Kingfisher
SYSTEMS, INC.

WHITE PAPER

Artificial Intelligence and Data Distillation

At Kingfisher, Machine Learning (ML) is viewed as a subset of Artificial Intelligence (AI) where — with our principal partner PermissionBit — there has been a focus on deep learning for time series and cybersecurity. Our approach to Natural Language Processing (NLP), described in a separate NLP White Paper on our website, is another focused component of our overall AI approach.

Prepared By:

Kingfisher Systems, Inc.

3110 Fairview Park Drive, Suite 1250

Falls Church, VA 22042

Phone: (703) 635-2951

Fax: (703) 820-7976

Email: varyss@kingfishersys.com

At Kingfisher, Machine Learning (ML) is viewed as a subset of Artificial Intelligence (AI) where — with our principal partner PermissionBit — there has been a focus on deep learning for time series and cybersecurity. Our approach to Natural Language Processing (NLP), described in a separate NLP White Paper on our website, is another focused component of our overall AI approach.

AI is often promoted as a solution to national security problems, but the suggestion assumes that solutions to the many machine learning and data science tasks that must be completed are known, or that the learning process will automatically solve them. Representation learning has been successful in highly structured domains such NLP and image recognition, but more poorly structured problems often require subject matter expertise in conjunction with machine learning experiences to constrain the problem. Limiting goals on less well specified problems to augmenting and accelerating existing decision procedures is often a prerequisite for success.

Time series, including those obtained by repeatedly measuring something of intelligence interest, are not well handled by deep neural networks. Advances in audio understanding with recurrent networks do not generalize to other domains. An example is cybersecurity, where large collections of machine data have not led to better network defense. Our partners developed better methods for encoding ultrafast time series, providing techniques that can be applied to social, political, and economic time series. In the cyber domain, the techniques yielded over 10-fold improvements on behavioral analytics and malware detection tasks.

Kingfisher Systems also brings broad and deep capabilities to quantitative social science. As part of this effort, Kingfisher has developed an interrelated set of machine learning and AI competencies — in addition to core deep learning methods such as convolutional neural nets, our competencies include managing exceptionally large unstructured and unlabeled datasets, the use of stochastic generative models for regularization, kernel methods for metric space embedding, and methods derived from dynamical systems research for analysis of high dimensional, nonlinear and ultrafast time series. Massive open source datasets in many domains have transformed intelligence collection and analysis. We combine over a decade of experience exploiting this data with the tools and expertise necessary to work with it, and we process hundreds of thousands to millions of articles and posts daily.

Political and social behavior is an emergent phenomenon defined by millions of individual articles and posts. Statistically robust models that reproduce the observed dynamics exist, but most mine the data for interesting observations or produce aggregate statistics that do not enable faster or better decisions. The dynamics is the source of complexity in time series and is the phenomenon at interest in most decision problems. Modeling the behavior as a dynamical system lets us apply statistical tools developed for dynamical systems to the entire dataset. We refer to this approach as *data distillation*, in contrast to the data mining approaches used by others.

Our collection system and distributed computing capabilities were specifically built to apply our techniques at scale. Interdisciplinary work in quantitative politics and dynamical systems led us to develop robust statistical measures to exploit it. Our data pipelines support continuous delivery, and with our expertise we can provide decision support services at the speed of the modern web.

Our staff and advisors combine leading data scientists and software engineers with former intelligence community and military personnel with backgrounds in technology and counterintelligence. AI is not just applied machine learning. It requires a deep domain understanding that allows the computational results to transform operations. With the domain expertise of our mission partners we can address all aspects of AI in the national security space.