimously holding that (1) claims to diagnostic methods of “comparing” or “analyzing” DNA sequences from a tissue sample with reference sequences were not patent eligible because they are only “abstract mental steps” and that (2) claims to drug-screening methods that compare the growth rate of cells transformed with an altered BRCA1 gene in the presence or absence of a potential cancer therapy were patent-eligible chemical processes.[**32**](http://www.yalelawjournal.org/the-yale-law-journal-pocket-part/intellectual-property/prometheus-rebound:-diagnostics,-nature,-and-mathematical-algorithms/#) One might have expected the recent decision of the Supreme Court in *Mayo v. Prometheus* to play a significant role in the Federal Circuit’s analysis of the method claims—especially given that the Court explicitly remanded for reconsideration in light of that decision—yet its teachings had little apparent impact on the analysis of the claims. Judge Lourie acknowledged that his analysis followed the Court’s *holding* in *Mayo v. Prometheus*,but not its *reasoning*.[**33**](http://www.yalelawjournal.org/the-yale-law-journal-pocket-part/intellectual-property/prometheus-rebound:-diagnostics,-nature,-and-mathematical-algorithms/#) He did not seek to resolve whether the diagnostic method claims covered “laws of nature,” instead holding the claims invalid because “comparing” or “analyzing” two gene sequences is an abstract mental process.[**34**](http://www.yalelawjournal.org/the-yale-law-journal-pocket-part/intellectual-property/prometheus-rebound:-diagnostics,-nature,-and-mathematical-algorithms/#) In upholding the patent eligibility of the drug-screening method claims, Judge Lourie focused on the fact that the method used human-modified bacteria that had been transformed with an altered BRCA1 gene. He considered it irrelevant to the patentable-subject-matter analysis that the other process steps of comparing growth rates in the cells were conventional.[**35**](http://www.yalelawjournal.org/the-yale-law-journal-pocket-part/intellectual-property/prometheus-rebound:-diagnostics,-nature,-and-mathematical-algorithms/#) Responding to the Court’s emphasis in *Mayo v. Prometheus* on the insufficiency of conventional steps to establish patentable subject matter, Judge Lourie noted that most chemical processes involve the use of known process steps and reactions and that this should not defeat patent eligibility where the process makes use of novel materials that are not naturally occurring.[**36**](http://www.yalelawjournal.org/the-yale-law-journal-pocket-part/intellectual-property/prometheus-rebound:-diagnostics,-nature,-and-mathematical-algorithms/#) The Federal Circuit’s analysis of those claims was straightforward, although it did not even purport to follow the reasoning of *Mayo v. Prometheus.* On the other hand, although the Federal Circuit’s reasoning was different, its ultimate decision on the patent eligibility of the method claims was broadly consistent with the Court’s reinforcement of traditional exclusions from patentable subject matter. The Supreme Court declined to review the Federal Circuit’s decision on the method claims, granting certiorari solely on the question whether human genes are patentable.[**37**](http://www.yalelawjournal.org/the-yale-law-journal-pocket-part/intellectual-property/prometheus-rebound:-diagnostics,-nature,-and-mathematical-algorithms/#) Perhaps the Court’s acquiescence in the Federal Circuit’s alternative analytical approach indicates that it is willing to defer to that court’s expertise in patent matters so long as it seems to be vigorously policing the subject-matter boundaries of the patent system.  It remains to be seen how the Court will review the Federal Circuit panel’s split decision affirming the patent eligibility of claims to isolated DNA molecules. The three separate opinions of the panel members on this question found limited guidance in the Supreme Court’s decision in *Mayo v. Prometheus.*[**38**](http://www.yalelawjournal.org/the-yale-law-journal-pocket-part/intellectual-property/prometheus-rebound:-diagnostics,-nature,-and-mathematical-algorithms/#) Each opinion drew primarily upon prior cases on the exclusion for natural products and processes that the Court had largely ignored.[**39**](http://www.yalelawjournal.org/the-yale-law-journal-pocket-part/intellectual-property/prometheus-rebound:-diagnostics,-nature,-and-mathematical-algorithms/#) Those cases are hardly a model of consistency and clarity, yet they provide a more coherent baseline than the cases on the exclusion for mathematical algorithms. In reviewing the latest disposition of *Association for Molecular Pathology*, the Supreme Court could begin to restore predictability to the rules of patentable subject matter by setting aside *Parker v. Flook* and *Diamond v. Diehr* and turning instead to these more pertinent authorities on the patentability of natural products and processes.  Perhaps the most important contribution of *Diamond v. Diehr* to the jurisprudence of patentable subject matter was that for the first time the Court identified claims to computer-implemented inventions that it considered patent eligible, thereby providing patent applicants with a model for patenting computer-implemented inventions properly. The Court now has an opportunity to make a similar contribution to the understanding of patentable subject matter for diagnostic inventions in *Association for Molecular Pathology*. The plaintiffs in that case have challenged the validity of multiple claims to DNA inventions that reflect different approaches to defining what constitutes a patentable invention. The three opinions from the panel offer distinct approaches to the patent-eligibility issues raised by these different claims, providing the Court with an unusually rich record for clarifying the distinction between unpatentable natural phenomena and patentable human inventions. If the Court can identify claims that it considers patent eligible, it will bring greater clarity to the issue than if it only identifies claims that lie outside the boundaries of patentable subject matter. In the absence of such clarity, the Court can expect many more appeals as the Patent and Trademark Office and the courts try to determine what the Supreme Court will deem patentable.    *Rebecca S. Eisenberg is the Robert and Barbara Luciano Professor of Law at the University of Michigan Law School.*    Preferred citation: Rebecca S. Eisenberg, Prometheus *Rebound: Diagnostics, Nature, and Mathematical Algorithms*, 122 YALE L.J. ONLINE 341 (2013), http://yalelawjournal.org/2013/04/01/eisenberg.html. |

**Comparing Human and Algorithm-Based Curation of Legal Data**

Rate This

 Susan Nevelow Mart and Jeffrey Lutig posted a paper to [**SSRN**](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2112574) titled “”The Case for Curation: The Relevance of Digest and Citator Results in Westlaw and Lexis” that looks at how Westlaw, which relies primarily on human curators, and Lexis, which relies primarily on algorithm-based curators compare.

*Humans and machines are both involved in the creation of legal research resources. For legal information retrieval systems, the human-curated finding aid is being overtaken by the computer algorithm. But human-curated finding aids still exist. One of them is the West Key Number system. The Key Number system’s headnote classification of case law, started back in the nineteenth century, was and is the creation of humans. The retrospective headnote classification of the cases in Lexis’s case databases, started in 1999, was created primarily although not exclusively with computer algorithms. So how do these two very different systems deal with a similar headnote from the same case, when they link the headnote to the digesting and citator functions in their respective databases? This paper continues an investigation into this question, looking at the relevance of results from digest and citator search run on matching headnotes in ninety important federal and state cases, to see how each performs. For digests, where the results are curated – where a human has made a judgment about the meaning of a case and placed it in a classification system – humans still have an advantage. For citators, where algorithm is battling algorithm to find relevant results, it is a matter of the better algorithm winning. But no one algorithm is doing a very good job of finding all the relevant results; the overlap between the two citator systems is not that large. The lesson for researchers: know how your legal research system was created, what involvement, if any, humans had in the curation of the system, and what a researcher can and cannot expect from the system you are using.*

I cannot cite from the article without the author’s permission, so I won’t.

**Thursday, September 20, 2012**

**The algorithm decided not to hire you: is that legal?**

I spend a lot of time thinking about privacy and algorithms.

The Wall Street Journal carried an interesting story "[Meet the New Boss:  Big Data](http://online.wsj.com/article/SB10000872396390443890304578006252019616768.html?mod=WSJ_hpp_MIDDLENexttoWhatsNewsThird)", about how algorithms are now being widely used to make human resources decisions, like hiring and promotion.  The article pointed out that such algorithms could run into legal problems, if they intentionally or unintentionally filter out protected categories of employees, like older employees, under US anti-discrimination laws.  But the article didn't discuss a more fundamental legal issue, at least in Europe.

In Europe, "automated individual decisions" are a violation of EU privacy laws.  Article 15 of the EU Privacy Directive guarantees:  "...the right to every person not to be subject to a decision which produces legal effects concerning him or significantly affects him and which is based solely on automated processing of data intended to evaluate certain personal aspects relating to him, such as his performance at work, creditworthiness, reliability, conduct, etc".

Well, that's about as clear as a law can get.  In our age of Big Data, we all know algorithms are being refined and used more and more widely to make decisions about hiring and promotion, and many other topics.  But when these decisions are made solely by algorithms, they are violating EU privacy laws.  Period.  The only way such algorithms can be used legally is to supplement them with certain other measures to safeguard the legitimate interests of the person being evaluated, e.g., by allowing him to put his point of view.

I'm a great believer that algorithms can help all of us (governments, businesses, individuals) make better decisions.  But when a computer program is making key decisions by itself about whom to hire or fire, or whether or not to extend credit to someone, it's fair to ask for additional safeguards.  The privacy laws in Europe require it.  I'm agnostic about whether algorithms are more or less fair than humans at making a lot of such decisions.  In any case, companies using such algorithms need to consider how to make them comply with European privacy laws.  When algorithms are used to supplement other evaluation tools, they should be legal.  When algorithms are used to make these decisions by themselves, there's a serious risk they would be considered illegal in Europe.  Use with care.

Posted by [Peter Fleischer](https://plus.google.com/103464834326367538629) at [1:53 PM](http://peterfleischer.blogspot.com/2012/09/the-algorithm-decided-not-to-hire-you.html) [](http://www.blogger.com/email-post.g?blogID=6974997875021040765&postID=1250365585264894546)

**2 comments:**

****

**Anonymous said...**

Peter, see my 2 cents on this here: http://www.concurringopinions.com/archives/2012/09/big-data-for-all.html

[September 22, 2012 at 9:14 AM](http://peterfleischer.blogspot.com/2012/09/the-algorithm-decided-not-to-hire-you.html?showComment=1348301656610#c8998949886928799681)

**said...**

Well said. Noticed that under the proposed EU Data Privacy Directive, automated decision making is now under the name "profiling."

[September 27, 2012 at 3:29 PM](http://peterfleischer.blogspot.com/2012/09/the-algorithm-decided-not-to-hire-you.html?showComment=1348756181944#c6545733918661448558)

MANAGEMENT

September 20, 2012, 11:16 a.m. ET

Meet the New Boss: Big Data

*Companies Trade In Hunch-Based Hiring for Computer Modeling*

By

JOSEPH WALKER

At an increasing number of companies, the hiring boss for rank and file jobs is now an algorithm -- and the computers are considering factors that are very different than what applicants have come to expect. Joseph Walker has details on The News Hub. Photo: Bloomberg.

When looking for workers to staff its call centers, [Xerox](http://online.wsj.com/public/quotes/main.html?type=djn&symbol=XRX) Corp. [XRX -0.99%](http://online.wsj.com/public/quotes/main.html?type=djn&symbol=XRX?mod=inlineTicker" \t ") used to pay lots of attention to applicants who had done the job before. Then, a computer program told the printer and outsourcing company that experience doesn't matter.

The software said that what *does* matter in a good call-center worker—one who won't quit before the company recoups its $5,000 investment in training—is personality. Data show that creative types tend to stick around for the necessary six months. Inquisitive people often don't.

Enlarge Image



Mikey Burton

**More:**

[**Big Data Upends the Way Workers Are Paid**](http://online.wsj.com/article/SB10000872396390444433504577651741900453730.html)

[**Do New Job Tests Foster Bias?**](http://online.wsj.com/article/SB10000872396390443890304578006283936708970.html)

"Some of the assumptions we had weren't valid," said Connie Harvey, Xerox's chief operating officer of commercial services.

After a half-year trial that cut attrition by a fifth, Xerox now leaves all hiring for its 48,700 call-center jobs to software that asks applicants to choose between statements like: "I ask more questions than most people do" and "People tend to trust what I say."

For more and more companies, the hiring boss is an algorithm. The factors they consider are different than what applicants have come to expect. Jobs that were once filled on the basis of work history and interviews are left to personality tests and data analysis, as employers aim for more than just a hunch that a person will do the job well. Under pressure to cut costs and boost productivity, employers are trying to predict specific outcomes, such as whether a prospective hire will quit too soon, file disability claims or steal.

