

SPARC 2022: Efficient Machine Learning

June 20 - August 26
GCHQ Manchester

The Big Picture

The development of modern Deep Learning has changed the nature and pace of technological innovation across the world. Demand for fast, high performance models is growing. Demand for performance improvements has been met at the expense of ever increasing computational cost; In 2012 AlexNet won the ImageNet Large Scale Visual Recognition Challenge with 62.3 million parameters. The current ImageNet leader has 2.4 billion parameters. The best language models use over 100 billion parameters.

Using these models is computationally expensive. This computational expense further induces **environmental**, **financial** and **technical** cost. That's where you come in. At SPARC 2022 we want you to help us investigate **sustainable and efficient machine learning**; we want to know what we can do to control these expenses, and what impact this will have on our mission.

Headline Challenges

The world of deep learning is pervasive and ever growing. We are setting specific challenges. These challenges are community driven. We are aiming for results that improve the efficiency of deep learning. The topics below are our focus, but we are a broad church; **expect more from the problem book**.

Speech

Compression for Automatic Speech Recognition (ASR). What do we lose? How do we lose it?

Speech Recognition at GCHQ is key to supporting our mission. The emergence of neural tooling for ASR has dramatically improved the benchmark for what good performance looks like against classic challenges in speech processing. Modern libraries like ESPnet even include large language models in Speech-to-Text (STT) decoders. All in all, the speech ecosystem represents a complex interaction between a wide array of different models, of different types. We want to know **what we can cut** and **where we can cut it**. Without your help, it's a long road. Bring your expertise to the table!

Text

Compression for Transformers? Delete Millions of Parameters!

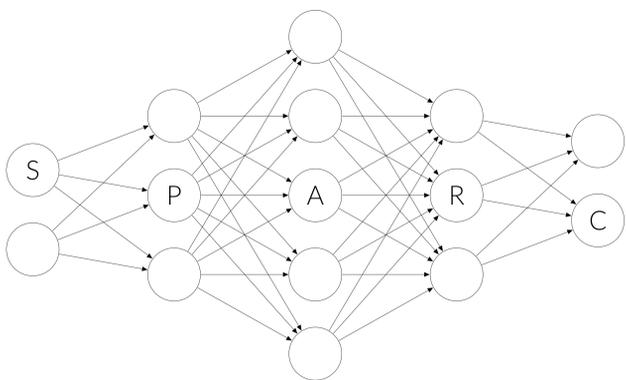
In state-of-the-art Natural Language Processing (NLP), deep transformer models with many billions of parameters lead the way. Whether its few-shot-learning, text generation or online Natural Language Inference (NLI), if you want the best results you'll have to pay the highest computational price. Can you **reduce this cost** whilst **maintaining performance** and measuring the effects?

TinyML

Efficiency and Model Security; What's the link?

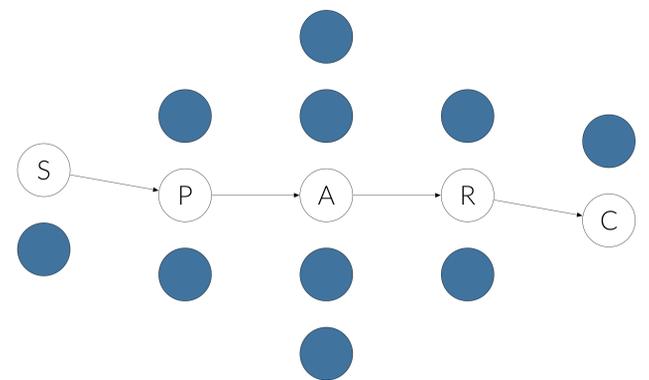
GCHQ and our partners are often limited by the devices on which we can deploy models; low resource and embedded devices are often all we have access to in mission applications. We want to know what differs between standard Machine Learning (ML) and ML on low resource devices. How do these differences due to model miniaturisation affect **performance** and **security**? Are these TinyML models inherently more or less robust to attack? What about adversarial transferability? Come and find out for yourself!

We Want You In Our Network



SPARC

Manchester



Get in touch

If this sounds like something you're interested in, email us at

sparc@gchq.gov.uk

and include your **name**, **organisation**, and some provisional **dates** you'd like to participate. We will prioritise people who want to commit the most time.

Save the Date

June 20 - August 26

Come and work with us. We're keen to have you here in person with us in **Manchester**, but if you can't come in person please join us **virtually** or balance the two. We will look to support your choice of working hours.

Share Your Problems

Above we've just given you a flavour of the types of problems we're interested in. We welcome your input. We want to build efficient machine learning solutions for a **faster**, **cheaper**, **greener** approach to deep learning. We want people from varying backgrounds to help us achieve this. **We need your perspective**.