

Jan Brabec: University of Cambridge internship report

I am writing to report on my one-month internship experience at the University of Cambridge, which took place in May 2023. I was visiting the CBL group at the Department of Engineering. During my time there, I had the opportunity to engage in various academic and cultural activities, as well as gain valuable insights into my research area.

My primary focus during the internship was to work on aggregating email detection data to determine whether the emails are phishing or benign using probabilistic programming. This task is closely related to my PhD. studies where I focus on the topic of class-imbalanced data classification in cybersecurity.

During my stay I was using the Turing.jl framework for Bayesian inference. Turing.jl is the most popular such framework in Julia and was developed in the group I joined and my host, Hong Ge, is the main author of the framework. I spent a significant amount of time designing the evaluation process and immersing myself in Bayesian modeling using Turing. I tested various modeling approaches focusing on combination of labels of varying certainty. I had enough time to study the inner workings of the framework deeply and overall gain a solid understanding of Bayesian methods such as various MCMC algorithms and relationships between various distributions useful for modeling of the problem.

Bayesian inference is a promising and useful approach to study because it offers a principled and flexible framework for incorporating prior knowledge and updating beliefs based on observed data. In this concrete problem we possess a baseline classifier that is partially hand-tuned and not fully learned from data. In addition, we possess a very limited and biased dataset of labeled data. Bayesian inference allows using the baseline classifier as a prior that regularizes the model learned on the small and biased dataset.

I developed a promising PoC that I am currently testing more deeply after my return back to Prague with the hope that if the experiments continue successfully this may result in a publication at a security conference.

Throughout the internship, I attended various talks and events that showcased the University of Cambridge's research culture. Some highlights included a lecture from a student of the late Stephen Hawking on the Big Bang theory or a talk by Geoffrey Hinton on the dangers of Artificial General Intelligence. These experiences helped me better understand the importance of addressing large-scale, complex research problems and the value of interdisciplinary collaboration.

During my stay, I had the chance to meet and interact with students from different research groups, which provided an opportunity to exchange ideas and learn about their work. The group was mostly focusing on biology and environmental science in its research. It was also focusing on basic research in Bayesian methods.

My time at the University of Cambridge has been an enriching experience that has broadened my knowledge and skills in my research area. I have gained a better understanding of Bayesian modeling and the potential applications of my work. Furthermore, my exposure to the university's research culture has inspired me to think more strategically and ambitiously about my

own work. I am confident that the knowledge and experience gained during my time at the University of Cambridge will be invaluable to my future academic endeavors.