Personality tests have a long history in hiring. What's new is the scale. Powerful computers and more sophisticated software have made it possible to evaluate more candidates, amass more data and peer more deeply into applicants' personal lives and interests.

Some companies are screening for such variables as attitudes toward alcohol use or the distance an applicant lives from the job. The process could get companies into legal trouble if it ends up excluding minorities or the disabled. Even if it doesn't, it might come off as unfair, or even creepy.

"The public gets less comfortable when you're using extrinsic or personal factors," said Dennis Doverspike, a professor of industrial and organizational psychology at the University of Akron in Ohio.

The new hiring tools are part of a broader effort to gather and analyze employee data. Globally, spending on so-called talent-management software rose to $3.8 billion in 2011, up 15% from 2010, according to research firm Gartner.

Big tech companies are jockeying to serve the growing market. Last month,[International Business Machines](http://online.wsj.com/public/quotes/main.html?type=djn&symbol=IBM) Corp. [IBM -2.31%](http://online.wsj.com/public/quotes/main.html?type=djn&symbol=IBM?mod=inlineTicker" \t ") agreed to pay $1.3 billion for [Kenexa](http://online.wsj.com/public/quotes/main.html?type=djn&symbol=KNXA) Corp., which uses data analysis to help companies recruit and retain workers. [Oracle](http://online.wsj.com/public/quotes/main.html?type=djn&symbol=ORCL) Corp. [ORCL +0.67%](http://online.wsj.com/public/quotes/main.html?type=djn&symbol=ORCL?mod=inlineTicker" \t ") acquired job-applicant tracking system company Taleo for $1.9 billion in February, and Germany's[SAP](http://online.wsj.com/public/quotes/main.html?type=djn&symbol=SAP.XE) AG [SAP.XE +0.74%](http://online.wsj.com/public/quotes/main.html?type=djn&symbol=SAP.XE?mod=inlineTicker" \t ")bought SuccessFactors, which specializes in performance tracking, recruiting and compensation, for $3.4 billion in December.

Xerox is being advised by Evolv Inc., a San Francisco start-up that helps companies hire and manage hourly workers. By putting applicants through a battery of tests and then tracking their job performance, Evolv has developed a model for the ideal call-center worker. The data say that person lives near the job, has reliable transportation and uses one or more social networks, but not more than four. He or she tends not to be overly inquisitive or empathetic, but is creative.

Applicants for the job take a 30-minute test that screens them for personality traits and puts them through scenarios they might encounter on the job. Then the program spits out a score: red for low potential, yellow for medium potential or green for high potential. Xerox accepts some yellows if it thinks it can train them, but mostly hires greens.

Though hiring is a crucial business function, conventional methods are remarkably short on rigor, experts say. Depending on who decides, what gets candidates hired can vary wildly—from academic achievement to work experience to appearance. Managers who go with their gut might get it right sometimes, but their hunches generally have little value in predicting how someone will perform on the job. Companies peddling a statistical approach to hiring say they can improve results by reducing the influence of a manager's biases.

Heather Ainsworth for The Wall Street Journal

Xerox leaves hiring for its 48,700 call-center jobs to software. Above, Xerox service agent Lance LaRosa in Rochester, N.Y.

Laszlo Bock, a senior vice president at[Google](http://online.wsj.com/public/quotes/main.html?type=djn&symbol=GOOG) Inc. [GOOG -0.93%](http://online.wsj.com/public/quotes/main.html?type=djn&symbol=GOOG?mod=inlineTicker" \t ") and an Evolv director, said software will supplement, if not supplant, many of the personnel decisions long made by instinct and intuition. "The initial thing companies like Evolv are looking at is people as they get hired, but over the years this can help companies pick who to advance, who to promote," he said. "Even at the best companies there's still a lot of guessing."

It isn't just big companies that are turning to software for hiring help. Richfield Management LLC, a Flint, Mich., waste-disposal firm that employs 200 garbage collectors, was looking for ways to screen out applicants who were likely to get hurt and abuse workers' compensation.

About a year and a half ago, Richfield turned to an online test developed by a small firm called Exemplar Research Group. It asks applicants to pick between statements like "When I'm working for a company I take pride in making it as profitable as possible" and "I'm only concerned with how well I can do financially in my job," then rate how strongly they agree or disagree.

The goal is to gauge an applicant's emotional stability, work ethic and attitude toward drug and alcohol. Those who score poorly are considered high disability risks. Richfield said its workers' comp claims have fallen 68% since it has used the test, and it now requires managers to use it to eliminate unsuitable applicants."If the person scores low on the test, we don't make an exception for that person," said General Manager Fred Vezzetti. "

Data-based hiring can expose companies to legal risk. Practices that even unintentionally filter out older or minority applicants can be illegal under federal equal opportunity laws. If a hiring practice is challenged in court as discriminatory, a company must show the criteria it is using are proven to predict success in the job.

Matthew Camardella, a partner at employment law firm Jackson Lewis LLP, specializes in determining whether companies are in compliance with equal opportunity laws. He said an increasing number of them are asking him to evaluate new software tools.

Bigger data sets can raise the risks of violating the law by increasing the number of statistical relationships that could unwittingly screen out protected groups, he said. But they also could make it easier for companies to prove that the factors being measured are relevant to the job being filled.

Evolv is cautious about exploiting some of the relationships it turns up for fear of violating equal opportunity laws. While it has found employees who live farther from call-center jobs are more likely to quit, it doesn't use that information in its scoring in the U.S. because it could be linked to race.

But Kenexa, which tested 30 million applicants last year for thousands of clients, has found that a lengthy commute raises the risk of attrition in call-center and fast-food jobs. It asks applicants for call-center and fast-food jobs to describe their commute by picking options ranging from "less than 10 minutes" to "more than 45 minutes." The longer the commute, the lower their recommendation score for these jobs, says Jeff Weekley, who oversees the assessments.

Applicants also can be asked how long they have been at their current address and how many times they have moved. People who move more frequently "have a higher likelihood of leaving," Mr. Weekley said.

[Bon-Ton Stores](http://online.wsj.com/public/quotes/main.html?type=djn&symbol=BONT) Inc. [BONT -5.71%](http://online.wsj.com/public/quotes/main.html?type=djn&symbol=BONT?mod=inlineTicker" \t ") began using Kenexa in 2009 to screen applicants for hourly wage jobs in its stores and distribution centers. The retailer was looking in part to weed out candidates who might have a tendency to steal or miss work due to injury, said Denise Domian, its senior vice president for human resources. Kenexa's roughly 40-minute survey, she said, asks questions that the software scores for honesty. "People who are trying to fool the system are going to get tripped up," she said.

—Emily Glazer contributed to this article.

**Write to**Joseph Walker at [joseph.walker@dowjones.com](mailto:joseph.walker@dowjones.com)

**We Robot Conference: 2. Law as algorithm**

by [**Kate Darling**](http://robohub.org/author/katedarling/)   |   May 3, 2013

[**make a comment**](http://robohub.org/we-robot-conference-2-law-as-algorithm/#disqus_thread)

On April 8-9, Stanford Law School held the second annual robotics and law conference, [We Robot](http://blogs.law.stanford.edu/werobot/). This year’s event focused on near-term policy issues in robotics and featured panels and papers by scholars, practitioners, and engineers on topics like intellectual property, tort liability, legal ethics, and privacy. The full program is [here](http://blogs.law.stanford.edu/werobot/agenda/).

***This post is part of*** [***Robohub’s We Robot coverage***](http://robohub.org/tag/werobot-2013/)***.***

Woodrow Hartzog and Greg Conti presented their paper (co-authored with Lisa Shay and John Nelson of West Point) on implementing law in computer systems.

**Paper:** [**Do Robots Dream of Electric Laws? An Experiment in Law as Algorithm**](http://blogs.law.stanford.edu/werobot/files/2013/04/Shay-et-al_Lisa.pdf)**Author:** [**Greg Conti**](http://www.rumint.org/gregconti/)**,** [**Woodrow Hartzog**](http://cumberland.samford.edu/faculty/woodrow-n-hartzog)**Moderator:** [**Harry Surden**](http://lawweb.colorado.edu/profiles/profile.jsp?id=316)

We Robot 2013 | Do Robots Dream of Electric Law...

0:00 / 56:46

This was an especially interesting presentation, given that last year at [We Robot in Miami](http://robots.law.miami.edu/), the same authors presented a paper on the problems of removing humans from the loop and turning law enforcement over to a computer ([Confronting Automated Law Enforcement](http://robots.law.miami.edu/wp-content/uploads/2012/01/Shay-EtAl-ConfrontingAutomatedLawEnf.pdf)). This year, the authors took the question a step further. In an experiment, they look at what happens when you convert laws into algorithms. It turns out that even simple rules (like speeding limits) require unexpected consideration.

In the experiment, 52 programers were asked to automate the enforcement of traffic speed limits. They were given a set of driving data and each wrote a program that measured the number of speed limit violations and issued traffic tickets accordingly. Despite having exact data for both vehicle speed and speed limits, the number of issued tickets varied among the programs. The authors attribute the variance to the fact that the programers were faced with having to make assumptions and legal interpretations, for example whether to code according to the letter of the law (100% enforcement of every violation) or the intent of the law (tolerating minor infractions). The study indicates that there can be unanticipated degrees of freedom in the design of enforcement algorithms, even when dealing with seemingly straightforward legal rules.

As for the broader implications, the authors drew attention to the following problem areas: *Culpability* We may not be able to automate laws that include culpability, because this is too difficult to determine automatically. *Objectivity of the wrongful conduct* E.g. speed does not necessarily equal recklessness. *Identification* With automated enforcement we would have to make particularly sure that someone’s identity can be verified. *Accessibility* To what extent does this lead to a surveillance society? What kind of information is and should be accessible?

The results of the study also caution against outsourcing coding of the law to third-parties. They indicate that seemingly minor details can make a huge difference, creating more variance than assumed and requiring decisions by someone with actual rule-making authority. The authors also mentioned the need to consider potential second and third order effects, such as on traffic flow, etc.

Surden pointed out that there are already examples of laws being translated into computer code today that many people, including law-makers, are not fully aware of. For example, tax preparation software [Turbotax](http://en.wikipedia.org/wiki/TurboTax) inherently involves a lot of judgment and design decisions that have gone under the radar and are accepted more or less unquestioned by the Internal Revenue Service.

To my (personal, IP-nerdy) delight, an audience member mentioned the problems with content ID-ing and automatic copyright takedown notices on platforms like YouTube. Often these programs will remove completely legitimate content, causing troubles for fair-use activity. Surden said that one take-away for automated enforcement from this is that the appeal process should be just as easy as the take-down process.

The discussion also turned to the question of socially desirable enforcement. If we can embed laws in systems, we could potentially have a system of perfect monitoring and perfect enforcement. But are there costs to that perfection? Both authors argued that *imperfect* enforcement allows for a healthy amount of discretion and flexibility, and that bureaucracy left to its own devices could be dangerous. Even more problematic than perfect enforcement, said Hartzog, is perfect prevention. As we’ve seen with red light camera systems, citizens can become outraged by what they perceive as a lack of value judgment. A disconnect between law and social perception can go so far as to be counterproductive. Conti also postulated that the type of person who is willing to take on some risk to achieve a certain gain would be beat out by perfect law enforcement – behavior that we might want to otherwise maintain in our society.

The audience was also interested in the question whether our current system of human error, bias, and corruption in law enforcement is better or worse than a system that leaves no flexibility for human judgment. Can we build room for discretion and community norms into systems? What happens if a human bias becomes systematized? Interestingly, when the programmers were asked after the experiment whether they would want to drive on the roads with the programs they had built, their answer was uniformly “no.” (Although one programer said yes, conditioned on creating a backdoor exception for herself.)

In terms of policy recommendations and reducing the uncertainty of coding law, the authors suggested creating a committee or organization to set standards for automated enforcement systems, and also that the code of such systems be kept transparent and open to examination.

