Attacking The Source

Introduction

<1 min

It is easy to forget the underlying technology behind AI systems: the software and carefully crafted algorithms. Unfortunately, as with many software, some vulnerabilities can impact them. Though vulnerability impact may look different in an AI system, the nature of vulnerabilities in those systems can generally be boiled down to input handling and logic flaws deep in them.

Suppose we can interact with these systems at critical points in their lifespan. In that case, we can exploit several flaws to modify, extract, and negatively impact the underlying AI system, which can lead to information disclosures or unexpected behaviors.