DARE
ARC Training Centre in
Data Analytics for Resources & Environments

Australian Government
Australian Research Council

Australia's science and research priorities

This is the response from DARE to: https://consult.industry.gov.au/sciencepriorities1

What are Australia's greatest:

a. challenges that science could help to address?
b. opportunities we should seize?
c. strengths we should maintain or build?

General response

We live in extraordinary times. Never before has the world been able to collect so much data about so many different things in so many different locations, using a multitude of sensors from the nano scale to the satellites observing the total earth. However, collecting data is only one part of the puzzle. **The major challenge** is to use this exponentially growing amount of data to improve policy and management on local and international scales.

The value of the collected data and the opportunities for a Fourth Industrial Revolution and Digital innovation are well recognized[1]. Digital innovation was estimated to have the potential to add $315bn to Australian economy[2] and $15.4Tr to the global economy[3].

Alongside this, there has been a sharp increase in the advancement of different machine learning and artificial intelligence techniques, resulting in development of several large-scale language models (LLM) relying on significant computing power and databases.

Despite all these exciting developments, there is still a long way to go before productivity gains from the increased data streams are actualised. There are still large gaps in the capacity of Australia to capitalise on the opportunities offered by the increased volumes of data. In DARE's opinion, Data Science, still has **major challenges in three major areas**:

1. Data Science currently relies on heavy computing to be able to capture uncertainties and to build large machine learning models. As such, the scalability of Data Science solutions is still limited by the availability of significant computing resources outside very large enterprises, and limited scalable methods, which may require less computing.
2. Many current Data Science solutions are limited with respect to scenario modelling and causal feedback loops. Machine learning and AI models predict within the vicinity of the training data but often fail to provide reliable predictions outside existing data boundaries.
3. Data Science is limited in achieving its multidisciplinary potential, the development of novel Data Science has been mainly focused within existing disciplines, such as Computer Science and

[1] https://www3.weforum.org/docs/WEF_Data_Science_In_the_New_Economy.pdf
[2] https://www.csiro.au/en/News/News-releases/2018/Australias-315bn-opportunity
[3] https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html

Mathematics and Statistics, with limited cross over into multi-disciplinary fields. To be able to solve the complex problems faced by modern society, Data Science needs to be more informed by other disciplines and develop methods that can solve complex ever-changing real-world problems. This also includes limits in human and AI interactions, where AI is often seen as a black box and incomprehensible process.

There are several **opportunities** to strengthen Australia's capacity in this area:

- Interest and recognition: Due to global developments in AI, there is increasing interest in Data Science, and increased recognition by business and the public of the *value* of Data and Data Science in a multitude of disciplines.
- Scenario capable models: There is a strong trend towards "interpretable" and "physics based" AI and machine learning, which promises to be more adaptable to complex and changing challenges in real world systems. Such approaches can be queried under changing conditions and can be used for scenario analyses. Making AI more human interaction friendly can be seen as part of this trend.
- Emerging investment: The emerging developments and investments in computing potentially provide increased computing capacity for large AI and Machine Learning models.
- Automation potential: With Australia's economy more and more underpinned by the service industry and lacking real opportunities for productivity gains, AI and Machine Learning open new avenues for automation of tasks (and thus, productivity gains). Using interpretable Data Science to solve complex multi-disciplinary problems will strengthen gains that are currently unachievable.
- Training: to ensure that data science remains responsive to emerging challenges and opportunities, it is important to prioritise ongoing education and training for researchers and creating the next generation of data science graduates.

The DARE Training Centre is already leveraging the benefits of data science methodologies applied to natural resource and environmental challenges in recent simulation and modelling reports with the NSW Smart Sensing Network "Where is The Water?"[4] and The Minderoo Foundations Flourishing Oceans initiative, the Global Fishing Index [5]. Including a National Research Priority based on multi-disciplinary applications of Data Science will strengthen these emerging trends.

For Australia to capitalize on the momentous opportunities that Data Science offers, there is an essential need to maintain the strengths in the areas of Mathematics and Statistics and Computer Science. However, this needs to be matched with a critical focus on maintaining multi-disciplinary teams and research capacity. The real gains in applications of Data Science can only be achieved in broader multi-

[4] https://static1.squarespace.com/static/5b623bcfda02bce646ae3f10/t/62d0a9e2a110f4190b825a6d/1657842161855/Where+is+All+the+Water+-+Final+Report.pdf

[5] https://www.minderoo.org/global-fishing-index/

disciplinary applications that build on foundational strengths in the more theoretical disciplines. In short, a strengthening of cross-STEM multi-disciplinary research and teaching is crucial.

Does Australia have the capability and capacity needed to address these challenges, opportunities and strengths? If not, how could we build this?

Currently, there are many Data Science initiatives at Universities, in government (such as CSIRO's Data61) and in the private sector. Many of these are focused on specific areas or topics.

To address these challenges there is a need to **build capacity and capability** in the following themes:

1. Underpinning theoretical Data Science to develop interpretable AI and Machine Learning as an interdisciplinary Mathematics, Statistics and Computer Science topic to develop novel ways to predict and forecast under rapidly changing external factors.
2. Fundamental research in Quantum computing, computing science and other computing technologies to rapidly increase the capacity to deal with massive scale problems.
3. Expanding research in Human AI interactions and interpretable Data Science to increase the opportunities to query, understand and interact with AI and data science tools.
4. Multi-disciplinary Data Science applications for complex problems in environment, minerals and manufacturing, business and finance, and social science areas.

Most of these themes have a strong STEM basis in programming, mathematics and statistics and physical processes (ranging from basic science to the environment). A need to strengthen the STEM areas to respond to the future is well recognised[6].

The first two themes listed above, are needed to drive innovation within the third and fourth theme of developing user-friendly, interpretable, multi-disciplinary Data Science applications for complex problems facing our nation at relevant scales.

Are the principles the right principles to shape the priorities?

We support the principles, and specifically:

- Be community informed to allow input beyond the government and research community.
- Be evidence-based, as we believe that priorities should be based on evidence of research gaps.
- Be bounded, to make the priorities clearly achievable and not open-ended.

[6] https://www.chiefscientist.gov.au/sites/default/files/STEMstrategy290713FINALweb.pdf