[***See all the We Robot coverage on Robohub* →**](http://robohub.org/tag/werobot-2013/)

[**Lectures**](http://robohub.org/category/learn-2/lectures/) - [**Views**](http://robohub.org/category/views/)

**Posted by** [**Kate Darling**](http://robohub.org/author/katedarling/)

[Liability for Algorithm Design & Big Data (Google Auto-complete)](http://www.husovec.eu/2013/06/liability-for-algorithm-design-big-data.html)

The sixth senate of the German Federal Supreme Court (BGH) in March decided a case against Google ([VI ZR 269/12](https://www.telemedicus.info/urteile/Internetrecht/Suchmaschinen/1401-BGH-Az-VI-ZR-26912-Autocomplete.html)) involving a question of its liability for Autocomplete Tool. The case was widely reported around various websites ([IPKat](http://ipkitten.blogspot.de/2013/05/autocomplete-can-google-turn-bad-news.html), [DW](http://www.dw.de/german-federal-court-raps-google-on-the-knuckles-over-autocomplete-function/a-16813363), [BBC](http://www.bbc.co.uk/news/technology-22529357), etc.) as *'Germany tells Google to tidy up auto-complete'*. Few weeks ago, the decision full-text became available and provoked interesting discussions among some German bloggers (e.g. [Adrian Schneider](http://www.telemedicus.info/article/2580-BGH-Die-Autocomplete-Entscheidung-im-Detail.html?pk_campaign=feed&utm_source=feedly), [Niko Härting](http://www.cr-online.de/blog/2013/05/24/begrundung-des-autocomplete-urteils-wer-soll-dies-denn-noch-verstehen/), [Thomas Stadler](http://www.internet-law.de/2013/05/urteil-des-bgh-zur-haftung-von-google-fur-die-suchwortvervollstandigung-im-volltext.html)). Kay Oberbeck from Google reacted that *"We fail to understand the federal court's ruling - that Google should be responsible for the search terms used by its users"*.

So what did the court say and what are the implications for the future of Big Data analysis and liability of companies for their algorithm designs?

The commentators generally share certain feeling of confusion over the decision. It might be because the decision was handed by the sixth senate, which uses certain crucial terms (e.g. *Störerhaftung*) in a slightly different context/meaning and/or less clearly than a first BGH senate, which is focusing on the intellectual property issues. Plus it is true that it is sometimes not entirely clear how the court arrived at the end result. I also have feeling that the court sometimes contradicts itself (e.g. § 26 last two sentences).

First of all, this case is *not* about secondary, but direct liability. Court is very explicit about the fact that auto-complete is to be treated as *own content of Google*. The court says:

DE:  Diese Beeinträchtigung des Persönlichkeitsrechts der Kläger ist der Beklagten auch unmittelbar zuzurechnen. Sie hat mit dem von ihr geschaffenen Computerprogramm das Nutzerverhalten ausgewertet und den Benutzern der Suchmaschine die entsprechenden Vorschläge unterbreitet. Die Verknüpfungen der Begriffe werden von der Suchmaschine der Beklagten und nicht von einem Dritten hergestellt. Sie werden von der Beklagten im Netz zum Abruf bereitgehalten und stammen deshalb unmittelbar von ihr.

EN [edited Google Translate]: 'This interference with the personal rights of the plaintiff is also directly attributable to the defendant. He has analyzed the user behavior using a self created computer program and presented the suggestions to the users of the search engine.  The linkage of the terms are made by the search engine of the defendant and not by a third party. They are provided for retrieval on network by the defendant and therefore come directly from him.'

And then further that:

'Zwar ist die Beklagte nicht bereits nach § 10 Telemediengesetz (künftig: TMG) von der Verantwortlichkeit für den Inhalt der von ihr betriebenen Website befreit. Das Berufungsgericht hat die Beklagte zutreffend als Diensteanbieter (§ 2 Satz 1 Nr. 1 TMG) qualifiziert, der eigene Informationen zur Nutzung bereit hält und deshalb gemäß § 7 Abs. 1 TMG nach den allgemeinen Gesetzen - mithin auch nach §§ 823 Abs. 1, 1004 BGB - verantwortlich ist (vgl. Senatsurteil vom 23. Juni 2009 - VI ZR 196/08, BGHZ 181, 328 Rn. 13 f. s. auch Heckmann, aaO; a.A. Brosch, aaO). Die Kläger nehmen die Beklagte nicht wegen der Durchleitung, Zwischenspeicherung oder Speicherung fremder Informationen, sondern wegen einer eigenen Information in Anspruch, konkret wegen der als Ergebnisse ihres Autocomplete-Hilfsprogramms dem Nutzer ihrer Internet-Suchmaschine angezeigten Suchwortergänzungsvorschläge. Es geht mithin um einen von der Suchmaschine der Beklagten angebotenen "eigenen" Inhalt und nicht um das Zugänglichmachen und/oder Präsentieren von Fremdinhalten, für die der Diensteanbieter gemäß §§ 8 bis 10 TMG nur eingeschränkt verantwortlich ist.'

The BGH here confirms that auto-compete constitutes own content of Google and that mere conduit, caching or hosting safe harbor therefore can not apply. This is of course one very important point. If you analyze the behavior data from your users and then present them further, according to BGH, those data are your own content. *The very fact of processing of data using your own algorithm thus turns presented results into your own*. Quite far reaching consequence, which BGH is trying to mitigate by other means later on (see below). The court further argues that hosting safe harbor can *not* apply also because the activity of auto-complete is *not* mere technical, automatic and of passive nature (yes, the unfortunate misreading of eCommerce Directive from § 114, *Google France*).

Having said that, the court stresses that from the fact that it is own content of Google, it does *not* automatically follow that Google is liable for all personality rights infringements (*'Daraus folgt allerdings noch nicht, dass die Beklagte für jede Persönlichkeitsrechtsbeeinträchtigung durch Suchvorschläge haftet'*). It is important to note that court discusses here not a liability for damages, but obligation to serve injunctions. In German law, as explained [in my paper here](http://ssrn.com/abstract=2257232), it is possible that injunctions are to be served also by persons that are innocent from the tort law perspective. But this case is about *own content*, so it logically should be about obligation to serve injunctions as *a wrongdoer*. Is that really the case?

Well, the sixth senate discusses legal basis of injunctions (§ 1004 BGB) and the concept of *'Störerhaftung'* (liability as a disturber), which is little bit misleading for IP lawyers, who hear this term mostly in it's reduced meaning from the first senate, in cases where narrow German intent-based secondary liability standard (*Teilnehmerschaft*) can not be established. However in it's original sense, the term defines all persons (disturbers) who have to serve injunctions, thus including both *tort feasors* and *innocent parties*. The BGH in my reading of the decision deals with obligation of Google to serve injunctions *as a direct tort feasor*. At the same time the court eventually arrives at the conclusion that Google has to serve injunction as a rule only upon receipt of knowledge about infringing terms in its Auto-complete Tool (as an exception the court mentions possible prevention obligation in respect to certain fields such as child pornography). How is that possible?

It seems that BGH does *not* put equation between *own content* and *own acts*. The court literary says that Google should not be blamed for development and use of the suggestion tool software. It explains it's position as follows:

Das Entwickeln und die Verwendung der die Suchvorschläge erarbeitenden Software ist der Beklagten nicht vorzuwerfen; hierbei handelt es sich vielmehr um eine durch Artt. 2, 14 GG geschützte wirtschaftliche Tätigkeit. Das Suchmaschinenangebot der Beklagten zielt auch nicht von vornherein auf eine Rechtsverletzung durch eine gegen eine bestimmte Person gerichtete unwahre Tatsachenbehauptung ab. Nur durch das Hinzutreten eines bestimmten Nutzerverhaltens können ehrverletzende Begriffsverbindungen entstehen. Die Tätigkeit der Beklagten ist andererseits aber nicht nur rein technischer, automatischer und passiver Art.

Bei Beeinträchtigungen, die eine pflichtwidrige Unterlassung als (Mit-)Ursache haben, ist zur Vermeidung einer zu weitgehenden Haftung eine fallweise wertende Betrachtung erforderlich. Die Verantwortlichkeit des Unterlassenden wird durch die Kriterien der Möglichkeit und Zumutbarkeit der Erfolgsverhinderung begrenzt.

The court here basically says that because Google's liability is dependent on searching behavior of its users, Google may only co-cause infringement of personality rights by it's *wrongful omission*. To assess whether the omission is wrongful, one has to take into account all the circumstances of the case, including criteria of feasibility and reasonableness of the success of prevention. My reading is therefore as follows.

*Own content* does not automatically establish *own acts* of data processor. *Own acts* in respect to own content that is a result of processing of third party data will arise *only if the processor omitted to act where he should have acted*.

This is rather interesting development when one considers the case-law of the first senate of the BGH. This senate seemed to be saying to us that direct liability arises when intermediaries act as own-content providers. And that intermediaries are not only directly liable for content which they have *created* but are for the content they have *adopted*. Direct liability for adopted content is based on § 7 TMG, which distinguishes between content providers as providers of their “own contents” and intermediaries as providers of “contents of third parties”. This distinction led to a court praxis according to which an intermediary can adopt the contents of third parties and become an own-content provider itself (German Federal Supreme Court, marions-kochbuch.de, 12.11.2009, Case No. I ZR 166/07). Now it seems that there are there possible types of third party content

a) submitted third party content within safe harbors,

b) submitted and adopted third party content (own content = own acts)

c) submitted and processed third party content (own content = own acts only upon wrongful omission)

What is however interesting is that arguably different jurisdictions might have different opinions on whether such algorithm-processed-data are already own content/acts of the processor. Also, Union law might have different opinion on what constitutes own content. Let's assume for the sake of argument that hosting safe harbor would otherwise apply to Google auto-complete tool. If national law can deny safe harbor by saying that Google itself *acts* and thus its service does not 'consists of the storage of information provided by a recipient of the service' as a consequence of the *national law*, then safe harbors are deeply flawed. It is exactly for this reason, why I believe that difference between *own content* and *submitted content* must be an issue of the Union law. Of course, German law can still keep viewing auto-complete results as own content of Google for personality rights analyzes, but for the purposes of applying hosting safe harbors, it has to rely on Union meaning. This further shows that safe harbor can in fact shield also from direct liability.

Let me demonstrate my point on some other example. Arguably, something similar could have happened also in the context of trade mark law not that long time ago. If CJEU in *Google France* would consider enabling of keywords registration within AdWords system a trade mark use by Google, then hosting safe harbor could potentially shield Google from *direct liability* as well. Because if Google would be *trade mark user* (hence potentially directly liable) and *hosting service* (hence having safe harbor) at the same time, it would be shielded from damages liability until it learns about the infringing keywords (and looses safe harbor), despite being direct infringer. Injunctions would be possible regardless of such knowledge of course.

I think the case shows an interesting trend. It is getting more and more difficult for courts to distinguish between own content and submitted content. The safe harbors we have today foresee only services that are rather passive in respect to third party content (e.g. passive hosting). The things get more and more complicated when the content is processed via various algorithms and only then served to users. This BGH case shows that courts may *vis-à-vis* own content arrive at similar conclusions as if they would apply the hosting safe harbor. But it also shows that more sophisticated services can not really rely on some Union law framework, because current reading of *'information provided by a recipient of the service'* (art. 14 of eCommerce Directive) is arguably narrow one. Maybe it's time to reconsider our safe harbors and especially focus on the role of injunctions that today operate outside of this "harboring" system.

Posted by Huťko at [9:39 AM](http://www.husovec.eu/2013/06/liability-for-algorithm-design-big-data.html) [](http://www.blogger.com/email-post.g?blogID=1167274025593955725&postID=1132629390153335263)

Labels: [Liability of ISPs](http://www.husovec.eu/search/label/Liability%20of%20ISPs), [personality rights](http://www.husovec.eu/search/label/personality%20rights), [search engines](http://www.husovec.eu/search/label/search%20engines)

4 comments:

[](http://www.cedricmanara.com/english)

[Cédric](http://www.cedricmanara.com/english) said...

Thanks Martin, very useful post (in particular for someone in my position: I do not read German and was unable to understand the ruling!).  FYI, the French supreme court has issued a very important ruling over Google Suggest two days ago. It finds Google cannot be found liable for the predictions that appear when users type words, as they do not reflect the will of the search tool. Without any intent from the search tool to express contentious sentences or to give a meaning to them, there cannot be any violation of French press law (note there was no reference to hosting liability).

[June 21, 2013 at 6:49 PM](http://www.husovec.eu/2013/06/liability-for-algorithm-design-big-data.html?showComment=1371833390925#c2386165366504144325)

[](http://www.blogger.com/profile/17385257405994833051)

[Huťko](http://www.blogger.com/profile/17385257405994833051) said...

