A NetAI Manifesto (Part II): Less Hubris, more Humility

Walter Willinger
NIKSUN, Inc.

Arpit Gupta
UCSB

Roman Beltiukov
UCSB

Wenbo Guo
Purdue Univ.

ABSTRACT

The application of the latest techniques from artificial intelligence (AI) and machine learning (ML) to improve and automate the decision-making required for solving real-world network security and performance problems (NetAI, for short) has generated great excitement among networking researchers. However, network operators have remained very reluctant when it comes to deploying NetAI-based solutions in their production networks. In Part I of this manifesto, we argue that to gain the operators' trust, researchers will have to pursue a more scientific approach towards NetAI than in the past that endeavors the development of explainable and generalizable learning models. In this paper, we go one step further and posit that this "opening up of NetAI research" will require that the largely self-assured hubris about NetAI gives way to a healthy dose humility. Rather than continuing to extol the virtues and "magic" of black-box models that largely obfuscate the critical role of the utilized data play in training these models, concerted research efforts will be needed to design NetAI-driven agents or systems that can be expected to perform well when deployed in production settings and are also required to exhibit strong robustness properties when faced with ambiguous situations and real-world uncertainties. We describe one such effort that is aimed at developing a new ML pipeline for generating trained models that strive to meet these expectations and requirements.

1. INTRODUCTION

In Part I of our NetAI manifesto [1], we criticize the use of the standard ML pipeline that is popular with NetAI researchers and discuss why relying on this widely-adopted ML workflow is fraught with problems that question the scientific foundations of the artifacts it produces. We argue for abandoning it altogether in favor of a new generation of ML pipelines that are capable of generating explainable ML models that can effectively be examined with respect to their generalizability and safety, describe an initial attempt at designing and implementing such a new ML pipeline, and comment on some exciting new opportunities that arise as result of this proposed paradigm shift in ML model development and evaluation.

In this Part II of the manifesto, we discuss how these proposed efforts towards "an opening up of NetAI research" relate to ongoing developments in the area of human-centered computing where the emergence of agents with increasingly autonomous capabilities are a major concern that dovetails with our calls for explainability and generalizability in everyday ML model development. Specifically, instead of further promoting the much-touted benefits of increasingly autonomous technologies, we elaborate on the urgent need to address the new risks and challenges these technologies pose and relate them to the popular view that the wide-spread adoption of NetAI-based autonomous capabilities will ultimately eliminate the need for human involvement.

In effect, we posit that responding to the new risks and challenges in a constructive manner demands a careful reappraisal that recognizes the existence of a natural "division of labor" between autonomous agents and humans that is counter to widely-held beliefs about the impact of increasingly autonomous technologies in general and NetAI-driven network automation in particular. To this end, we describe a novel ML pipeline that demonstrates the type of division of labor that can ensure a future where we can have the best of both worlds — the full benefits of autonomous capabilities without their possibly debilitating side effects such as "future automation surprises" [2].

2. NETWORK AUTOMATION AND AUTONOMOUS CAPABILITIES

The growing popularity of networked devices and applications imposes increasingly stringent security- and performance-related requirements on the underlying communication infrastructure. Satisfying these ever more demanding and complex requirements in an efficient and scalable manner with limited infrastructure resources and shrinking operational budgets poses significant challenges for today's network operators. A promising approach to address these challenges is to automate some of the real-time decision-making that satisfying these requirements necessitates. To this end, for more than a decade, networking researchers have been busy demonstrating the potential of NetAI, have developed NetAI-based solutions aimed at supporting network automation, and have been envisioning a future where NetAI will be critical for realizing the vision of "self-driving networks." Many of these efforts are, however, based on widely-held beliefs about the impact of increasingly autonomous technologies in general and NetAI-driven network automation in particular, namely that these developments will ensure that human involvement in the decision-making processes required for operating and managing future networks can be reduced to the point where it will eventually become unnecessary.

Tellingly, these beliefs are collectively referred to as the "seven deadly myths" of autonomous systems in [3] where the authors systematically bust these "myths" of autonomy and provide reasons why each of them should be called out and cast aside. Although the authors of [3] take a broad view of autonomous systems and autonomous capabilities, their paper should be required reading for networking researchers, especially because the authors' observations and are directly applicable to and highly relevant for current efforts that focus on NetAI-driven network automation as a stepping stone towards realizing the future vision of self-driving networks. In the NetAI domain, the autonomous capabilities derive from running trained ML models in the data plane (e.g., on programmable switches) and relying on them to make real-time inference decisions. Here, each model can be viewed as an autonomous agent that has been designed with a specific networking task in mind (e.g., detecting the onset of an amplification-type DDoS attack), has been shown to be performant (according to some specified evaluation procedure), and the human operator feels comfortable relinquishing control to the model, thus automating a task that previously was performed by the operator.

