

Artificial Intelligence and Ethics in Practice – Deep Analysis

Alan Pelz Sharpe¹

¹ Deep Analysis, Nashua Technology Park, 200 Innovative Way, Nashua, NH, USA

Deep Analysis has been a research and advisory firm focused on the ethical use of AI in the enterprise for several years. We have advised technology firms large and small, published a book on AI's practical use, and are currently building a framework and toolset to test and monitor AI systems. In short, I like to think we know what we are talking about; the trouble is though many organizations both in the public and private sectors want to talk about AI, few want to deal with the myriad of ethical issues that surround the use of AI in the workplace. In our work at Deep Analysis, these issues can typically be categorized into two key areas, though there are many others.

- Concerns over AI Bias
- Automation of Work

When we think about AI and ethics, most commonly, concerns about bias are raised; for example, when AI is used in recruitment, is it biased against people of color, certain nationalities, sexual preferences, gender, etc.? The answer is yes; pretty much every AI system contains biases as they are trained on historic HR data that reflects the human tendencies of the recruiters. Sometimes these biases can become apparent and be corrected; they are often more subtle and go undetected. This problem exists not just in HR; such biases usually exist in loan or housing application systems; they live in software used in policing and healthcare. In our experience, unless these biases are blatantly obvious, few ever address them or know how to. That's a concern to us, and it's a situation that will only worsen over the coming years.

AI and ML are most commonly used in the enterprise to automate and replace human work activities. That is not something most technology vendors want to acknowledge. Often go out of their way to market their software as being designed to improve efficiencies and improve the employee experience.

However, this is seldom the case as in our professional experience; the goal of most AI projects is to reduce the number of FTE (full-time employees) in the workplace, thus, in theory, though less often in practice, reducing costs. Since the move to remote working over the past couple of years, more companies have started using AI to monitor employees under the banner of employee wellness. Though different from bias, the use of AI to automate human work at scale and monitor workers raises multiple ethical issues that few wish to acknowledge. If you have read this far and not fallen into a mental pit of doom and despair, then well done you, and thankfully there is some light on the horizon. In contrast to everything stated so far, what we have learned in our work is that organizations are unaware of the challenges surrounding AI most of the time. Once made aware of the challenges and how they can effectively face them, many managers and executives begin to tread far more cautiously.

The fact is, and as industry analysts, we can say this with absolute certainty, AI and its value in the enterprise have been oversold, and its effectiveness overrated. AI is not half as intelligent as the sellers of AI technology would like their customers to believe, and early adopters are already becoming wise to this. Before the pandemic Deep Analysis used to run in-person workshops on the practical use of AI in the enterprise; these were attended in the main by senior managers and executives, rather than technologists. As is often the case in such settings, we learned as much from the attendees as they did from the workshop. The key lesson we came away from these workshops was that most AI enterprise projects are poorly conceived and poorly executed. Further, just a little understanding of how AI works and how to set up and run an AI project, without a need to understand the underlying math, went a very long way.

This understanding includes Black Box versus Transparently AI not at a technical level, to show people that there are many different flavours of AI that do different things and fit other use cases such as a spending a little time explaining the key differences between various AI aspects such a predictive versus prescriptive AI, supervised versus unsupervised, etc. Then we compare and contrast Linear Progression, Naïve Bayes, Hidden Markov models, etc.

But most of the time is spent explaining how to staff and plan an AI project and how to support and maintain it once it goes live. AI projects are not the same as traditional IT projects. They run pretty differently, yet this is often the first time these buyers of AI learn that. Though

we try to reign it in a bit, the truth is the element in our training and book that gets the most attention is the section on the dark side of AI, providing examples and explanations as to how AI can go spectacularly wrong and be misused.

Despite the negativity I have expressed here, Deep Analysis is, oddly enough, a proponent for the use of AI in the enterprise. Deep Analysis is and always will be a small firm, and although we do have some influence in the tech world, we can only do so much.

We need much more education to be available to business leaders. Should they wish to use AI in their organizations, they can ensure that they do so safely, ethically, and effectively. We also need tools and frameworks to evolve to ensure that AI systems are subject to audits for accuracy and fairness. Once in operation, they can quickly become black boxes that cannot be unlocked later. These tools can open up new opportunities and undertake tasks at a scale and speed previously impossible.

Used wisely, AI can be transformative, and deliver beneficial, affordable change. Our concern is that too little is known about the technology by those wanting to use AI, too much hype, and AI can be destructive when used unwisely. There is not much that one can do to curb the use of AI to automate human work; automation and job losses have always gone hand in hand.

That said, more honesty in the conversation would be a good thing. But there is much to be done to improve people's understanding of how AI works, where it fits, where it does not, and how to use it effectively.