Dear Cédric,  many thanks for your comment. That's very interesting. May I ask why did the Cour de Cassation require intent?   If you have any more detailed comment on the case, I would be interested in reading it.  Best, Martin

[June 22, 2013 at 10:29 AM](http://www.husovec.eu/2013/06/liability-for-algorithm-design-big-data.html?showComment=1371889775196#c6596174382576447019)

[](http://ispliability.wordpress.com/)

[Miquel](http://ispliability.wordpress.com/) said...

Hi Martin, very useful post! I’m not sure whether the notion of hosting “information provided by a recipient of the service” is being narrowly construed. In any event, though, it is interesting to see that in some instances, even rejecting the applicability of the safe harbour, courts still find it necessary some kind of knowledge on the part of the provider for liability to arise, which leads to a somewhat equivalent result. I’m thinking for instance of the Italian case Mediaset v. Yahoo! Video (Sentenza del Tribunale di Milano 19 May 2011). There the court held that the safe harbor didn't apply because Yahoo! Video was carrying out a so-called “active hosting” (instead of a passive one), but still it held that Yahoo could not be obliged to monitor its content, as this would jeopardize other rights (including freedom of expression). Yahoo was eventually held liable because it was notified of the infringing content and didn’t remove it. But it would have been held not liable in the absence of such a notification, which ends up being a similar result than if the safe harbour would have been considered applicable. Thanks again for this great post! Miquel

[June 22, 2013 at 2:12 PM](http://www.husovec.eu/2013/06/liability-for-algorithm-design-big-data.html?showComment=1371903125220#c1792075070340675672)

[](http://www.blogger.com/profile/17385257405994833051)

[Huťko](http://www.blogger.com/profile/17385257405994833051) said...

Hi Miquel,  thanks for taking time to comment.  I agree that is very interesting to see how courts construe different solutions based on their domestic tort law similarly to safe harbors.   What I meant by narrow interpretation of “information provided by a recipient of the service” is that if you think of autocompete as a separate piece of activity of Google, you can argue that search terms that appear in search suggestions are provided and stored by users, who did previous searches. And hence that this stored information per se are hosting activity, of course under broader reading of first part of Art. 14 eCommerce Directive ("an information society service is provided that consists of the storage of information provided by a recipient of the service, Member States shall ensure that the service provider is not liable for the information stored at the request of a recipient of the service"). By activity I mean not the entire algorithm, but only appearance of third party information within such algorithm. Think of Google Adwords that are also mixed in nature, but one substantial part of activity can be seen as hosting.   Sure, one can then argue that harbor benefit would not help Google because of more than mere technical, automatic and of passive nature of autocomplete or because a recipient of the service is acting under the control of the provider. I don't dispute that. But I feel that there are much more cases that could fall under “information provided by a recipient of the service” as we currently apply it to.  I think that distinctive feature of Mediaset v. Yahoo! when compared to this case is that in Mediaset v. Yahoo! Italian court still treated those videos as third party content (although outside of safe harbor), rather than own content.  What do you think?  Best, Martin

[June 22, 2013 at 3:00 PM](http://www.husovec.eu/2013/06/liability-for-algorithm-design-big-data.html?showComment=1371906010409#c7541509067730776393)

[Post a Comment](http://www.blogger.com/comment.g?blogID=1167274025593955725&postID=1132629390153335263&isPopup=true)

Links to this post

[Create a Link](http://www.blogger.com/blog-this.g)

Legal Machines

This site was created by, and is maintained by, [Selmer Bringsjord](http://www.rpi.edu/~brings), Director of RPI's [Minds & Machines Program](http://www.rpi.edu/dept/ppcs/MM/mm.uc.html). Copyright 1997, Selmer Bringsjord.

[Overview](http://www.rpi.edu/dept/cogsci/AILAW/manifesto2.htm#Overview)|| [Project Leaders](http://www.rpi.edu/dept/cogsci/AILAW/manifesto2.htm#Project Leaders) || [Team Members](http://www.rpi.edu/dept/cogsci/AILAW/manifesto2.htm#Team Members) || [Specific Sub-Projects](http://www.rpi.edu/dept/cogsci/AILAW/manifesto2.htm#Specific Sub-Projects) || [Relevant Links ||](http://www.rpi.edu/dept/cogsci/AILAW/manifesto2.htm#Relevant Links)[Contact Info](http://www.rpi.edu/dept/cogsci/AILAW/manifesto2.htm#Contact Info)

**Overview**

AI can be used to create both things with which we battle and things with which we work. The latter purpose, in the end, is probably a good deal more productive than the former (unless the former is used as a training device). Legal expert systems are intended to be as smart as Deep Blue, but rather more helpful. Imagine this scenario:

The senior associate at the prestigious Wall St. law firm looks at the desk clock and smiles. Four o'clock on Friday and a blissful work-free weekend awaits.... Just one more hour until freedom. Then, as is typical of his Friday afternoon reveries, he is rudely interrupted, this time by his computer terminal beeping to notify him of new e-mail in his inbox. He reads the subject line and fears the worst. As he pores over the attached files, he realizes that his most horrific fears have been realized. The firm's biggest client, "Guldmen Sechs", is asking him to document a transaction that makes the Travelers--Salomon Brothers acquisition/merger look like the purchase of a pack of chewing gum. As always, investment bankers' sense of time seems to be governed only by their "needs", without any reference to physical laws of nature. Apparently, the economy of our great nation, and indeed the world, will collapse if the documents are not on their desks when they come in Monday morning.

A year ago, the attorney would have picked up the telephone and cancelled his weekend plans. The revised weekend "plan" would have been to live in the office with a team of two junior associates, two secretaries and one or two paralegals. But today, the attorney doesn't pick up the phone to cancel his weekend plans, nor to mobilize the troops. He looks at his clock again -- 4:20 -- rolls his sleeves back up and gets to work.

He begins an interactive conversation with his terminal, working with the artificially intelligent agent in much the same manner that he might instruct a junior associate. He tells the agent a little bit about the nature of the transaction and its high-level structure. The agent's embedded expertise begins to draft the requisite documentation even as it continues to interact with the attorney. As new information comes in, the documentation becomes increasingly refined and tightly tailored to the particularities of this transaction. Unlike the human junior associate, the artificial agent will not make careless mistakes in spelling or grammar, and will not forget a single item on the checklist of requisite legal provisions. Furthermore, the artificial agent will ensure that all of the documentation is legally consistent and complete. Finally, the agent produces a model of the documentation for you to review. Add serveral final details, have the agent draft a cover email for the atached document files and send it back over the ether to the investment banker. Time ... 5:00 ... shut the lights and out the door.

Two blocks away in his office high atop the World Trade Center, the investment banker put on his coat and hears his terminal beep. He innocently checks to see who has sent him e-mail. Bewildered at how the attorney was able to turn around the documents so quickly, he slumps down in his ergonomic chair and picks up the phone to cancel his weekend plans.

This is a neat story, but do we have the "Right Stuff"? This may be a question not about our intelligence, but about the legal domain: Is it ripe for the advent of computer-based legal counselors? Well, when it comes to expert systems you have the right stuff when two general points ring true:

The knowledge in the domain must be complex (if the knowledge is too simple, no need for an expert system), but not *too* complex, for then knowledge articulation becomes impossible.

The knowledge must be representable in a way that transforms most "seat of the pants" human judgment into explicit rules or guidelines suitable for use in computation.

The legal domain, by our lights, fits the bill perfectly.

Of course, even when the domain seems appropriate, external considerations can discourage the use (and subsequent development of) such systems. For example, the medical domain seems to have the right stuff, but the public (patients) don't seem ready to accept the notion that a machine should assist a doctor not in the measurement of some vital statistic, but in the formulation of a medical opinion. Perhaps the gut reaction is that if consultation with a machine is required, the doctor must not know what he or she is doing (much the same way that a patient about to undergo surgery might feel somewhat uncomfortable if he saw his surgeon studying up on the procedure in a book minutes before - or during - the operation).

Our intuition is that people won't mind in the least if their lawyers are empowered by artificial colleagues--quite the contrary, if they are the beneficiaries of quicker turnaround time, lower legal fees and higher quality work product. If the case is won, if the deal is done, complaints will be few and far between. At least so it seems to us. Indeed, we believe that use of such systems will become *de rigeur* in law offices across the country, as ubiquitous as the telephone, often demanded by the clients. [The case for Law as the perfect AI domain](http://www.rpi.edu/dept/ppcs/AILAW/Lawdomain.htm).

**Project Leaders**

**Dave Ferrucci** is a Senior Scientist at IBM's T.J. Watson Research Center. In addition to working on a book on automatic language generation, Dr. Ferrucci has worked as a system architect designing distributed computing infrastructures for the banking and retail industries. He received his doctorate at Rensselaer Polytechnic Institute (RPI). Dr. Ferrucci's doctoral research was supported by a number of IBM fellowships. His thesis focused on the theory and practice of Artificial Intelligence (AI) and Logic Programming to configuration and design problems. On other projects, Dr. Ferrucci has applied this work to the intelligent generation of English narratives. He has taught several graduate-level courses in Knowledge Representation. For IBM Research Dr. Ferrucci has a pplied AI technologies to problems in computer integrated manufacturing, computer system configuration, and distributed computing. He has consulted as a Knowledge Engineer on various industrial projects through RPI's Center for Industrial Innovation. He has been invited to speak on Knowledge Engineering, Knowledge Representation and OO Programming at IBM Research, General Electric R&D, Digital Equipment Corporation, Brown University and Rensselaer Polytechnic Institute. Dr. Ferrucci's research and experience are particularly suited to his role as technical leader for a new generation of legal software to assist the practitioner in applying language and configuring contractual documents with greater ease and higher quality.

**Allan Silver** is president of [Legal Knowledge Systems](http://www.lksinc.com/), an enterprise formed in 1995 to design and develop sophisticated software systems for the legal industry. Among a variety of software offerings, Legal Knowledge Systems produces a constraint-based, interactive configuration system which assists attorneys in the drafting of complex legal contracts. Prior to this venture, Mr. Silver was a practicing attorney at a large Wall Street law firm, specializing in international project finance. He has led teams of lawyers in the representation of commercial and investment banks, governmental lenders, private developers, and utilities in a variety of projects, including power generation, mining, telecommunications, papermill, and oil and gas projects. Before becoming involved in project finance, he worked in the legal department of a German commercial bank, and before beginning his legal career, he worked as an account executive in the commercial financing and factoring industry, supervising a multimillion dollar loan portfolio. Mr. Silver received a B.S. degree in Economics and an M.B.A. in Finance from New York University and his J.D. degree from Fordham University. He is co-author of an article entitled "Thais Welcome IPPs," (with G. Wigmore) which appeared in the September 3, 1993 issue of Project Finance International and an article entitled "In Support of Private Power," (with J. Green) which appeared in the Winter 1993 issue of Infrastructure Finance. Mr. Silver was a panelist speaking on state support at the September 1993 EXNET conference on Financing International Power Projects.

[Selmer Bringsjord](http://www.rpi.edu/~brings)specializes in the logico-mathematical and philosophical foundations of artificial intelligence (AI), and, on the applied side, in the intersection of AI and creativity. He received his bachelor's degree from the University of Pennsylvania, and the PhD in philosophy and logic from Brown University in 1987. Since then he has been on faculty at Rensselaer, where in the Departments of Philosophy, Psychology and Cognitive Science, and Computer Science, he teaches AI, logic, and philosophy of mind. His pedagogy is in large part computation-based: All of his courses make intensive use of the Web, and of courseware of various types (e.g., Hyperproof). The materials thereon for his courses *Introduction to Logic* and *Computability and Logic* are in particular demand; they are used by publishers of logic courseware (e.g., Cambridge University Press and Stanford's CSLI). Bringsjord was on Rensselaer's team that won the prestigious Hesburgh Award (1995) for excellence in undergraduate education (for technology-based interactive learning). He was also a Lilly Fellow in 1989, during which time he designed and implemented an electronic textbook for introducing cognitive and computer science. He is co-director, with David Porush, of the Creative Agents Project, which has its roots in a project known as *Autopoeisis*, launched by a generous gift of $300,000 from the Luce Foundation and grants from Apple Computer. Bringsjord is author of the critically acclaimed *What Robots Can \& Can't Be* (1992, Kluwer; ISBN 0-7923-1662-2), which is concerned with the future of attempts to create robots that behave as humans. Two new technical books, *Super-Minds*, and *Artificial Intelligence and Literary Creativity*, are forthcoming this year (Kluwer Academic/Lawrence Erlbaum). The book *Abortion: A Dialogue* will also be published this Fall by Hackett. Dr. Bringsjord is also the author of a novel (*Soft Wars*}; Penguin, 1991), and papers ranging in approach from the mathematical to the informal, and covering such areas as AI, logic, natural theology, and ethics. He has lectured and interviewed in person and on radio and television across the United States, and in England, France, Ireland, Australia, Germany, Thailand, Japan and Canada.

