Thesis Project Portfolio

Using Computer Vision for Analysis of Collapsed Bose-Einstein Condensate in Dual Sagnac Interferometer

(Technical Report)

The Rise of Generative AI: Emergent Intelligence, Extreme Risks, and the Race Toward AGI

(STS Research Paper)

An Undergraduate Thesis

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Peter Sailer

Spring, 2024

Department of Computer Science

Table of Contents

Sociotechnical Synthesis

The rapid advancement and widespread deployment of increasingly powerful machine learning architectures have great potential to facilitate the discovery of new knowledge and inventions. However, as with the introduction of any powerful technology, there are also significant risks associated with increasing widespread access. This thesis will explore both sides of this nuanced tradeoff. The technical report explores a constructive scientific use case for computer vision in extracting phase angles from an atom interferometer gyroscope (AIG) while the thesis paper navigates the potential risks associated with the development of generative AI, specifically large language models (LLMs). Both pieces offer a unique perspective on the dual nature of machine learning as a tool for innovation and a source of societal concern.

Efficiently being able to accurately record changes in orientation is integral to the success of high-altitude travel, where satellite navigation systems such as GPS have sparse coverage. As such, spacecraft are often dependent on internal gyroscopes which track changes in angular acceleration and movement through higher altitudes and space. One such highly precise and relatively compact option for navigating space flight is the atom interferometer gyroscope. In particular, the Dual Sagnac Interferometer, an AIG developed by the uva cold atom lab, works by splitting a Bose-Einstein Condensate (BEC) into two or more coherent matter waves that follow different paths before recombining, causing interference patterns that can be used to measure properties such as rotation.

My technical report presents a method for predicting a vessel's change in orientation based solely on images of the collapse BEC within the interferometer. To accomplish this I first perform a data reorganization technique called principal component analysis (PCA) which

generates a dozen basis images capturing different independent factors affecting the appearance of the collapsed particles. I then construct a linear combination of these basis images and perform a multivariate linear regression across them optimizing such that the difference between the test images dot product with the linear combination and the true angle of deflection is minimized. This results in a single optimal basis image which can simply be multiplied by the input images to calculate their corresponding angles of deflection. The report demonstrates feasibility of this low resource approach in extracting derived features from images of BECs using computer vision.

While the technical report focuses largely on a benign practical application of machine learning in scientific research, the STS research paper investigates the potentially dangerous capabilities of modern GPT architectures. Between hackathons and safety conferences, researchers have shown that both proprietary and open source algorithms are capable of generating new bioweapons when modified to do so. Companies such as Palantir have already found ways of weaponizing current surveillance technology by generating attack option recommendations and drone target assessment through the use of LLMs. What's more, OpenAI admits that GPT-4 is capable of generating long term misinformation campaigns favorable to autocratic regimes. The paper also speaks on the emergent capabilities of LLMs. Models such as GPT-4 which have exceeded a certain model size, appear to be suddenly endowed with hard to forecast capabilities such as the ability to use tools, self reflect, and perform multi step reasoning. The presence of these emergent capabilities make demonstrably clear how little is known about these architectures, especially when all the major AI labs such as DeepMind, Anthropic, and OpenAI shy away from an explanation. While considerable action has been taken to raise awareness of the potential dangers at bay, the sentiment has mostly remained privately held

within the AI community. Lawmakers are starting to catch on, but with the pace of progress set so steep, any actions that have been taken appear to be ineffective. Hence, the development of evermore powerful and unrestricted language models seems inevitable at the current time.

In conclusion, while the technical report and the STS research paper focus on different aspects of AI technology, they are linked by the overarching theme of transformative potential. The technical report demonstrates the beneficial impact that AI can have on scientific research and technological advancement, while the STS paper highlights ethical concerns associated with the development of advanced GPTs. Together, these papers emphasize the importance of balancing the benefits and risks of advancing AI technology and the need for ongoing efforts to ensure that AI is developed ethically. As AI continues to progress at an unprecedented pace, it is crucial that researchers, policymakers, and the public work together to navigate the complex landscape of AI development and deployment.