In the context of NetAI, of particular interest is the first myth discussed in [3] that views "autonomy" as a unidimensional concept. Instead, the authors of [3] argue that it is more useful to describe autonomous agents at least in terms of the two dimensions referred to as *self-directedness* and *self-sufficiency*. Here, self-directedness is defined as the independence of an agent from its physical environment and reflects a notion of autonomy that is synonymous with independence from outside control. Self-sufficiency, on the other hand, is meant to capture the idea of self-generation of goals and reflects a view that equates autonomy with the capability of an agent to take care of itself. Slightly paraphrasing [3], a main motivation for autonomous capabilities is to reduce the burden on humans by increasing an agent's self-sufficiency to the point that it can be trusted to operate in a self-directed manner. However, when the self-sufficiency of the agent capabilities is seen as inadequate for performing the task the agent was designed for (e.g., in situations where the consequences of errors may be disastrous), it is common to limit the self-directedness of the agent, either by humans taking control manually or falling back to an automated control that is known to prevent the system from doing harm to itself or others through faulty actions (i.e., low self-directedness and low self-sufficiency).

When self-directedness is reduced to the point where the agent is prevented from fully exercising its capabilities (i.e., low self-directedness, high self-sufficiency), the result is an under-reliance on the technology — although the agent may be sufficiently competent to perform a set of actions in the current situation, human-imposed manual controls or policies may prevent it from doing so. The flip side of this aspect is over-trust (i.e., high self-directedness, low self-sufficiency) where an agent is allowed to operate too freely in situations that outstrips its capabilities. The challenge then faced by designers of autonomous agents or systems capabilities is striving to maintain an effective balance between self-directedness and self-sufficiency which in turn imposes the additional challenge on the designers to make the agent or system understandable. In NetAI parlance, these challenges are all-too-familiar. Making agents understandable is synonymous with developing explainable ML models, and model explainability is paramount for assessing model generalizability (i.e., assessing when the model works and doesn't wok (and why not)) and model safety (i.e., identify and quantify harmful and unintended model behavior).

3. AUTONOMOUS CAPABILITIES: RISKS AND CHALLENGES

As discussed in [3], like the mentioned first myth, most of the seven deadly myths or beliefs exist because they ignore or downplay in one way or another the new challenges and risks that materialize with increasingly autonomous capabilities and often take the form of "surprises" or "unintended consequences" that can reduce or even wipe out apparent benefits that may result from increased autonomy. The type of new challenges and risks is highlighted in [4] and has been termed "Doyle's catch" in [5]. It states that

> *"Computer-based simulations and rapid prototyping tools are now broadly available and powerful enough that it is relatively easy to demonstrate almost anything, provided that conditions are made sufficiently idealized. However, the real world is typically far from idealized, and thus a system must have enough robustness in order to close the gap between demonstration and the real world."*

Although not expressed in NetAI language, we recognize Doyle's catch to be yet another formulation of the generalizability problem in ML — the failure of black-box models trained in the confines of an idealized setting (e.g., simple testbed) to maintain their performance when used in the real world (e.g., an actual production network). Thus, whether it is designing autonomous agents that are not prone to "surprises" or "unintended consequences", or developing autonomous capabilities without falling into Doyle's catch, or generating generalizable ML models, the technical challenge faced by researchers interested in developing ML-based solutions for networking problems is to define new ML pipelines that output trained models that are capable of "closing the gap between the demonstration and the real thing".

Since a majority of trained models that have been developed to date in the different application domains of ML are the result of applications of the standard ML pipeline, they are neither able to address nor resolve this challenging task. For one, since the output of the standard ML pipeline are in general black-box models, they provide little to no insights into the models' inner workings and instead continue to bolster the popular view that ML models are able to perform some "magic". Importantly, being black-box in nature, they are by and large unable to yield useful information for researchers interested in ascertaining the models' ability to generalize and there are currently no readily available ML pipelines that facilitate both the identification and remediation of underspecification issues in trained black-box models, a critical step in assessing the models' generalizability. Moreover, using the standard ML pipeline to obtain trained models has the effect of obfuscating the nature of the training data; that is, intentionally or unintentionally blurring critical information about the what, how and who of the collected data and thus about the data's quality.