**Specific Sub-Projects**

**Building an Artificial Agent that Excels on the LSAT**

Aspiring lawyers in the United States must take the LSAT (Law School Admissions Test) in order to enter Law School. We are building an artificial agent capable of taking (and scoring high on) the LSAT. (If it's true, as many hold, that lawyerly talent is required to excel on this test, then our agent will provide us with a foundation from which to build toward various technologies in the intersection of logic, AI, and Law.) The LSAT is divided up into different areas; for example, *Logical Reasoning*, *Reading Comprehension*, and *Logic Games*. We are tackling the *Logic Games* section first. (This section is sometimes called 'Analytical Reasoning.') Here are some sample problems from this section:

"The Dream" ([html](http://www.rpi.edu/~faheyj2/SB/LSAT/dream/dream.html) || [dvi](http://www.rpi.edu/~faheyj2/SB/LSAT/dream.dvi) || [postscript](http://www.rpi.edu/~faheyj2/SB/LSAT/dream.ps))

"Chess" ([html](http://www.rpi.edu/~faheyj2/SB/LSAT/chess/chess.html) || [dvi](http://www.rpi.edu/~faheyj2/SB/LSAT/chess.dvi) || [postscript](http://www.rpi.edu/~faheyj2/SB/LSAT/chess.ps))

"Cavies" ([html](http://www.rpi.edu/~faheyj2/SB/LSAT/cavies/cavies.html) || [dvi](http://www.rpi.edu/~faheyj2/SB/LSAT/cavies.dvi) || [postscript](http://www.rpi.edu/~faheyj2/SB/LSAT/cavies.ps))

"Crates" ([html](http://www.rpi.edu/~faheyj2/SB/LSAT/crates/crates.html) || [dvi](http://www.rpi.edu/~faheyj2/SB/LSAT/crates.dvi) || [postscript](http://www.rpi.edu/~faheyj2/SB/LSAT/crates.ps))

"Candidates" ([html](http://www.rpi.edu/~faheyj2/SB/LSAT/candidates/candidates.html) || [dvi](http://www.rpi.edu/~faheyj2/SB/LSAT/candidates.dvi) || [postscript](http://www.rpi.edu/~faheyj2/SB/LSAT/candidates.ps))

The Menagerie Problem

[Micah Clark's analyses and solutions in OTTER and Hyperproof](http://www.rpi.edu/~clarkm4/MM/legal/Menagerie/)

[html version of Menagerie Problem](http://www.rpi.edu/~brings/LOGARG/menagerie/menagerie.html)

**Relevant Links**

[A Guide To Artificial Intelligence and Law Resources](http://www.dur.ac.uk/~dla0www/centre/ail_rsrc.html)

[Sixth International Conference on ARTIFICIAL INTELLIGENCE and LAW (ICAIL '97)](http://www.cs.latrobe.edu.au/conferences/ICAIL97/) || [International Association for Artificial Intelligence and Law](http://nathan.gmd.de/iaail/iaail.html)

[Web Pages About Artificial Intelligence and Law](http://nathan.gmd.de/iaail/related.html)

[Foundation for Legal Knowledge Systems (JURIX '97)](http://www.rechten.vu.nl/jurix97.htm)

[Logic Programming at Department of Computing, Imperial College, London](http://www.doc.ic.ac.uk/report/lp_s10.html) || [[Research in Legal Reasoning]](http://www-lp.doc.ic.ac.uk/ProjectPages/legal.html)

[Seminar on Artificial Intelligence and Law](http://www.slaw.neu.edu/public/home/clinics/ai_law/sylsp95.htm)(Professors Donald Berman & [Carole Hafner](http://www.ccs.neu.edu/home/hafner/), Northeastern University)

[Legal Reasoning in the Helic-II](http://www.icot.or.jp/ICOT/Outline/FGCS-Proc/helic2.html~)

[SHYSTER](http://cs.anu.edu.au/software/shyster/)

**Contact Info**

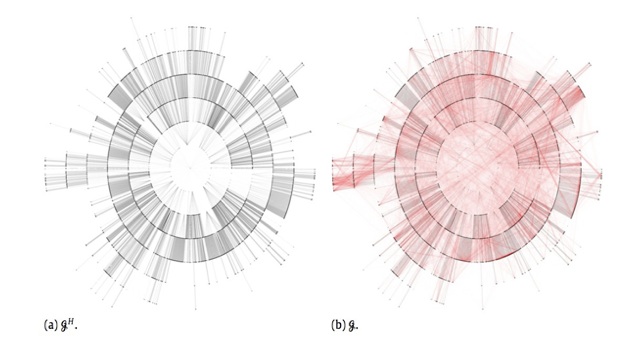
*Allan Silver /*[*silvera@lksinc.com*](mailto:silvera@lksinc.com)

*Dave Ferrucci /*[*ferrucci@watson.ibm.com*](mailto:ferrucci@watson.ibm.com)

[Selmer Bringsjord](http://www.rpi.edu/~brings)

**Measuring the Complexity of the Law : The United States Code [Repost]**

Posted on [October 8, 2010](http://computationallegalstudies.com/2010/10/08/measuring-the-complexity-of-the-law-the-united-states-code-repost/) by [Daniel Martin Katz](http://computationallegalstudies.com/author/dmartink/)

[](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1578094)

Understanding the sources of complexity in legal systems is a matter long considered by legal commentators. In tackling the question, scholars have applied various approaches including descriptive, theoretical and, in some cases, empirical analysis. The list is long but would certainly include work such as Long & Swingen (1987), Schuck (1992), White (1992), Kaplow (1995), Epstein (1997), Kades (1997), Wright (2000), Holz (2007) and Bourcier & Mazzega (2007). Notwithstanding the significant contributions made by these and other scholars, we argue that an extensive empirical inquiry into the complexity of the law still remains to be undertaken.

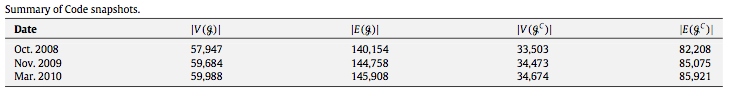
While certainly just a slice of the broader legal universe, the [**United States Code**](http://en.wikipedia.org/wiki/United_States_Code)represents a substantively important body of law familiar to both legal scholars and laypersons. In published form, the Code spans many volumes. Those volumes feature hundreds of thousands of provisions and tens of millions of words. The United States Code is obviously complicated, however, measuring its size and complexity has proven be non-trivial.

In our paper entitled, [**A Mathematical Approach to the Study of the United States Code**](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1578094) we hope to contribute to the effort by formalizing the United States Code as a mathematical object with ahierarchical structure, a citation network and an associated text function that projects language onto specific [**vertices**](http://en.wikipedia.org/wiki/Vertex_(graph_theory)).

In the visualization above, Figure (a) is the full United States Code visualized to the section level. In other words, each ring is a layer of a [**hierarchical tree**](http://en.wikipedia.org/wiki/Tree_structure) that halts at the section level. Of course, many sections feature a variety of nested sub-sections, etc. For example, the well known 26 U.S.C. 501(c)(3) is only shown above at the depth of Section 501.  If we added all of these layers there would simply be additional rings. For those interested in the visualization of specific Titles of the United States Code … we have previously created fully zoomable visualizations of [**Title 17 (Copyright)**](http://computationallegalstudies.com/2009/08/05/copyright-%E2%86%92-title-17-u-s-code-w-sea-dragon-from-microsoft-labs/), [**Title 11 (Bankruptcy)**](http://computationallegalstudies.com/2009/04/01/with-bankruptcy-on-our-minds-the-structure-of-title-11-usc/), [**Title 26 (Tax) [at section depth]**](http://computationallegalstudies.com/2009/04/13/visualizing-26-usc-___-at-the-section-depth/), [**Title 26 (Tax) [Capital Gains & Losses]**](http://computationallegalstudies.com/2010/03/25/visualizing-a-subset-of-the-tax-code-capital-gains-losses-at-full-depth/) as well as specific pieces of legislation such as the original Health Care Bill – [**HR 3962**](http://computationallegalstudies.com/2009/11/09/visualizing-the-structure-of-h-r-3962-the-health-care-bill/).

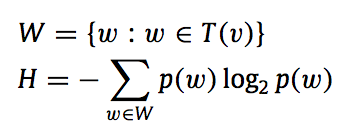
In the visualization above, Figure (b) combines this hierarchical structure together with a citation network.  We have previously [**visualized the United States Code citation network**](http://computationallegalstudies.com/2009/09/14/the-structure-of-the-united-states-code/) and have a working paper entitled [**Properties of the United States Code Citation Network**](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1502927). Figure (b) is thus a realization of the full United States Code through the section level.

With this representation in place, it is possible to measure the size of the Code using its various structural features such as [**vertices**](http://en.wikipedia.org/wiki/Vertex_(graph_theory)) *V* and its [**edges**](http://en.wikipedia.org/wiki/Graph_theory) *E*.  It is possible to measure the full Code at various time snapshots and consider whether the Code is growing or shrinking. Using a limited window of data, we observe growth not only in the size of the code but also its network of dependancies (i.e. its citation network).

[](http://ec2-107-21-222-181.compute-1.amazonaws.com/wp-content/uploads/2010/08/Picture-62.png)

Of course, growth in the size United States Code alone is not necessarily analogous to an increase in complexity.  Indeed, while we believe in general the size of the code tends to contribute to “complexity,” some additional measures are needed.  Thus, our paper features structural measurements such as number of sections, section sizes, etc.

In addition, we apply the well known [**Shannon Entropy**](http://en.wikipedia.org/wiki/Entropy_(information_theory)) measure (borrowed from [**Information Theory**](http://en.wikipedia.org/wiki/Information_theory)) to evaluate the “complexity” of the message passing / language contained therein. [**Shannon Entropy**](http://en.wikipedia.org/wiki/Entropy_(information_theory)) has a long intellectual history and has been used as a measure of complexity by many scholars.  Here is the formula for Shannon entropy:

[](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1578094)

For those interested in reviewing the full paper, it is forthcoming in [**Physica A: Statistical Mechanics and its Applications**](http://www.elsevier.com/wps/find/journaldescription.cws_home/505702/description#description). For those not familiar, Physica A is a journal published by Elsevier and is a popular outlet for [**Econophysics**](http://en.wikipedia.org/wiki/Econophysics)and [**Quantitative Finance**](http://en.wikipedia.org/wiki/Mathematical_finance). A current draft of the paper is available on the [**SSRN**](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1578094) and the [**physics arXiv**](http://arxiv.org/abs/1003.4146).

We are currently working on a follow up paper that is longer, more detailed and designed for a general audience.  Even if you have little or no interest in the analysis of the United States Code, we hope principles such as entropy, structure, etc. will prove useful in the measurement of other classes of legal documents including contracts, treaties, administrative regulations, etc.

This entry was posted in [Uncategorized](http://computationallegalstudies.com/category/uncategorized/). Bookmark the [permalink](http://computationallegalstudies.com/2010/10/08/measuring-the-complexity-of-the-law-the-united-states-code-repost/).

[← Sebastian Seung: I Am My Connectome [ TED 2010 ]](http://computationallegalstudies.com/2010/10/03/sebastian-seung-i-am-my-connectome-ted-2010/)

[Twitter Mood Predicts the Stock Market →](http://computationallegalstudies.com/2010/10/17/twitter-mood-predicts-the-stock-market/)

**One Response to *Measuring the Complexity of the Law : The United States Code [Repost]***

http://0.gravatar.com/avatar/4e4c2dfb9b73125b70348a0559550f1f?s=40&d=http%3A%2F%2F0.gravatar.com%2Favatar%2Fad516503a11cd5ca435acc9bb6523536%3Fs%3D40&r=G**Alan E. Dunne** *says:*

[August 20, 2010 at 18:12](http://computationallegalstudies.com/2010/10/08/measuring-the-complexity-of-the-law-the-united-states-code-repost/#comment-1708)

With Respect

Have you thought of doing, or does any reader of this post intend to do, a parallel analysis of constitutions? In particular, of the constitution of India, said to be more detailed as well as more favourable to the federal government versus those of Australia, Canada and the United States?

