Artificial Intelligence–Based Cybercrime

Bogdan Hoanca
University of Alaska, Anchorage, USA

Kenrick J. Mock
University of Alaska, Anchorage, USA

INTRODUCTION

The field of Artificial Intelligence (AI) has made swift progress in recent years, becoming much more pervasive in the lives of ordinary citizens. Many of the advances in AI have already led to exciting capabilities: intelligent agents on smartphones (Apple's Siri and Microsoft's Cortana), intelligent voice interaction devices in the home (Amazon's Echo) and countless interactive toys for children. Other upcoming advances are even more impactful, from self-driving cars to learning robots (Pinto & Gupta, 2015) and intelligent systems that automate the jobs of white-collar professionals. In fact, much has been written about the dangers of automation in terms of job losses (McKinsey&Company, 2017). Others fear even more wholesale threats to society, a dystopian future where super-intelligent robots enslave the relatively inferior humans (Barrat, 2013). Much of the AI-fear is driven by the prospects of super-intelligent AI agents (ASI – artificial superintelligence), agents so intelligent that humans will be as powerless against them as are ants against a human farmer. The literature on AI dangers focuses mostly on the unintentional dangers of AI developing into an ASI agent of destruction or danger, not because it develops a goal to kill us, but because it sees no problem in killing us in pursuit of a goal we designed it for (Bostrom, 2014).

Although fears of super-intelligent AI might be justified at some point, and although humanity needs to plan for the time and the manner in which ASI will be deployed (Bostrom, 2014), researchers have widely divergent views of when ASI agents will become reality (Walsh, 2018). Many experts are not even willing to speculate when ASI will arrive (Ford, 2018). In the meantime, a more immediate threat arises from not-yet-super-intelligent but already available AI: the ability of human attackers to use non-ASI systems to automate, enable and enhance cybercrime as we know it, as well as the ability to open totally new channels for cybercrime. Whether the long-term threat of ASI will materialize, the immediate threat of criminals using AI for cybercrime today needs to be considered – and is the focus of this article.

While focusing only on current AI capabilities, not on super-intelligent agents that are likely to emerge in the more distant future, we also focus only on intentional malicious uses of AI, not on dangers arising from unintentional consequences or on malfunction of AI systems. Neither do we discuss the threat of AI weapons, including autonomous or intelligent ones. Such systems are intended to create harm, and they can obviously be used for that purpose, whether by a legitimate authority or by a nefarious group.

Even after limiting the scope as described above, the range of capabilities of AI for achieving nefarious purposes is vast, at least as extensive if not more so than the range of capabilities for beneficial ones. Within this vast range of possibilities, we classify AI cybercrime into three general and loosely overlapping areas: using AI to commit cybercrime online, using AI via new cybercrime channels that

DOI: 10.4018/978-1-5225-9715-5.ch003

reach into physical space, and using AI or knowledge of AI to strike at the core of other AI systems, by corrupting data or algorithms. These are not three separated areas: they largely overlap, and the extent of their overlap will continue to increase. After providing a brief overview of AI history, the status of cybercrime in general, the article will delve into the three areas, highlighting the ways in which they interact presently as well as in the foreseeable future.

BACKGROUND

AI-based cybercrime is driven by two relentless forces: the development of AI, which has seen incredible strides in recent years, and the increasing volume and diversity of cybercrime, driven by more powerful technologies and more widespread use of the Internet. As background for the main topic of the article, this section will review recent trends in AI development as well as the status of cybercrime in general. The body of the article will focus on the convergence of AI and cybercrime.

Artificial Intelligence (AI)

A research field that emerged in 1956 at Dartmouth University, AI has undergone periods of excitement and growth, interspersed with periods of "AI winter" when excitement and funding dried out. The initial excitement was driven by the hope that major progress could be made swiftly in teaching computers to carry out intelligent activities. As researchers engaged with what they thought would be the most challenging tasks, the surprising discovery was that what was thought as most challenging would turn out to be easy, while some of the most effortless human skills proved most difficult to duplicate by computers. For example, computers were able to run highly complex calculations, to carry out abstract manipulations, create art (Charlesworth, 2018) and to get good at the most complex games, over time defeating the human world champions at backgammon (Berliner, 1980), chess (Weber, 1996) and most recently at Go (Hern, 2017). While the earlier successes were primarily examples of brute force computing overcoming human limited cognitive abilities, the Go victory was based on the system teaching itself to play by playing multiple games against itself (programmed with no strategy, but only with the rules of the game). To prove the point, after defeating the human champion, the Go playing AI proceeded to teach itself to play chess and defeated the chess program Stockfish 8 in a 100-game matchup (Silver, et al., 2017). Outside the field of games, AI systems are making similar progress, although at a somewhat slower pace: understanding images, natural language and achieving "common sense" turned out to be rather challenging for computers, even though such tasks were literally "child's play" for humans.

Part of the problem with the initial failures in AI was that the approach was top-down: attempting to formalize and codify learning and knowledge, and to impart a corpus of this knowledge to the computer. This was overwhelming in terms of assembling such a corpus, as well as in terms of maintaining and debugging it. Starting in the twenty first century, the vast computing power made available by Moore's Law, further made accessible by cloud computing and the Internet, revitalized some of the earlier attempts at using a bottom-up approach, where computers were given rules or examples and taught how to learn. Although neural networks with small numbers of neurons have been used almost from the earliest days of AI, the vast amounts of power and training data available have opened new possibilities when using much larger and deeper neural networks, an approach termed "deep learning." While small networks are unable to improve in performance beyond a certain volume of training data, it became apparent that increasingly complex networks were able to continue to learn and to take advantage of the