

The Red Herring of Defensibility and Predictive Coding™

By Craig Carpenter, VP & General Counsel, Recommind



Predictive Coding™ has been getting a lot of press lately, and not just in legal-specific publications like the [ABA Journal](#) and [Law.com](#) but in major, “mainstream” publications like the [Wall Street Journal](#) as well. This seemingly newfound interest has actually been building for several years, with the tip of the proverbial iceberg only breaching the surface in the last nine months – once customers like [Fulbright & Jaworski](#), [Morgan Lewis](#), [Pfizer](#), and [WilmerHale](#) began touting Predictive Coding’s real-world benefits publicly.

As Recommind first [defined](#) and [popularised](#), Predictive Coding™ automates the majority of the review process by using powerful concept search and auto-categorisation technology to 1) find key documents quickly, automatically and irrespective of keyword used in a search or a document, 2) automatically prioritise all documents for review (from most relevant or important to least), and 3) provide a computer-generated “review” of most documents in a collection (typically 60-90% of a collection). The resultant benefits are twofold: ECA (Early Case Assessment) which is automated, keyword-agnostic and incredibly insightful even before review has started, and a far faster and less expensive review – as in 60-90% faster and less expensive. And all with far superior quality and consistency.

One interesting aspect of the skyrocketing interest in Predictive Coding™ are the people and organisations who are most vocal on the topic – positively or negatively – which can essentially be divided into two distinct groups. The first group are industry practitioners, namely in-house counsel, government regulators and outside counsel who are either already using Predictive Coding™ or interested in knowing more about it. There are plenty of skeptics in this group, which is not at all surprising given that 1) the legal industry has always been and will always be risk- and change-averse, and 2) such skepticism is de rigeur when a technology or methodology is moving from the “early adopter” phase to the mainstream. This is a sober, serious group who are keenly interested in the promise of Predictive Coding™ but wanting to know about its efficacy and requirements from peers first-hand before investing heavily in it. This group has been interacting regularly with the early adopter crowd, including customers like [Fulbright & Jaworski](#), [Morgan Lewis](#), [WilmerHale](#) and many more Fortune 500 companies and very large government regulators who feel obligated (for obvious reasons) to not advertise the advanced technology they use.

The second group are those who see any advanced review technology, including Predictive Coding™, as competitive threats to their installed base of customers using their antiquated workflows and technology. This groups is overwhelmingly populated by legacy technology vendors, including purveyors of [woefully outdated linear review](#) and those espousing a dangerously simplistic “[pizza box](#)” approach to eDiscovery. This group reacts to the Predictive Coding™ conversation in one of two ways: either seeking to co-opt “predictive coding” in aggressive marketing messages claiming to have such capabilities, or by attempting to cast doubt on the efficacy of Predictive Coding™ by raising uneducated [red herrings](#) like defensibility. Neither contributes anything of value to the conversation about Predictive Coding™ – or to the advancement of eDiscovery technology generally – as their efforts simply seek to confuse customers about what Predictive Coding™ is and which vendor(s) are actually capable of delivering on its promises, which is of course the intent of this group.

As such, we will ignore this latter group and instead focus on those participants in the conversation who **really** matter: those eDiscovery practitioners who have genuine interest in and questions about Predictive Coding™ and what it might be able to do to address their acute issues with legacy linear review.

Defensibility is a topic that rarely gets much attention, but when it does it can be particularly tricky – mainly because it sounds scary, especially to those who don’t deal with it often (including vendors who toss around legal terms opportunistically with no real understanding of them...but don’t get me started). In reality, legal defensibility is a simple concept in both theory and practice: was a party’s behaviour sufficiently reasonable such that it did not prejudice the court or the opposing side? In the context of Predictive Coding™, the question is simply whether the methodology followed (including

technology used) was reasonable. Of particular note here is the word “reasonable”, which does not mean “perfect”, “infallible”, or “flawless”; as we will see with linear review, the reasonableness standard simply means “good enough”, which in many – perhaps most – cases is far less effective than one might think.

In eDiscovery, and more specifically document review, what is considered reasonable is the current “standard” approach followed by the majority of litigants, namely linear review in its several flavours (explained in more detail [here](#)). Thus, for a review methodology and technology – any methodology or technology, including Predictive Coding™ – to be “defensible” it must simply be able to be shown to be at least as accurate as linear review. While a detailed examination of the true accuracy of linear review is beyond the scope of this post (but will be detailed in future posts), from working with the majority of the AmLaw 100 for nearly a decade we at Recommind know a “typical” first pass linear review accuracy rate (i.e. correct coding decisions less false positives and false negatives, for relevance, privilege and perhaps issue-relation) tends to be in the 50-70% range, perhaps even as high as 75% under optimal circumstances. Put simply, document reviewers typically code 1 in 3 documents incorrectly.

So the defensibility of Predictive Coding™ comes down to one simple question: is Predictive Coding™ more accurate than linear review, i.e. are Predictive Coding™ calls more than 75% accurate? While we can’t speak to the efficacy of solutions from other vendors purporting to have “predictive coding” capabilities (and neither can they), we can say definitively and with absolute confidence that Predictive Coding™ has been shown repeatedly to have accuracy in the mid-high nineties percentile, with accuracy often well over 99%. In fact, one of the many benefits of Predictive Coding™ is that one can actually dictate the level of accuracy one wants with Predictive Coding™...but we’ll need to leave that discussion for a future post.

As noted above, definitively proving the defensibility of Predictive Coding™ is actually quite simple – especially when such proof is in the form of real numbers from real lawsuits, investigations and regulatory proceedings defended and prosecuted by many of the world’s largest law firms, regulatory bodies and corporations.