Yours Sincerely

[Reply](http://computationallegalstudies.com/2010/10/08/measuring-the-complexity-of-the-law-the-united-states-code-repost/?replytocom=1708#respond)

**Leave a Reply**

[**The State Decoded**](http://www.statedecoded.com/)

**State codes, for humans.**

[](http://www.statedecoded.com/)

Top of Form

Search

Bottom of Form

**Main menu**

[Skip to primary content](http://www.statedecoded.com/#content)

[Skip to secondary content](http://www.statedecoded.com/#secondary)

[**Home**](http://www.statedecoded.com/)

[About](http://www.statedecoded.com/about/)

[States](http://www.statedecoded.com/states/)

[Downloads](https://github.com/statedecoded/statedecoded/tags)

[Contact](http://www.statedecoded.com/contact/)

[**Version 0.7 Released**](http://www.statedecoded.com/2013/06/release-v07/)

Posted on [**June 20, 2013**](http://www.statedecoded.com/2013/06/release-v07/)

[Version 0.7 of The State Decoded is now available on GitHub](https://github.com/statedecoded/statedecoded/tree/v0.7). This is a really meaty release, dedicated entirely to optimizations: it’s faster, more efficient, easier to extend, easier to contribute to, easier to deploy, and easier to navigate. This release is comprised of a whopping 353 Git commits—that’s more than every commit that went into versions 0.1 through 0.6, *combined.* Here are some of the major changes:

**Tuning**

Every line of code was reviewed to see how it could be made faster—in tiny ways (instances of stristr() replaced with strstr()—or, better, strpos()) and in large ways (tossing out whole methods and starting again). The indices in MySQL were evaluated and revised, and any PHP error of level E\_NOTICE and above was quieted. And the parser’s memory usage has been reduced substantially, making it faster and more efficient to process large legal codes that previously might have strained (or broken) Apache’s per-process memory limit.

**Caching**

There are now hooks for both [APC](http://php.net/manual/en/book.apc.php) and [Varnish](https://www.varnish-cache.org/), so that folks running either of those popular caching applications (or, better, both of them) can reap the speed benefits. API keys, all constants, and templates are cached in APC now, with more caching on tap in upcoming releases.

**Development Environment**

This version was developed substantially within a [Vagrant](http://www.vagrantup.com/) staging environment, resulting in the inevitable optimization of The State Decoded to run within Vagrant. That involved a lot of tiny changes (e.g., respecting port numbers in URLs) that collectively create a smooth experience when developing locally. We have [the under-development Vagrant configuration for The State Decoded in its own repository.](https://github.com/statedecoded/statedecoded-vagrant) This version and all future versions will have a Vagrant machine image available for download, to make it trivial to get started. Vagrant is working on a path to deploy Vagrant machine configurations as[AWS](http://aws.amazon.com/) instances, which is why this is a bandwagon worth hopping on now.

**Standardization**

Out with [HTML Purifier](http://htmlpurifier.org/), in with [HTML Tidy](http://tidy.sourceforge.net/). Out with [MDB2](http://pear.php.net/package/MDB2), in with [PDO](http://php.net/manual/en/book.pdo.php). Out with late-nineties-style commenting and code formatting, in with PEAR-style commenting and code formatting. There was nothing *wrong* with any of those prior approaches, but it’s best to establish an environment that contributors expect—that makes it easier for folks to contribute code to the project, or customize it for their own website. Also, HTML Tidy and PDO are already installed by default on a great many systems, which simplifies the setup process.

**Extensibility**

Several steps have been made to facilitate customization. All non-obvious database columns are now commented, there’s infrastructure for inline help text (stored and distributed as JSON), and there’s support for importing, storing, and display arbitrary metadata fields alongside the standard data about each law.

**New Features**

And, of course, we couldn’t resist a few new features. There’s now keyboard navigation within laws and structures, for those power-users who want to flip through laws quickly. There’s baked-in support for [Disqus](http://disqus.com/)-based commenting on each law page—just enter your site’s Disqus shortname in config.inc.php and you’re up and running. And, finally, there’s bulk generation of both plain text and JSON versions of laws. To what end? Dunno—that’s for you to figure out.

This release involved a lot of work over the course of four months. [Bill Hunt](http://krues8dr.com/) has scrubbed in to help as a core contributor, and he’s responsible for a lot of these improvements. Some very helpful bug reports, wiki edits, and pull requests came from [Chris Birk](https://github.com/cmbirk) and[Daniel Trebbien](https://github.com/dtrebbien).

Next up: versions 0.8 and 0.9, which will be released soon, and close together. Version 0.8 will be comprised of [the UI/UX branch](https://github.com/statedecoded/statedecoded/tree/UI), which only has [a handful of small tasks remaining](https://github.com/statedecoded/statedecoded/issues?milestone=10), thanks to months of work by [Meticulous](http://meticulous.com/) ([John Athayde](http://boboroshi.com/) and [Lynn Wallenstein](http://lynn.io/)). And Version 0.9 will be comprised of [the Solr branch](https://github.com/statedecoded/statedecoded/tree/Solr), and is dedicated to baking [Apache Solr](http://lucene.apache.org/solr/) into The State Decoded. That’s based on several months of work by the team at [OpenSource Connections](http://www.opensourceconnections.com/), who finished up earlier this week—there are only about [a dozen outstanding issues](https://github.com/statedecoded/statedecoded/issues?milestone=7), all of which build on OSC’s work to add some valuable new features to The State Decoded.

Posted in [**General**](http://www.statedecoded.com/category/general/) | [**Leave a reply**](http://www.statedecoded.com/2013/06/release-v07/#respond)

[**Bulk Downloads of Five State Codes**](http://www.statedecoded.com/2013/05/5-bulk-downloads/)

Posted on [**May 30, 2013**](http://www.statedecoded.com/2013/05/5-bulk-downloads/)

A very real obstacle to putting up state code websites is getting a copy of that state’s laws. For example, there’s a New Jersey group that wants to set up The State Decoded for their state. But, like most states, New Jersey doesn’t provide bulk downloads—it’s not possible to simply get a raw copy of the files. The backup option is what’s known as “screen-scraping”—having software load every single law on the official state law website, one by way, and copy the laws from there. This is a terrible solution, but it’s all that’s available in most U.S. states. [The New Jersey statutes website](http://lis.njleg.state.nj.us/cgi-bin/om_isapi.dll?clientID=597227&depth=2&expandheadings=off&headingswithhits=on&infobase=statutes.nfo&softpage=TOC_Frame_Pg42) is distinctly un-scrapeable. I don’t know that it’s *impossible*, but it would be an unpleasant task.

Today, Carl Malamud of Public.Resource.org tweeted the news that he’s got five new state codes online as bulk data:

In addition to [bulk machine-readable files](https://law.resource.org/pub/us/code/), they’re also available in [a variety of file formats on Archive.org](http://archive.org/details/govlaw). They join the Maryland and Washington D.C. codes that he’s already made available as bulk downloads. ([Maryland Decoded](http://www.statedecoded.com/2013/05/maryland-decoded/) is up now, and [the Open Law DC project](http://dccode.org/) has a great site for their code, with [a State Decoded implementation under development](https://github.com/krues8dr/statedecoded/tree/dccode) that’ll be the subject of [a hackathon](http://dchackforchange.eventbrite.com/) on Saturday’s [National Day of Civic Hacking](http://hackforchange.org/).)

Now the onus is on folks in Arkansas, Colorado, Georgia, Idaho, and Mississippi to set up to the plate and put this data to work. Who’s going to implement The State Decoded in these states?

Posted in [**General**](http://www.statedecoded.com/category/general/)

[**Maryland Decoded**](http://www.statedecoded.com/2013/05/maryland-decoded/)

Posted on [**May 10, 2013**](http://www.statedecoded.com/2013/05/maryland-decoded/)

A new State Decoded site launched today: [Maryland Decoded](http://marylandcode.org/). A project of the [OpenGov Foundation](http://opengovfoundation.org/), they’re doing some innovative stuff on the still-under-development platform. For instance, they’re crowd-sourcing “catch lines”—the titles that most states apply to their laws. Maryland does not have catch lines, so instead of having a law titled “Murder in the First Degree,” they simply have [GCR § 2-201](http://marylandcode.org/gcr-2-201/). Solution? Anybody can suggest a catch line, and they’ll build up their own catch lines, gradually.

Every state presents its own set of challenges and opportunities. The OpenGov Foundation is capitalizing on the opportunities to overcome the challenges and helping to improve The State Decoded for those who will follow in their path.

Posted in [**Member Site**](http://www.statedecoded.com/category/member-site/)

[**Washington D.C. and the Work Ahead**](http://www.statedecoded.com/2013/03/dc-work-ahead/)

Posted on [**March 29, 2013**](http://www.statedecoded.com/2013/03/dc-work-ahead/)

On *Greater Greater Washington,* Tom MacWright recently wrote [a blog entry highlighting the problems of access to the Washington D.C. Code](http://greatergreaterwashington.org/post/18132/dcs-laws-arent-yours/). There is, first, a legal obstacle: Washington D.C. claims copyright over their laws, which is to say that it is illegal to reproduce them without permission of the city. Then, second, what is perhaps a more significant obstacle: they outsource the maintenance of their legal code.

The city of Washington D.C. long ago started paying WestLaw—and now LexisNexis—to[turn the D.C. Council’s bills into laws](http://macwright.org/2013/02/13/the-code-compiled.html). As a result, they now have neither the knowledge nor the infrastructure to maintain their own laws. The only way that D.C. can find out what their laws say is to pay LexisNexis to tell them. This is consequently true for the public, as well. If a resident of D.C.—like MacWright—wants to know what the law says, there’s no sense in asking (or FOIAing) the city, because the city has outsourced the process so completely that they know nothing.

MacWright has a few options to know what the law says. The first is to travel to a library on each occasion that he wants to know something (assuming he can find one that has a current copy of the DC Code), and read it there. The second is that [he can buy a copy, for $867.00](http://legalsolutions.thomsonreuters.com/law-products/Statutes/District-of-Columbia-Official-Code/p/100000736). And the third is that he can use [the DC Code website](http://government.westlaw.com/linkedslice/default.asp?SP=DCC-1000), maintained by WestLaw, which is every bit as awful as any other state code website.

So how is the D.C. Code to get the State Decoded treatment? How can a digital copy be imported into the software, for the general public benefit? It can’t be FOIAed from D.C. Council, since they don’t have it. It’s clearly impractical to scan in 25 volumes of hardbound books. Normally that would leave [scraping the website](http://en.wikipedia.org/wiki/Web_scraping), but [WestLaw’s website has a EULA that prohibits copying material off of the website](http://government.westlaw.com/linkedslice/default.asp?Action=Copyright&RS=GVT1.0&VR=2.0&SP=DCC-1000). WestLaw has been hired to do for the Washington D.C. government what they cannot or will not do for themselves—post laws to the web—and because they choose to impose copyright restrictions, that is a legal barrier preventing that material from being reused.

Normally, this would be the end of the road—Washington D.C. would have cut off their code from being improved (or even reused) in any way by third parties. In this case, though, the story has a different ending. [Public.Resource.Org](http://public.resource.org/) has taken the surprising and admirable tack of purchasing all of the volumes of the D.C. Code, slicing them up, scanning them in, OCRing them, and distributing them for free.

[](http://www.flickr.com/photos/publicresourceorg/8554646423/)

Deglued bundles on the bottom right awaiting their turn on the two high-speed scanners. When they are completed, the bundles are put upper right to await a QA pass.

All of the volumes can be downloaded as PDFs or, [via the Internet Archive](http://archive.org/search.php?query=subject%3Acode.dc.gov), in nearly any other file format one can think of. For those who would like a print copy for a more reasonable price, print-on-demand service [Lulu sells each volume for just $12](http://www.lulu.com/shop/district-of-columbia/dc-code-volume-1/paperback/product-20943270.html).

