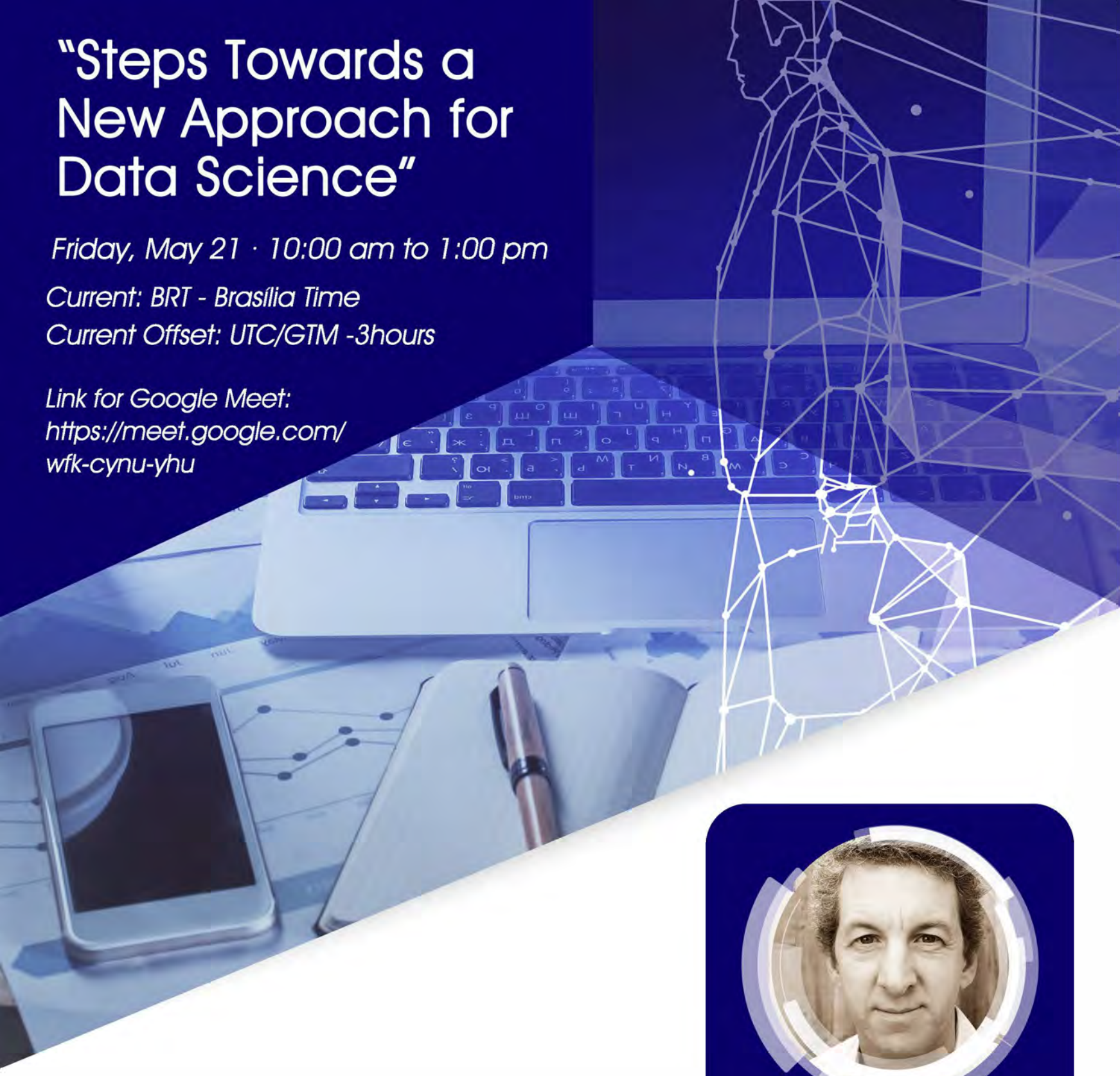
"Steps Towards a New Approach for Data Science"

Friday, May 21 · 10:00 am to 1:00 pm

Current: BRT - Brasília Time
Current Offset: UTC/GTM -3hours

Link for Google Meet:
https://meet.google.com/wfk-cynu-yhu



Abstract: The use of data science is becoming increasingly prevalent in the context of scientific progress. This appears to stem from a combination of: accessibility of data; availability of computing resources; useability of tools; ambition.

Some of the progress relates to the use of popular tools, specifically those employing Deep Learning, that have the key advantage that they can be configured relatively straightforwardly to exploit huge computational resources and can be readily applied to large datasets. The result in commercial settings is that the number of transistors being used for Deep Learning is doubling every four months; commercial ambition is dramatically outpacing Moore's law.

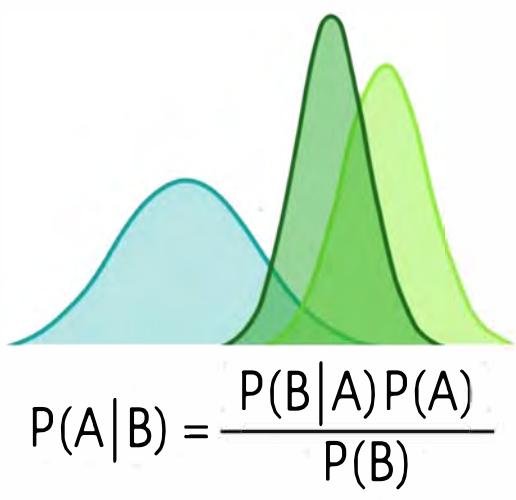
These techniques have the key disadvantage that they do not fully capitalise on pre-existing scientific understanding and do not work well in the context of small data which is expensive to collect (in terms of finances, time or ethical considerations). This situation has led to the development and adoption of probabilistic programming languages (PPLs), tools that enable users to describe scientific hypotheses and the relationship of these hypotheses to data. The tools use Bayesian statistics to make inferences from the data and the use of PPLs is currently growing exponentially. One popular PPL is ***Stan***.

In this talk, Simon will describe this context and then explain how the team's recent research (***Big Hypotheses***) relates to the development of a family of algorithms and techniques (SMC Stan, Streaming Stan) that can exploit large-scale parallel computational resources. The overarching aim of this research is to make it possible for us to generate accurate results in the context of problems that have only recently been considered impossible to solve. To bring this idea to life (and with reference to ***CoDatMo***), the problem of analysing data pertinent to the prevalence of COVID (eg in Brazil) will be discussed with a focus what has been done, what is happening now and what might now be possible in the future.

Biography: ***Simon Maskell*** is a Professor of Autonomous Systems, Director of the **Centre for *Doctoral Training in Distributed Algorithms*** and leads the ***Signal Processing research group*** comprising of 50 + individuals with strengths in in Bayesian computational methodology, autonomy, decision support, data fusion, tracking, image processing, radar processing, acoustic analysis, text analytics, machine learning, behavioural analytics, simulation and energy-efficient hardware implementation. Simon is trained as an engineer and as such, he sees the world as a set of problems to solve and tries to anticipate the problems of tomorrow through state-of-the-art Bayesian techniques.

We will have time for questions and answers and networking at the end.

Relevant References:
https://arxiv.org/abs/2004.12838
https://ieeexplore.ieee.org/document/9158397



$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

UNIVERSITY OF LIVERPOOL

UNINOVE
Universidade Nove de Julho