Specifically, this data quality issue can be attributed to two factors. First, many publicly available datasets are unrealistic in the sense that they have been collected from environments that have little to nothing in common with the "real thing" (i.e., the model's target environment). Second, existing data-collection efforts are fragmented; that is, they only apply to a specific learning problem and/or network environment. In particular, how to extend them to collect representative (labeled) data for a new learning problem and/or from a different target environment is a largely unresolved problem. Together, the obfuscating effect that the use of standard ML pipeline has with respect to the data employed for model training and the fact that the root causes of most model underspecification issues can be traced to problems with the quality of the training data [7] severely limit the options that researchers have to tackle the generalizability problem in ML. On the one hand, these issues highlight how the prolonged and widespread use of the standard ML pipeline in the different application domains of ML has resulted in a self-assured hubris among researchers in general and NetAI researchers in particular about the autonomous capabilities of ML-based agents. On the other hand, they also argue that in the face of the complexities and uncertainties experienced in the real world, some amount of humility is required to realize and accept that designing ML-based agents that are explainable and can be assessed with respect to their generalizability and safety cannot be accomplished by means of established methods such as the standard ML pipeline but demands implementing new ML workflows or pipelines that are radical departures from how ML models have been developed to date.

4. TREATING TRAINING DATASETS AS FIRST-CLASS CITIZENS

In the NetAI domain, the described tension between hubris and humility when researchers are faced with designing ML-based solutions for networking problems is greatly complicated by the fact that, in general, collecting data from the "real thing" to train ML models is, for privacy-related or other reasons, often not possible. In fact, the question of how to develop ML models that maintain their excellent performance even in the "unseen data" case (i.e., without an ability to collect data from "the real thing") while exhibiting the required balance between self-directedness (being robust to the uncertainties in the environment) and self-sufficiency (being able to perform safely despite the inherent fragilities that the complexity of autonomous capabilities entails) has largely stymied researchers in the past but deserves their full attention going forward.

In an initial attempt to resolving this taunting challenge, we recently incorporated Trustee [7], our latest ML pipeline for developing explainable ML models into the design of a radically new closed-loop ML workflow that we call netUnicorn [6]. netUnicorn highlights two innovative and original concepts. First, it leverages a novel data-collection platform that enables the collection of different datasets for any given learning problem from one or more physical or virtual network infrastructures, accurately emulating different target environments with high fidelity. Second, netUnicorn uses Trustee-generated feedback about the latest trained model to iteratively collect new training datasets from some flexibly configurable idealized environment such that the models trained with these new datasets exhibit improved generalizability and have a better chance to maintain their good performance in the "real thing" (i.e., in the actual production network where collecting training data is ruled out). The premise is that the models that netUnicorn outputs will instill greater trust among network operators for production deployments, thereby driving widespread adoption of ML in the field of networking in the future.

5. CONCLUSION

The novelty of netUnicorn prevents us from reporting initial experiences from researchers who recognize the need for new ML pipelines that strive for outputting generalizable ML models. However, based on our own experience to date that admittedly encompasses only a small sample of different learning tasks and different target environments, the ML artifacts that result from applying our new closed-loop ML pipeline can be shown to be performant and to exhibit improved generalizability. However, only time will tell whether abandoning the standard ML pipeline with its deliberate tendency to rely on the "magic" of black-box models and implicit attempts at obfuscating the critical role of the underlying training data and replacing it with radically different ML pipelines such as netUnicorn will have the intended consequences — the routine development of ML models that are both explainable and generalizable and where the role that the utilized data plays with respect to model training can be assessed explicitly. However, we posit that it is by developing new ML pipelines such as Trustee and netUnicorn that the NetAI domain can pave the way towards a future where ML models will be both recognized as a means for scientific discovery and appreciated for being of inherently practical value for achieving the autonomous capabilities required by ongoing efforts that see NetAI-based network automation as a stepping stone towards realizing the vision of self-driving networks.

6. REFERENCES

[1] W. Willinger et al.A NetAI Manifest (Part I): Less Explorimentation, More Science. *Performonce Evaluation Review, this issue (2023).*

[2] D. D. Woods. Automation Surprises. *In: Joint Cognitive Systems: Patterns in Cognitive Systems Engineering, 113–142, Taylor & Francis, 2006.*

[3] J. M. Bradshaw et al.The Seven Deadly Myths of 'Autonomous Systems'. *IEEE Intelligent Systems 28(3), 2–8, 2013.*

[4] D. L. Alderson and J. C. Doyle. Contrasting views of complexity and their implications for network-centric infrastructures. *IEEE SMC-Part A, 40(4), 839–852 (2010).*

[5] D. D. Woods. The Risks of Autonomy: Doyle's Catch. *Journal of Cognitive Engineering and Decision Making, 10(2), 131–133 (2016).*

[6] R. Beltiukov, W. Guo, A. Gupta, and W. Willinger. In Search of netUnicorn: A Data-Collection Platform to Develop Generalizable ML Models for Network Security Problems. *https://arxiv.org/abs/2306.08853 (2023).*

[7] A. S. Jacobs et al.AI/ML for network security: The emperor has no clothes. *Proc. ACM CCS'22 (2022).*