Lest the motivations of Public.Resource.Org be unclear, [a “Proclamation of Digitization” accompanies the release](https://law.resource.org/pub/us/code/dc/dc.proclamation.pdf), citing a pair of Supreme Court rulings (“the authentic exposition and interpretation of the law, which, binding every citizen, is free for publication to all, whether it is a declaration of unwritten law, or an interpretation of a constitution or a statute”) and declaring that “any assertion of copyright by the District of Columbia or other parties on the District of Columbia Code is declared to be NULL AND VOID as a matter of law and public policy as it is the right of every person to read, know, and speak the laws that bind them.” The organization mailed out elaborate packages, containing portions of the D.C. Code, to announce its availability. (You can [see my own unboxing photos](http://www.flickr.com/photos/waldoj/sets/72157633102174393/).)

All that remains is for somebody to marry this source of data with The State Decoded as, indeed, somebody is already talking about doing. The D.C. Council or WestLaw may not be happy about this—it’s quite possible that one or both entities will take legal action to halt this—but I’m confident that it will be found that the law supports making those very laws public.

Posted in [**General**](http://www.statedecoded.com/category/general/)

[**Two Mini-Projects: Subsection Identifier and Definition Scraper**](http://www.statedecoded.com/2013/02/subsection-identifier-definition-scraper/)

Posted on [**February 27, 2013**](http://www.statedecoded.com/2013/02/subsection-identifier-definition-scraper/)

The State Decoded project has spun off a couple of sub-projects, components of the larger project that can be useful for other purposes, and that deserve to stand alone. (Both are found on [our GitHub repository](https://github.com/statedecoded/).)

The first is [Subsection Identifier](https://github.com/statedecoded/subsection-identifier), which turns *theoretically* structured text into *actually*structured text. It is common for documents in outline form (contracts, laws, and other documents that need to be able to cross-reference specific passages) to be provided in a format in which the structural labels flow into the text. For example:

A. The agency may appoint a negotiated rulemaking panel (NRP) if a regulatory action is expected to be controversial. B. An NRP that has been appointed by the agency may be dissolved by the agency when: 1. There is no longer controversy associated with the development of the regulation; 2. The agency determines that the regulatory action is either exempt or excluded from the requirements of the Administrative Process Act; or 3. The agency determines that resolution of a controversy is unlikely.

One of the helpful features of The State Decoded is that it breaks up this text, understanding not just that every labelled line can stand alone, but also that the final line, despite being labelled “3,” is actually “B3,” since “3″ is a subset of “B.” That functionality has been forked from The State Decoded, and now stands alone as Subsection Identifier, which accepts passages of text, and turns them into well-structured text, like such:

( [0] => stdClass Object ( [prefix\_hierarchy] => stdClass Object ( [0] => A ) [prefix] => A. [text] => The agency may appoint a negotiated rulemaking panel (NRP) if a regulatory action is expected to be controversial. ) [1] => stdClass Object ( [prefix\_hierarchy] => stdClass Object ( [0] => B ) [prefix] => B. [text] => An NRP that has been appointed by the agency may be dissolved by the agency when: ) [2] => stdClass Object ( [prefix\_hierarchy] => stdClass Object ( [0] => B [1] => 1 ) [prefix] => B.1. [text] => There is no longer controversy associated with the development of the regulation; ) [3] => stdClass Object ( [prefix\_hierarchy] => stdClass Object ( [0] => B [1] => 2 ) [prefix] => B.2. [text] => The agency determines that the regulatory action is either exempt or excluded from the requirements of the Administrative Process Act; or ) [4] => stdClass Object ( [prefix\_hierarchy] => stdClass Object ( [0] => B [1] => 3 ) [prefix] => B.3. [text] => The agency determines that resolution of a controversy is unlikely. ) )

The second mini-project is [Definition Scraper](https://github.com/statedecoded/definition-scraper), which extracts defined terms from passages of text. Many legal documents begin by defining words that are then used throughout the document, and knowing those definitions can be crucial to understanding that document. So it can be helpful to be able to extract a list of terms and their definitions. Definition Scraper needs only be handed a passage of text, and it will determine whether it contains defined terms and, if it does, it will return a dictionary of those terms and their definitions.

Running this passage through Definition Scraper:

*“The Program” refers to any copyrightable work licensed under this License.  
A “covered work” means either the unmodified Program or a work based on the Program.*

Yields the following two-entry dictionary:

( [program] => “The Program” refers to any copyrightable work licensed under this License. Each licensee is addressed as “you”. “Licensees” and “recipients” may be individuals or organizations. [covered work] => “covered work” means either the unmodified Program or a work based on the Program. )

Definition Scraper is also a core function of The State Decoded, but warrants becoming its own project because it is so clearly useful for applications outside of the framework of The State Decoded.

The decision to spin off these projects was prompted by [a report by the John S. and James L. Knight Foundation](http://www.knightfoundation.org/publications/knight-news-challenge-year-3-evaluation-report), the organization that funds The State Decoded, which evaluated the success of their News Challenge winners. They found several common attributes among the more successful funded projects, including this:

*Projects that achieved strong use and adoption of their code often built and released their software in individual components, knowing that certain elements had value and a wide range of uses beyond the main focus of their project.*

As development on The State Decoded continues, we may well spin off more mini-projects, if it becomes clear that more components of the overall project could be useful stand-alone tools.

Posted in [**News**](http://www.statedecoded.com/category/news/)

[**Version 0.6 Released**](http://www.statedecoded.com/2013/02/release-v06/)

Posted on [**February 7, 2013**](http://www.statedecoded.com/2013/02/release-v06/)

[Version 0.6 of The State Decoded is now available on GitHub](https://github.com/statedecoded/statedecoded/tree/v0.6). This release is a really exciting one—it establishes a public API for State Decoded sites and creates a standard XML format for importing laws! This is an important release of The State Decoded, one that stands to increase significantly the accessibility of the project to developers, both within the software and without. A total of [23 issues were resolved](https://github.com/statedecoded/statedecoded/issues?milestone=8&state=closed), nearly all of which are towards those two goals.

**Public API**

The State Decoded now has a fully fleshed out RESTful, JSON-based API. It has three methods: Law, Structure, and Dictionary. Law provides all available information about a given law. Structure provides all available information about a given structural unit (the various organizational units of legal codes—”titles,” “chapters,” “parts,” etc.). And Dictionary provides the definition (or definition*s*) for a term within a legal code. The data for these comes directly from the internal API that drives the site—what’s available publicly is what drives the site privately. In fact, I’m toying with the idea of having the site consume its own API, using internal APIs solely to serve data to the external API, and having every other part of the site get its data from that external API.

For a quick start trying this out on the Virginia site, you can use the trial key,4l6dd9c124ddamq3 (though don’t build any applications using that key, or they will break when it expires), and [see the API documentation to put it to work](https://github.com/statedecoded/statedecoded/wiki/API-Documentation). For a *really* quick start, you can just [browse the Code of Virginia via the API](http://vacode.org/api/structure/?key=4l6dd9c124ddamq3), [check out a list of definitions](http://vacode.org/api/dictionary/person?key=4l6dd9c124ddamq3), or [read the text of a law](http://vacode.org/api/0.1/law/2.2-3704?key=4l6dd9c124ddamq3). If you decide that you like what you see, [register for a key](http://vacode.org/api-key/) and put this API to work.

Personally, this is the release that I’ve been waiting for. There’s an extent to which the purpose of The State Decoded project is really just to provide an API for legal codes; the fact that there’s a pretty website atop that API is just icing on the cake.

**XML Format**

A significant obstacle to implementing The State Decoded has been the need to customize the parser for each installation. Every legal code is different—there are no standards—with some all in one big SGML file, others stored in thousands of XML files, and other still needing to be scraped off of webpages. That necessitated modifying [the State Decoded parser](https://github.com/waldoj/statedecoded/blob/v0.6/includes/state-sample.inc.php) to interface the data from the source files with the internal API. That’s really not an obstacle to people and organizations who are serious about implementing The State Decoded. But plenty of people might be serious if they could just try it out first. There’s a huge gradient between “huh, looks interesting” and “I must get my city/state/country laws online!” It’s foolish to assume that people won’t just want to try it out first. After all, that’s how *I* prefer to get started with software and projects.

The solution was to establish an XML standard for importing legal codes into The State Decoded, to provide a low-barrier-to-entry path in addition to the more complex path. To be clear, this is not an attempt to create an XML standard for legal codes. This is a loosely typed standard, used solely as an import format for The State Decoded. Many legal codes are already stored as XML—that’s the most common file format—so getting those codes into The State Decoded now only requires writing a bit of [XSLT](http://en.wikipedia.org/wiki/XSLT). This is a much lower barrier to entry.

The XML looks like this:

<?xml version="1.0" encoding="utf-8"?> <law> <structure> <unit label="" identifier="" order\_by="" level=""></unit> </structure> <section\_number></section\_number> <catch\_line></catch\_line> <order\_by></order\_by> <text> <section prefix=""></section> </text> <history></history> </law>

Several of those fields are optional, too. There will certainly be legal codes and organizations for which this won’t do the trick—they’ll need to modify the parser to handle some unusual types of data, fields, third-party data sources, etc. But for most people, this will be a big improvement.

[The Code of Virginia is available as State Decoded XML](http://vacode.org/downloads/), so if you’ve been considering playing with The State Decoded, it just got a whole lot easier to deploy a test site. Just download that XML and [follow the installation instructions](https://github.com/statedecoded/statedecoded/wiki/Installation-Instructions).

Thanks to Tom MacWright, Andrew Nacin, Daniel Trebbien, and Chad Robinson for their pull requests, wiki edits, and trouble tickets.

Posted in [**News**](http://www.statedecoded.com/category/news/)

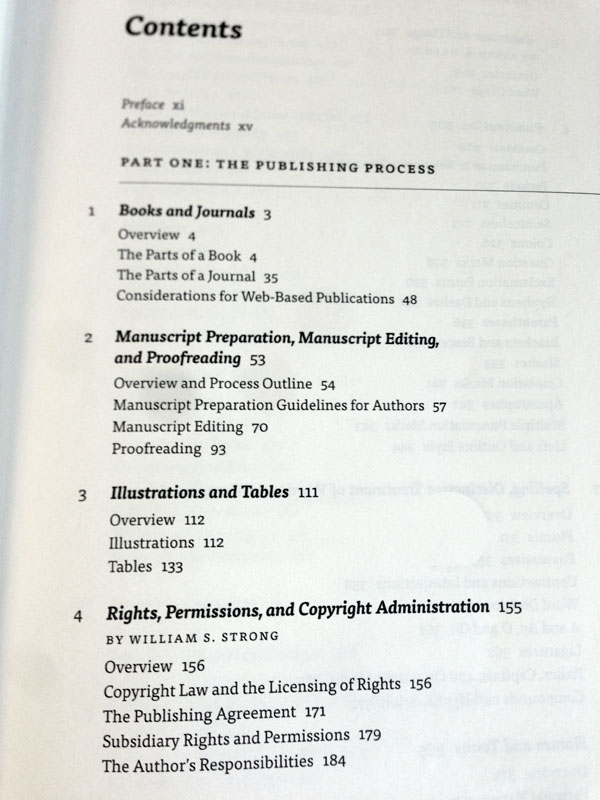
[**No Love from LexisNexis**](http://www.statedecoded.com/2013/01/lexisnexis-love/)

Posted on [**January 3, 2013**](http://www.statedecoded.com/2013/01/lexisnexis-love/)

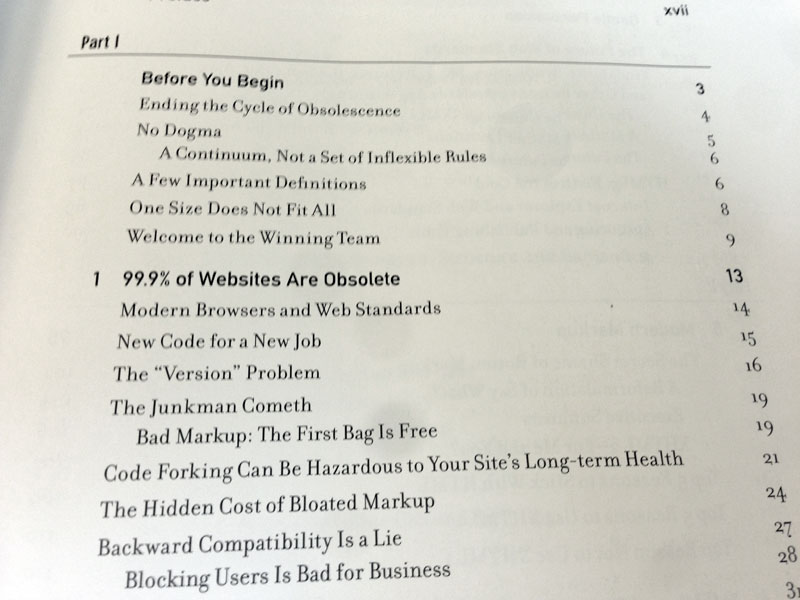
This is what the table of contents looks like in LexisNexis’s printed edition of the Code of Virginia:

I was a bit stunned the first time I saw this. It’s just word soup. There’s simply no effort to make it legible. No thought has gone into this. There is, in short, no love.

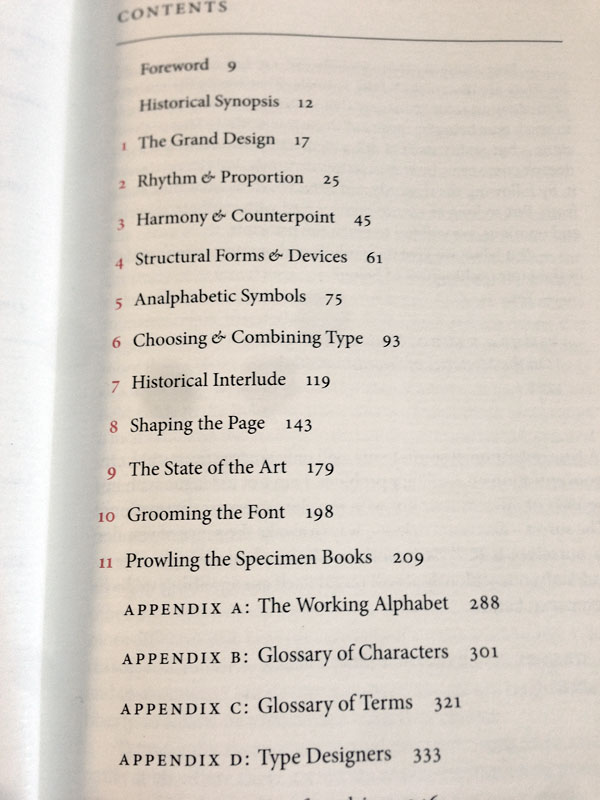
I feel like, as a culture, we basically understand how to make tables of contents. Right? Grabbing a few books off my desk, more or less at random, I thought I’d compare Lexis’s table of contents to those of othersHere’s *The Chicago Manual of Style*:

[](http://www.statedecoded.com/wp-content/uploads/2013/01/chicago-toc.jpg)

*Designing with Web Standards,* by Jeffrey Zeldman and Ethan Marcotte:

[](http://www.statedecoded.com/wp-content/uploads/2013/01/zeldman-toc.jpg)

And Robert Bringhurst’s *The Elements of Typographic Style:*

[](http://www.statedecoded.com/wp-content/uploads/2013/01/bringhurst-toc.jpg)These are all different, but via various small design cues they all manage to accomplish the same thing: they make it easy for somebody to browse through the contents of the text and locate the specific section that they need. Microsoft Word, right out of the box, will happily render a table of contents in styles reminiscent of all of these, with minimal effort.

LexisNexis isn’t even *trying.* I can’t pretend to know why. But with this as the current state of affairs in the presentation of legal information, it’s trivial for The State Decoded—or anybody with a copy of Word—to improve upon it.

Posted in [**General**](http://www.statedecoded.com/category/general/)

[**Version 0.5 Released**](http://www.statedecoded.com/2012/12/release-v05/)

Posted on [**December 21, 2012**](http://www.statedecoded.com/2012/12/release-v05/)

[Version 0.5 of The State Decoded is now available on GitHub](https://github.com/statedecoded/statedecoded/tree/v0.5). This release is full of general enhancements, and some of them are significant. [Twenty-four issues were resolved](https://github.com/statedecoded/statedecoded/issues?milestone=6&state=closed) with this release, including some new features, some significant optimizations, some standardization, and further abstraction of functionality to make it easier to implement.

Here are the most interesting changes:

All functionality likely to require customization with each implementation now resides in a state-specific file, rather than being mixed in with core functionality.

The beginnings of a templating system are in place, allowing images, CSS, and HTML to be packaged together, in the general direction of how WordPress works.

A new method has been added to the Law class, that simply verifies that a given law exists. This has led to a 350% improvement in page rendering times (with [the benchmark law](http://vacode.org/46.2-100/), 2,142 milliseconds reduced to 610 milliseconds), a result of the need to verify that every law mentioned in a section actually exists.

Several files have been renamed, in order to prevent customizations from being overwritten with upgrades. This is an important step towards providing an upgrade path between versions.

Two bulk download files are automatically generated each time the parser is run—a JSON version of the custom dictionary, and a JSON version of the entire legal code.

Much has been done towards standardization generally, so that the project adheres to best practices in PHP and MySQL. While this is of little benefit to the end user, for anybody actually getting their hands dirty with code, it should make things much simpler. There’s a lot more to be done to comply with [PEAR coding standards](http://pear.php.net/manual/en/standards.php), but that’s underway.

[Virginia attorney James Steele](http://jamessteelelaw.com/) created a print stylesheet to format laws nicely when he printed them out. He was kind enough to contribute that to the project, and printouts of laws are now vastly improved.

Most of these changes are, in one way or another, moving the project towards standardization, automation, and normalization, to make it easier to deploy, maintain, and use. It should all be a lot easier to understand for a programmer diving into it for the first time.

The next release is version 0.6, dedicated to API improvements. [That will be comprised of a relatively small number of issues](https://github.com/waldoj/statedecoded/issues?milestone=8), but they’re big ones: creating a RESTful JSON-based API, and supporting a crudely typed XML input format to simplify the process of parsing new codes. The latter is important, because the present arrangement requires that one know enough PHP to modify the parser to suit their own code’s unique storage and formatting. The idea here is that you can, alternately, use the tools of your choice to create an XML version of that code, and as long as that XML is of the style expected by the parser, it can be imported without having to edit a line of PHP in The State Decoded. Note that v0.6 was supposed to be the release in which the Solr search engine was integrated deeply into the software. That has now been pushed back—it’ll probably be v0.9—in order to accommodate a vendor’s schedule.

Posted in [**News**](http://www.statedecoded.com/category/news/)

[**Version 0.4 Released**](http://www.statedecoded.com/2012/09/release-v04/)

Posted on [**September 14, 2012**](http://www.statedecoded.com/2012/09/release-v04/)

Today, [version 0.4 of The State Decoded](https://github.com/statedecoded/statedecoded) was tagged on GitHub and [bundled up for download](https://github.com/statedecoded/statedecoded/downloads), the result of six weeks of work. This release is dedicated (almost) exclusively to enhancements to the dictionary system. [Eighteen issues](https://github.com/statedecoded/statedecoded/issues?milestone=5&state=closed) comprise the changes in this release, sixteen of which pertain to the built-in automatic, custom dictionary system, which finds defined terms within legal codes and stores them in a dictionary, using that data to embed contextual definitions that are relevant to each law.

There are a few big changes:

The State Decoded comes with a built-in dictionary of general legal terms. Using several different non-copyrighted, government-created legal dictionaries, a collection of nearly 500 terms have been put together, which will help people to understand common legal terms that are rarely defined within legal codes, such as “mutatis mutandis,” “tort,” “pro tem,” and “cause of action.”

Dictionary terms are now identified more aggressively, which means that for many states, the size and scope of the custom dictionary is going to expand substantially. In the case of Virginia there was a *49%* increase (a leap from 7,681 to 11,504 definitions), a striking difference that could be observed immediately when browsing the site.

The problem of nested/overlapping definitions has been solved. When one definition was nested within another (e.g., if we have definitions for both “robbery” and “armed robbery”), then mousing over “robbery” would yield a pair of pop-up definitions, one obscuring the other. Now only the definition for the longest term is defined under those circumstances.

Internal terminology has been standardized. In various places the dictionary and its components were all called different things (glossary, definitions, dictionary, terms, etc.) in different places. Now the collection of words is called a “dictionary,” each defined word is a “term,” and the description of that that term means is a “definition.”)

The retrieval and display of definitions is substantially faster—they take about half the time that they used to. This is a result of optimizing and simplifying the structure of the database table in which definitions are stored.

[A list of all closed issues is available](https://github.com/statedecoded/statedecoded/issues?milestone=5&state=closed) for those who want specifics. And for those who are suckers for details, this is the first release for which [a detailed Git commit log](https://github.com/statedecoded/statedecoded/commits/master) is available, with relatively detailed comments for all 68 commits that comprise this release.

This release is two weeks late, almost entirely because of time spent on [a pernicious and difficult parsing bug](https://github.com/statedecoded/statedecoded/issues/113) that, it only occurred to me today, shouldn’t have blocked this release because, while an important problem, it has absolutely nothing to do with definitions. (The problem that is being wrestled with is how to handle subsections of laws that span paragraphs. Easy to describe, difficult to solve, at least for those state codes that pretend that a paragraph and a section are one and the same. I’m looking at you, Virginia.) That issue has been moved back to v0.5, and I’ll go right back to wrestling with it on Monday.

Next up, [version 0.5 will be another general-enhancements release](https://github.com/statedecoded/statedecoded/issues?milestone=6&state=open). Version 0.6 will be the[Solr](http://lucene.apache.org/solr/) release—the version in which the popular search software becomes integrated deeply into the project. Version 0.7 will be the API release, where the nascent API gets built out to full functionality and documented properly. Version 0.8 will be the user interface release, in which the design will be overhauled, a responsive design will be implemented, serious work will go into the typography, an intercode navigation system will be implemented, contextual help and explanations will be embedded throughout, and the results of some light UI testing will be incorporated. Version 0.9 will be dedicated to optimizations—making everything go faster and be more fault-tolerant, both through improving the code base and supporting the APC and Varnish caching systems. And, finally, version 1.0 will be the first release in which State Decoded becomes a platform that facilitates the sort of analysis and data exchange that makes this project so full of possibility—things like flexible content export, visualizations, user portfolios of interesting laws, and surely lots of other things.

Posted in [**News**](http://www.statedecoded.com/category/news/)

[**Typeface Authority**](http://www.statedecoded.com/2012/08/typeface-authority/)

Posted on [**August 13, 2012**](http://www.statedecoded.com/2012/08/typeface-authority/)

With the design process for The State Decoded underway, we’re putting a lot of thought into typography. Helpful to this process has been both Ruth Anne Robbins’ “[Painting with print: Incorporating concepts of typographic and layout design into the text of legal writing documents](http://www.ca7.uscourts.gov/rules/painting_with_print.pdf)” and Derek H. Kiernan-Johnson’s “[Telling Through Type: Typography and Narrative in Legal Briefs](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1646105).”

Both of those papers are conceptual in nature, so they’re complemented nicely by Errol Morris’ two-part series [[1](http://opinionator.blogs.nytimes.com/2012/08/08/hear-all-ye-people-hearken-o-earth/), [2](http://opinionator.blogs.nytimes.com/2012/08/09/hear-all-ye-people-hearken-o-earth-part-2/)] about the results of a quiz that he ran on the *New York Times*website, ostensibly measuring readers’ optimism. In fact, he was measuring the impact of different typefaces on readers’ responses. Those who doubt that a typeface could have much of an impact on the credulity of a reader should consider the effect of [Comic Sans](http://en.wikipedia.org/wiki/Comic_Sans), which Morris discovered (unsurprisingly) correlated strongly with incredulity on the part of readers. Of the six typefaces that he tested (Baskerville, Comic Sans, Computer Modern, Georgia, Helvetica, and Trebuchet), Baskerville proved the most persuasive. The effect was small, but significant.

This is the sort of consideration that is clearly lacking in the present rendering of laws, both online and in print. (Typographically, LexisNexis’s printed state codes are a train wreck.) It’s also precisely the consideration that will set apart those sites based on The State Decoded, or anybody who cares to employ the project’s stylesheets. There will be more news about this ongoing design work in the weeks ahead.